scientific reports

Check for updates

OPEN

Sharpen data-driven prediction rules of individual large earthquakes with aid of Fourier and Gauss

In Ho Cho

Predicting individual large earthquakes (EQs)' locations, magnitudes, and timing remains unreachable. The author's prior study shows that individual large EQs have unique signatures obtained from multi-layered data transformations. Via spatio-temporal convolutions, decades-long EQ catalog data are transformed into pseudo-physics quantities (e.g., energy, power, vorticity, and Laplacian), which turn into surface-like information via Gauss curvatures. Using these new features, a rule-learning machine learning approach unravels promising prediction rules. This paper suggests further data transformation via Fourier transformation (FT). Results show that FT-based new feature can help sharpen the prediction rules. Feasibility tests of large EQs ($M \geq 6.5$) over the past 40 years in the western U.S. show promise, shedding light on data-driven prediction of individual large EQs. The handshake among ML methods, Fourier, and Gauss may help answer the long-standing enigma of seismogenesis.

Large earthquakes (EQs) remain one of the most difficult physics phenomena, attracting the newest technologies. Recently, researchers actively leverage machine learning (ML) methods[1,2]. For instance[3], adopted deep neural networks (DNNs) to forecast aftershock locations without relying upon fault orientation. Their DNNs take co-seismically generated static elastic tensor's change as input and produce binary prediction of whether each refined grid cubic cell (5 km each dimension) will contain aftershocks or not[4]. Reconstructed the time series data in EQ catalog of Southern California to a sequence of two-dimensional (2D) images, and they combined an autoencoder and temporal convolutional neural networks to the new data to predict the probability of extreme events. Reference[5] adopts the long short-term memory networks to learn EQ's spatio-temporal relationship. At the same time, deep learning helps to engender denser and deeper data sets of EQs[6,7], which can enable unsupervised deep learning-driven exploration and discovery of hitherto unseen behaviors and patterns of EQ[8]. By regarding EQs as spatio-temporal point process, a combination of reinforcement learning and neural networks is used[9] for enabling data-driven fitting and learning of heterogeneous Gaussian diffusion kernels to improve predictions of the point processes. Still, these ML-based approaches are at the burgeoning phase and require systematic validation and comparison against existing approaches. Reference[10] conducted a comprehensive comparative study on neural networks-based EQ forecasting and prediction methods of the past three decades and found that these new methods call for broader and systematic validations since simple empirical methods may exhibit equivalent or even better performance.

Reference[11] gives a comprehensive overview of recent physics-based EQ forecasting methods, for which a well-established performance evaluation framework is available by community[12,13]. Diverse ML methods play an important role in understanding EQs' complex behaviors and patterns, e.g., Refs.[14–16].

Despite the notable advances empowered by ML-driven approaches, the long-sought capability of predicting "individual" large EQs' locations and magnitudes within a short time frame (days or weeks ahead) remains unreachable. Th s study seeks to add a new dimension to this daunting question. The author's prior study[17] shows that, after multi-layered data transformations, individual large EQs appear to have unique signatures that can be represented by new high-dimensional features. In particular, the observed EQ catalog data are transformed via spatio-temporal convolution, and then further transformed into a number of pseudo physics quantities (i.e., energy, power, vorticity, and Laplacian). They later turn into smooth surface-like information via Gauss curvatures, giving rise to new high-dimensional features. The new features of pseudo physics quantities are used to build a customized prediction model by the Bayesian evolutionary algorithm in conjunction with flexible base

CCEE Department, Iowa State University, Ames, IA 50011, USA. email: icho@iastate.edu

functions. Validations with the past 40-year EQs catalog data of the western U.S. region show that the Gauss curvature-based coordinates appear to hold uniqueness for individual large EQs ($M_w \geq 7.0$) demonstrating a promising reproduction of individual large EQs' locations and magnitudes 30 days before the event.

Here, this study expanded the study region from (longitude, latitude, depth) = $[-130°, -110°] \cup [30°, 45°] \cup [-5\,\text{km}, 20\,\text{km}]$ to $[-132.5°, -110°] \cup [30°, 52.5°] \cup [-5\,\text{km}, 20\,\text{km}]$ and the magnitude range from $M_w \geq 7.0$ to $M_w \geq 6.5$. As a result, the total number of target large EQs within the 40 years (1980 through 2019) increases from 8 to 17. Importantly, this study adds further sophistication to previous multi-layered data transformations via Fourier transformation of the Gauss curvature-based features.

The key equations and formulas of the proposed approach are summarized in Table S1 of Supplementary Information.

Results

Multi-layered data transformation

The critical novelty of the proposed approach lies in the multi-layered data transformation to generate physics-infused ML-friendly new features. The overall architecture of data transformation is summarized in Fig. 1

