Khawaja, M. A., Schorlemmer, D., Hainzl, S., Iturrieta, P. C., 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.

https://doi.org/10.1785/0120220028

Multi-resolution grids in earthquake forecasting: the Quadtree approach

Khawaja M. Asim[1,2], Danijel Schorlemmer[1], Sebastian Hainzl[1], Pablo Iturrieta[1,2], William H. Savran[3], José A. Bayona[4], and Maximilian J. Werner[4]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany

[2]Institute of Geosciences, University of Potsdam, Potsdam, Germany

[3]University of Southern California, CA, USA

[4]School of Earth Sciences, University of Bristol, Bristol, UK

Corresponding Author

Khawaja M. Asim, German Research Center for Geosciences, Telegrafenberg, 14473, Potsdam, Germany, khawaja@gfz-potsdam.de

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.

ABSTRACT

The Collaboratory for the Study of Earthquake Predictability (CSEP) is an international effort to evaluate probabilistic earthquake forecasting models. CSEP provides the cyberinfrastructure and testing methods needed to evaluate earthquake forecasts. The most common way to represent a probabilistic earthquake forecast involves specifying the average rate of earthquakes within discrete spatial cells, subdivided into magnitude bins. Typically, the spatial component uses a single-resolution Cartesian grid with spatial cell dimensions of $0.1^\circ \times 0.1^\circ$ in latitude and longitude, leading to 6.48 million spatial cells for the global testing region. However, the quantity of data (e. g. number of earthquakes) available to generate and test a forecast model is usually several orders of magnitude less than the millions of spatial cells, leading to a huge disparity in the number of earthquakes and the number of cells in the grid. In this study, we propose the Quadtree to create multi-resolution grids, locally adjusted mirroring the available data for forecast generation and testing, thus providing a data-driven resolution of forecasts. The Quadtree is a hierarchical tree-based data structure used in combination with the Mercator projection to generate spatial grids. It is easy to implement and has numerous scientific and technological applications. To facilitate its application to end-users, we integrated codes handling Quadtrees into pyCSEP, an open-source Python package containing tools for evaluating earthquake forecasts. Using a sample model, we demonstrate how forecast model generation can be improved significantly in terms of information gain if constrained on a multi-resolution grid instead of a high-resolution uniform grid. In addition, we demonstrate that multi-resolution Quadtree grids lead to reduced computational costs. Thus, we anitcipate that Quadtree grids will be useful for developing and evaluating earthquake forecasts.

INTRODUCTION

Earthquake forecasts are an important ingredient of Probabilistic Seismic Hazard Analysis (PSHA), describing the expected magnitude and spatio-temporal distribution of future earthquakes. In the last two decades, it has become a common practice to assess the reliability of forecast models through rigorous testing against independent, future data, referred to as prospective testing. The Collaboratory for the Study of Earthquake Predictability (CSEP) was established as an international collaboration that conducts experiments to prospectively evaluate earthquake forecast models, using a set of rules and metrics common to all competing models (Schorlemmer et al., 2007). CSEP Testing Centers were set up in California at the Southern California Earthquake Center (SCEC) Schorlemmer and Gerstenberger (2007), in Japan at the Earthquake Research Institute of the University of Tokyo (Tsuruoka et al., 2012), in New Zealand at GNS Science (Gerstenberger and Rhoades, 2010), and in Europe at ETH Zurich (Marzocchi et al., 2010).

As the first step of an earthquake forecast experiment, CSEP characterizes a specific region as a testing region. This decision is mainly based on the purpose of the experiment but also constrained by the seismic activity and the seismic network coverage for the region (Schorlemmer et al., 2010). CSEP forecasts are expressed as Poisson rates for a forecasting period (usually 1 day, 3 months, or 5 years). The testing region is divided into spatial cells in longitude/latitude scale with a grid-spacing of $0.1^\circ \times 0.1^\circ$. The forecasts in these cells are further subdivided into rates within 0.1 magnitude units. Thus, a complete forecast is provided in the form of earthquake rates distributed over space-magnitude bins, called a grid-based forecast. To evaluate earthquake forecasts against observed data, an unambiguous procedure is specified on how these observations are going to be processed and used, referred to as an authoritative dataset within CSEP (Zechar et al., 2010b). The testing metrics are also based on a community consensus. An overview of the Poisson consistency testing metrics used to evaluate earthquakes forecasts is available in Section CSEP overview.

The standard $0.1^\circ \times 0.1^\circ$ spatial resolution for grid-based forecasts lead to 7682 and 6.48 million spatial cells for the Californian (Schorlemmer and Gerstenberger, 2007) and global testing regions (Strader et al., 2018), respectively. A single-resolution grid has been convenient in forecast experiments for the following reasons. (i) The 0.1° cell dimension matches approximately the accuracy of earthquake locations, making smaller cells not useful (Bakun et al., 2011). (ii) They can be easily stored due to their regularity and can be easily used by others without complex format descriptions or difficult parsing. (iii) They are convenient in terms of programming as common libraries for managing this type of data are ubiquitous. (iv) The global grid perfectly aligns in this case with various regional grids, e. g. in California, New Zealand, and Italy, allowing global forecast models to be comparable regionally by simply masking the cell outside the local testing region.

However, such grids also come with disadvantages: the single-resolution grid defined for the global experiment comes with 6.48 million spatial cells, which is further increased by the cell subdivision into magnitude bins to more than 200 million space-magnitude bins (Taroni et al., 2014; Bayona et al., 2021). This amount requires considerable computational resources for forecast model generation, processing, and storage. The efficiency of the forecast generation and evaluation process is of high importance, especially when dealing with 1-day forecasts that are generated and tested daily. A single Poisson consistency test of one forecast with 200 million space-magnitude bins can easily take up to several hours, depending on the computer used and the efficiency of the codes.

The second disadvantage of single-resolution grids is that seismicity mostly occurs on major faults and plate boundaries, leading to a spatially heterogeneous, fractal distribution on both global and regional scales (Kagan, 2007), as shown in Figure S1 of the supplemental material to this article. The global distribution of earthquakes is inhomogeneous, and only a fraction of cells received any notable event. The smaller magnitude events on regional scales also exhibit the same property. Figure 1(a) shows the number of recorded earthquakes in $0.1^\circ \times 0.1^\circ$ spatial cells for the California testing region. To quantify the spatial distribution, Figure 1(b) shows the percentage of cells having zero, one, two, or more events per grid cells for the whole globe and California, based on 28465 $M \geq 5.15$ earthquakes, covering the 37 years from 1976 to 2013 for the globe, and 25227 earthquakes with $M \geq 2.5$ in the period between 2000 and 2015 for California, respectively. In the global grid all earthquakes are contained in less than 1% of cells leaving 99% of cells without any earthquake. Even on a regional scale, e. g. in California, 70% of the spatial cells are without any earthquake.

This huge disparity in the number of earthquakes available in the catalogs and the number of cells indicate that such a high-resolution grid is unwarranted for generating forecast models. Generally, the statistical power of tests used to evaluate forecast models is directly related to the quantity of data available for testing (Bezeau and Graves, 2001). Thus, the extreme imbalance between the number of spatial cells and the quantity of earthquake data (low sample size) can result in a low statistical power of the applied tests (Button et al., 2013). Furthermore, empty cells during the training require special attention because during model evaluation the occurrence of a single event in a cell with a predicted zero rate would lead to an immediate rejection of the whole model (Schorlemmer and Gerstenberger, 2007). Therefore, to provide forecasts for areas without past earthquakes, forecast modelers need to either employ spatial seismicity smoothing kernels (Akinci et al., 2018; Helmstetter and Werner, 2014) or allocate a baseline forecast rate to the cells, referred to as water level (Bird and Kreemer, 2015; Kagan and Jackson, 2011; Bird et al., 2010). Using an adaptive grid resolution to reflect the available data for generating a forecast model can offer a better alternative to such decisions.

Simply decreasing the resolution of the spatial grid by increasing the cell size everywhere would reduce the number of cells but would also lead to a loss of resolution in regions of high activity. Therefore, it is desirable to use a multi-resolution grid with high resolution (small cells) in seismically active regions and low resolution (large cells) in regions with low to no seismicity. This saves computational resources that would otherwise be needed to analyze many spatial cells without earthquakes.

Section CSEP overview provides a more detailed overview of previous and ongoing efforts carried out by CSEP. In Section Alternative spatial forecast descriptions, we discuss different possibilities for acquiring multi-resolution grids and then present our proposed approach. Section Model Testing and comparison shows the results and provides answers to potential questions regarding using multi-resolution grids in place of conventional grids.

CSEP OVERVIEW

CSEP accepts forecasts in a pre-defined grid format, upon which different statistical and mathematical testing procedures are performed for evaluation. The testing suite contains multiple tests to evaluate the different aspects of forecasts assuming a Poisson process, e. g. N-test, CL-test, S-test and M-test (Schorlemmer et al., 2007; Werner et al., 2011; Zechar et al., 2010a). The N-test measures the agreement of the total number of events between forecast and observation. The CL-test provides an overall measure of the consistency of the forecast's spatial and magnitude distributions with observations. Because it simultaneously compares the spatial and magnitude distribution of the forecast, the CL test cannot evaluate the individual distributions. Therefore, two new tests have been introduced to evaluate the spatial and magnitude distributions separately, known as the S-test and the M-test, respectively (Zechar et al., 2010a). The performance of multiple models can be compared using the paired T-test and W-test based on information gain per earthquake (IGPE), which determine if the difference between the model scores is statistically significant (Rhoades et al., 2011).

CSEP testing experiments have evolved over the years in terms of testing methods and software design (Gordon et al., 2015; Bayona et al., 2022). Recently, the software structure of CSEP has been redesigned from a monolithic code base to an object-oriented and open-source framework in Python, known as pyCSEP (Savran et al., 2022a,b). It provides modules for accessing and processing earthquake catalogs, representation of forecasts, community-agreed statistical tests to evaluate earthquake forecasts, and routines for visualizing the results. To avoid the assumption of a Poisson process, a new type of experiments has been introduced that represents earthquake forecasts in the form of synthetic catalogs (Savran et al., 2020; Field et al., 2021). The generation of an exhaustive set of synthetic earthquake catalogs tends to provide coverage of the earthquake probabilities without assuming the independence of the different cells, but may or may not be able to cover the low-probability regions successfully.

