nature communications

Article

https://doi.org/10.1038/s41467-024-51596-z

Abnormal low-magnitude seismicity preceding large-magnitude earthquakes

Received: 3 October 2023

Accepted: 13 August 2024

Published online: 28 August 2024

Check for updates

Társilo Girona [1] & Kyriaki Drymoni [2]

Unraveling the precursory signals of potentially destructive earthquakes is crucial to understand the Earth's crust dynamics and to provide reliable seismic warnings. Earthquake precursors are ambiguous, but recent experimental studies suggest that robust warning signs may precede large seismic events in the short (day-to-months) term. Here, we show that the M6.4-M7.1 2019 Ridgecrest sequence (California) and the M7.1 2018 Anchorage earthquake (Alaska) were preceded by up to ~3 months of tectonic unrest on regional scales, as evidenced by abnormal low-magnitude seismicity spreading over the ~15-25% of Southern California and Southcentral Alaska. This precursory unrest has been discovered with an algorithm that integrates an innovative random forest machine learning approach and statistical features built from earthquake catalogs. Supported by a novel suite of finite element solid mechanics models, we propose that precursory, abnormal, low-magnitude seismicity arises if the pore fluid pressure within large fault segments escalates significantly as they approach failure, which leads to major uneven changes in the regional stress field. Our findings and method may open up new perspectives for surveillance agencies to anticipate when a region approaches an earthquake of great magnitude weeks to months before it occurs.

Enhancing our ability to forecast the spatial and temporal distribution of large magnitude, and thus potentially destructive earthquakes, is a formidable challenge in modern seismology[1–3]. Earthquake forecasting has been explored since the last century via a plethora of different methods aiming to detect geophysical, geochemical, and even biological precursors[4–8]. Some of these possible precursors include thermal anomalies on the Earth's surface[9–11]; emissions of acoustic[12] and acoustic-gravity[13] waves; changes in groundwater level and composition[14,15]; release of radon and other gases[16,17]; electric, magnetic, and ionospheric perturbations[18,19]; crustal deformation[20]; the emergence of slow slip events[21,22]; and even anomalous behavior of farm animals[23]. Earthquake precursors are still unreliable despite recent advances in data acquisition and analysis, and some of those precursors have been categorically questioned[24,25]. Hence, the development of alert level systems for seismic activity, similar to those successfully implemented by surveillance agencies to track volcanic unrest[26], remain a chimera.

Precursory activity to large-magnitude earthquakes has also been investigated by tracking variations in the slope of the Gutenberg–Richter relationship (or *b*-value[27–29]) and by exploring other anomalous statistical changes in the low-magnitude seismicity[30–32]. In particular, low-magnitude seismicity has been proposed to experience changes over wide areas prior to large earthquakes[33], although previous analyses are inconclusive because the spatiotemporal distribution of seismic events is generally complex and specific precursory patterns are challenging to recognize and generalize[1,24]. This might be alleviated thanks to the advent of openly available, high-quality, earthquake catalogs; and to new supervised and/or unsupervised machine learning-based frameworks[34], which may be able to identify subtle and nonlinear hidden precursory patterns from previous experience[35,36]. Supervised machine learning-based frameworks have already been successfully applied to accurately anticipate laboratory earthquakes from pre-failure acoustic signals[37] and from the spatiotemporal deformation of analog tectonic

[1]Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA. [2]Earth and Environmental Sciences, Ludwig-Maximilians-Universität in Munich, Munich, Germany. e-mail: tarsilo.girona@alaska.edu

https://doi.org/10.1038/s41467-024-51596-z

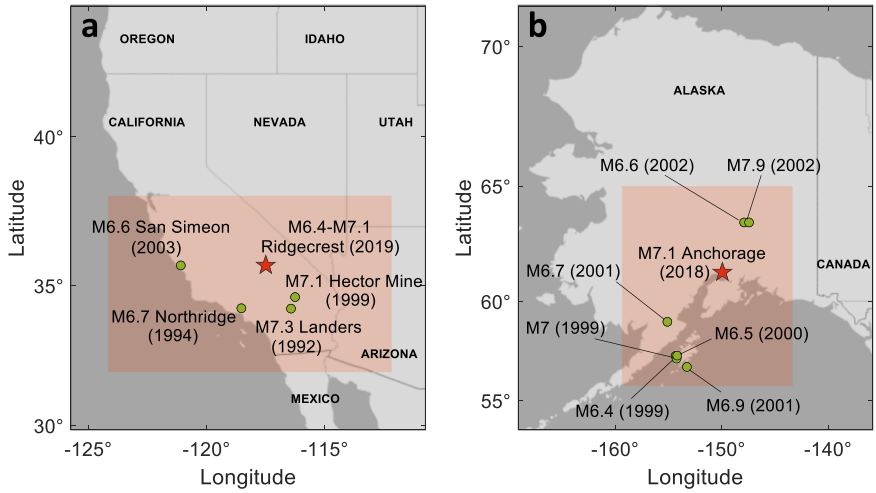

Fig. 1 | Areas explored in this project. a West Coast of the United States, with focus on Southern California. **b** Alaska, with focus on the Southcentral Alaska region. The red stars represent the two target seismic events (2019 Ridgecrest sequence and 2018 Anchorage earthquake) used to test our machine learning algorithm, whereas the green circles show the large-magnitude earthquakes (M ≥ 6.4) used in the training step. The low-magnitude events used for training (not shown in the panels) are those occurring within the red-shaded rectangles. Geographical maps were created using MATLAB R2020b (Mapping Toolbox).

plates[38], thus suggesting that similar approaches might eventually be able to anticipate real earthquakes from seismic, acoustic, and space geodesy data.

In this manuscript, we demonstrate through a new multivariate, supervised, machine learning-based algorithm that low-magnitude seismicity can alert about impending large-magnitude earthquakes in Southern California and Southcentral Alaska. Unlike previous studies focused on providing specific forecasts of the timing, magnitude, and/or epicentral location of large earthquakes and aftershocks[35,39,40], we exploit machine learning to explore the emergence of robust, non-linear, precursory patterns that may be hidden in low-magnitude seismicity. This concept is applied to two target events: The 2019 Ridgecrest earthquake sequence and the 2018 Anchorage earthquake (Fig. 1). The former was the strongest seismic sequence occurring in Southern California in two decades[41] and included an M6.4 earthquake occurring on July 4 and an M7.1 earthquakes occurring on July 5. The latter, the most societally significant seismic event occurring in Southcentral Alaska in half a century[42], was an M7.1 earthquake occurring on November 30. In particular, our machine learning-based algorithm is trained to determine whether the aforementioned events were preceded by anomalies in the low-magnitude seismicity (defined here as earthquakes with magnitude between $M_{Min} = 1$ and $M_{Max} = 6$), akin to the activity preceding other large-magnitude earthquakes (defined here as earthquakes with magnitude ≥ 6.4, i.e., the magnitude of the first event of the Ridgecrest sequence) of Southern California and Southcentral Alaska. A detailed description of our algorithm is provided in "Methods".

Results

Below we show the results obtained when running our machine learning-based algorithm at the epicentral location of the first event of the 2019 Ridgecrest sequence and at the epicentral location of the 2018 Anchorage earthquake (Fig. 2). These results are expressed in terms of the degree or probability of unrest (P_{un}), which is defined as the ratio of decision trees classifying as anomalous a set of statistical features obtained from the earthquake catalog. Note that P_{un} can also be interpreted as the probability of a large-magnitude earthquake (M ≥ 6.4) occurring within the following 30 days (see "Methods"). Our model reveals that the epicenter of the target earthquakes showed signs of unrest for several weeks before the occurrence of the events. In particular, the degree of unrest is very low, with small spikes never exceeding ~5%, until just ~40 days before the Ridgecrest earthquakes

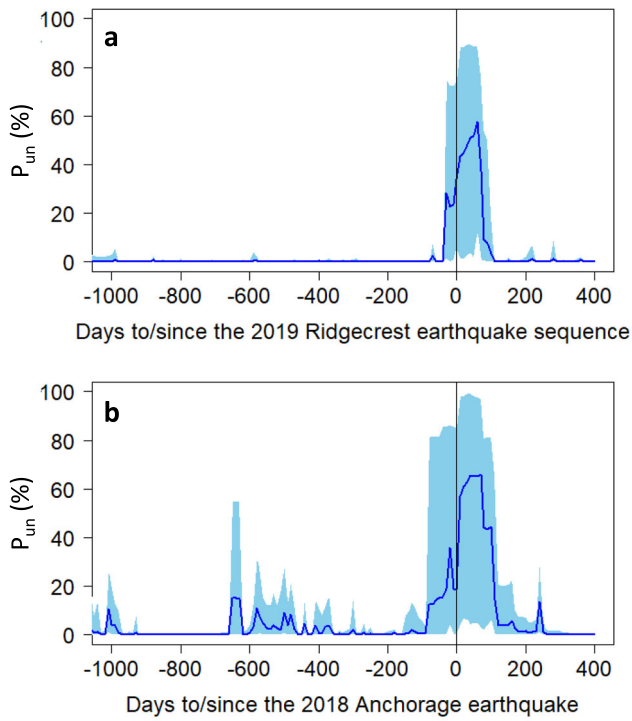

Fig. 2 | Temporal evolution of the probability of unrest (P_{un}; or probability that a large-magnitude earthquake happens in 30 days or less), from 1000 days before to 400 days after the events. a Results obtained when running our algorithm at the epicentral location of the first earthquake (M6.4) of the 2019 Ridgecrest sequence (California). **b** Results obtained when running our algorithm at the epicentral location of the M7.1 2018 Anchorage earthquake (Alaska). Blue lines show the mean value of the predictions obtained through an ensemble of 200 random forest models. Blue-shaded areas show the range of variation of the predictions, with the lower and upper bounds representing minimum and maximum, respectively. Vertical black lines show the occurrence of the earthquakes. See more details in "Methods".