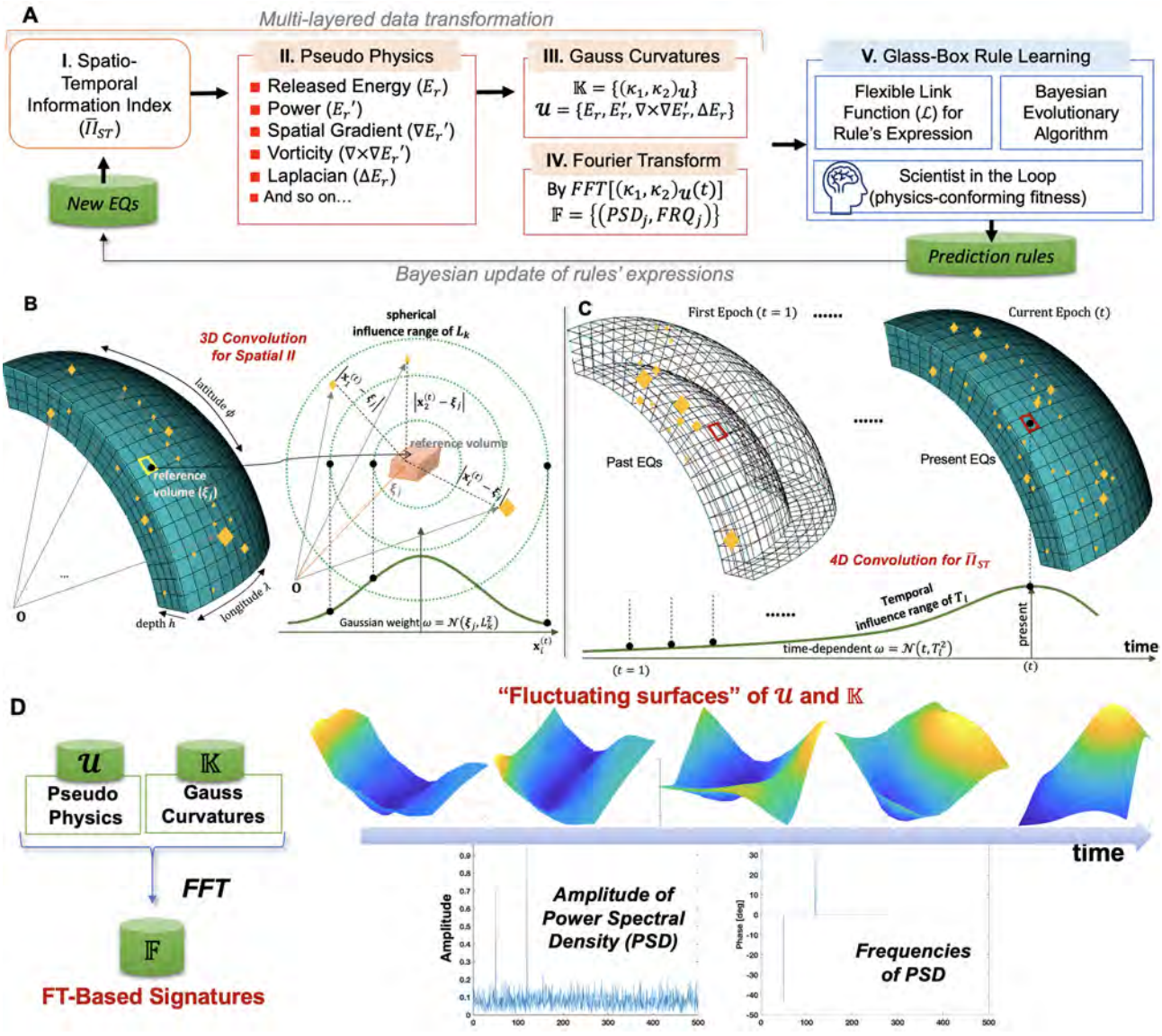

Figure 1. Overall architecture and conceptual illustrations of central steps of the proposed approach: (**A**) Multi-layered data transformation from raw EQs catalog data to new features in terms of pseudo physics quantities, Gauss curvatures, and Fourier bases (I–IV). Transparent rule-learning machine learning (ML) method, denoted "glass-box" ML (V), to unravel prediction rules' expressions of individual large EQs; (**B**) 3D convolution for generating spatial information index; (**C**) Temporal convolution for spatio-temporal information index; (**D**) Fast Fourier transform (FT) to generate FT-based new feature that can quantify time-varying information about "fluctuating surfaces" of the pseudo physics and Gauss curvature-based features.

The fi st data transformation (Figs. 1B,C) converts the raw EQ catalog data (USGS[18]) into new scalar features, denoted as spatio-temporal information index ($\overline{II}_{ST}$). The reference volume is defined as a discretized volume of the Earth lithosphere with increment of (longitude, latitude, depth), $(\Delta\lambda, \Delta\phi, \Delta h) = (0.1°, 0.1°, 5\,\text{km})$. For the convolution process, the geodetic coordinates, $(\lambda, \phi, -h)_i^{(t)}$ are transformed into the earth-centered rectilinear coordinate $(x, y, z)_i^{(t)}$[17]. Each observed EQ's moment magnitude $M_i^{(t)} \in \mathbb{R}[0, 10)$ are assumed to reside at $\mathbf{x}_i^{(t)} = (x, y, z)_i^{(t)}$, following the "point source" concept. Physically, $\overline{II}_{ST}(\boldsymbol{\xi}_j)^{(t)}$ quantifies the accumulated influences of the adjacent EQs close to the jth reference volume center and of the past EQs up to the present time (t). The details of the data transformation from raw EQ catalog data into spatio-temporal information index generation are presented in "Methods".