So far, almost all forecast experiments have used $0.1^\circ \times 0.1^\circ$ grid cells despite the limitations mentioned previously. Here, we propose to define data-driven multi-resolution grids and use them as alternatives to the conventional spatial grids for forecast model development and evaluation, evolving the CSEP experiments to address the issues highlighted in Section Introduction.

ALTERNATIVE SPATIAL FORECAST DESCRIPTIONS

Multi-resolution and data-driven grids can be acquired through various techniques. However, the desired approach should satisfy a series of requirements:

Fewer cells The grids should have considerably fewer cells than the corresponding classical CSEP grids for the same region.

Simple The definition of grid cells should be simple and easy to understand.

Ease of use The implementation of software codes using the new grid should be straightforward and rely on standard libraries.

Index Each cell in the grid should be unambiguously identifiable and can be indexed.

Global coverage The grid should completely cover the global testing region and should have the ability to fully represent the regional testing areas.

Multi-resolution grid The grid should offer the flexibility to be assembled with higher resolution in regions with more data and lower resolution in regions with less data.

CSEP test compatibility The grid must be compatible with CSEP tests already designed for evaluating forecasts described by conventional grids to offer seamless integration into CSEP experiments.

Multi-resolution comparability Modelers will make different choices, resulting in grids with different resolutions at the same locations. Thus, the corresponding testing procedure needs to allow for comparative testing of two models even if the grids do not match.

We explore different grid types with various advantages and disadvantages and, based on the aforementioned requirements, assess their suitability as a replacement for the classical CSEP grids. The fundamental requirements for a forecast grid to fulfill are *fewer cells* and *multi-resolution grid*. Furthermore, a grid must offer *CSEP testing capability* in order to be used for CSEP experiments in the first place. Once a grid comes with the *multi-resolution grid* property, it may lead to different model-dependent grids. Thus, the grid must fulfill the requirement of *multi-resolution comparability* for impartial and independent CSEP experiments. Of the competing grid types that satisfy these basic requirements, we shall favor the approach that satisfies most of the remaining properties.

Simple grid with larger cells A Cartesian grid like the classical CSEP grid but with uniformly reduced resolution. In this case, latitude and longitude values are spaced uniformly, and codes can effectively handle cells by their center points. Therefore, such a grid is simple, easy to implement, and contains fewer cells. A globally lowered resolution will reduce the number of cells to a more meaningful number, given the amount of data available. However, testing in high-seismicity regions like Japan will suffer from the loss in precision. The classical CSEP regional grids were already the result of balancing the resolution needs for high-seismicity and low-seismicity regions. Thus, a considerable resolution reduction will not be useful for regional experiments anymore. Consequently, this grid definition has the potential to match all requirements except for the *multi-resolution grid* and *multi-resolution comparability* requirements.

Set of polygons A pre-defined set of polygons covering either the globe or a regional testing area such that small polygons are used in areas where high precision is warranted and large polygons in areas of sparse data. Such a set of polygons is computationally efficient similar to the *Simple grid with larger cells* but much more difficult to use due to possibly irregular polygons, thus not matching the *ease of use* requirement. Furthermore, due to its lacking flexibility in model-dependent resolutions, it does not match the *multi-resolution grid* and *multi-resolution comparability* requirements. Therefore, this type of grid is not suitable as a replacement.

Hexagonal grid Representing a spherical shape using a hexagonal grid is done using the Goldberg polyhedron (Goldberg, 1937) that, like a soccer ball, consists of 12 pentagons positioned at the center points of the dodecahedron face centers and a large number of hexagons, depending on the resolution of the grid. Such a grid can represent the global testing region by hexagons and pentagons, thereby matching the property of *global coverage*. The global coverage can be achieved using either smaller cells at high-resolution or bigger cells at a lower resolution like the *simple grid with larger cells* but does not provide the *multi-resolution grid*. Furthermore, the handling of hexagonal cells (including 12 pentagonal cells) is computationally expensive and does not match the requirements of *simple* and *easy of use*.

Triangular grid The testing region can also be represented using a triangular grid, where each cell is an equilateral triangle. This type of grid is an icosahedron, the

Platonic solid with 20 equilateral triangular faces (Sadourny et al., 1968). The triangular grid has the ability to represent the global testing region and regional testing areas with triangles of equal areas, thereby fulfilling the property *global coverage*. It also offers the possibility to increase or decrease the resolution by replacing a triangle with its four embedded equilateral triangles. This matches the property of *fewer cells*, *multi-resolution*, and *multi-resolution comparability*. However, implementing a triangular grid for the global testing region is a challenging task along with a difficult indexing procedure. Therefore, this approach does not match the *simple*, *ease of use*, and *index* requirements.

Voronoi cells A grid based on Voronoi cells is the result of partitioning a plane into polygons or cells based on generating points on a plane in such a way that each cell contains exactly one generating point, and every point in each cell is closer to its generating point than to any other generating point (Aurenhammer, 1991). For Voronoi cells, a multi-resolution grid can be acquired by the density of generating data points, such as earthquake locations (Gordon et al., 2015). Thus, it can match the requirements of *global coverage*, *fewer cells*, and *multi-resolution grid*. However, the resolution depends on the chosen generating points, resulting in one point per cell, and thus the cells have irregular shapes depending on the input data. This approach does not offer the flexibility to change the resolution locally. Therefore, such a grid does not match the requirements *ease of use*, *simple*, and *multi-resolution comparability*.

Coarse simple grid Another possible choice is the use of grid coarsening to alter the grid resolution. Grid coarsening means the combining of adjacent smaller cells, such as $0.1^\circ \times 0.1^\circ$ cells, into a single bigger cell (Chen et al., 2012). The highest resolution grid before coarsening is called the fine grid, while the grid after coarsening is called the coarse grid. The data-driven coarsening can be applied to acquire a multi-resolution grid with fewer but larger cells. In this approach, one first generates a fine grid with the highest resolution and then reduces the resolution wherever required. This choice of a grid can potentially be a solution because it is easy to understand, similar to the conventional Cartesian grid in terms of rectangular-shaped cells. It can also provide fewer cells, with the possibility of adjusting the resolution locally, thereby matching all but the *multi-resolution comparability* requirement. To achieve comparability across different grid resolutions, the grid creation process should be governed by a well-defined mechanism. There can be many alternative possibilities to combine adjacent smaller cells to lower the resolution. Thus, changes in grid resolution must be controlled by a transformation technique, i. e. a governing mechanism to determine which specific smaller cells should be combined into a single cell to get a lower-resolution (bigger) cell and vice versa.

Quadtree grid The Quadtree is a tree-based hierarchy for dividing a region into four quadrants and then dividing all or some quadrants recursively into four quadrants again until a final grid is achieved, referred to as Quadtree grid (Samet, 1984). This approach can be used to create grids that are *simple* and offer *ease of use* and are *CSEP test compatible*. A data-driven recursive division of quadrants can be used to create a *multi-resolution grid*. It offers a simple and efficient procedure for organizing spatial cells in a specific order due to the tree-based hierarchical structure, thus providing a unique *index* to every cell. The Quadtree provides a well-defined strategy for handling the grid resolution by ensuring that an increase in the grid resolution can only be achieved by recursively dividing a cell into four (pre-defined) smaller cells. Similarly, it does not allow for random combinations of any cells to decrease the grid's resolution. Instead, it defines the four specific cells that can be combined into a bigger cell. It also matches the *multi-resolution comparability* requirement by making the comparison of different grids convenient at any resolution. The pre-defined cells' boundaries prevent possible mishandling or unintended changes of the cells' boundaries. However, it comes with disadvantages as well, i. e. it is unable to represent the global testing region beyond 85.05° north and south, thereby not matching the *global coverage* requirement. It matches the global coverage requirement for all practical purposes because there is no earthquake hazard in the regions very close to the poles. We can find Python-based open-source libraries which provide implementations of the *Quadtree grid* including the management of the tree-based hierarchy, indexing, and translating these indexes to longitude/latitude coordinates of cells. Given the fact that the CSEP software is also developed in Python, this can help to integrate the *Quadtree grid* with the CSEP software.

Table S1 summarizes the advantages and disadvantages of aforementioned grid types. The *Triangular grid* and the *Quadtree grid* fulfill important requirements as compared to the other grid types. The *Triangular grid* offers an advantage over the *Quadtree grid* in terms of *global coverage*, while the *Quadtree grid* is matching the *simple*, *index* and *easy to use* requirements. The lack of coverage on poles by the *Quadtree grid* does not have any great impact on the CSEP global experiments, as only 0.1% of global seismicity occurs on the poles beyond the coverage of the *Quadtree grid*. Consequently, we select the *Quadtree grid* as the best alternative choice for representing grid based seismicity models.

Quadtree spatial grid

The Quadtree is a hierarchical tree structure in which each node is allowed to have either zero or four child nodes, hence the name. The starting node of the tree structure is referred to as the root node. Thus, we refer to the globe as the root node and the four quadrants as four child nodes called *tiles*. The Quadtree implementation for a global map requires a suitable projection, of which many are available, such as sinusoidal (Snyder, 1987), Equirectangular (Snyder, 1997), Mercator (Snyder, 1987), Robinson (Robinson, 1974). Of these, the Mercator projection has two main properties that make it a suitable choice for its use with the Quadtree. It considers the Earth as a flat surface with north and south as straight up and down, respectively (cylindrical property). Furthermore, it preserves all curves that cross each other on Earth, therefore not changing the shapes of small objects (conformal property). A slight variant of the Mercator projection, known as Pseudo Mercator or Web Mercator projection, was adapted by Google Maps in 2005 for a square representation of the global map (Battersby et al., 2014). It has become a standard that is followed by most online web-map service providers and applications for efficient display of maps, e. g. Mapbox, OpenStreetMap, and Microsoft's Bing Maps. The Quadtree works in combination with Web Mercator projection to generate square tiles at different zoom levels to store and render global maps. In the Web Mercator projection, the polar regions beyond 85.05° north and south are excluded due to their large area inflation.