(Fig. 2a). Then, mean and maximum values of P_{un} increased quickly to ~20–25% and ~75%, respectively, and remained that high up to the first sequence shock. After the M6.4 event, the maximum probability of unrest increased to ~90%, probably controlled by aftershocks, before it

https://doi.org/10.1038/s41467-024-51596-z

decreased again to background levels (≲5%) around three months after the sequence. On the epicenter of the 2018 Anchorage earthquake, the maximum probability of unrest increased abruptly to ~80% around three months before the event, with mean and maximum P_{un} reaching up to ~35% and ~85%, respectively, just a few days before the earthquake (Fig. 2b). Our model also reveals an anomalous phase from ~700 days to ~300 days before the Anchorage earthquake, although mean and maximum values of the probability of unrest are much lower (<15% and <55%, respectively) than those observed before the Anchorage shock. Similar to the Ridgecrest sequence, our model reveals an increase of P_{un} after large-magnitude events in Alaska, reaching up to ~99% over the three months following the Anchorage earthquake. Around 170 and 250 days after the 2018 Anchorage event, the probability of unrest decreased below ~20% and ~5%, respectively.

Epicentral locations are not known before an earthquake, so a more realistic and useful analysis consists of running our machine learning-based algorithm on the geographical coordinates of grids covering all of Southern California and Southcentral Alaska (we used grids with 0.1° latitude × 0.1° longitude spacing). This approach allows us to build probability maps and thus explore the spatiotemporal evolution of abnormal low-magnitude seismicity in the run-up to large-magnitude earthquakes. Our analysis reveals that anomalous low-magnitude seismicity is not constrained to the epicentral area of large-magnitude events, but spreads over multiple fault zones of the regions explored (Figs. 3 and 4 and Supplementary Movies 1 and 2). In particular, around 90 days before the 2019 Ridgecrest sequence, unrest first emerged in the southern tip of San Joaquin Valley (e.g., Mt. Poso fault, and Wheeler Ridge and Pleito fault zones). Then, ~60 days before the sequence, abnormal low-magnitude seismicity spread over the California–Nevada-Arizona border region, including the Indian Springs Valley faults and Las Vegas area (e.g., Eglington and Frenchman Mountains faults) (Fig. 3a, b). Later, ~30 days before the sequence, unrest kept spreading in the California–Nevada-Arizona border region and emerged in the epicentral area of the Ridgecrest events (e.g., Little Lake and Panamint Valley—southern section—fault zones); southwest flank of Sierra Nevada mountains (e.g., Ken Canyon fault); Antelope Valley (e.g., Lenwood–Lockhart and Helendale–South Lockhart fault zones; San Andreas fault –Mojave section-); San Gabriel Mountains (e.g., San Gabriel, San Jacinto, Sierra Madre fault zones); Los Angeles county (e.g., Raymond, Santa Monica, and Elsinore fault zones); and the Southern Channel Islands (up to San Clemente Island; e.g., Palos Verdes, San Pedro Basin, Catalina fault zones). All these areas continued showing anomalous seismic behavior until the occurrence of the Ridgecrest events (Fig. 3c). In the run-up to the 2018 Anchorage earthquake, unrest emerged in two main areas of Southcentral Alaska: Center-northeast (e.g., Cook Inlet Folds, Castle Mountain faults, northern Kenai Peninsula, Copper River Basin, Denali fault –west Muldrow–Alsek section–, Billy Creek faults; Fig. 4a, b) and south-southeast (e.g., Alaska–Aleutians subduction zone –Aleutian Megathrust–; Kodiak shelf fault zone; Fig. 4c) around 80 and 10 days before the earthquake, respectively. More information on the main fault zones present in the areas where unrest is detected can be found in the US Geological Survey's interactive Quaternary faults database (https://doi.org/10.5066/F7S75FJM).

The analysis of the total area experiencing tectonic unrest (i.e., area with the maximum probability of unrest ≥50%) reveals three major outcomes (Fig. 5). First, small-scale, isolated patches of tectonic unrest emerged sporadically in the target regions (Figs. 3a and 4a). These patches constitute a background noise level that do not warn of large-magnitude earthquakes, and that might be associated with failed large-magnitude events or with local anomalous patterns produced by the swarms of low-to-mid size earthquakes. For example, we find that ~1.5–7% of Southern California's total area experimented unrest from ~1000 days to ~100 days before the 2019 Ridgecrest sequence (Fig. 5). This analysis also reveals that the anomalous phase detected on the epicenter of the Anchorage earthquake from ~700 to ~300 days before its occurrence (Fig. 2b) is just a spurious signal associated with spatial background noise. Second and most important, the area experiencing unrest spread significantly in the run-up to large

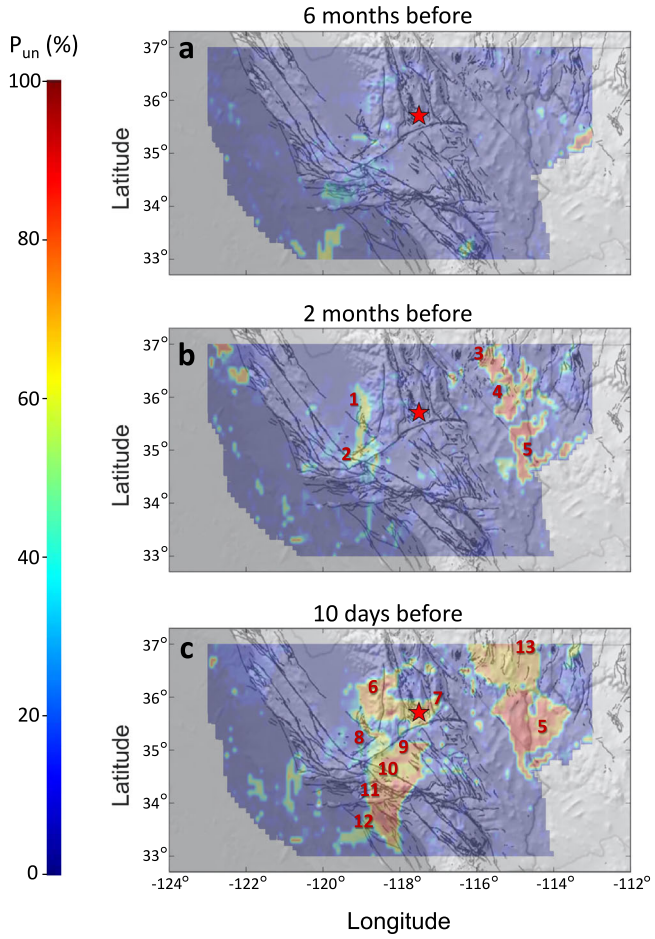

Fig. 3 | Spatiotemporal evolution of the maximum probability of unrest (P_{un}; or probability that a large-magnitude earthquake happens in 30 days or less) in Southern California. Results obtained for **a** 6 months, **b** 2 months, and **c** 10 days before the first event of the 2019 Ridgecrest sequence (red star). Uncolored areas are unexplored locations, or locations with no earthquakes in the region and time interval of interest (see "Methods"). Black lines show the major fault lines of Southern California, Southwest Nevada, and Northwest Arizona (US Geological Survey's interactive Quaternary faults database; https://doi.org/10.5066/F7S75FJM); and warm colors and red numbers highlight areas with anomalous low-magnitude seismicity. (1) Southern tip of San Joaquin Valley (e.g., unnamed fault zone, Mt. Poso fault). (2) Southern tip of San Joaquin Valley (Wheeler Ridge and Pleito fault zones). (3) California–Nevada–Arizona border region (Spotted Range, Mercury Ridge, West Pintwater Range, and Indian Springs Valley faults). (4) California–Nevada–Arizona border region (Las Vegas faults, e.g., Eglington and Frenchman Mountains faults). (5) California–Nevada–Arizona border region (small unnamed faults). (6) Southwest flank of Sierra Nevada mountains (e.g., Ken Canyon fault). (7) Epicentral area of Ridgecrest events (Southern Sierra Nevada, Little Lake, and Panamint Valley—southern section—fault zones). (8) Southern tip of San Joaquin Valley (Poso Creek, Edison, and other unnamed faults). (9) Antelope Valley (Lenwood–Lockhart, Mirage Valley, and Kramer Hills fault zones; Blake Ranch fault). (10) Antelope Valley (San Andreas fault—Mojave Section). (11) San Gabriel Mountains and Los Angeles County (San Gabriel and Sierra Madre fault zone, Los Angeles faults, Compton thrust fault, Palos Verdes fault zone). (12) Southern Channel Islands (San Pedro Basin fault, junction between San Diego Through fault and Catalina Fault zone). (13) California–Nevada–Arizona border region (Maynard Lake, Sheep Basin, Sheep-East Desert Ranges, Sheep Range, Kane Spring Wash, Wildcat Wash, Arrow Canyon Range, Carp Road, and California Wash faults). Geographical maps were created using MATLAB R2020b (Mapping Toolbox).

https://doi.org/10.1038/s41467-024-51596-z

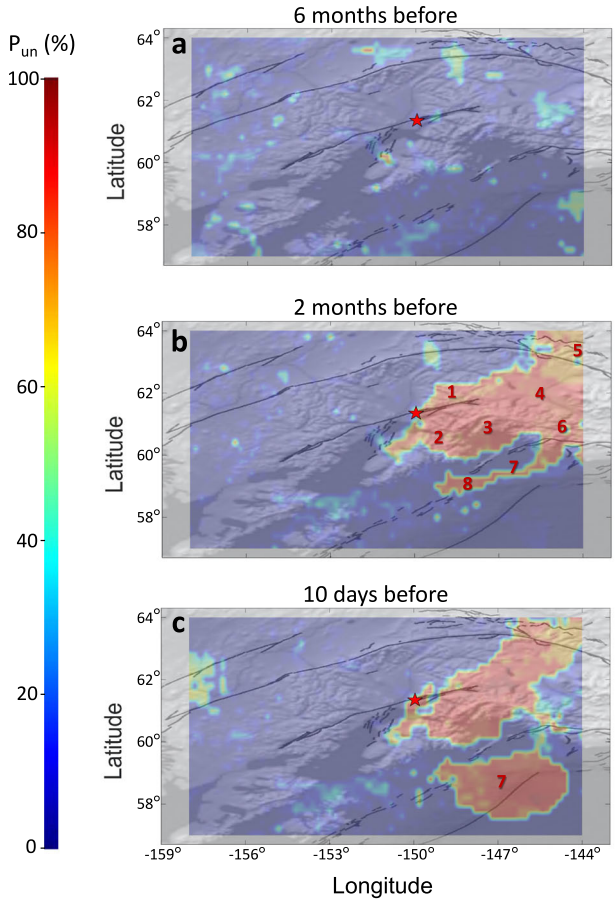

Fig. 4 | Spatiotemporal evolution of the maximum probability of unrest (P_{un}; or probability that a large-magnitude earthquake happens in 30 days or less) in Southcentral Alaska. Results obtained for **a** 6 months, **b** 2 months, and **c** 10 days before the 2018 Anchorage earthquake (red star). Uncolored areas are unexplored locations, or locations with no earthquakes in the region and time interval of interest (see "Methods"). Black lines show major fault lines identified in Southcentral Alaska (US Geological Survey's interactive Quaternary faults database; https://doi.org/10.5066/F7S75FJM), and warm colors and red numbers highlight areas with anomalous low-magnitude seismicity. (1) Cook Inlet Folds (Big Lake North, Pittman, Wasilla St. No. 1-Needham Anticlines); Castle Mountain, Caribou, and East Boulder Creek faults. (2) Northern Kenai Peninsula. (3) Chugach Mountains, Prince William Sound. (4) Copper River Basin, Wrangell Mountains. (5) Denali fault (west Muldrow–Alsek section), Northern Foothills fold and thrust belt (Macomb Plateau, Cathedral Rapids, Dot "T" Johnson, Granite Mountain, and Billy Creek faults), McCallum Creek fault. (6) Chugach-St. Elias fold and thrust belt (Bagley fault). (7) Alaska–Aleutians subduction zone (Aleutian Megathrust). (8) Kodiak Shelf fault zone. Geographical maps were created using MATLAB R2020b (Mapping Toolbox).