The second data transformation is to convert the spatio-temporal IIs into the pseudo physics quantities. Amongst many physics quantities, the best-so-far set of pseudo physics quantities are identifi d as {released energy, power, vorticity, Laplacian}[17]. To be purely data-driven, no pre-defined statistical or empirical laws are used. Instead, a flexible function, called "link function (LF)", is used to learn expressions of the pseudo physics quantities. The best-so-far form of the pseudo released energy $E_r^{(t)}(\boldsymbol{\xi}_j)$ is identified as

$$E_r^{(t)}(\boldsymbol{\xi}_j) = \sum_{k=1}^{n_L=2} \sum_{l=1}^{n_T=2} \mathcal{L}^{(k,l)}(\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l);\ \boldsymbol{\theta}^{(k,l)}), \tag{1}$$

where $\boldsymbol{\theta}^{(k,l)}$ is the best-so-far free parameters of the associated LF $\mathcal{L}^{(k,l)}$. Here, $\mathcal{L}^{(k,l)}$ takes $\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l)$ as input and produces a smooth, nonlinear output.

$L_k(k = 1, \ldots, n_L)$ is the spatial influence range that means the spatial proximity-dependent importance in the spatial convolution process. Similarly, $T_l(l = 1, \ldots, n_T)$ is the temporal influence range meaning the temporal proximity-dependent importance in the temporal convolution process (see "Methods" for details).

Any mathematical form can be used as LF, and a simple yet general exponential form LF works well[17], i.e., for the pseudo released energy $\mathcal{L}^{(k,l)}(\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l)\ \boldsymbol{\theta}^{(k,l)}) = \exp\left(a^{(k,l)}\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l)^{b^{(k,l)}}\right) - 1$ where $\boldsymbol{\theta}^{(k,l)} = \{a^{(k,l)}, b^{(k,l)}; k = 1, \ldots, n_L, l = 1, \ldots, n_T\}$. The pseudo "vorticity" $\boldsymbol{\omega} = (\omega_\lambda, \omega_\phi, \omega_h)$ is generated by $\boldsymbol{\omega} := \nabla_g \times \left(\nabla_g \frac{\partial E_r^{(t)}(\boldsymbol{\xi}_j)}{\partial t}\right)$ where $\frac{\partial E_r^{(t)}(\boldsymbol{\xi}_j)}{\partial t}$ corresponds to the pseudo "power" and the pseudo "Laplacian" is calculated as $\nabla_g^2 E_r^{(t)}(\boldsymbol{\xi}_j) = \frac{\partial^2 E_r'}{\partial \lambda^2} + \frac{\partial^2 E_r'}{\partial \phi^2} + \frac{\partial^2 E_r'}{\partial h^2}$, where $\nabla_g(.)$ means the spatial gradient with respect to the geodetic coordinate system (λ, ϕ, h). As shown in Ref.[17], amongst many pseudo physics quantities and their combinations, ML selected out the four quantities—the released energy, power, the fi st vorticity term, and the first Laplacian term, $(E_r^{(t)}, \frac{\partial E_r^{(t)}(\boldsymbol{\xi}_j)}{\partial t}, \omega_\lambda, \frac{\partial^2 E_r'}{\partial \lambda^2})$, at least for the western U.S. region. Again, this selection is purely data-driven since ML simply seeks to find the best combination that can outperform other cases without any prejudice. The third data transformation is to convert the pseudo physics quantities into Gauss curvatures. At each depth, the distributions of the pseudo physics quantities constitute smooth yet complex surfaces. To effectively inform ML with surface-like information, the next data transformation focuses on Gauss curvatures[19]—consisting of two principal curvatures κ_1 and κ_2 (detailed calculation procedures are presented in Ref.[17]). Using the Gauss curvatures near EQs, it is easy to quantify the distributions' shapes of the pseudo physics quantities. In Ref.[17], these Gauss curvature-based coordinates may serve as a unique signature of individual extreme EQs. The coordinate vector K consists of the principal Gauss curvatures (κ_1, κ_2) of four pseudo physics quantities at time t at a reference volume $\boldsymbol{\xi}_j$ as

$$K(t; \boldsymbol{\xi}_j) := ((\kappa_1, \kappa_2)_E, (\kappa_1, \kappa_2)_P, (\kappa_1, \kappa_2)_V, (\kappa_1, \kappa_2)_L)), \tag{2}$$

where E, P, V and L stand for the pseudo released energy, the pseudo power, the pseudo vorticity's fi st term, and the pseudo Laplacian's fi st term, respectively, all being calculated at time t and the reference volume $\boldsymbol{\xi}_j$.

The fourth data transformation is to convert the time histories of Gauss curvatures into Fourier transform (FT)-based features (Fig. 1D). The inclusion of the FT-based new features has two reasons. First, FT-based features can easily convey temporal information of other features in terms of amplitudes and frequencies. The second reason is to leverage the strength of the Fourier series in representing general, complex functions. As shall be demonstrated in this paper, the inclusion of trigonometric functions (inspired by the Fourier series) in the prediction rules appears to improve and sharpen the prediction accuracy substantially.

Table 1 summarizes the key procedures of FT-based new feature generation using the time history of Gauss curvatures of pseudo physics quantities. The set of Gauss curvature-based coordinates at $\boldsymbol{\xi}_j$ up to the present time $t_n = n \times \Delta t$ is given by

$$\mathbb{K}(n; \boldsymbol{\xi}_j) := \{K(t; \xi_j) \in \mathbb{R}^8 |\ t = 1, \ldots, t_n\} \in \mathbb{R}^{(n\times 8)} \tag{3}$$

where Δt is the sampling interval, one day in this paper.