The implementation of the Quadtree in combination with the Web Mercator projection of the globe is shown in Figure 2. The root Quadtree tile is a square representing the whole globe, excluding the polar regions from 85.05° latitude north and south. The root tile has no Quadkey assigned to it. In the first step, the root tile is divided into four square subtiles, the NE, NW, SW, and SE regions. These tiles are indexed using numbers of the base-four system `0`, `1`, `2`, and `3`, respectively. These numbers are called the *Quadkey* of each tile. The dividing lines are the prime meridian and the Equator. Each of these four tiles can be further divided into four square subtiles. The Quadkeys of these subtiles are generated from the Quadkey of the parent tile by adding the relative Quadkey (`0`, `1`, `2`, or `3`), e. g. the subtiles of tile `2` are `20`, `21`, `22`, and `23`. The number of times a tile is divided is called zoom level (L). Thus, the number of digits in the Quadkey represents the zoom level of the decomposition of the root tile. This way, the entire globe can recursively be divided into as many tiles as desired (Samet, 1984). This indexing process can go on for any number of zoom levels, providing a unique Quadkey for every potential subtile. After the desired decomposition is achieved, we refer to it as a Quadtree-based grid, and each tile is referred to as a spatial grid cell to remain consistent with the naming convention of the CSEP experiments.

The geographical distance between two meridians decreases towards the poles. Therefore, the area of the $0.1^\circ \times 0.1^\circ$ cells from a single-resolution conventional CSEP grid decreases along latitude towards the poles. Although the size of the square-shaped Quadtree cells in the figure appears to be the same everywhere, the area decrease is pronounced due to the Mercator projection's cylindrical property because each cell's latitudinal dimension is also reduced towards the poles. This phenomenon is visualized in Figure S2, showing a single-resolution Quadtree grid at zoom level 3 ($L = 3$), which results in a total number of $N = 4^L = 64$ cells, each cell indicating its area in units of $10^6\,\mathrm{km}^2$.

Quadtrees are already used in numerous fields of science and technology, such as image processing (Liu et al., 2017), computer vision (Chung et al., 2015), fluid dynamics (Panfilov et al., 2021), aerospace (Xue and Wei, 2021), and indexing of spatial databases (Hussain and Hassan, 2020). Here, for the first time, we propose its use in earthquake forecasting by representing the CSEP testing regions in the form of Quadtree grids. Quadtrees can be used to generate single-resolution grids like the conventional $0.1^\circ \times 0.1^\circ$ grids as well as multi-resolution data-driven grids. Single-resolution grids can be acquired by fixing the zoom level for each tile when generating the grid. Alternatively, the earthquake density can be applied to determine the grid-resolution locally introduced in Section Seismic density-based spatial grid.

Seismic density-based spatial grid

Here, we show how to generate a multi-resolution Quadtree grid constrained by the observed seismicity. To generate such a grid, firstly, we define a threshold for the maximum number of earthquakes allowed per cell, $N_{\max}$. If the earthquake count in a cell exceeds $N_{\max}$, then that cell is further divided into four sub-cells by locally increasing the zoom level by one step. The resulting four sub-cells receive their share of earthquakes depending on the locations of the earthquakes within the cell. This cell division repeats until no cell contains more than $N_{\max}$ earthquakes.

The above-described single criterion $N_{\max}$ might lead to a very high grid resolution in highly active and well-monitored seismic areas, leading to cell sizes becoming even smaller than the location error of earthquakes. To avoid such cases, one can introduce an additional criterion such as a minimum cell area or maximum zoom level ($L_{\max}$) allowed for a cell. In this study, we analyze different criteria to generate alternative single- and multi-resolution grids. For differentiation, we name all the grids based on the input criteria.

Here, we demonstrate the use of a training earthquake catalog to create multi-resolution grids for the global and Japanese testing regions (Figure 3). Figure 3a and Figure 3b show grids generated using the global CMT catalog with 28465 $M \geq 5.15$ events recorded between 1976 and 2013. The two subdivision criteria, $N_{\max} = 100$ and $L_{\max} = 11$, have been selected for the grid shown in Figure 3a, named *N100L11*. This choice means that the cell division stops if zoom-level 11 is reached, even if some cells contain more than $N_{\max}$ earthquakes. When a cell is divided into four sub-cells, some child cells may be empty if there is no earthquake located in those cells. However, the overall proportion of cells without earthquakes is much less as compared to the conventional $0.1^\circ \times 0.1^\circ$ grid. The global grid *N100L11* contains 922 cells, all of them with zoom levels smaller than 11. Thus, the $L_{\max} = 11$ criterion was not applied, as all grid cells contain fewer than 100 earthquakes. The grid shown in Figure 3b is generated by choosing $N_{\max} = 10$ and $L_{\max} = 11$ named *N10L11*. This grid contains 8089 spatial cells, with 72 cells containing more than ten earthquakes. This implies that the cell division stopped in those 72 cells due to reaching the maximum allowed zoom level of $L_{\max} = 11$. Furthermore, out of the 28465 earthquakes in the selected global earthquake catalog, 28 events are located outside of the Quadtree limits, i. e. their latitude was beyond $\pm 85.05^\circ$. Therefore, these 28 events ($< 0.1\%$) cannot be considered for generating and testing earthquake forecast models while working with Quadtree-based grids. Figure 3c and Figure 3d show the multi-resolution Quadtree grids for the Japanese region using a regional Japan Meteorological Agency (JMA) catalog with 167073 $M \geq 1.0$ earthquakes recorded between 2000 and 2007. The dense catalog offers the capability to generate high-resolution spatial grids, where the cell area can be as small as $1\,\mathrm{km}^2$. The spatial grid at zoom level 14 leads to cells with a surface area from 1 to $2\,\mathrm{km}^2$ for the Japanese testing region. Using the regional catalog with $N_{\max} = 10$ and $L_{\max} = 14$ leads to a multi-resolution grid with small cells difficult to visualize in the figure. Therefore, Figure 3c shows a spatial grid with $N_{\max} = 1000$ and $L_{\max} = 14$ (*N1000L14*) and Figure 3d shows a grid with $N_{\max} = 400$ and $L_{\max} = 14$ (*N400L14*).

Sample forecast model for a given Quadtree-based grid

In this section, we demonstrate the use of the Quadtree multi-resolution grid to generate a simple earthquake forecast model. The model is based on the simple assumption that the past seismicity will be the predictor of future seismicity. Specifically, the model assumes that the earthquake rate observed in every cell during the training period remains constant, thereby preserving the spatial distribution of seismicity. This sample forecast is generated from the input catalog, in which the grid resoluton is adaptive and the past earthquake rate is assumed to be well constrained in most cells by the number of observed events and can be simply re-scaled to the forecast period. However, cells without recorded seismicity in the learning period might still have earthquakes in the future, but the rate is too low to be estimated. Setting the predicted rate to zero in such cells leads to an immediate rejection of the whole model if only one earthquake occurs in any of these cells during the testing period (Schorlemmer and Gerstenberger, 2007). To address this well-known issue, an established strategy is to assign a baseline rate to these cells referred to as water-level value (Bird and Kreemer, 2015; Kagan and Jackson, 2011; Bird et al., 2010). A constant earthquake density, R_0, is assumed in such cells leading to earthquake forecast rates proportional to the area of those cells, A. In particular, the algorithm to create the model forecast is as follows:

1. Choose a value of the water-level density R_0.

2. Calculate the non-normalized forecast rates for cells indexed by i

$$\tilde{R}_i = \begin{cases} N_i & \text{if } N_i > 0 \\ A_i \times R_0 & \text{otherwise} \end{cases} \tag{1}$$

   where N_i is the observed number of earthquakes in the i^{th} cell during the learning period.

3. Scale the forecast to N

$$\lambda_i = \tilde{R}_i \times \frac{\sum N_i}{\sum \tilde{R}_i} \times \frac{T_{\text{test}}}{T_{\text{learning}}}, \tag{2}$$

where T_{learning} corresponds to the duration of the data used to generate forecasts, T_{test} is the intended duration of the forecast period and λ_i is the forecast rate in each cell i of the grid. The forecast rates in each cell can further be extended to forecasts across magnitude bins using the Gutenberg-Richter relation. However, we focus here only on spatial forecasts. Figure 4 shows the sample forecast acquired using the Quadtree grid shown in Figure 3(b), *N10L11*. The same catalog used to generate the grid here is also used to determine the forecast rates. The figure shows the forecast rate in each cell computed for one year, shown on a logarithmic scale for better visibility.

Model Testing and comparison

In previous CSEP experiments, especially in the global forecast experiment, only a few hundred earthquakes were available in the test catalogs to evaluate the performance of models having forecast rates for millions of spatial cells. We need to explore how many earthquakes are required in the test catalog to carry out meaningful testing of earthquake forecast models and how this quantity may change if we change the resolution of the grid.

Another potential topic is to find an optimal model-specific grid resolution for generating forecasts based on the amount of available data (e. g. number of earthquakes or strain-data points). Using Quadtree grids as an integral part of CSEP enable modelers and testers to explore such scientific questions. As an example, we conduct an experiment in the next section using different multi-resolution grids to find the optimal resolution for the forecast generation in the case of the sample forecast model introduced above. Having too few events in grid cells leads to forecasts with large relative errors degrading the forecasts. Similarly, inappropriate forecasts also result from averaging over too large cells that cannot resolve strong spatial variations. This problem of finding an appropriate grid resolution for generating an earthquake forecast model is similar to the problem of overfitting vs. underfitting or bias error vs. variance error, like in all statistical and machine-learning-based modeling (Belkin et al., 2019). Decreasing the resolution of a grid can oversimplify the model by capturing less spatial information, which causes underfitting and introduces a bias error. In contrast, increasing the resolution can increase the complexity of the model by capturing random fluctuations, which causes overfitting and increases the variance error. Thus finding such a balance in the grid resolution can be explored by using the capability to adjust the resolution locally based on data.

In CSEP experiments using Quadtree grids, we expect different resolutions of the grids for every model depending on the data used to generate that model. One way to evaluate the competing forecast models defined for different grids is to choose a single testing grid to compare all models fairly. Comparing forecast on a single test grid requires aggregation or de-aggregation from one grid resolution to another. Figure S3 explains the process of forecast aggregation and de-aggregation from a cell of one grid (model grid) to another (test grid). Forecast aggregation can be done by summing the forecast rates of all the child cells to generate the rate of a parent cell. Similarly, the forecast rate of a parent cell is de-aggregated into the child cells by assuming uniform earthquake distribution within the parent cell and distributing rates to the child cells based on the area of each cell, i. e. the forecast is assigned to each child cell proportionately according to their area by dividing the rate among child cells based on the area of each child cell. Certainly, the assumption of uniform density for each cell is a simplification because the spatial earthquake activity is known to be inhomogeneous on any scale. Because the forecast model has no further resolution-increasing capability, a uniform distribution is the simplest assumption reflecting the level of knowledge.