earthquakes (Figs. 3b, c and 4b, c). In particular, we find that the area with abnormal low-magnitude seismicity in Southern California and Southcentral Alaska increased by up to one order of magnitude (from ~2% to ~17%, and from ~2% to ~22%, respectively) during the three months preceding the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake (Fig. 5). Third, the area and probability of unrest around the epicenters suddenly increased after large earthquakes, but the regions returned back to background levels shortly after (Fig. 5; Supplementary Movie 1 and 2). Our model also shows a gradual long-term (years) decreasing trend of the unrest area from ~1000 days to ~500 days before the 2018 Anchorage earthquake (Fig. 5), which may reflect the return to background levels after the occurrence of two other large earthquakes in the region (M6.4 earthquake occurring 70 km SSW of Redoubt volcano on July 20, 2015; and

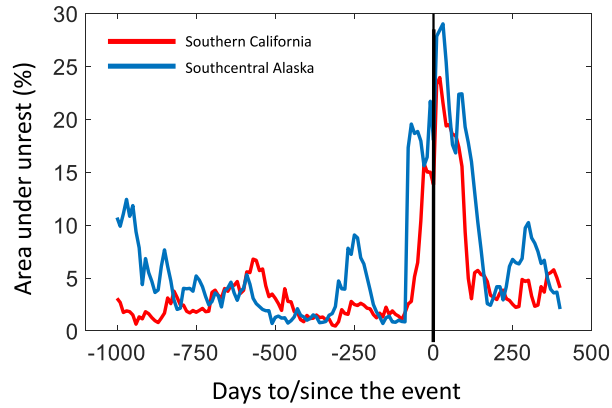

Fig. 5 | Temporal evolution of the area (in percentage) of Southern California (red) and Southcentral Alaska (blue) that is subject to unrest. The time series show the results obtained from 1000 days before to 400 days after the first event of the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake. Unrest means that the maximum degree or probability of unrest (P_{un}; as obtained from an ensemble of 200 random forest models) is greater than 50%. The vertical line shows the occurrence of the earthquakes. See more details in "Methods".

M7.1 earthquake occurring 86 km east of Old Iliamna on January 24, 2016; see Supplementary Movie 2). It is worth noting that we used the same algorithm both for Southern California and Southcentral Alaska.

The sensitivity of our results to the key hyperparameters and statistical features of our machine learning-based algorithm is also explored through a parametric analysis. In particular, we run our algorithm for a total of 20 different scenarios varying the definition of low-magnitude seismicity (i.e., using different values for M_{Min} and M_{Max}), the unrest period assumed for training (from $T_{unrest} = 10$ days to $T_{unrest} = 100$ days), and the radius of the region of interest where the statistical features are calculated (from R = 20 km to R = 150 km) (Supplementary Table 1). In addition, we run five other scenarios in which our algorithm is trained with each of the statistical features separately (Supplementary Table 1). For simplicity, the rest of parameters of our model are kept the same for all the runs (time series length T = 2 years, 1-year backward sliding windows, 1-day time steps, and $N = 50$ random nodes used for training); more details on the different hyperparameters and statistical features used in the model can be found in Methods. The parametric analysis reveals that: (i) Our results are little sensitive to the hyperparameters, thus demonstrating the robustness of the method; (ii) precursory patterns emerge more significantly when low-magnitude seismicity is broadly defined (i.e., when using $M_{Min} = 1$ and $M_{Max} = 6$), when a short unrest period is assumed for training ($T_{unrest} \lesssim 30$ days), and for large regions of interest ($R \gtrsim 120$ km); (iii) precursory patterns vanish if using $M_{Min} \gtrsim 1.5$, thus suggesting that regional unrest preceding large-magnitude earthquakes is mostly reflected in the very low-magnitude seismicity; and (iv) precursory patterns mostly reflect anomalous variations in the depth of the low-magnitude earthquakes, while the statistical feature reflecting anomalous variations in magnitude is the least relevant. A detailed description of the parametric exploration and results can be found in Supplementary Tables 1 and 2.

Discussion

Interpretation of our results: dissimilar accumulation of stresses

The fact that the observed anomalies mostly reflect abnormal, very low-magnitude ($1 \lesssim M \lesssim 1.5$) seismicity suggests that the probability of unrest may elevate due to temporal variations of the magnitude of completeness in the catalog, which might be produced by mid-magnitude foreshocks preceding the main shocks[43]. To test this hypothesis, we adopted the approach outlined by Zhuang et al.[43] and

https://doi.org/10.1038/s41467-024-51596-z

plotted event magnitudes against the seismic event order across multiple sectors in Southern California and Southcentral Alaska, which coincide with areas with and without precursory anomalies (Supplementary Figs. 1–3). None of these plots reveal any obvious variation in catalog completeness in the weeks to months before the Ridgecrest and Anchorage earthquakes, whereas incompleteness clearly manifests after the occurrence of the main shocks (Supplementary Figs. 2a and 3a).

To further test the catalog completeness hypothesis, we plotted the Gutenberg–Richter (GR) relationship using the low-magnitude seismicity occurring around the epicenters of the target events (within a 120 km radius, as used in our method) and for each moving window used to calculate the statistical features feeding our machine learning-based algorithm (Supplementary Figs. 4 and 5). In addition, we calculated the temporal evolution of the absolute value of the slope of the GR relationship (m_{GR}) for magnitudes between M = 1 and M = 3 (Supplementary Fig. 6); note that m_{GR} is the so-called *b*-value[27] only if the magnitude of completeness is below M = 1, and that m_{GR} is expected to decrease if the magnitude of completeness is above M = 1 and if it increases in the run-up to the target events. Our analysis reveals that, in the run-up to the Ridgecrest sequence and the Anchorage earthquake, the GR law holds for low magnitudes as well (i.e, in the range from M = 1 to M = 3–4, and thus the magnitude of completeness is below M = 1); and that m_{GR} does not show any precursory decreasing trend or anomaly. All this together confirms that the precursory abnormal seismicity reported in this work is not an artifact produced by a decline in catalog quality and, consequently, suggests that our method does indirectly identify regional stress alterations taking place in the shallow crust during the preparatory phase of major earthquakes[44].

Interestingly, there was a noteworthy swarm of low-magnitude earthquakes approximately 30 days prior to the Ridgecrest events around the epicentral area of the sequence (Sector 1; see Supplementary Fig. 1a) and in the area spanning from Antelope Valley to Southern Chanel Islands (Sector 2; Supplementary Fig. 1a) (Fig. 6). This swarm coincides with a big spread of anomalous seismicity, as detected by our algorithm; however, the first anomalous behavior in the aforementioned areas arises ~90 days before the sequence, i.e., around two months before the occurrence of the swarm (Fig. 6a, b). In addition, the probability of unrest increases in the California–Nevada-Arizona border region (Sector 3; see Supplementary Fig. 1a) around 60 days before the Ridgecrest sequence, yet this surge does not align with any swarm of small-scale seismic events in that area (Fig. 6c), same as in other areas where no anomaly is detected (Fig. 6d, e). Moreover, the area subject to unrest in Southern California starts to increase ~70 days before the occurrence of the swarm (Fig. 5), and precursory unrest without noticeable swarms of small-scale seismic events is also observed for Alaska (Fig. 7). Collectively, these observations suggest that while precursory unrest may be influenced to some extent by local swarms of low-magnitude earthquakes, it is not unequivocally linked to such swarms. Instead, precursory unrest seems to be associated with nonlinear, multivariate, subtle statistical fluctuations in the spatiotemporal distribution of low-magnitude seismicity; this phenomenon demands further investigation and analysis, and a potential explanation is provided in the following.

Several studies have hypothesized that low-magnitude seismicity should undergo changes over vast areas prior to large-magnitude earthquakes if the regional stress field varies when large, locked fault segments get closer to failure. This hypothesis has been explored by analyzing the energy release and accelerating seismicity patterns before large earthquakes, and by coupling Coulomb stress analyses with fault modeling[33,45]. Faults are usually modeled in seismology as dislocations and cracks without internal structures[46]; whereas in other disciplines, such as volcano tectonics, faults are commonly modeled as elastic inclusions with an internal mechanical structure that allows accounting for the shear stress accumulation inside and around them, their slip potential, or permeability[47]. The latter approach allows considering faults as shear fractures composed of two hydromechanical units[47,48]: The damage zone, primarily composed of fractures and veins (fractured media); and the fault core, which consists of breccia and gouge (porous media).