Regarding $\mathbb{K}(n; \boldsymbol{\xi}_j) \in \mathbb{R}^{(n\times 8)}$ as a matrix, the m_{th} column, denoted as $\mathbf{k}^{(m)}(n; \boldsymbol{\xi}_j)\,(m = 1, \ldots, 8)$, corresponds to the time series of a principal Gauss curvature of a pseudo physics quantity. For instance, $\mathbf{k}^{(1)}(n; \boldsymbol{\xi}_j)$ the 1st column of $\mathbb{K}(n; \boldsymbol{\xi}_j)$, means the time series of κ_1 of the pseudo released energy up to this time t_n at a reference volume $\boldsymbol{\xi}_j$ whereas the 8th column $\mathbf{k}^{(8)}(n; \boldsymbol{\xi}_j)$ means the time series of κ_2 of the pseudo Laplacian's first term.

[Step-I] Time-history of gauss curvatures of pseudo physics quantities
Loop: $\forall \xi_j \in \mathrm{V}$
Loop $:t = 1, ..., t_n$
Calculate$K(t; \xi_j) = ((\kappa_1, \kappa_2)_E, (\kappa_1, \kappa_2)_P, (\kappa_1, \kappa_2)_L, (\kappa_1, \kappa_2)_V)_j^{(t)}$
EndLoop
$\mathbb{K}(n; \xi_j) = \begin{bmatrix} ((\kappa_1, \kappa_2)_E, (\kappa_1, \kappa_2)_P, (\kappa_1, \kappa_2)_L, (\kappa_1, \kappa_2)_V)_j^{(1)} \\ \vdots \\ ((\kappa_1, \kappa_2)_E, (\kappa_1, \kappa_2)_P, (\kappa_1, \kappa_2)_L, (\kappa_1, \kappa_2)_V)_j^{(t_n)} \end{bmatrix} \in \mathbb{R}^{n \times 8}$
End Loop
[Step-II] Column-wise Fast Fourier Transform (FFT)
Loop: $(\forall \xi_j \in \mathrm{V})$
Loop $:(i = 1, ..., 8)$
$(\mathbf{p}_{PSD}^{(i)}, \mathbf{f}^{(i)}) = \mathrm{FFT}[\mathbf{k}^{(i)}(n; \xi_j) = i_{th}$ column of $\mathbb{K}(n; \xi_j)]$
$(\overline{\mathbf{p}}_{PSD}^{(i)}, \overline{\mathbf{f}}^{(i)}) =$ Sorted $(\mathbf{p}_{PSD}^{(i)}, \mathbf{f}^{(i)})$ by PSD in descending order
$(\overline{\mathbf{p}}_{top}^{(i)}, \overline{\mathbf{f}}_{top}^{(i)}) =$ Top 10 largest $(\overline{\mathbf{p}}_{PSD}^{(i)}, \overline{\mathbf{f}}^{(i)})$ by PSD
EndLoop
$\mathbb{F}(n; \xi_j) = \left[\overline{\mathbf{p}}_{top}^{(1)} \ \overline{\mathbf{f}}_{top}^{(1)} \ \cdots \overline{\mathbf{p}}_{top}^{(8)} \ \overline{\mathbf{f}}_{top}^{(8)}\right] \in \mathbb{R}^{10 \times 16}$
End Loop

Table 1. Algorithm—Fourier transform (FT)-based new feature generation.

To generate the Fourier transform-based new features, this study performed the fast Fourier transform (FFT). The well-established library *FFTW*[20,21] is used to carry out FFT of the discrete time series of the Gauss curvature-based features of the pseudo physics quantities, i.e., each column of $\mathbb{K}(n; \boldsymbol{\xi}_j)$. The FFT generates the resultant set F_{PSD} consisting of the power spectral densities $\mathbf{p}_{PSD} \in \mathbb{R}^n$ and the associated frequencies $\mathbf{f} \in \mathbb{R}^n$. In short, $\mathbb{K}(n; \boldsymbol{\xi}_j) \rightarrow F(n; \boldsymbol{\xi}_j) := \{\mathbf{p}_{PSD}^{(1)}, \mathbf{f}^{(1)}, \ldots, \mathbf{p}_{PSD}^{(8)}, \mathbf{f}^{(8)}\}$. Then, for each column, we can remove the DC component and sort the column vectors in descending order with respect to the magnitude of PSD. The resulting sorted set $(\overline{F})$ from $F(n; \boldsymbol{\xi}_j) \rightarrow \overline{F}(n; \boldsymbol{\xi}_j) := \{\overline{\mathbf{p}}_{PSD}^{(1)}, \overline{\mathbf{f}}^{(1)}, \ldots, \overline{\mathbf{p}}_{PSD}^{(8)}, \overline{\mathbf{f}}^{(8)} \mid \overline{p}_{i-1}^{(m)} \geq \overline{p}_i^{(m)}, \forall i \in [1, n]\}$ where $\overline{\mathbf{p}}_{PSD}^{(m)} = (\overline{p}_1^{(m)}, \ldots, \overline{p}_n^{(m)})^{\mathrm{T}}$. Thus, $\overline{p}_i^{(m)}$ is the i_{th} entity of the sorted column vector $\overline{\mathbf{p}}_{PSD}^{(m)}$ in descending order. $\overline{\mathbf{f}}^{(m)}$ is the sorted frequency vector according to the $\overline{\mathbf{p}}_{PSD}^{(m)}$.