Any forecast which is correct and thus passes the Poisson consistency tests on a high-resolution grid will also pass the same tests after aggregation to a lower resolution. The reason is that the sum of Poisson processes, e. g. related to the four sub-cells, is again a Poisson process with a rate equal to the sum of the rates of the individual processes. Thus, consistency tests can be performed on a lower resolution Quadtree grid without introducing any bias. In contrast, comparative tests should consider that aggregating forecasts on a lower-resolution grid leads to a loss of the models' spatial information. Thus, any aggregation should be avoided, and the competitive forecast tests should be performed at the highest-resolution grid. Otherwise, some models may lose their advantage of using high-resolution input datasets to provide a high-resolution forecast. One possible way for comparative analysis of forecasts submitted on different grids is to de-aggregate all forecasts to the locally highest resolution in all the grids and compare the models using the T-test provided by CSEP (Rhoades et al., 2011; Savran et al., 2022a).

In CSEP tests, the forecast evaluations use Poisson joint log-likelihood (POLL) value shown in Equation 3

$$POLL = \sum_{i=1}^{N_{\text{bin}}} (-\lambda_i + \omega_i \ln(\lambda_i) - \ln(\omega_i!)) \tag{3}$$

where N_{bin} refers to the total number of spatio-magnitude bins, ω_i is the number of observed earthquakes and λ_i is the expected number of earthquakes in the spatio-magnitude bins $i = 1, \ldots N_{\text{bin}}$ (Schorlemmer et al., 2007; Zechar et al., 2010a; Werner et al., 2011;

Bayona et al., 2022). Alternatively, we can avoid the definition of any particular test grid by using the log-likelihood defined for point-processes, L (Daley and Vere-Jones, 2003; Rhoades et al., 2011). This value equals POLL in the limit case of infinitesimal small cell sizes. Thus, the result for the point-process log-likelihood function is practically the same as POLL for high-resolution grids but avoids the computational costs of de-aggregation. The point-process log-likelihood is widely used in evaluating forecasts of Epidemic Type Aftershock Sequence (ETAS) models (Zhuang et al., 2011; Bray and Schoenberg, 2013). L is defined for N_{eq} observed events that occurred at epicenter locations x_i with $i = 1, \ldots, N_{\text{eq}}$ according to

$$L = \sum_{i=1}^{N_{\text{eq}}} \ln\left(R(x_i)\right) - \int_A R(x)dx \tag{4}$$

where R is the forecasted earthquake rate density. The larger L, the better the model's ability to explain the data. Assuming a uniform earthquake distribution within each cell , the rate density R at an earthquake epicenter is simply defined by $R(x_i) = \lambda_i / A_i$, where A_i and λ_i refers to the area and forecasted event rate of the cell in which the earthquake happened. Thus, for the model's forecast, the log-likelihood is calculated by

$$L = \sum_{i=1}^{N_{\text{eq}}} \ln\left(\frac{\lambda_i}{A_i}\right) - \sum_{i=1}^{N_{\text{bin}}} \lambda_i \tag{5}$$

It directly uses the point information (epicenters) of the earthquakes and the forecast density for that location without (de-)aggregation of the forecasts onto another grid. Using this L-value, the comparative tests can also be performed.

CASE STUDY

The Quadtree enables us to constrain forecast models to the available information.

For demonstration, we firstly analyze the data available for different data-based multi-resolution grids, introduced in Section Seismic density-based spatial grid. Secondly, we apply the sample forecast discussed in Section Sample forecast model for a given Quadtree-based grid and analyze its forecast ability if it is constrained to the alternative grids.

We acquire multiple spatial grids with different resolutions by using different criteria. We determine the percentage of cells without any earthquake and the total number of cells for each grid. Table 1 lists different grids and provides the total number of spatial cells and the percentage of cells without any earthquake. It shows that seismicity is contained in only a small fraction of the cells in single-resolution grids. Decreasing the resolution of spatial grids reduces the number of cells, but it also fails to capture the spatial information about the seismicity distribution. On the contrary, the multi-resolution grids capture the distribution of seismicity better by increasing the resolution only in seismically dense regions. It leads to a higher percentage of cells with earthquakes in the learning period, thereby enabling forecast modelers to provide high-resolution forecasts for the areas that are more important in terms of seismic hazard. Figure S4 shows the histogram of number of earthquakes per cell in different Quadtree grids, indicating relatively even distribution of earthquakes across grid cells, as compare to the single-resolution grids as observed in Figure 1.

In most cells of single-resolution grids, no earthquake is observed during the learning period. In contrast, the multi-resolution grids contain significantly fewer cells without earthquakes than the single-resolution grids, resulting in better-constrained forecasts. For illustration, we create a forecast on the single-resolution grid *L11* with 4.2 million cells and the multi-resolution grid *N10L11* with 8089 cells. These forecasts are based on events of $M \geq 5.15$ using a learning period of 37 years from 1976 to 2013. Both grids have a significantly different percentage of cells with no recorded earthquakes. We assign a water level to those empty cells (see Section Sample forecast model for a given Quadtree-based grid) and compare the resulting forecasts for different choices of the water-level. For each case, we calculate the log-likelihood value based on point-processes L (Equation 5) for the target events in the test period between 2014 and 2019, containing $N = 4869$ events with $M \geq 5.15$. The larger the values of the forecasted event density (λ/A) at the epicenters of the target events, the larger the L value.

This analysis is repeated for both spatial grids, and the results are plotted in Figure 5. The joint log-likelihood values are negative numbers, with maximum values indicating the best agreement between forecast and observation. The figure shows that the forecast based on the multi-resolution grid with fewer cells and a smaller fraction of cells without recorded earthquakes provides higher log-likelihood values than the forecast generated on the single-resolution grid with a large fraction of such cells. Distributing water-level values to the spatial cells without any earthquake also contributes to the total number of earthquakes yielded by the forecast model during the testing period, referred to as background seismicity. If the water-level value increases, then the total count of background seismicity caused by the water level also increases. Furthermore, the quality (L-value) of the forecast based on the multi-resolution grid does not strongly depend on the value of the water level, indicated by a broad maximum. These results indicate that the multi-resolution grid forecast is superior because it is significantly better constrained and less influenced by the unconstrained value of the water level.

Now we analyze the quality of the forecast of the sample model dependent on the grid type, which is used for model generation. In particular, we explore the optimal grid resolution to constrain the forecast given the available training catalog. We analyze the seven multi-resolution grids mentioned in Table 1. We generate the forecasts for each multi-resolution grid by fixing the water level to $10^{-7}/\text{km}^2/\text{year}$, which is close to the maximum found in Figure 5. Then we conduct a pair-wise test to compare the performance of the forecast against the model with the highest resolution, i. e. the grid *N1L11*. We use the T-test to evaluate the relative performance of the different grid-based forecasts. It measures the IGPE of one forecast over another, including its confidence interval. One forecast is considered more informative than the other if the confidence interval is above 0 (Rhoades et al., 2011; Bayona et al., 2021). We conduct the T-test relative to the forecast generated on the grid *N1L11* in two ways. One is conducted after explicitly de-aggregating the forecasts on the test grid *N1L11* and using the implementation of Rhoades et al. (2011) provided in pyCSEP (Savran et al., 2022b,a). In the other implementation of the T-test, we compute the IGPE by the L values defined in Equation 6 without any definition of a test grid.

$$IGPE = \frac{L_a - L_b}{N_{\text{total}}} \tag{6}$$

Figure 6 shows the performance of the sample forecast models generated on different grids, which are evaluated against the forecast generated on the *N1L11*-grid. Both methods yield the same results, demonstrating that the pairwise comparative tests can be calculated without even defining a common testing grid, using the point-process log-likelihood, L. The same forecast model created with different definitions of spatial grids results in different IGPE-values, suggesting that the choice of grid resolution for creating the forecasts affects the performance of the model. The grid *N10L11* stands out with the highest IGPE-score among all the seven multi-resolution grids explored in this experiment. A similar result is obtained if the water level is varied from 10^{-6} to 10^{-8} per km^2. This result indicates that the *N10L11*-grid optimally uses the available training data and leads to the best-constrained forecast. In contrast, grids with higher resolution lead to less-constrained forecasts involving larger uncertainties, thus worse forecasts. On the other hand, grids with lower resolution cannot resolve the spatial variability of the real earthquake distribution. The optimal grid may vary with the model and the information content.

Quadtree grids with fewer cells can yield better-constrained forecasts and additionally reduce the burden on computational resources. We carry out a performance evaluation to compare the run-time of CSEP tests in the case of the conventional grid of $0.1^\circ \times 0.1^\circ$

and the multi-resolution grid *N10L11*. We used the sample forecast model to run the tests and measure the time taken for the CL-test for both grids. We used the standard implementation of those CSEP tests provided in the pyCSEP (Savran et al., 2022a,b). The test was carried out on a computer with i7 (9^{th} generation) processor. The CL-test for the conventional grid with 6.48 million cells took 20.6 minutes, while it took only 2.6 seconds for the multi-resolution grid *N10L11* with 8089 cells. The testing time for the forecast with grid *N10L11* is almost 475 times faster as compared to the forecast with the conventional grid. Apparently, 20.6 minutes for the CL-test on the regular grid may not seem too much for a time-independent long-term forecast in the case of yearly performance evaluation of one model. However, this would matter a lot in the scenario of daily or weekly forecasts yielded by numerous time-dependent forecast models. Evaluating time-dependent or short-term forecasts on conventional grids would lead to an enormous computation time. Furthermore, storing single high-resolution forecasts also costs gigabytes of space. In such scenarios, the short-term or time-dependent model forecasts would also overrun the storage space. Thus, using multi-resolution grids for forecast generation and evaluation will save computational resources.