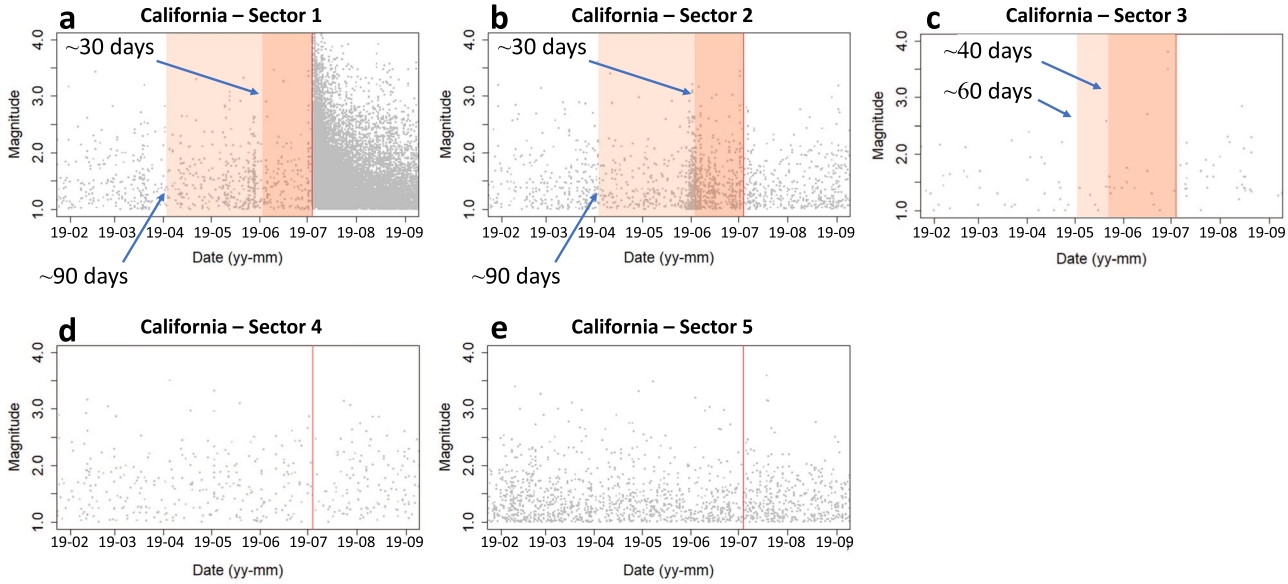

Fig. 6 | Exploration of the possible link between abnormal low-magnitude seismicity and the occurrence of seismic swarms in Southern California. **a**–**e** Magnitude of the events versus time for Sector 1 (latitude range 35N–36.5 N and longitude range 120W–117W), Sector 2 (latitude range 33N–35.2 N and longitude range 119W–117W), Sector 3 (latitude range 34.5N–37N and longitude range 116W–113.5 W), Sector 4 (latitude range 33N–34N and longitude range 117W–114W), and Sector 5 (latitude range 34.5N–37N and longitude range 121W–119.5 W). The vertical red line shows the occurrence of the first event of the Ridgecrest sequence. The lighter red-shaded rectangle shows the emergence of the first anomalous area in the sector, and the darker red-shaded rectangle shows a second spread of the anomaly in the sector. See Supplementary Fig. 1 for more details on the location of the sectors.

https://doi.org/10.1038/s41467-024-51596-z

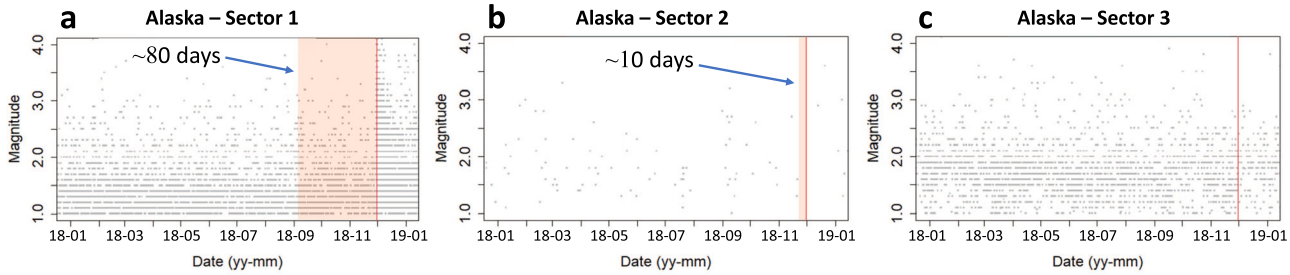

Fig. 7 | Exploration of the possible link between abnormal low-magnitude seismicity and the occurrence of seismic swarms in Southcentral Alaska. **a**–**c** Magnitude of the events versus time for Sector 1 (latitude range 60N–63N and longitude range 150W–144W), Sector 2 (latitude range 57.5N–60N and longitude range 150W–144W), and Sector 3 (latitude range 58N-64N and longitude range 158W–153W). The vertical red line shows the occurrence of the Anchorage earthquake. The shaded rectangles show the emergence of anomalous low-magnitude seismicity in the sector. See Supplementary Fig. 1 for more details on the location of the sectors.

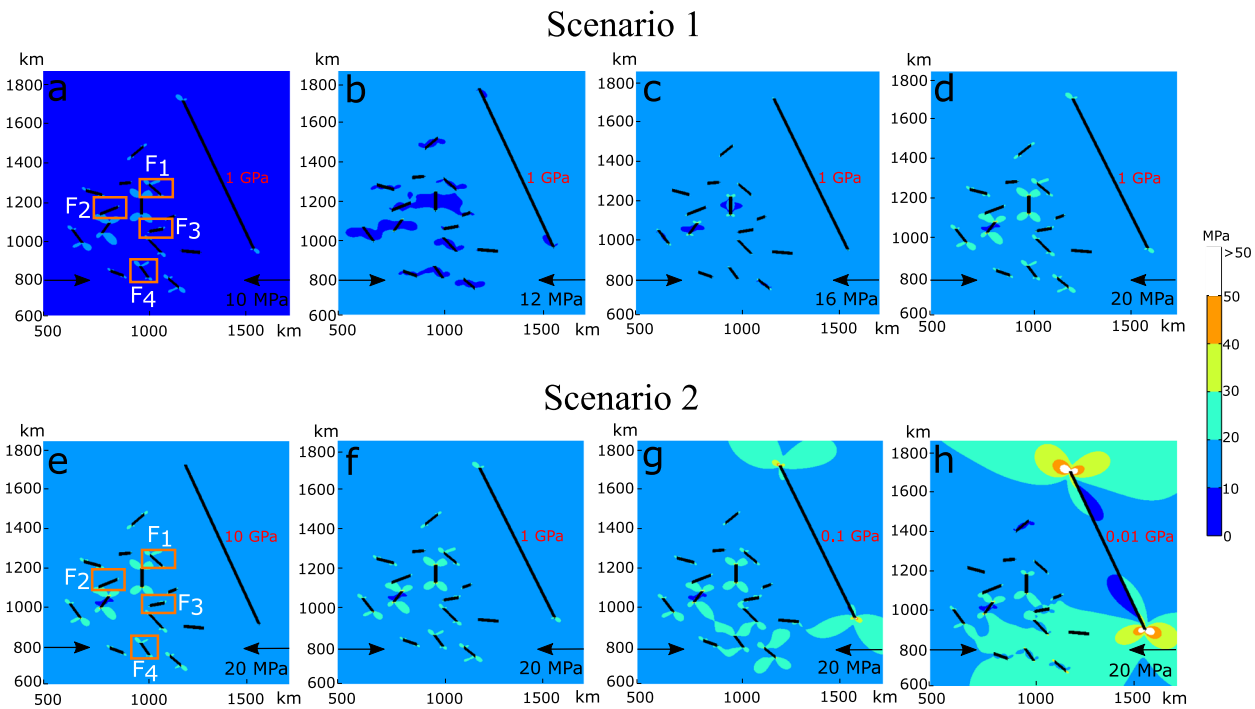

Fig. 8 | 2D FEM models of a faulted domain that gradually reaches the conditions for failure. **a**–**d** Progressive horizontal compression (from 10 to 20 MPa), assuming constant stiffness for the large fault segment (E_{LF} = 1 GPa) and for the small faults (E_{SF} = 0.03–0.92 GPa). **e**–**h** Constant horizontal compression (20 MPa) with variable large fault stiffness (E_{LF} = 0.01–10 GPa) and constant small faults stiffness (E_{SF} = 0.03–0.92 GPa). Models are developed with COMSOL Multiphysics (v6.2).

Studies on fault architecture in earthquake mechanics have shown that active faults concentrate stresses around them before failure[49–51]. Stress concentration, in turn, generates interconnected fractures in the damage zone, which increase fault permeability and porosity[47]. The latter induces the circulation of crustal fluids, which in turn increases the fluid pressure and enhances the aperture of other fractures in the damage zone[47]. This positive feedback also enhances a reduced cohesion and a friction decrease, which lead to a decreasing fault stiffness before failure[52,53]. For that reason, active faults can be characterized by higher fluid circulation and lower stiffness in the run-up to failure.

Below, we use the elastic inclusion fault modeling approach to explore the regional stress changes that may occur in the run-up to large-magnitude earthquakes, depending on fault stiffness and regional loading conditions. This exploration is performed by running a suite of finite element method (FEM) solid mechanics 2D models built with the COMSOL Multiphysics (v6.2) software (a detailed description of our FEM models' buildup is provided in Methods). In particular, we investigated how the existence of a large (850 km), locked fault (mimicking a fault segment responsible for large-magnitude seismicity; LF from now on) can affect the von Mises stress accumulation in an adjacent (~250 km away) network of small (40–97 km), randomly distributed, locked fault lines (mimicking fault segments responsible for low-magnitude seismicity) upon two geometrical settings and two end-member conditions that are expected to precede large-magnitude earthquakes (Fig. 8): (1) Progressive loading of the region with constant LF stiffness (0.01 GPa), for which we applied horizontal compression in the numerical domain from 10 MPa to 20 MPa, consistent with borehole data and field and geodetic analysis in California[54–56]; and (2) constant loading (20 MPa) in the region with a progressive decrease of the LF stiffness of up to four orders of magnitude, ranging from 10 GPa to 0.01 GPa. For this analysis, we assume constant stiffness for the small faults of the network for simplicity.