To generate practically meaningful features, amongst many peaks in the power spectra, this paper extracted the top 10 amplitudes and the associated frequencies. FT-based new feature set is denoted as $\mathbb{F}(n; \boldsymbol{\xi}_i) \in \mathbb{R}^{10 \times 16}$, $\overline{F}(n; \boldsymbol{\xi}_j) \rightarrow \mathbb{F}(n; \boldsymbol{\xi}_i) := \{\overline{\mathbf{p}}_{top}^{(1)}, \overline{\mathbf{f}}_{top}^{(1)}, \ldots, \overline{\mathbf{p}}_{top}^{(8)}, \overline{\mathbf{f}}_{top}^{(8)} \mid \overline{\mathbf{p}}_{top}^{(m)} \subset \overline{\mathbf{p}}_{PSD}^{(m)}, \overline{\mathbf{f}}_{top}^{(m)} \subset \overline{\mathbf{f}}^{(m)}\}$.

There is no strict restriction to how many top amplitudes are selected. Th s paper adopts up to top 10 since it can encompass sufficient energy of the total energy of the input signal. For instance, Fig. 2A shows that the top 10 peaks sufficiently large energy level and that beyond the ten peaks, the energy level decreases to 20% of that of the largest peak. Figure 2B shows exceptions when the 10th peak's energy level does not decrease to below 50% of that of the largest peak. The results of this paper support that the inclusion of the top 10 amplitudes and their frequencies in the FT-based features is successful to distinguish and learn hidden rules of the imminent

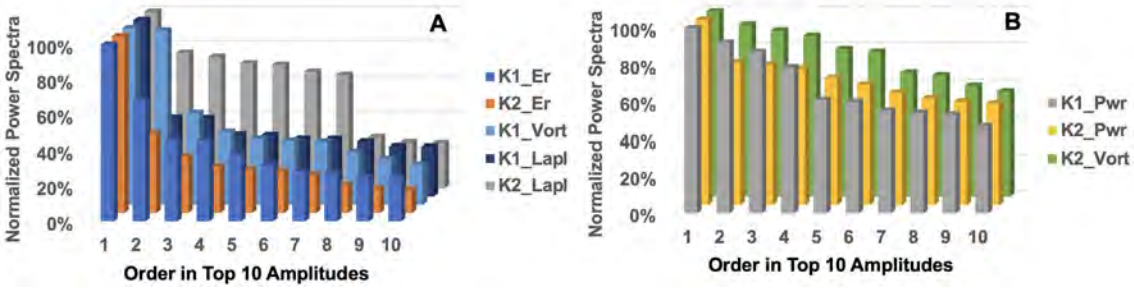

Figure 2. Normalized power spectra of the top 10 amplitudes and their frequencies after sorting the FFT results. (**A**) Rapidly decreasing relative energy levels of top 10 peaks from FFT of Gauss curvatures (denoted as K1 and K2). (**B**) Slowing decreasing relative energy levels. Er: The pseudo released energy, Vort: the pseudo vorticity, Lapl: the pseudo Laplacian, and Pwr: the pseudo power.

extreme EQs. Including more peaks (thus more energy) will be straightforward, and investigation into their impacts shall be a future research topic.

Data-driven prediction rules for individual large EQs

This paper pursues the purely data-driven prediction rules that are customized for individual large EQs, being independent of existing magnitude prediction models[22–25], or earthquake forecasting methods[26–33]. Overall architecture of the adopted hidden rule-learning ML algorithm is illustrated in Fig. 1A(V). The generality of the hidden rule-learning approach shown in Fig. 1 has been demonstrated with complex physics phenomena at diverse scales from nano[34], to micro[35], to composite structures[36,37], and to the Earth lithosphere[17].

Prediction rules are unraveled by the Glass-Box Rule-Learning algorithm that uses the multi-layered data (Fig. 1A(V)). The role of "Scientist in the Loop" is to monitor the rule-based predictions and help decide whether to recommend appending the identifi d best-so-far rule into the storage (i.e., global memory for future inheritance and predictions) or not. For instance, some predictions may have good fitness scores numerically, but their prediction plots may not satisfy the domain expert's knowledge. Then, the "Scientist in the Loop" may queue for additional rule-learning by changing ML-control parameters or expanding search spaces. Since each large EQ has its own prediction rule in the proposed approach, such a re-learning can be done separately and multiple times, specifically for the EQ. This scientist's role augments the data-driven rule-learning process to better comply with domain science.

In the previous work of the author[17], the best-so-far prediction rule was identified by a multiplicative combination of cubic regression spline (CRS)-based LFs of (i) the pseudo released energy (its LF is denoted $\mathcal{L}_E$), (ii) the pseudo power ($\mathcal{L}_P$), (iii) the pseudo vorticity ($\mathcal{L}_\omega$), and (iv) the pseudo Laplacian ($\mathcal{L}_L$). The CRS-based LF can leverage its high flexibility[38,39], and its general form is given in "Methods". CRS-based LFs can embrace constant shift, linear, and nonlinear curves[38,39]. Thus, the best-so-far data-driven prediction rule without the Fourier transform-based features (denoted $M_{CRS}^{(t+1)}$) is given by

$$M_{CRS}^{(t+1)}(\boldsymbol{\xi}_j) = \mathcal{L}_E(E_r^{*(t)})\mathcal{L}_P\left(\mathrm{Sg}\left(e^2\frac{\partial E_r^{*(t)}}{\partial t}\right)\right)\mathcal{L}_\omega\left(\mathrm{Sg}\left(e^2\omega_\lambda\right)\right)\mathcal{L}_L\left(\mathrm{Sg}\left(10^{-4}\frac{\partial^2 E_r^{*(t)}}{\partial\lambda^2}\right)\right) \tag{4}$$

where $E_r^{*(t)}$ is the best-so-far pseudo released energy at epoch t and at the reference volume $\boldsymbol{\xi}_j$. The free parameters associated with the best-so-far CRS LFs $\mathcal{L}_E$, $\mathcal{L}_P$, $\mathcal{L}_\omega$, and $\mathcal{L}_L$ are denoted by $\boldsymbol{\theta}_E$, $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_\omega$, and $\boldsymbol{\theta}_L$, respectively. Sg(.) stands for a typical sigmoid function. The detailed rationales for the data-driven rules are presented in Ref.[17].