We have recently initiated a prospective global forecast experiment using Quadtree grids. It is basically the continuation of the same experiment discussed in Bayona et al. (2021), but the forecast models have been aggregated on various Quadtree grids. The details of the experiment, relevant data, interim results and the codes to run the experiment are publicly available on Zenodo and Gitlab (see Section Data and resources).

Quadtree provides a series of benefits such as a compact representation of forecasts by focusing on the regions most important for seismic hazard, a better ability to capture the spatial information, and above all, it provides an opportunity to optimize forecasts. With the use of the Quadtree approach for earthquake forecasting studies, we foresee a number of studies to explore solutions to some other problems: (i) the effect of grid resolution on the statistical power of CSEP consistency tests, (ii) global models are to be generated directly using the Quadtree grids with an optimal resolution of the grid adapted to the availability of input data, (iii) physics-based models are often tested as binary forecasts, using contingency matrix-based evaluation measures, such as receiver operating curves (ROC) (DeVries et al., 2018; Sharma et al., 2020) which are affected by the unbalanced nature of the datasets, therefore, testing of physics-based models for aftershock prediction can benefit from Quadtree multi-resolution grids, (iv) the multi-resolution grids for earthquake forecast models will be used to explore the necessary information content in a forecast to investigate what level of forecasting detail is warranted by the input data. This approach will help understand the limits of predictability and provide insights into the limits of precision in forecasting. The Quadtree has the potential to become a norm in earthquake forecast studies in the future. It can be used for generating time-dependent earthquake forecast models. In such scenarios, the data-based resolution of models can dynamically update and optimize the forecasts by incorporating the incoming data and changing the grid resolution accordingly. With this happening, we can also envision the use of optimizing capabilities of computational intelligence to play an important role in helping to automatically optimize grid resolutions for incorporating new incoming data for time-dependent forecast models without the need of intervention from a modeler.

CONCLUSION

Earthquake forecasts modeling and testing is currently performed on a uniform high-resolution $0.1^\circ \times 0.1^\circ$ grid not well suited for the generally sparse and highly heterogeneous spatial distribution of seismicity. A global and regional seismicity analysis shows that the conventional grid leads to a huge number of spatial cells, requiring massive computational resources for forecast storing and testing. Furthermore, the grid includes a high fraction of cells without any recorded earthquake in the past, thereby leading to the need of generating and testing forecasts for millions of spatial cells with only a few hundred or thousand earthquakes recorded. Such a disparity can have implications to constrain forecast models and can potentially lead to low statistical power of the tests used for the evaluation of forecasts. Thus, we explore alternatives for CSEP experiments allowing for easy implementation of multi-resolution grids with fewer cells and comparability with the existing CSEP tests. Our screening of different possible solutions yields the Quadtree as the best choice fulfilling the requirements for designing data-driven multi-resolution spatial grids for CSEP experiments.

The Quadtree is a tree-based structure used in combination with the Mercator projection to acquire a spatial grid at different resolutions. We demonstrate the use of the Quadtree for improved model development and testing for a sample model assuming stationary seismicity. Compared to single-resolution grids, the model is better constrained on density-based multi-resolution grids adapted to the seismicity density in the training period. As a result, the forecasts are significantly improved, and the computational time is reduced. Thus, we provide the Quadtree as a technical enhancement for CSEP and propose the use of multi-resolution grids for modeling and testing earthquake forecasts. In the future, we intend to use Quadtree grids for CSEP experiments. In pyCSEP, we have provided sufficient help (including examples) for the modelers to generate a data-based Quadtree grid and use it for modeling forecasts in place of using conventional grids. Meanwhile, in the forthcoming studies, we intend to explore the statistical power of CSEP consistency tests associated with the choice of grid resolution and finding the optimal multi-resolution grid for improving the forecast models.

DATA AND RESOURCES

We acquired the earthquake catalog for Japan from the Japan Meteorological Agency (JMA, `http://www.jma.go.jp`, last accessed September 2021), the catalog for California from the Advanced National Seismic System (ANSS, `https://earthquake.usgs.gov/data/comcat/`, last accessed September 2021) and the global catalog from the Centroid Moment Tensor webpage (globalCMT, `https://www.globalcmt.org/`, last accessed December 2021). The Quadtree approach has been integrated as a part of extensive software package developed for CSEP tests known as pyCSEP. The codes, including documentation and examples, are available here: https://github.com/SCECcode/pycsep. The data and documentation for Quadtree global forecast experiment can be found here: https://doi.org/10.5281/zenodo.6305669. The results and code to run the experiment are available here: https://git.gfz-potsdam.de/csep-group/gefe-quadtree. The supplementary material provided with this article consists of four figures and a table to further elaborate the motivation and application of the Quadtree for earthquake forecasting research.

ACKNOWLEDGMENT

The authors are grateful to Warner Marzocchi, Chris Rollins, and two anonymous reviewers for reviewing and providing constructive feedback to improve the manuscript. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement Number 821115, Real-Time Earthquake Risk Reduction for a Resilient Europe (RISE). This research was also supported by the Southern California Earthquake Center (Contribution No. 11797). SCEC is funded by NSF Cooperative Agreement EAR-1600087 & USGS Cooperative Agreement G17AC00047. In particular, we want to thank the open-source community for the Linux operating system and the many programs used in this study, e. g. Python (Van Rossum and Drake Jr, 1995), Spyder (Raybaut, 2009) and PyCharm (https://www.jetbrains.com/pycharm), OpenStreetMap (https://www.openstreetmap.org) and QGIS (https://qgis.org/).

REFERENCES

Akinci, A., Moschetti, M. P., and Taroni, M. (2018). Ensemble smoothed seismicity models for the new Italian probabilistic seismic hazard map. *Seismological Research Letters*, 89(4):1277–1287.

Aurenhammer, F. (1991). Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405.

Bakun, W., Gómez Capera, A., and Stucchi, M. (2011). Epistemic uncertainty in the location and magnitude of earthquakes in Italy from macroseismic data. *Bulletin of the Seismological Society of America*, 101(6):2712–2725.

Battersby, S. E., Finn, M. P., Usery, E. L., and Yamamoto, K. H. (2014). Implications of web Mercator and its use in online mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 49(2):85–101.

Bayona, J., Savran, W., Strader, A., Hainzl, S., Cotton, F., and 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.

Bayona, J. A., Savran, W. H., Rhoades, D. A., and Werner, M. (2022). Prospective evaluation of multiplicative hybrid earthquake forecasting models in california. *Geophysical Journal International*, 229(3):1736–1753.

Belkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854.

Bezeau, S. and Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. *Journal of clinical and experimental neuropsychology*, 23(3):399–406.

Bird, P. and Kreemer, C. (2015). Revised tectonic forecast of global shallow seismicity based on version 2.1 of the Global Strain Rate Map. *Bulletin of the Seismological*

Society of America, 105(1):152–166.

Bird, P., Kreemer, C., and Holt, W. E. (2010). A long-term forecast of shallow seismicity based on the Global Strain Rate Map. *Seismological Research Letters*, 81(2):184–194.

Bray, A. and Schoenberg, F. P. (2013). Assessment of point process models for earthquake forecasting. *Statistical science*, 28(4):510–520.

Button, K. S., Ioannidis, J., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, 14(5):365–376.

Chen, A. S., Evans, B., Djordjević, S., and Savić, D. A. (2012). Multi-layered coarse grid modelling in 2D urban flood simulations. *Journal of Hydrology*, 470:1–11.

Chung, K.-L., Huang, Y.-H., Lin, C.-H., and Fang, J.-P. (2015). Novel bitrate saving and fast coding for depth videos in 3D-HEVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(10):1859–1869.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer.

DeVries, P. M., Viégas, F., Wattenberg, M., and Meade, B. J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature*, 560(7720):632–634.

Ekström, G., Nettles, M., and Dziewoński, A. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200:1–9.

Field, E. H., Milner, K. R., Page, M. T., Savran, W. H., and van der Elst, N. (2021). Improvements to the third Uniform California Earthquake Rupture Forecast ETAS model (UCERF3-ETAS). *The Seismic Record*, 1(2):117–125.

Gerstenberger, M. C. and Rhoades, D. A. (2010). New Zealand earthquake forecast testing centre. In *Seismogenesis and Earthquake Forecasting: The Frank Evison Volume II*, pages 23–38. Springer.

Goldberg, M. (1937). A class of multi-symmetric polyhedra. *Tohoku Mathematical Journal, First Series*, 43:104–108.

Gordon, J. S., Clements, R. A., Schoenberg, F. P., and Schorlemmer, D. (2015). Voronoi residuals and other residual analyses applied to CSEP earthquake forecasts. *Spatial Statistics*, 14:133–150.

Guy, M. R., Patton, J. M., Fee, J., Hearne, M., Martinez, E. M., Ketchum, D. C., Worden, C. B., Quitoriano, V., Hunter, E. J., Smoczyk, G. M., et al. (2015). *National Earthquake Information Center systems overview and integration*. US Department of the Interior, US Geological Survey Reston, VA.

Helmstetter, A. and Werner, M. J. (2014). Adaptive smoothing of seismicity in time, space, and magnitude for time-dependent earthquake forecasts for California. *Bulletin of the Seismological Society of America*, 104(2):809–822.

Hussain, A. A. and Hassan, R. F. (2020). Enhancing Quad tree for spatial index using space filling curves. *Engineering and Technology Journal*, 38(1B):15–25.

Kagan, Y. Y. (2007). Simplified algorithms for calculating double-couple rotation. *Geophysical Journal International*, 171(1):411–418.

Kagan, Y. Y. and Jackson, D. D. (2011). Global earthquake forecasts. *Geophysical Journal International*, 184(2):759–776.

Liu, H., Huang, K.-K., Ren, C.-X., Yu, Y.-F., and Lai, Z.-R. (2017). Quadtree coding with adaptive scanning order for space-borne image compression. *Signal Processing: Image Communication*, 55:1–9.

Marzocchi, W., Schorlemmer, D., and Wiemer, S. (2010). Preface to the special volume "An earthquake forecast experiment in Italy". *Ann. Geophys*, 53(3).

Panfilov, M., Popinet, S., Vostrikov, V., Baishemirov, Z., and Berdyshev, A. (2021). Numerical modeling of fluid flow through multiscale fractured-porous media by quadtrees. *Journal of Computational Physics*, 444:110566.