Under Scenario 1, the von Mises stress (further details in "Methods") linearly increases by a factor of two throughout the domain

https://doi.org/10.1038/s41467-024-51596-z

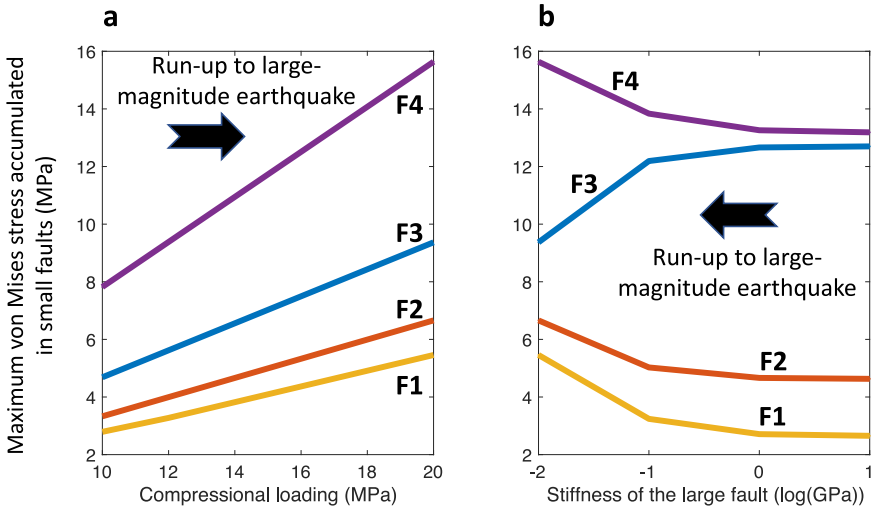

Fig. 9 | Results of our 2D FEM numerical models. a Maximum von Mises stress accumulated in four random small faults upon compressional loading. **b** Maximum von Mises stress accumulated in four random small faults for different values of the stiffness of the large fault segment. Black arrows show the expected direction of variation of the compressional loading and stiffness in the run-up to a large-magnitude earthquake. Models are performed with COMSOL Multiphysics v6.2.

(Fig. 8a–d). For instance, around Faults 1 to 4 of the network, the maximum von Mises stress rises from ~2.7 to ~5.5 MPa, from ~3.3 to ~6.7 MPa, from ~4.7 to ~9.4 MPa, and from ~7.8 to ~15.6 MPa, respectively (Fig. 9a). This indicates that as a large-magnitude earthquake approaches, compression propels all the small faults within the region toward failure. While this may elucidate a heightened rate of low-magnitude seismicity in the region, as reported prior to some large-magnitude earthquakes[33], it remains unclear how this mechanism could induce the precursory, nonlinear, statistical anomaly found in this study.

Conversely, when we impose a constant compression (20 MPa) but progressively decrease the LF stiffness (Scenario 2), a dissimilar and nonlinear concentration of stresses emerges in the region (Fig. 8e–h). Specifically, the reduction in LF stiffness leads to varying increases and decreases in the accumulated von Mises stress at different small faults of the network (Fig. 9b). For instance, the maximum von Mises stress around Fault 3 declines non-linearly from ~12.7 MPa to ~9.4 MPa as the LF stiffness drops by four orders of magnitude, from 10 GPa to 0.01 GPa. In contrast, the von Mises stress around Faults 1, 2, and 4 increases non-linearly by ~2.8 Mpa, ~2.1 MPa, and ~2.4 Mpa, respectively. The magnitude and sign of this variation are subject to: (i) The stiffness of the LF, with more significant stress changes occurring when the LF becomes very soft ($\lesssim$0.1 GPa); (ii) the arrangement of small faults within the network, as interactions between closely situated soft faults lead to more pronounced stress distribution changes; and (iii) the distance from the network to the LF, since greater stress accumulation occurs closer to the LF. In short, the run-up to large-magnitude earthquakes can lead certain small faults in the region closer to failure, while simultaneously pushing others further away from failure. This uneven behavior is a consequence of the softening of large regional faults, characterized by a reduction in stiffness due to heightened fluid circulation.

As part of our parametric analysis, our numerical study was extended to additional 2D and 3D geometrical settings to provide further insights on how the strike and dip of faults may affect the distribution of von Mises stresses (Supplementary Fig. 7). Our findings indicate that varying strike and dip angles influence fault interactions and the accumulation of von Mises stresses in the region; nevertheless, the models confirm the main conclusions drawn from the simpler 2D scenario outlined above (Supplementary Fig. 8): A decrease in LF stiffness leads to a regional, dissimilar concentration of stresses. It is worth noting that the different 2D and 3D scenarios analyzed, which incorporate various strike and dip configurations, simulate conditions expected for earthquakes with varied focal mechanisms (e.g., normal or thrust faults with dip <90° and strike-slip faults with dip = 90°).

In essence, our FEM-based numerical models show from a mechanical standpoint that irregular accumulation of stresses can appear in small fault networks when the stiffness of large faults declines—a circumstance expected as failure approaches[52,53]. Based on these findings, we conjecture that the emergence of anomalous low-magnitude seismicity before major earthquakes, as observed prior to the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake, could be ultimately driven by the reduction in stiffness of major faults rather than by gradual regional loading. The parametric analysis of our numerical simulations also reveals that the differential buildup of stresses in nearby minor faults becomes more significant when very low stiffness is applied to the major fault ($\lesssim$0.1 GPa), implying that abnormal, precursory low-magnitude seismicity can only arise when fault segments prone to major earthquakes are highly fractured and/or contain permeable cores (such as breccias and cataclastic rocks) susceptible to elevated pore fluid pressure[57,58]. It is worth noting that other mechanisms might also cause uneven stress accumulations in a fault network, such as dissimilar fluid transport in faults of the network; transitions from distributed deformation to progressive shear localization; stress transfer between faults; and fault friction, cohesion, strength, and weaknesses[32]. However, these reflect local processes and properties, and do not allow establishing an obvious cause-effect relationship between abnormal low-magnitude seismicity and the conditions leading up to the failure of a large fault segment. In other words, it remains uncertain that the aforementioned local processes and properties would translate into a regional phenomenon driven by the precursory conditions of a large fault segment that is close to failure and produce a major earthquake. In the future, geodetic data might provide further insights into the regional, dissimilar accumulation of stresses driven by decreasing stiffness in large faults, potentially enhancing our ability to anticipate large-magnitude earthquakes weeks to months in advance[38].

Forecasting potential of our machine learning-based approach

Earthquake forecasting involves anticipating the timing, magnitude, and/or epicenter of major earthquakes and/or aftershocks[35,39,40,59]. Here, we have delved into this topic by devising a method capable of

uncovering potential, hidden precursory patterns concealed within earthquake catalogs, with our primary focus being on the timing aspect. However, it is worth highlighting that our methodology also provides valuable insights into refining the size and location of major earthquakes; this contrasts with prior studies, which often concentrated on providing forecasts of one of the aforementioned observables only[35,39,40].

Specifically, our machine learning-based analysis reveals that the preparatory phase of the Ridgecrest and Anchorage earthquakes shared temporal similarities with previous seismic events in Southern California and Southcentral Alaska, allowing for the identification of precursory activity manifesting over timescales spanning from weeks to months. This timescale aligns with the preparatory phase reported for five large strike-slip earthquakes in Southern California since 1960, inferred from the combination of renormalized patterns related to increasing low-magnitude seismic activity and earthquake correlation range[30]. Our results, however, contrast to longer and shorter lead times reported in other studies on Southern California and elsewhere. For instance, Keilis-Borok et al.[60] documented low-magnitude seismicity changes occurring 2–3 years before strong earthquakes across various tectonic settings, suggesting a more protracted preparatory phase. Interestingly, Ben-Zion and Zaliapin[32] examined localization processes of low-magnitude seismicity and found evidence of rock damage accumulation around rupture zones for the 2019 Ridgecrest sequence and other large earthquakes in Southern and Baja California during the preceding decade. This was followed by seismic localization and event clustering around 2–3 years and 1–2 years, respectively, before the main shock. Huang et al.[61] found a foreshock sequence around 30 min before the initiation of the 2019 Ridgecrest sequence. These previous studies, combined with our findings, highlight a multifaceted process involving the accumulation and release of tectonic stresses over different timescales.

In addition, while our methodology considers magnitude as an input parameter, the inherent structure of our algorithm yields the probability of earthquakes surpassing a predefined magnitude threshold, which can be adjusted to suit specific requirements. This serves as the foundation for calculating probabilities of earthquakes of specific magnitude or greater. Moreover, although the precursory anomalies identified by our method hinder precise pinpointing of exact earthquake locations, we have noted that the epicenters of the two case studies fall within well-defined unrest areas spanning spatial scales ranging from tens to hundreds of kilometers. Indeed, this adheres to the well-established approach of examining seismic precursors within large regions, encompassing several hundred kilometers, to account for potential long-range correlations[62]; and aligns with prior research suggesting that the preparatory phase of major earthquakes involves broad-scale processes affecting fault networks, resulting in non-local seismic indicators[30,62,63]. In any case, while unrest areas are broad, our results suggest that they notably restrict the potential location of major events in the target regions. For instance, before the Ridgecrest sequence, there were no signs of potential large-magnitude earthquakes in other highly seismic zones, such as the San Francisco Bay Area, but the precursory anomaly was confined to locations surrounding the sequence's epicenters. Similarly, our method did not reveal any hints of potential large-magnitude earthquakes in the western fault zones of Southcentral Alaska before the occurrence of the Anchorage event. All these findings together indicate that our approach shows potential for improving seismic monitoring and surveillance by constraining the timing (over temporal spans from days to months), magnitude (surpassing predefined thresholds), and location (across spatial scales from tens to hundreds of kilometers) of forthcoming major earthquakes.

A major significance of our method lies in its potential to be integrated into operational earthquake monitoring systems and forecasting schemes[64], enabling the swift detection of tectonic unrest. This, in turn, could facilitate the rapid deployment of additional instruments to advance our understanding of the short-term preparatory phase of major earthquakes[65]. The early identification of tectonic unrest also holds the promise of bolstering societal preparedness for the potential occurrence of a major earthquake in a specific region, aiding in the assessment of alert levels. This aligns with decision-making strategies that focus on acquiring reliable signals indicating an increasing probability of an impending disaster, rather than assigning specific probability values[66]. Similar alert level systems are employed in volcano observatories to communicate volcanic activity[26,67].

While our algorithm has been tested on two specific regions, it has the potential for broader applicability. It can be readily applied to other regions globally, provided robust, highly complete, and long-lasting (~decades) earthquake catalogs become available. However, it is essential to emphasize that our algorithm should not be employed in new regions without training the random forest models with the historical seismicity of that specific area. Moreover, to maintain consistency and robustness, our algorithm should be periodically updated, involving re-training each time a major earthquake occurs in the region where it is being implemented. For instance, we recommend incorporating the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake in the training process moving forward.