In this study, the Fourier bases are used to sharpen the prediction rule. The top 10 frequencies of the principal Gauss curvatures are used in the Fourier bases. The best-so-far data-driven prediction rules with the Fourier transform-based features in conjunction with CRS LFs is denoted by $M_{FT\&CRS}^{(t+1)}$.

$$\begin{aligned} M_{FT\&CRS}^{(t+1)}(\boldsymbol{\xi}_j) =& \mathcal{L}_E \times (\mathcal{L}_{FT}({\kappa_1}^{(E)}) + \mathcal{L}_{FT}({\kappa_2}^{(E)})) \\ & \times \mathcal{L}_P \times (\mathcal{L}_{FT}({\kappa_1}^{(P)}) + \mathcal{L}_{FT}({\kappa_2}^{(P)})) \\ & \times \mathcal{L}_\omega \times (\mathcal{L}_{FT}({\kappa_1}^{(\omega)}) + \mathcal{L}_{FT}({\kappa_2}^{(\omega)})) \\ & \times \mathcal{L}_L \times (\mathcal{L}_{FT}({\kappa_1}^{(L)}) + \mathcal{L}_{FT}({\kappa_2}^{(L)})) \end{aligned} \tag{5}$$

where $\mathcal{L}_E$, $\mathcal{L}_P$, $\mathcal{L}_\omega$, and $\mathcal{L}_L$ are from Eq. (4); the Fourier-based link functions are give as

$$\mathcal{L}_{FT}({\kappa_1}^{(E)}) = \sum_{i=1}^{10}(a_i\cos(2\pi\bar{f}_i^{(1)}{\kappa_1}^{(E)}) + b_i\sin(2\pi\bar{f}_i^{(1)}{\kappa_1}^{(E)})) \tag{6}$$

Similarly, all other Fourier-based LFs are given in Supplemental Material. Here, the sorted frequency $\bar{f}_i^{(1)}$ is the i_{th} entity of $\bar{\mathbf{f}}_{top}^{(1)}$. In particular, the Fourier frequencies of $(\kappa_1, \kappa_2)_E$ are used to augment $\mathcal{L}_E$ as shown in the first line of Eq. (5). The FT-based LFs appear to offer considerable flexibility to the prior best-so-far prediction rule (Eq. 4) and thus improve accuracy of the prediction rules.

It should be noted that all the LFs are customized for individual large EQs in this paper. Rather than a single set of LFs describing all EQs, each large EQ will have its own best-so-far LFs (i.e., specialized prediction rules). Th s customized approach will be helpful for a future expansion to a reinforcement learning-based evolution of this framework in which unsupervised ML methods can continue learning and improving the best prediction rules (i.e., LFs) for many different large EQs, without human interventions.

Feasibility test results

Th s paper expanded the study region compared to the previous study[17] to wider reference events ($M_w \geq 6.5$) in the West U.S. region (i.e., longitude and latitude in (− 132.5, − 110) and (30, 52.5) [deg], respectively, and depth (− 5, 20) [km]) within the past 40 years from 1980 through 2019.

The ML-identifi d rule uses the observed 10-year data, 30 days before the event without any physics mechanisms or statistical laws. The best-so-far rule appears to be successful in reproducing the next-month earthquake's location and magnitude 30 days before the event as shown in Fig. 3. The ML-identified rule appears to reproduce the global peak event noticeably well. In addition to Fig. 3, all other ML-driven reproductions of 8 large events of magnitudes ($M_w \geq 7.0$) in the West U.S. region are shown in Figs. S2 and S3. Also, Figs. S4–S6 presents ML-driven reproductions of 9 large events of magnitudes ($M_w \in [6.5, 7.0)$) in the western U.S. region.