Raybaut, P. (2009). Spyder-documentation. *Available online at: pythonhosted. org*.

Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., and Imoto, M. (2011). Efficient testing of earthquake forecasting models. *Acta Geophysica*, 59(4):728–747.

Robinson, A. H. (1974). A new map projection: Its development and characteristics. *International yearbook of cartography*, 14(1974):145–155.

Sadourny, R., Arakawa, A., and Mintz, Y. (1968). Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere. *Monthly Weather Review*, 96(6):351–356.

Samet, H. (1984). The quadtree and related hierarchical data structures. *ACM Computing Surveys (CSUR)*, 16(2):187–260.

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

Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., and Michael, A. (2020). Pseudoprospective evaluation of UCERF3-ETAS forecasts during the 2019 Ridgecrest sequence. *Bulletin of the Seismological Society of America*, 110(4):1799–1817.

Savran, W. H., Werner, M. J., Schorlemmer, D., and Maechling, P. J. (2022b). pyCSEP: A Python Toolkit For Earthquake Forecast Developers. *Journal of Open Source Software*, 7(69):3658.

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

Schorlemmer, D. and Gerstenberger, M. (2007). RELM testing center. *Seismological Research Letters*, 78(1):30–36.

Schorlemmer, D., Gerstenberger, M., Wiemer, S., Jackson, D., and Rhoades, D. (2007). Earthquake likelihood model testing. *Seismological Research Letters*, 78(1):17–29.

Sharma, S., Hainzl, S., Zöeller, G., and Holschneider, M. (2020). Is Coulomb stress the best choice for aftershock forecasting? *Journal of Geophysical Research: Solid Earth*, 125(9):e2020JB019553.

Snyder, J. P. (1987). *Map projections–A working manual*, volume 1395. US Government Printing Office.

Snyder, J. P. (1997). *Flattening the earth: two thousand years of map projections*. University of Chicago Press.

Strader, A., Werner, M., Bayona, J., Maechling, P., Silva, F., Liukis, M., and 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.

Taroni, M., Zechar, J., and Marzocchi, W. (2014). Assessing annual global M 6+ seismicity forecasts. *Geophysical Journal International*, 196(1):422–431.

Tsuruoka, H., Hirata, N., Schorlemmer, D., Euchner, F., Nanjo, K. Z., and 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.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Werner, M. J., Helmstetter, A., Jackson, D. D., and Kagan, Y. Y. (2011). High-resolution long-term and short-term earthquake forecasts for California. *Bulletin of the Seismological Society of America*, 101(4):1630–1648.

Xue, M. and Wei, M. (2021). Small Unmanned Aerial Vehicle Flight Planning in Urban Environments. *Journal of Aerospace Information Systems*, pages 1–9.

Zechar, J. D., Gerstenberger, M. C., and Rhoades, D. A. (2010a). Likelihood-based tests for evaluating space–rate–magnitude earthquake forecasts. *Bulletin of the Seismological Society of America*, 100(3):1184–1195.

Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., and Jordan, T. H. (2010b). The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. *Concurrency and Computation: Practice and Experience*, 22(12):1836–1847.

Zhuang, J., Werner, M., Hainzl, S., Harte, D., and Zhou, S. (2011). Basic models of seismicity: spatiotemporal models. community online resource for statistical seismicity analysis, doi: 10.5078/corssa-07487583.

AUTHOR LIST

Khawaja M. Asim, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany. Email: khawaja@gfz-potsdam.de

Danijel Schorlemmer, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany.

Sebastian Hainzl, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany.

Pablo Iturrieta,GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany

William H. Savran, USC, Zumberge Hall of Science (ZHS), 3651 Trousdale Pkwy, Los Angeles, CA 90089-0740.

José A. Bayona, School of Earth Sciences, University of Bristol, Queens Road, BS81QU, Bristol, UK.

Maximilian J. Werner, School of Earth Sciences, University of Bristol, Queens Road, BS81QU, Bristol, UK.

List of Figures

Figure 3: 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, N_{max}, and the maximum zoom level, L_{max}, allowed for every cell. Grids of the global testing region: (a) $N_{\text{max}} = 100$ and $L_{\text{max}} = 11$ (*N100L11*), (b) $N_{\text{max}} = 10$ and $L_{\text{max}} = 11$ (*N10L11*). Japanese testing region: (c) N_{max}=1000 and $L_{\text{max}} = 14$ (*N1000L14*), (d) $N_{\text{max}} = 400$ and $L_{\text{max}} = 14$ (*N400L14*). In frames (a) and (b), the points refer to $M \geq 5.15$ earthquakes between 1976 and 2013 while in frames (c) and (d) the points refer to $M \geq 1.0$ events between 2000 and 2007. Map tiles by Carto under CC BY 3.0. Data by OpenStreetMap and contributors under ODbL. Modified from original.

Figure 4: Forecast rates of the sample model given in numbers of earthquakes per year for every cell, acquired for the global Quadtree grid *N10L11* shown in Figure 3(b). Map tiles by Carto under CC BY 3.0. Data by OpenStreetMap and contributors under ODbL. Modified from original.

Figure 5: Quality of the sample forecasts generated on the single-resolution grid L11 and the multi-resolution grid N10L11 as a function of the water level: the point-process log-likelihood values (Equation 5) are shown in red with the scale on the left. The corresponding percentages of the background seismicity related to cells without earthquakes in the training period are shown in blue with the scale on the right. The vertical and horizontal dotted lines indicate the *L*-maxima and corresponding background values, respectively. The observations are more consistent with the forecast based on the multi-resolution grid compared to the forecast based on the single-resolution grid.

Figure 6: T-test results for the forecasts of the sample model using different multi-resolution grids for forecast generation. For each forecast, the IGPE-value is calculated relative to the model's forecast using the *N1L11*-grid. The IGPE values are shown by circles and their 95% confidence intervals by error bars. The results calculated with the log-likelihood values defined for point-processes (Equation 5) are shown in blue, while the corresponding results, which are calculated by means of POLL-values (Equation 3) after de-aggregating all forecasts to the common *N1L11* test-grid, are shown in red. The test catalog contains all $M \geq 5.15$ earthquakes in the Global CMT that occurred in the period 2014–2019.

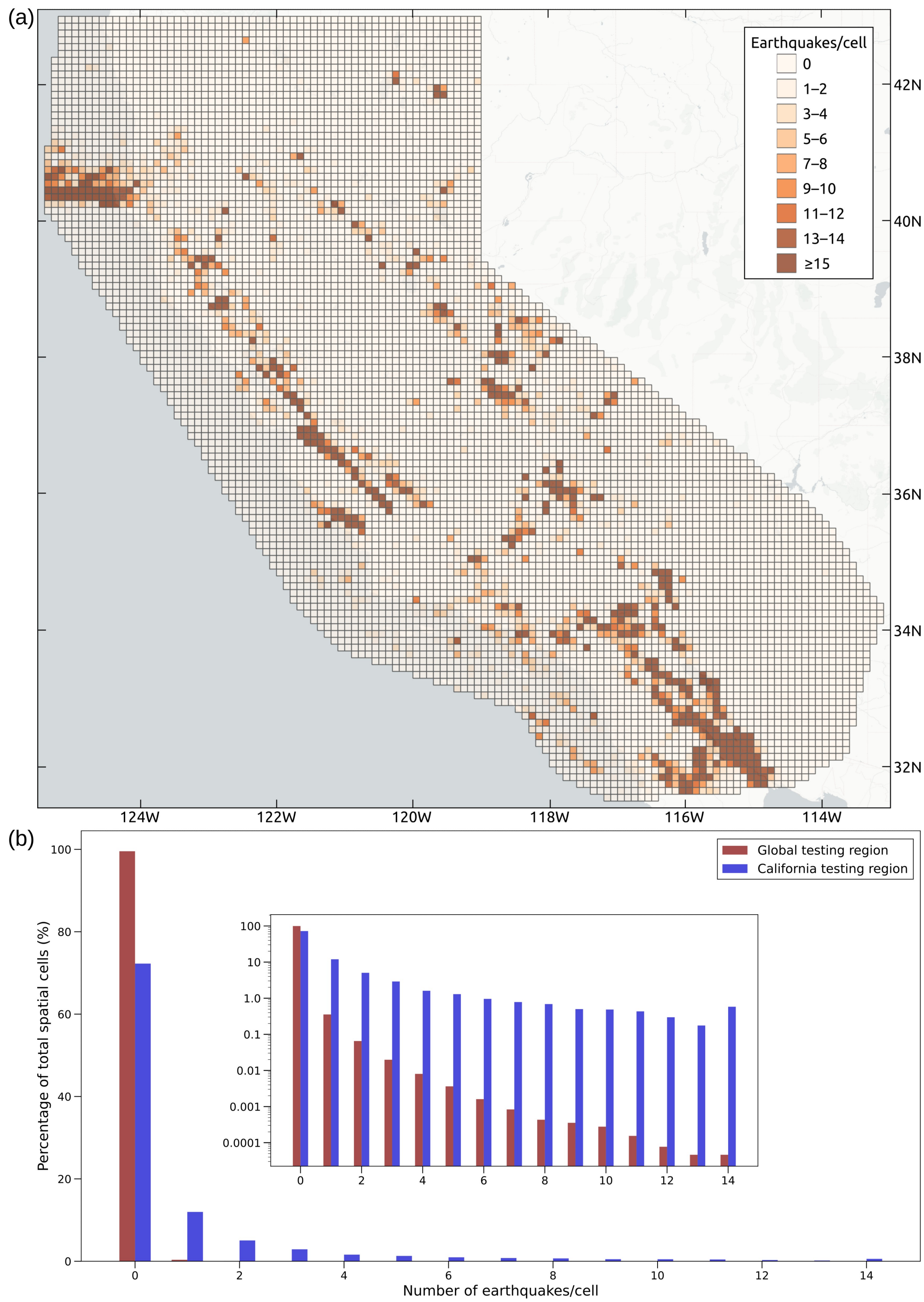

Figure 1. Quantification of the spatial earthquake distribution, highlighting that earthquakes occur only in a small fraction of spatial cells for a $0.1^\circ \times 0.1^\circ$ spatial grid at global and regional scales. (a) Earthquakes per cell in California testing region. (b) Frequency of cells with zero, one, two, or more earthquakes in the $0.1^\circ \times 0.1^\circ$ grid for California (blue) based on the Advanced National Seismic System (ANSS) catalog in California with 25227 earthquakes of $M \geq 2.5$ from 2000 to 2015 (Guy et al., 2015) and the globe (red) based on the Global CMT catalog containing 28465 earthquakes with $M \geq 5.15$ from 1976 to 2013 (Ekström et al., 2012). The inset shows the same result in a log-linear scale to increase visibility. Map tiles by Stamen Design under CC BY 3.0. Data by OpenStreetMap and contributors under ODbL. Modified from original.