As any other method, the algorithm presented in this paper also has limitations. In particular, it cannot assess the level of unrest in regions with minimal low-magnitude earthquake activity because the model statistical features cannot be reliably computed. In addition, our choice of Southern California and Southcentral Alaska for analysis was based on the availability of a sufficient number of earthquakes with magnitude M ≥6.4 over the past ~30 years (Supplementary Table 3). Finally, the success of our approach relies on the assumption that future seismicity will produce anomalies similar to those produced by the historical seismicity used for training. These limitations underscore the importance of ongoing research and earthquake forecasting adaptation as machine learning and monitoring techniques continue to evolve.

In summary, gaining insights into the multivariate temporal patterns of seismic activity leading up to major earthquakes is a vital stride in enhancing our capacity to anticipate devastating events. Our findings concerning the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake highlights the significance of identifying subtle, nonlinear anomaly trends in near-real time, and might serve as a foundation for crafting efficient operational forecasting algorithms[64]. This, in turn, plays a pivotal role in strengthening earthquake preparedness and thus in reducing the risks in communities susceptible to major seismic threats.

Closing remarks

Although machine learning has been successfully applied to anticipate laboratory earthquakes[37,38], the application of these methods to real earthquakes is more challenging, and their predictive power has been proposed not to be necessarily better than simpler and more traditional approaches[34]. However, this work shows that machine learning offers a lot of potential to classify low-magnitude seismicity, better identify precursory activity, and detect hidden patterns in the run-up to large-magnitude events. In this regard, we found that abnormal low-magnitude seismicity, or regional tectonic unrest, emerged in Southern California and Southcentral Alaska for weeks to months prior to the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake. This is identified with a new multivariate, random forest-based classification algorithm that detects when the statistical properties of low-magnitude earthquakes deviate from regular patterns. These deviations are challenging to identify by visual inspection of single variables, but they can be recognized with a nonlinear approach (random forest)

that considers the complex interplay between the different model variables. Nonlinear classification algorithms trained from previous experience are more efficient and versatile in detecting anomalous patterns of low-magnitude seismicity than classical methods consisting of tracking single variables (e.g., acceleration of seismicity rates[45]), whose temporal evolution may differ a lot from one case to another and thus results are difficult to generalize.

Our analysis provides observational evidence of the regional interaction between different fault zones during stress-loading to failure[68,69]; supports that different large-magnitude earthquakes have similar preparatory phases, at least for the two regions explored; and is consistent with previous laboratory experiments showing abnormal, large-scale signals that emerge over vast areas before failure[37,38]. Furthermore, our analysis and mechanistic finite element solid mechanics models support the long-held hypothesis that abnormal low-magnitude seismicity should appear over vast areas before large-magnitude earthquakes because largely locked fault sections cause alterations in the regional stress field[33,45]. In particular, our finite element solid mechanics models suggest that precursory variations in the regional stress field may occur because fault segments behave as elastic inclusions with dissimilar mechanical properties from those of the surrounding host crust, thus triggering an uneven response to the processes leading to major earthquakes and the accumulation of stresses both at local and regional scales[58].

Finally, it is worth highlighting that the machine learning-based approach presented in this paper only requires information that is currently being archived routinely in earthquake catalogs; could help to better understand the dynamics of fault networks and identify variations in the regional stress field; and can be easily implemented by surveillance agencies to monitor low-magnitude seismicity in near-real time. Eventually, our approach could help to design earthquake alert level strategies based on the detection of regional tectonic unrest, and to improve the forecast of large-magnitude earthquakes from weeks to months in advance in Southern California, Southcentral Alaska, and potentially elsewhere.

Methods

Machine learning-based model

Our machine learning-based algorithm is trained to determine whether the 2019 Ridgecrest seismic sequence and the 2018 Anchorage earthquake were preceded by anomalous low-magnitude seismicity (defined here as earthquakes with magnitude between $M_{Min} = 1$ and $M_{Max} = 6$). Training is performed with the low-magnitude seismicity that preceded other large seismic events (defined here as earthquakes with magnitude $M \geq 6.4$, i.e., the magnitude of the first event of the Ridgecrest sequence) that occurred in Southern California and Southcentral Alaska between January 1, 1989, and December 31, 2012. Our algorithm approach is based on two hypotheses: (i) The large-magnitude earthquakes used for training were preceded by tectonic unrest occurring for some weeks at least in the epicentral area. Hence, if the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake were preceded by analog tectonic unrest, this should emerge spontaneously when running our model on unseen data (from January 1, 2013, onwards). (ii) Anomalous tectonic unrest and quiescence can be recognized with a random forest classification model when monitoring the scattering or variability of different features of the low-magnitude seismicity, including the inter-event time, depth, magnitude, and latitude and longitude of the epicenters. Note that we utilize the standard deviation as a metric to represent the variability of the features, but it has no statistical meaning in this context. The analysis we undertake in this study consists of five main steps (data retrieval, algorithm setup, calculation of statistical features, labeling and training of the algorithm, and testing of the algorithm on unseen data) that are detailed in the next subsections.

Data retrieval

We retrieved data from the publicly available earthquake catalog of the United States Geological Survey (USGS) (https://earthquake.usgs.gov/earthquakes/search/). In particular, we downloaded all the information relative to the $M \geq 1.0$ earthquakes (on any magnitude scale) that occurred between January 1, 1989, and December 31, 2020, in Southern California (in the area delimited by [32.5°, 38°] latitude and [−124°, −112°] longitude) and in Southcentral Alaska (in the area delimited by [56°, 65°] latitude and [−159°, −143°] longitude). It is important to clarify that data between January 1, 1989, and December 31, 2012, were only used to train the algorithms. Also, we only retrieved $M \geq 1$ earthquakes to maximize the homogeneity and completeness of the catalog explored (e.g., by minimizing any possible bias produced by the increasing number of earthquakes detected with time due to the deployment of new seismic networks and better detection limits of modern seismic instruments). Finally, we extracted from the downloaded catalogs the date of occurrence, depth, and assigned magnitude of the earthquakes, as well as the latitude and longitude of their epicenters.

Algorithm setup

We set up our training algorithm by selecting a random distribution of N nodes with coordinates (i,j,t) within the spatiotemporal windows of analysis, where i, j, and t indicate latitude, longitude, and time, respectively. We also selected the nodes $(i_{LME}, j_{LME}, t_{LME})$, where i_{LME} and j_{LME} are the geographical coordinates of the epicenter of the large-magnitude earthquakes (i.e., earthquakes with magnitude above a given threshold; $M \geq M_{LME}$), and t_{LME} is their date of occurrence. For each node ((i, j, t) and $(i_{LME}, j_{LME}, t_{LME})$), we designed circular regions of radius R centered on the spatial coordinates ((i, j) and (i_{LME}, j_{LME})) and intervals of interest consisting of 1-year backward temporal windows sliding from $t - T$ and $t_{LME} - T$ years to t and t_{LME}, respectively, at time steps of 1 day (Supplementary Fig. 9). Note that a node (i, j, t) or $(i_{LME}, j_{LME}, t_{LME})$ is discarded if the circular region of interest does not contain low-magnitude seismicity. For this study, we chose $M_{LME} = 6.4$ (the magnitude of the first main shock of the Ridgecrest sequence), $T = 2$ years, $R = 120$ km, and only $N = 50$ nodes to maintain a reduced computational cost, although we also performed a parametric exploration to analyze the influence of several of the parameters (Supplementary Tables 1 and 2). For training purposes, we used four earthquakes of the Southern California region (M7.3 1992 Landers, M6.7 1994 Northridge, M7.1 1999 Hector Mine, and M6.6 2003 San Simeon earthquakes), seven earthquakes from Southcentral Alaska (M7.0 1999, M6.4 1999, M6.5 2000, and M6.9 2001 Kodiak Island region; M6.7 2001 Southern Alaska; and M6.6 2002 and M7.9 2002 Central Alaska earthquakes), and up to 50 random nodes distributed over the areas of analysis in Southern California and Southcentral Alaska (Supplementary Table 3). This algorithm step was repeated 200 times to generate different ensembles of up to $N = 50$ nodes, aiming to assess how the selection of the random nodes affects machine learning predictions.

Calculation of statistical features

Using the earthquake catalog downloaded, we calculated five statistical features from the low-magnitude seismicity (defined here as earthquakes with magnitude in the range $M = M_{Min} - M_{Max}$) that occurred within the circular regions and temporal sliding windows of the nodes (Supplementary Fig. 9). These statistical features are the standard deviation of the inter-event time (σ_{IET}), or time between low-magnitude earthquakes; and the standard deviation of their depth (σ_D), latitude (σ_{LAT}), longitude (σ_{LON}), and magnitude (σ_{MAG}). This approach yields T-year time series of σ_{IET}, σ_D, σ_{LAT}, σ_{LON}, and σ_{MAG} that are assigned to each node ((i, j, t) and $(i_{LME}, j_{LME}, t_{LME})$) of the spatiotemporal grid (Supplementary Figs. 10 and 11). Finally, the T-year time

series were standardized (i.e., mean-removed and divided by the standard deviation) to enhance the performance of machine learning-based classifications.

Labeling and training of the algorithm

We labeled the T-year time series in two classes: unrest (class 1) and quiescence (class 0). Unrest (class 1) was assigned to every time step of the T-year time series of the spatiotemporal nodes ($i_{LME}, j_{LME}, t_{LME}$) between $t_{LME} - T_{unrest}$ days and t_{LME}, i.e., over the T_{unrest} days preceding the large-magnitude earthquakes. In contrast, quiescence (class 0) was assigned to every time step of the time series of the nodes (i, j, t) between t − T years and t, and to every time step of the time series of the nodes ($i_{LME}, j_{LME}, t_{LME}$) between $t_{LME} - T$ years and $t_{LME} - T_{unrest}$ days (Supplementary Figs. 10 and 11). This means that unrest was assigned to the epicenter of the large-magnitude earthquakes during the T_{unrest} preceding days, and quiescence was assigned to the rest of spatiotemporal nodes. Analogue labeling strategies have been proposed for forecasting volcanic eruptions[70]. Then, we trained 200 random forest (RF) models, one for each of the ensembles of the aforementioned random nodes, to classify the time steps of the T-year time series as unrest (class 1) or quiescence (class 0) from the statistical features (σ_{IET}, σ_D, σ_{LAT}, σ_{LON}, σ_{MAG}). Each of our RF models contained 500 decision trees and was generated by randomly picking two features for each model iteration, whereas training and evaluation were performed through a tenfold cross-validation technique with three repeats. For example, using an unrest period of $T_{unrest} = 30$ days and a maximum low-magnitude seismicity of $M_{MAX} = 4$, we generated RF models for both Southern California and Southcentral Alaska that properly classified over 99.5% of the unrest and quiescence time steps of the training dataset. In particular, our models yield an increase from $P_{un} \approx 0$ to $P_{un} \approx 100\%$ within one month of the training events (Supplementary Fig. 12), where P_{un} is the degree or probability of tectonic unrest (defined as the ratio of decision trees classifying the statistical features as unrest). Note that P_{un} can be also interpreted as the probability of occurrence of a large-magnitude earthquake within the next T_{unrest} days.