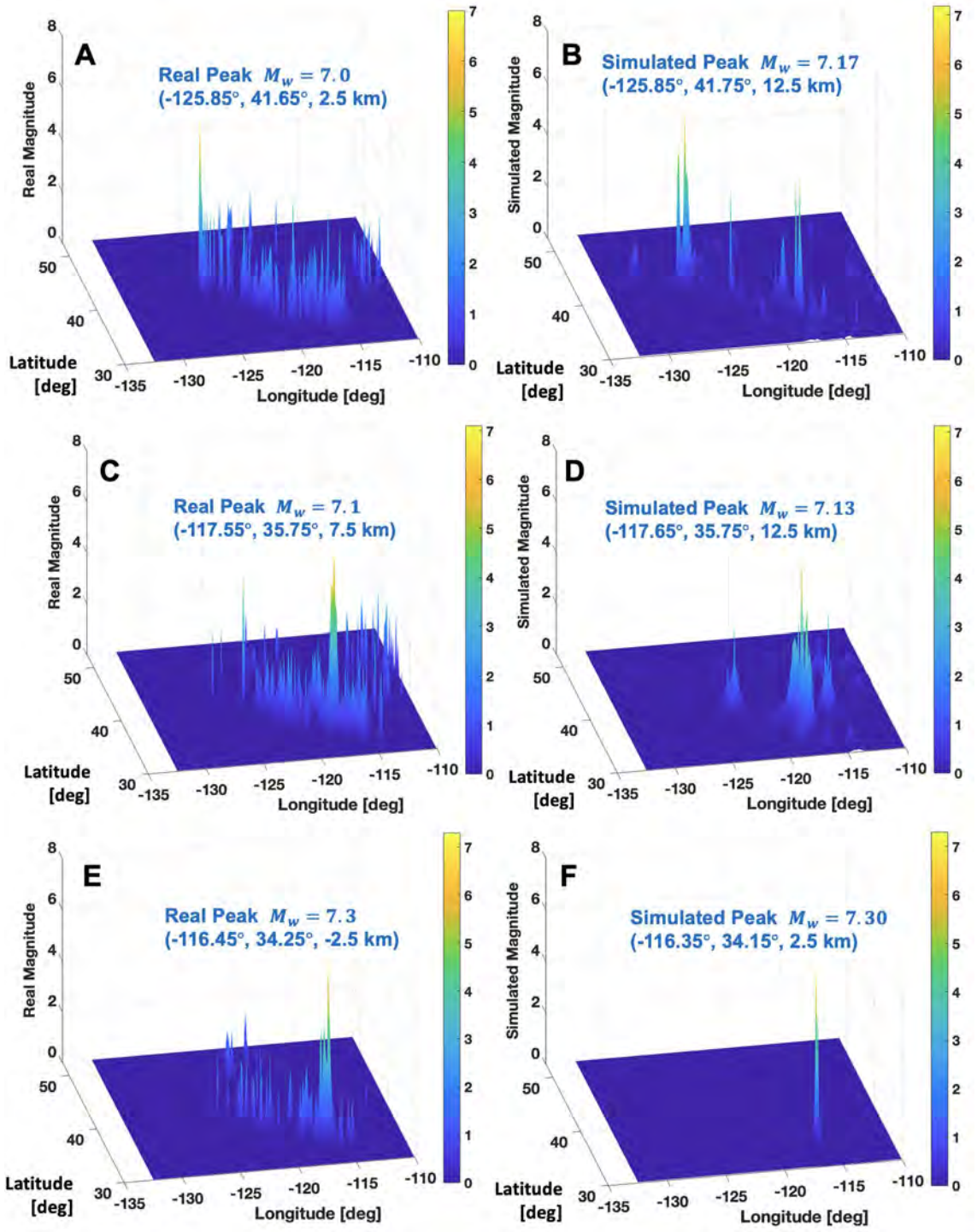

Figure 3. Reproduction of large magnitude events $M_w > 7.0$ by using the customized ML-identifi d data-driven prediction rules with FT-based new features: (**A**,**B**) Observed real and simulated earthquake events on 1991/8/17 (target day ID 1004246); (**C**,**D**) 2019/7/6 (1014431); (**E**,**F**) 1992/6/28 (1004562).

In some cases, the ML-identifi d rules reproduce reasonably the global peak's location and magnitude with a few false small peaks (e.g., Figs. S2A–D). Such false peak reproductions appear more noticeable in events of magnitudes ($M_w \in [6.5, 7.0)$) than events of magnitudes ($M_w \geq 7.0$). For instance, Figs. S5C–F appear to show the wrong reproductions of false peaks. It is related to the limit of the best-so-far ML-identifi d rules which shall be improved in the future extension. Still, the overall performance is promising since the largest peaks' locations and magnitudes are reasonably reproduced by the customized data-driven model. Table 2 summarizes the prediction results of individual 8 large earthquakes of magnitude ($M_w \geq 7.0$) and 9 large earthquakes of ($M_w \in [6.5, 7.0)$) using the best-so-far data-driven prediction model. For the 8 large earthquakes of magnitude ($M_w \geq 7.0$), the mean differences in (latitude, longitude, depth, magnitude) between real peak and