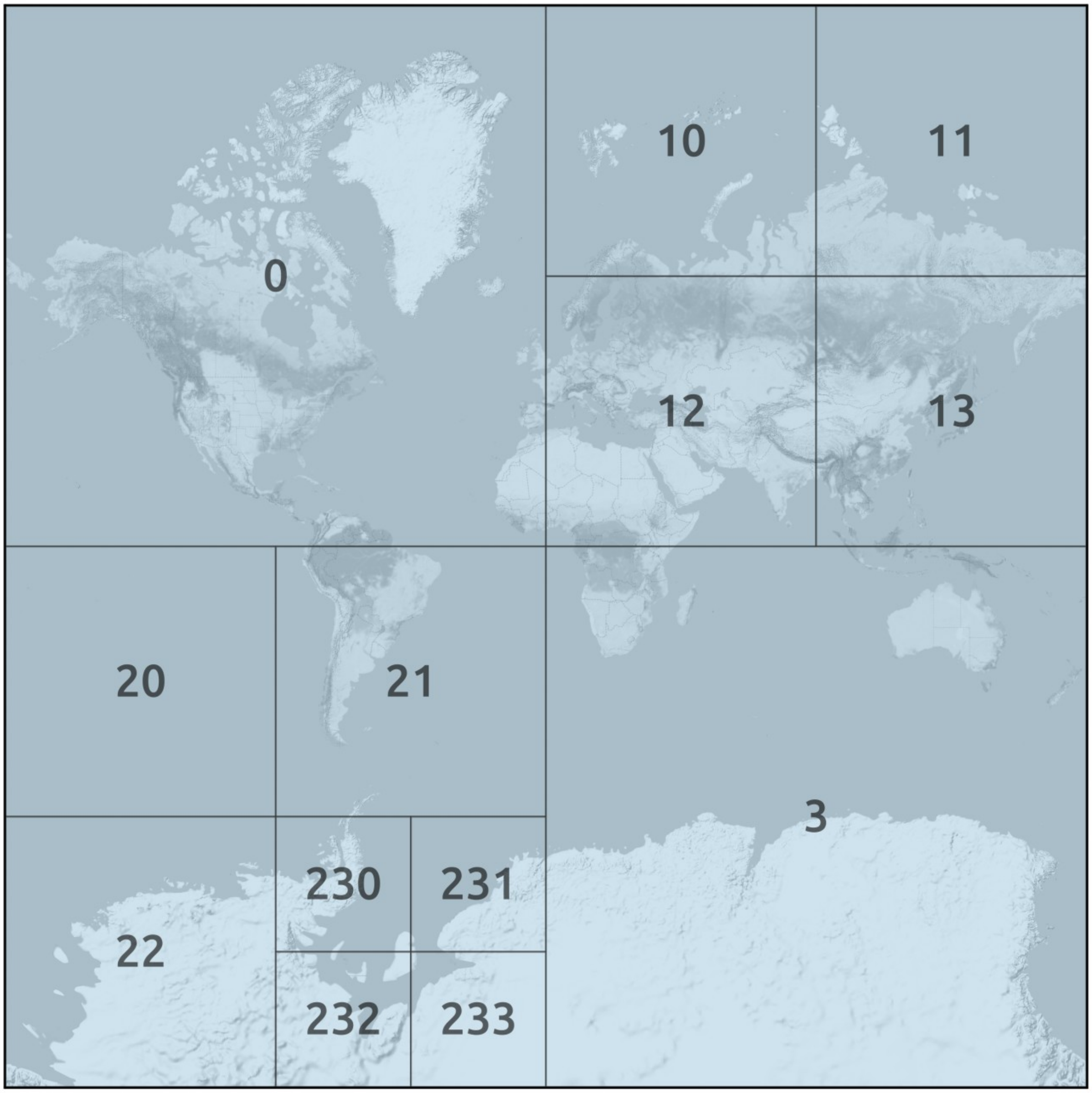

Figure 2. Quadtree-based recursive division of the globe into tiles at various zoom levels. Map tiles by Stamen Design under CC BY 3.0. Data by OpenStreetMap and contributors under ODbL. Modified from original.

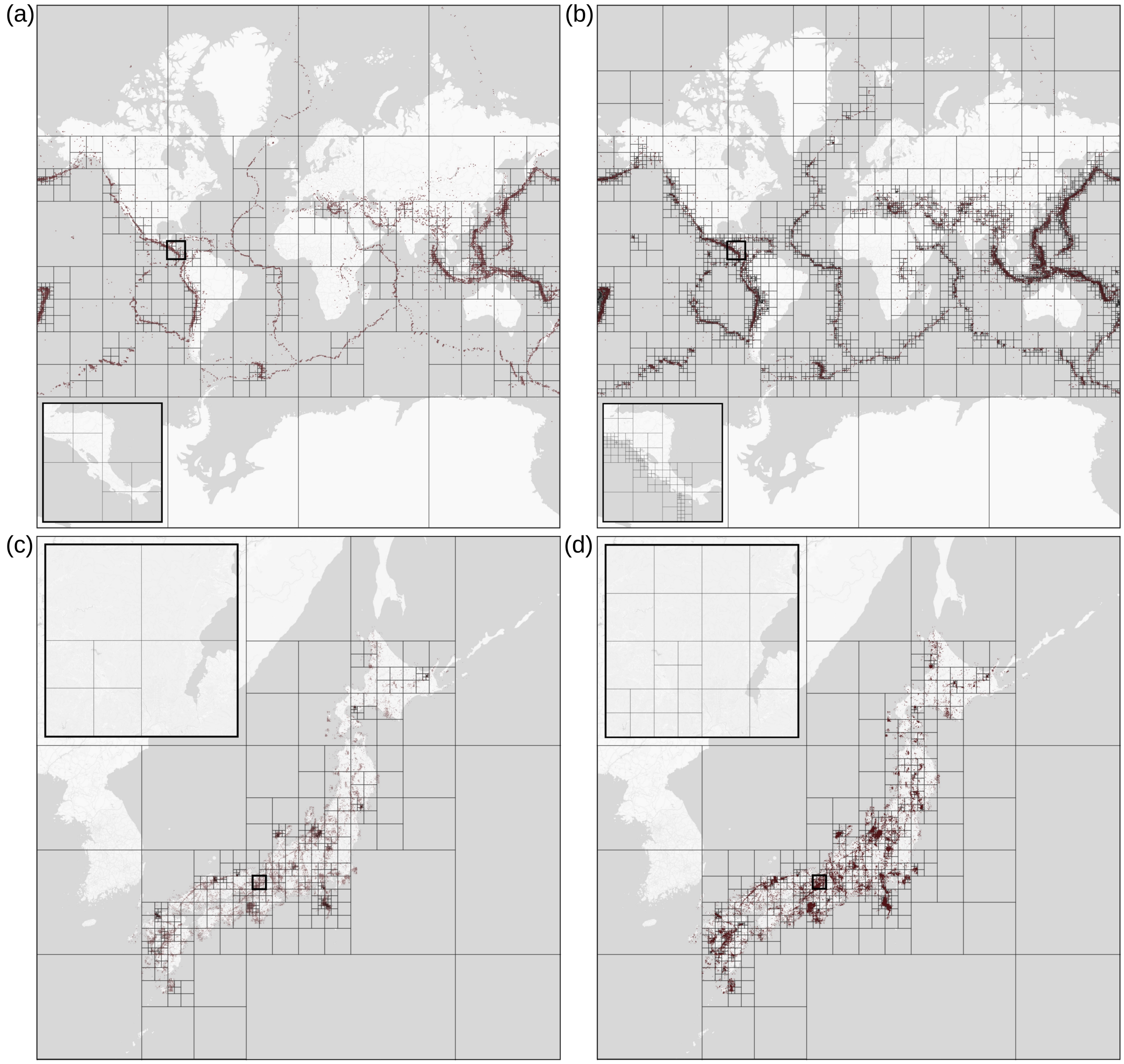

Figure 3. Data-driven generation of multi-resolution Quadtree grids based on earthquake catalogs. The grid resolution is determined by two conditions, the maximum number of earthquakes allowed per grid cell, $N_{\max}$, and the maximum zoom level, $L_{\max}$, allowed for every cell. Grids of the global testing region: (a) $N_{\max} = 100$ and $L_{\max} = 11$ (*N100L11*), (b) $N_{\max} = 10$ and $L_{\max} = 11$ (*N10L11*). Japanese testing region: (c) $N_{\max}$=1000 and $L_{\max} = 14$ (*N1000L14*), (d) $N_{\max} = 400$ and $L_{\max} = 14$ (*N400L14*). In frames (a) and (b), the points refer to $M \geq 5.15$ earthquakes between 1976 and 2013 while in frames (c) and (d) the points refer to $M \geq 1.0$ events between 2000 and 2007. Map tiles by Carto under CC BY 3.0. Data by OpenStreetMap and contributors under ODbL. Modified from original.