It is worth noting that, since we employ 1-year backward windows to generate 2-year time series of statistical features, the first data point of the time series in Supplementary Figs. 10 and 11 reflects seismic events occurring up to three years before the occurrence of each of the large-magnitude earthquakes ($M \geq 6.4$) used for training. Our specific setup ensures that no events with magnitude $M \geq 6.4$ occurred within the defined regions of interest (R = 120 km) during the three years preceding the large-magnitude earthquakes used for training. Expanding the training period would elevate the chances of encountering additional large-magnitude earthquakes during the timeframe labeled as quiescence, which might result in incorrect labeling (quiescence vs unrest) and subsequent inadequate training of our machine learning model. Labeling also requires special attention if enlarging the regions of interest and lowering the magnitude threshold for defining large-magnitude earthquakes.

Testing of the algorithm on unseen data

We run our machine learning models (trained using $T_{unrest} = 30$ days) on unseen data, i.e., the low-magnitude seismicity ($M = M_{Min} - M_{Max}$, with $M_{Min} = 1$ and $M_{Max} = 6$) occurred from January 1, 2013, onwards. To do this, we first calculated the T-year time series of statistical features (σ_{IET}, σ_D, σ_{LAT}, σ_{LON}, σ_{MAG}) for the spatiotemporal coordinates ($i_{tar}, j_{tar}, t_{tar}$), where i_{tar} and j_{tar} are the geographical coordinates of the target location (e.g., epicenters of the first main shock of the 2019 Ridgecrest sequence and the 2018 Anchorage earthquake, or any other location) and t_{tar} is the target prediction date (e.g., 365 days, 90 days, or 1 day before the occurrence of the earthquakes). The statistical features for each of the T-year time series were calculated following the same approach described above, and we also used the same hyperparameters utilized to train the RF models (T = 2 years, R = 120 km, 1-year backward sliding windows, and 1-day time steps). Finally, we run the 200 RF models ensemble over the standardized T-year time series of statistical features and retrieved the minimum, maximum, and mean values of P_{un}. To successfully mimic the appropriate use of our method during near-real-time monitoring, we saved only the prediction of P_{un} for the last time step of the T-year time series (i.e., one day before the target prediction date, t_{tar}), and repeated the process from $t_{tar} = 1000$ days before to $t_{tar} = 400$ days after the Ridgecrest sequence and the Anchorage earthquake. This approach also ensures that the temporal evolution of P_{un} (Figs. 2–5) does not contain artifacts produced by the standardization of the T-year time series, and corresponds to the day-to-day prediction of P_{un} at the target locations.

Finite element solid mechanics model

Our 2D finite element solid mechanics models aim to explore regional stress changes occurring in the run-up to large-magnitude earthquakes. We built these models with COMSOL Multiphysics (v6.2), solving the stationary momentum equation and applying the elastic inclusion fault modeling approach[58].

In particular, we designed a 2000 × 2000 km domain and imported fifteen small fault segments with 40–97 km length and one large fault (LF) segment with 850 km length as rectangular features. Furthermore, we designed two settings (different geometrical configurations) in 2D to investigate whether the orientation (strike) and geometry of the small fault segments (crosscuts or misaligned/non-planar faults) affect the stress distribution and accumulation produced by the large fault segment. That said, we modeled mainly two scenarios: (1) progressive loading (horizontal compression), and (2) constant loading with a progressive decrease of the large fault stiffness. In all runs the top and the bottom of the domain are free surfaces, and we calculated the Von Mises stress at the fault segments. Finally, these models were further explored in 3D to investigate how the strike and dip of the small faults influence the accumulation of stresses in the domain (Supplementary Figs. 7 and 8).

We assigned material properties to the host rock and the faults based on the common crustal values used in the literature, which reflect a linear elastic behavior. The host rock was modeled as a stiff homogenized domain with Young's modulus of 30 GPa, whereas the LF was given values from 10 GPa (stiff) to 0.01 GPa (soft) to mimic the progression from a less active fault segment to a more active one. Soft values represent aseismic creep or locked conditions of the fault segment, whereas stiff values reflect the conditions of a not recently active fault in the region[67]. The stiffness of the small faults is initally defined as random values between 0.03 to 0.92 GPa, and this value is kept constant for all the simulations. The host rock and all the faults have Poisson's ratios equal to 0.25 and density values of 2300 kg/m^3. Regarding boundary conditions, we impose a compressional stress regime between 10 MPa and 20 MPa, which mimics the realistic regional stress conditions for Southern California and Southcentral Alaska[54,55].

Data availability

The data used in this study are available in the earthquake catalog database of the United States Geological Survey (USGS) [https://earthquake.usgs.gov/earthquakes/search/]. The specific dataset used for training and testing can be downloaded from ref. 71 https://doi.org/10.5281/zenodo.13212238.

Code availability

The R script used for data processing and analysis; the Matlab script used to create Figs. 3, 4, and Supplementary Movies 1 and 2; the machine learning-based models generated; the specific dataset used

for training and testing; and a description of the files generated can be downloaded from ref. 71 https://doi.org/10.5281/zenodo.13212238.

References

1. Turcotte, D. L. Earthquake prediction. *Annu. Rev. Earth Planet. Sci.* https://doi.org/10.1146/annurev.ea.19.050191.001403 (1991).
2. Rundle, J. et al. The complex dynamics of earthquake fault systems: new approaches to forecasting and nowcasting of earthquakes. *Rep. Prog. Phys.* https://doi.org/10.1088/1361-6633/abf893 (2021).
3. Pei, W., Zhou, S., Zhuang, J., Xiong, Z. & Piao, J. Application and discussion of statistical seismology in probabilistic seismic hazard assessment studies. *Sci. China Earth Sci.* **65**, 1–12 (2021).
4. Kanamori, H. Earthquake prediction: an overview. In *International Handbook of Earthquake & Engineering Seismology* (eds Lee, W. H. K. et al.) 1205–1216 (Academic Press, Amsterdam, 2003).
5. Mogi, K. *Earthquake Prediction* (Academic Press, 1985).
6. Lomnitz, C. *Fundamentals of Earthquake Prediction* (John Wiley, 1994).
7. Ouzounov, D., Pulinets, S., Hattori, K. & Taylor, P. *Pre-earthquake Processes: A Multidisciplinary Approach to Earthquake Prediction Studies* Vol. 234 (John Wiley & Sons, 2018).
8. Conti, L., Picozza, P. & Sotgiu, A. A critical review of ground based observations of earthquake precursors. *Front. Earth Sci.* **9**, 676766 (2021).
9. Tramutoli, V. et al. Assessing the potential of thermal infrared satellite surveys for monitoring seismically active areas: The case of Kocaeli (Izmit) earthquake, August 17, 1999. *Remote Sens. Environ.* **96**, 409–426 (2005).
10. Pulinets, S. A., Ouzounov, D., Karelin, A. V., Boyarchuk, K. A. & Pokhmelnykh, L. A. The physical nature of thermal anomalies observed before strong earthquakes. *Phys. Chem. Earth* https://doi.org/10.1016/j.pce.2006.02.042 (2006).
11. Lu, X. et al. Thermal infrared anomalies associated with multi-year earthquakes in the Tibet region based on China's FY-2E satellite data. *Adv. Space Res.* **58**, 989–1001 (2016).
12. Dolgikh, G. I. et al. Deformation and acoustic precursors of earthquakes. *Dokl. Earth Sci.* **413**, 281–285 (2007).
13. Hayakawa, M. et al. On the precursory signature of Kobe earthquake on VLF subionospheric signals. In *IEEE International Symposium on Electromagnetic Compatibility*. (Institute of Electrical Engineers INC, 1997).
14. Koizumi, N. et al. Preseismic changes in groundwater level and volumetric strain associated with earthquake swarms off the east coast of the Izu Peninsula, Japan. *Geophys. Res. Lett.* **26**, 3509–3512 (1999).
15. Khilyuk, L. F., Robertson Jr, J. O., Endres, B. & Chilingarian, G. V. *Gas Migration: Events Preceding Earthquakes* (Elsevier, 2000).
16. Hwa Oh, Y. & Kim, G. A radon-thoron isotope pair as a reliable earthquake precursor. *Sci. Rep.* https://doi.org/10.1038/srep13084 (2015).
17. Pulinets, S. A. & Boyarchuk, K. A. *Ionospheric Precursors of Earthquakes* (Springer, Berlin, 2004).
18. Uyeda, S., Nagao, T. & Kamogawa, M. Short-term earthquake prediction: current status of seismo-electromagnetics. *Tectonophysics* **470**, 205–213 (2009).
19. Sorokin, V. M., Chmyrev, V. M. & Hayakawa, M. A review on electrodynamic influence of atmospheric processes to the ionosphere. *Open J. Earthq. Res.* **9**, 113–141 (2020).
20. Sun, Y., Niu, F., Liu, H., Chen, Y. & Liu, J. Crustal structure and deformation of the SE Tibetan plateau revealed by receiver function data. *Earth Planet. Sci. Lett.* **349**, 186–197 (2012).
21. Ito, Y. et al. Episodic slow slip events in the Japan subduction zone before the 2011 Tohoku-Oki earthquake. *Tectonophysics* **600**, 14–2 (2013).
22. Ruiz, S. et al. Intense foreshocks and a slow slip event preceded the 2014 Iquique Mw 8.1 earthquake. *Science* **345**, 1165–1169 (2014).
23. Wikelski, M. et al. Potential short-term earthquake forecasting by farm animal monitoring. *Ethology* **9**, 931–941 (2020).
24. Geller, R. J., Jackson, D. D., Kagan, Y. Y. & Mulargia, F. Earthquakes cannot be predicted. *Science* **5306**, 1616–1616 (1997).
25. Zöller, G., Hainzl, S., Tilmann, F., Woith, H. & Dahm, T. Comment on "Potential short-term earthquake forecasting by farm animal monitoring" by Wikelski, Mueller, Scocco, Catorci, Desinov, Belyaev, Keim, Pohlmeier, Fechteler, and Mai. *Ethology* **127**, 302–306 (2021).
26. Winson, A. E., Costa, F., Newhall, C. G. & Woo, G. An analysis of the issuance of volcanic alert levels during volcanic crises. *J. Appl. Volcanol.* **3**, 1–12 (2014).
27. Smith, W. D. The b-value as an earthquake precursor. *Nature* **289.5794**, 136–139 (1981).
28. El-Isa, Z. H. & Eaton, D. W. Spatiotemporal variations in the b-value of earthquake magnitude–frequency distributions: classification and causes. *Tectonophysics* **615**, 1–11 (2014).
29. Peng, C. et al. Performance evaluation of an earthquake early warning system in the 2019–2020 M 6.0 Changning, Sichuan, China, Seismic Sequence. *Front. Earth Sci.* **9**, 699941 (2021).
30. Keilis-Borok, V. I., Shebalin, P. N. & Zaliapin, I. V. Premonitory patterns of seismicity months before a large earthquake: five case histories in Southern California. *Proc. Natl. Acad. Sci. USA* https://doi.org/10.1073/pnas.202617199 (2002).
31. Marzochhi, W., Taroni, M. & Falcone, G. Earthquake forecasting during the complex Amatrice-Norcia seismic sequence. *Sci. Adv.* https://doi.org/10.1126/sciadv.1701239 (2017).
32. Ben-Zion, Y. & Zaliapin, I. Localization and coalescence of seismicity before large earthquakes. *Geophys. J. Int.* **223**, 561–583 (2020).
33. Jaumé, S. & Sykes, L. Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes. *Pure appl. geophys.* **155**, 279–305 (1999).
34. Mignan, A. & Broccardo, M. Neural network applications in earthquake prediction (1994–2019): meta-Analytic and statistical insights on their limitations. *Seismol. Res. Lett.* https://doi.org/10.1785/0220200021 (2020).
35. Asencio–Cortés, G., Morales–Esteban, A., Shang, X. & Martínez–Álvarez, F. Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. *Comput. Geosci.* https://doi.org/10.1016/j.cageo.2017.10.011 (2018).
36. Bergen, K. J., Johnson, P. A., de Hoop, M. V. & Beroza, G. C. Machine learning for data-driven discovery in solid earth geoscience. *Science* https://doi.org/10.1126/science.aau0323 (2019).
37. Rouet-Leduc, B. et al. Machine Learning Predicts Laboratory Earthquakes. *Geophys. Res. Lett.* https://doi.org/10.1002/2017GL074677 (2017).
38. Corbi, F. et al. Machine learning can predict the timing and size of analog earthquakes. *Geophys. Res. Lett.* https://doi.org/10.1029/2018GL081251 (2019).
39. Panakkat, A. & Adeli, H. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *Int. J. Neural. Syst.* https://doi.org/10.1142/S0129065707000890 (2007).
40. DeVries, P. M. R., Viégas, F., Wattenberg, M. & Meade, B. J. Deep learning of aftershock patterns following large earthquakes. *Nature* https://doi.org/10.1038/s41586-018-0438-y (2018).
41. Barnhart, W. D., Hayes, G. P. & Gold, R. D. The July 2019 Ridgecrest, California, earthquake sequence: Kinematics of slip and stressing in cross-fault ruptures. *Geophys. Res. Lett.* https://doi.org/10.1029/2019GL084741 (2019).

42. West, M. E. et al. The 30 November 2018 Mw 7.1 Anchorage Earthquake. *Seismol. Res. Lett.* **91**, 66–84 (2019).
43. Zhuang, J., Ogata, Y. & Wang, T. Data completeness of the Kumamoto earthquake sequence in the JMA catalog and its influence on the estimation of the ETAS parameters. *Earth Planets Space* **69**, 36 (2017).
44. Tape, C. et al. Earthquake nucleation and fault slip complexity in the lower crust of central Alaska. *Nat. Geosci.* **11**, 536–541 (2018).
45. Bowman, D. D. & King, G. C. P. Accelerating seismicity and stress accumulation before large earthquakes. *Geophys. Res. Lett.* https://doi.org/10.1029/2001GL013022 (2001).
46. Rowe, C. D. & Griffith, W. A. Do faults preserve a record of seismic slip: a second opinion. *J. Struct. Geol.* **78**, 1–26 (2015).
47. Gudmundsson, A. *Volcanotectonics: Understanding the Structure, Deformation and Dynamics of Volcanoes* (Cambridge University Press, 2020).
48. Caine, J. S., Evans, J. P. & Forster, C. B. Fault zone architecture and permeability structure. *Geology* **11**, 1025–1028 (1996).
49. Dieterich, J. H. Preseismic fault slip and earthquake prediction. *J. Geophys. Res. Solid Earth* **83**, 3940–3948 (1978).
50. Sibson, R. H. Structural permeability of fluid-driven fault-fracture meshes. *J. Struct. Geol.* **18**, 1031–1042 (1996).
51. Harris, R. A. Introduction to special section: stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res. Solid Earth* **103**, 24347–24358 (1998).
52. Yoshioka, N. & Iwasa, K. A laboratory experiment to monitor the contact state of a fault by transmission waves. *Tectonophysics* **413**, 221–238 (2006).
53. Warren-Smith, E. et al. Episodic stress and fluid pressure cycling in subducting oceanic crust during slow slip. *Nat. Geosci.* **12**, 475–481 (2019).
54. Zoback, M. D. et al. New evidence on the state of stress of the San Andreas fault system. *Science* **238**, 1105–1111 (1987).
55. Lund Snee, J. E. & Zoback, M. D. Multiscale variations of the crustal stress field throughout North America. *Nat. Commun.* **11**, 1951 (2020).
56. Fialko, Y. Interseismic strain accumulation and the earthquake potential on the southern San Andreas fault system. *Nature* **441**, 968–971 (2006).
57. McKenzie, D. & Brune, J. N. Melting on fault planes during large earthquakes. *Geophys. J. Int.* **29**, 65–78 (1972).
58. Drymoni, K., Browning, J. & Gudmundsson, A. Volcanotectonic interactions between inclined sheets, dykes, and faults at the Santorini Volcano, Greece. *J. Volcanol. Geothermal Res.* **416**, 107294 (2021).
59. Field, E. H. & Milner, K. R. Candidate products for operational earthquake forecasting illustrated using the HayWired planning scenario, including one very quick (and not-so-dirty) hazard-map option. *Seismol. Res. Lett.* **89**, 1420–1434 (2018).
60. Keilis-Borok, V. I. & Rotwain, I. M. Diagnosis of time of increased probability of strong earthquakes in different regions of the world: algorithm CN. *Phys. Earth Planet. Inter.* **61**, 57–72 (1990).
61. Huang, H., Meng, L., Bürgmann, R., Wang, W. & Wang, K. Spatiotemporal foreshock evolution of the 2019 M 6.4 and M 7.1 Ridgecrest, California earthquakes. *Earth Planet. Sci. Lett.* **551**, 116582 (2020).
62. Peresan, A., Kossobokov, V., Romashkova, L. & Panza, G. F. Intermediate-term middle-range earthquake predictions in Italy: a review. *Earth Sci. Rev.* **69**, 97–132 (2005).
63. Dobrovolsky, I. P., Zubkov, S. I. & Miachkin, V. I. Estimation of the size of earthquake preparation zones. *Pure Appl. Geophys.* **117**, 1025–1044 (1979).
64. Peresan, A., Kossobokov, V. G. & Panza, G. F. Operational earthquake forecast/prediction. *Rend. Lincei* **23**, 131–138 (2012).
65. Sugan, M., Kato, A., Miyake, H., Nakagawa, S. & Vuan, A. The preparatory phase of the 2009 Mw 6.3 L'Aquila earthquake by improving the detection capability of low-magnitude foreshocks. *Geophys. Res. Lett.* **41**, 6137–6144 (2014).
66. Iturrieta, P. et al. Evaluation of a Decade-Long Prospective Earthquake Forecasting Experiment in Italy. *Seismol. Res. Lett.* https://doi.org/10.1785/0220230247 (2024).
67. Cameron, C. E. et al. Alaska volcano observatory alert and forecasting timeliness: 1989–2017. *Front. Earth Sci.* **6**, 86 (2018).
68. Rundle, P. B. et al. Nonlinear network dynamics on earthquake fault systems. *Phys. Rev. Lett.* https://doi.org/10.1103/PhysRevLett.87.148501 (2001).
69. Dolan, J. F., Bowman, D. D. & Sammis, C. G. Long range and long-term fault interactions in Southern California. *Geology* https://doi.org/10.1130/G23789A.1 (2007).
70. Dempsey, D. E., Cronin, S. J., Mei, S. & Kempa-Liehr, A. W. Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand. *Nat. Commun.* **11**, 3562 (2020).
71. Girona, T. & Drymoni, K. Abnomal low-magnitude seismicity preceding large-magnitude earthquakes -- dataset, scripts, machine learning models, and results. (1.0). *Zenodo* https://doi.org/10.5281/zenodo.13212238 (2024).

Acknowledgements

Earthquake catalog courtesy of the U.S. Geological Survey. This research project and publication was supported by the University of Alaska Fairbanks Geophysical Institute. K.D. acknowledges financial support from the European Research Council (ERC) Consolidator Grant MODERATE (10100165). T.G. appreciates discussions on the matter of this paper with Vicente Girona Sánchez, María del Rosario Hernández Gil, Warner Marzocchi, and Patricia Martínez Garzón.

Author contributions

T.G. conceived the project and led the design and implementation of the machine learning-based algorithm. K.D. led the FEM numerical modeling. T.G. and K.D. wrote the manuscript, designed figures, and discussed the models, results, and conclusions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51596-z.

Correspondence and requests for materials should be addressed to Társilo Girona.

Peer review information *Nature Communications* thanks Miao Zhang and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024