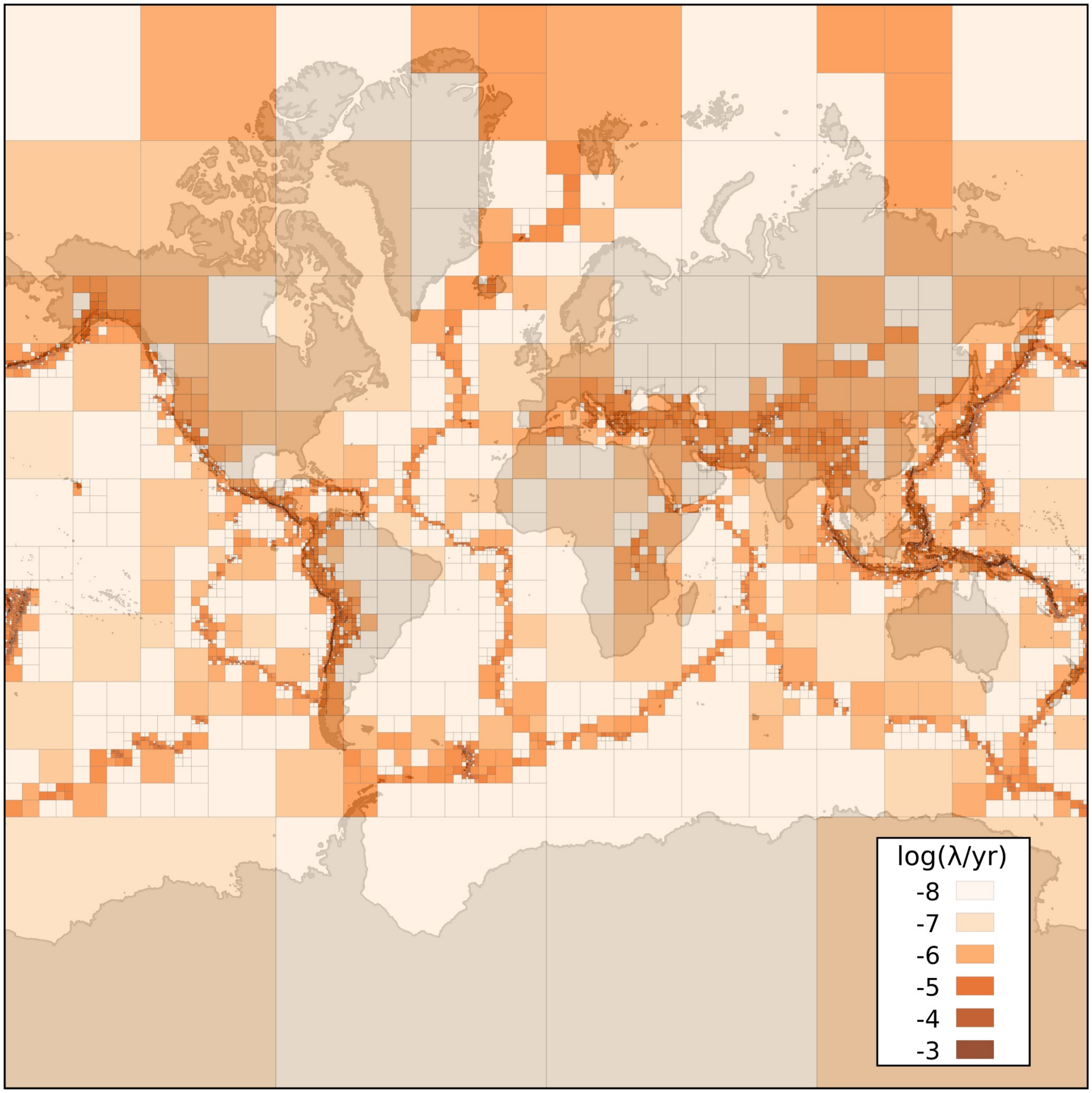

Figure 4. Forecast rates of the sample model given in numbers of earthquakes per year for every cell, acquired for the global Quadtree grid *N10L11* shown in Figure 3(b). Map tiles by Carto under CC BY 3.0. Data by OpenStreetMap and contributors under ODbL. Modified from original.

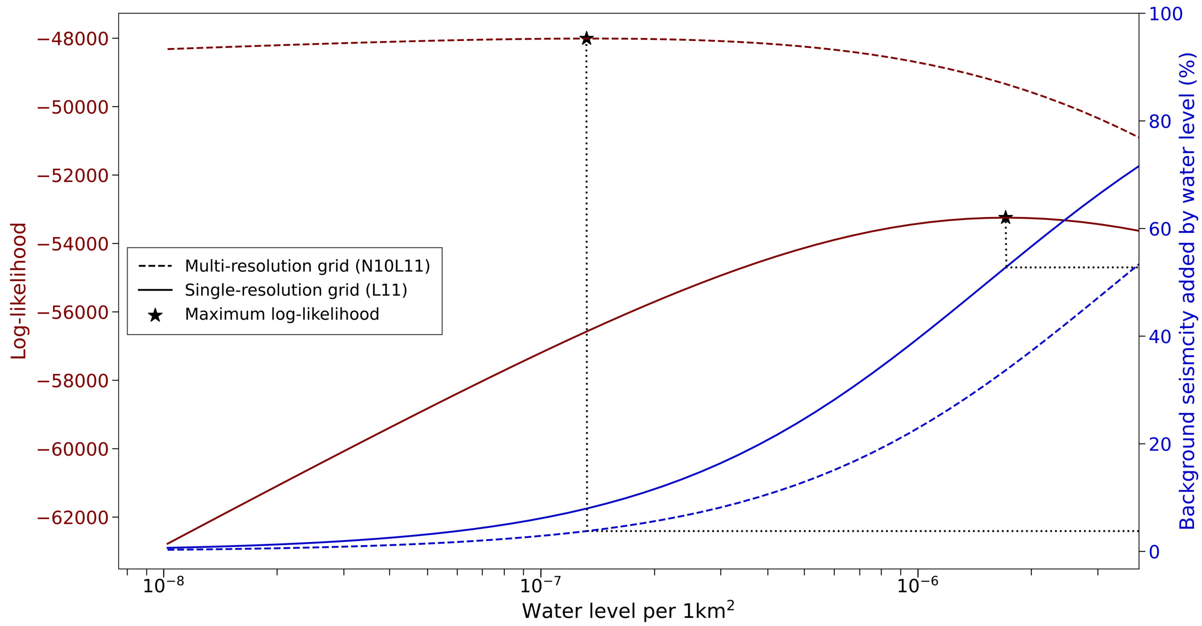

Figure 5. Quality of the sample forecasts generated on the single-resolution grid L11 and the multi-resolution grid N10L11 as a function of the water level: the point-process log-likelihood values (Equation 5) are shown in red with the scale on the left. The corresponding percentages of the background seismicity related to cells without earthquakes in the training period are shown in blue with the scale on the right. The vertical and horizontal dotted lines indicate the L-maxima and corresponding background values, respectively. The observations are more consistent with the forecast based on the multi-resolution grid compared to the forecast based on the single-resolution grid.

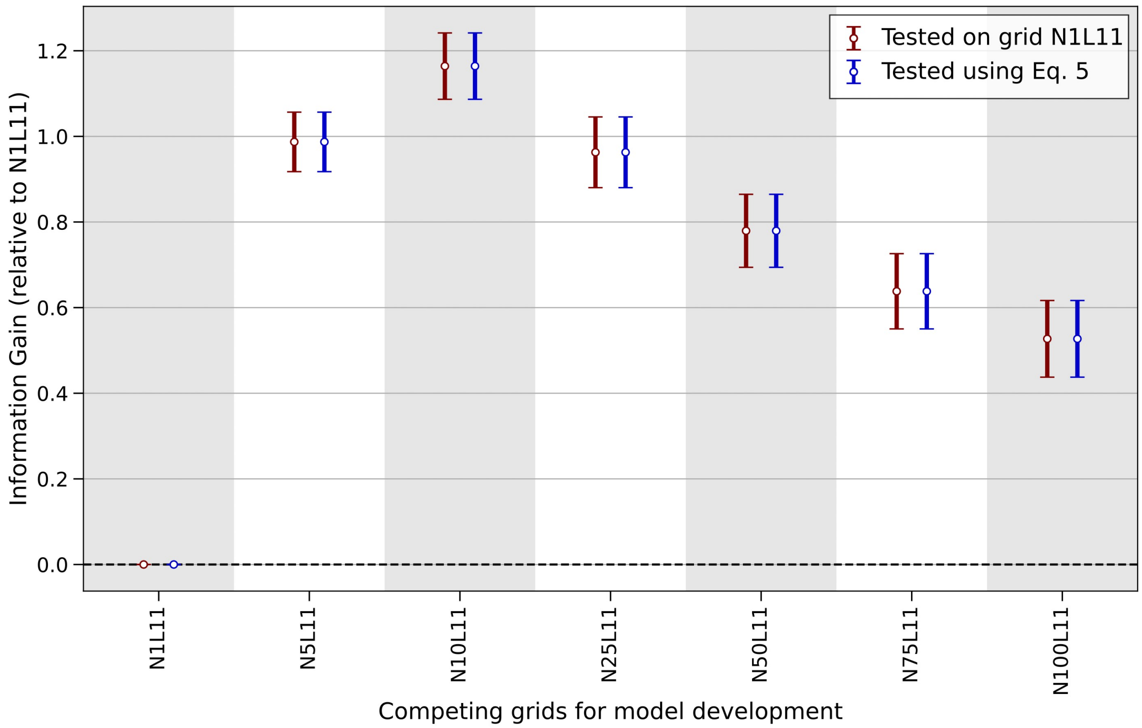

Figure 6. T-test results for the forecasts of the sample model using different multi-resolution grids for forecast generation. For each forecast, the IGPE-value is calculated relative to the model's forecast using the *N1L11*-grid. The IGPE values are shown by circles and their 95% confidence intervals by error bars. The results calculated with the log-likelihood values defined for point-processes (Equation 5) are shown in blue, while the corresponding results, which are calculated by means of POLL-values (Equation 3) after de-aggregating all forecasts to the common *N1L11* test-grid, are shown in red. The test catalog contains all $M \geq 5.15$ earthquakes in the Global CMT that occurred in the period 2014–2019.

Table 1. The total number of grid cells and the percentage of cells without any earthquake in the case of different single-resolution grids (left) and multi-resolution grids (right).

Grids	Total Cells	Cells without EQ (%)	Grids	Total Cells	Cells without EQ (%)
L5	1024	67.5	N100L11	922	12.9
L6	4096	81.7	N75L11	1243	13.19
L7	16384	90.4	N50L11	1780	14.1
L8	65536	94.9	N25L11	3502	15.9
L9	262144	97.5	N10L11	8089	21.9
L10	1048576	98.9	N5L11	14782	30.6
L11	4194304	99.6	N1L11	39811	55.5