Target day ID (date)			M_w	ΔM_w	$\lambda(^\circ)$	$\phi(^\circ)$	h (km)	$\Delta\xi$(km)
1004246	(1991/8/17)	Observed real	7		− 125.86	41.68	1.3	
		Predicted	7.17	0.17	− 125.85	41.75	12.5	13.7
1004498	(1992/4/25)	Observed real	7.2		− 124.23	40.33	9.86	
	Cape Mendocino	Predicted	6.83	0.37	− 124.35	40.45	12.5	19.0
1004562	(1992/6/28)	Observed real	7.3		− 116.44	34.2	− 0.097	
	Landers	Predicted	7.3	0	− 116.35	34.15	2.5	11.7
1005357	(1994/9/1)	Observed real	7		− 126.3	40.41	4.97	
		Predicted	6.81	0.19	− 126.85	40.25	7.5	63.7
1007228	(1999/10/16)	Observed real	7.1		− 116.25	34.6	13.73	
	Hector Mine	Predicted	6.9	0.2	− 116.25	34.95	7.5	39.3
1009297	(2005/6/15)	Observed real	7.2		− 125.95	41.29	16	
		Predicted	6.97	0.23	− 126.15	41.15	17.5	27.2
1011051	(2010/4/4)	Observed real	7.2		− 115.3	32.29	9.99	
	El Mayor–Cucapah	Predicted	6.99	0.21	− 115.15	32.35	12.5	18.1
1014431	(2019/7/6)	Observed real	7.1		− 117.6	35.77	8	
	Ridgecrest	Predicted	7.13	0.03	− 117.65	35.75	12.5	7.5
1004211	(1991/7/13)	Observed real	6.9		− 125.64	42.18	11	
		Predicted	6.47	0.43	− 126.85	41.75	2.5	142.9
1005130	(1994/1/17)	Observed real	6.7		− 118.54	34.21	18.2	
	Northridge	Predicted	6.63	0.07	− 118.25	34.35	7.5	37.3
1005528	(1995/2/19)	Observed real	6.6		− 125.76	40.59	4.62	
		Predicted	6.45	0.15	− 125.65	40.35	7.5	29.5
1008756	(2003/12/22)	Observed real	6.5		− 121.1	35.7	8.38	
	San Simeon	Predicted	6.54	0.04	− 121.05	35.85	12.5	18.0
1009072	(2004/11/2)	Observed real	6.7		− 128.77	49.28	10	
		Predicted	7.06	0.36	− 127.85	49.45	12.5	104.0
1012487	(2014/3/10)	Observed real	6.8		− 125.13	40.83	16.44	
	Ferndale	Predicted	6.73	0.07	− 125.55	40.85	2.5	48.8
1012532	(2014/4/24)	Observed real	6.5		− 127.73	49.64	10	
		Predicted	6.77	0.27	− 126.35	50.35	7.5	172.4
1013491	(2016/12/8)	Observed real	6.6		− 126.19	40.45	8.45	
		Predicted	6.77	0.17	− 126.35	40.95	7.5	58.3
1014174	(2018/10/22)	Observed real	6.8		− 129.29	49.33	10	
		Predicted	7.23	0.43	− 129.25	49.45	7.5	14.3

Table 2. Individual large earthquake reproductions using the best-so-far customized data-driven models with FT-based new features. 8 events of $M_w \geq 7.0$ and 9 events of $M_w \in [6.5, 7.0)$. ΔM_w and $\Delta\boldsymbol{\xi}$ are the absolute differences in magnitude and location between the real observation and the predicted peaks, respectively. Some common names of EQs are included. The coordinates (λ, ϕ, h) of all the observed real peaks in this table are the same as those in the USGS catalog database. All other comparison plots between predicted and observed peaks of this paper present the center coordinates of the reference volume ($\Delta\lambda = \Delta\phi = 0.1^\circ$; $\Delta h = 5$ km) that contains the peak.

ML-reproduced peak are (0.12°, 0.15°, 4.21 km, 0.18). The mean differences increase to (0.28°, 0.51°, 5.4 km, 0.22) for the 9 large earthquakes of magnitude ($M_w \in [6.5, 7.0)$). These difference underpins the overall accuracy of the best-so-far ML-identified rules in reproducing three-dimensional locations and magnitudes 30 days before the event. But, it also implies that the rule's accuracy appears to deteriorate for the second largest group of ($M_w \in [6.5, 7.0)$). Uncertainty also increases for this second largest group. For the 8 large earthquakes of magnitude ($M_w \geq 7.0$), the standard deviation of the differences in (latitude, longitude, depth, magnitude) between real peak and ML-reproduced peak are (0.1°, 0.17°, 3.19 km, 0.12). The standard deviation of differences increase to (0.22°, 0.52°, 4.52 km, 0.15) for the 9 large earthquakes of magnitude ($M_w \in [6.5, 7.0)$). Improvement of accuracy and underlying uncertainty shall be a natural future extension topic.

The preserved interpretability is noteworthy. The best-so-far prediction rules are remembered by storing all the free parameters of the LFs. By retrieving the parameters and plugging them into the corresponding LFs' expressions (e.g., Eqs. 6 or 11), one can investigate and interpret individual physical terms and their behavior (e.g., Fig. S8 of the author's prior work[17]). During the rule-learning process, the poor-performing combinations of features and their LFs are rejected. By doing so, this approach can help improve physical interpretation of the ML-identifi d rules. In particular, the best-performing prediction rule (Eq. 5) turns out to select the pseudo vorticity's fi st term ω_λ and the pseudo Laplacian's fi st term $\frac{\partial^2 E_r^{*(t)}}{\partial \lambda^2}$ out of many other feature terms (e.g.,

$\omega_\lambda, \omega_\phi, \omega_h, \sqrt{\omega_\lambda^2 + \omega_\phi^2}, \frac{\partial^2 E_r^{*(t)}}{\partial \phi^2}, \frac{\partial^2 E_r^{*(t)}}{\partial h^2}, \nabla_g^2 E_r^{(t)}$, and so on). Physically, ω_λ may describe the slow rotational motion about the longitudinal axis, and the directions of the western U.S. region's plate motions and the known major faults are roughly parallel or normal to the longitudinal axis. Therefore, any accurate data-driven prediction rules, if properly unraveled, should be able to highlight certain salient physical terms, and such favorable capabilities appear to be confi med. Purely based on data, this paper's best prediction rules pinpoint salient feature terms, underpinning the preservation of physical intepretability.

Conclusion

The inclusion of FT-based features in the prediction rules appears to be effective to improve accuracy of large EQ's location and magnitude 30 days before the event. Figure 4 compares the positive impact of the FT-based features. Also, the FT-based features appear to help the ML-identifi d best-so-far rules to sharpen the predicted magnitude distributions and remove the incorrect peaks. For instance, these positive roles of FT-based features can be clearly seen from comparison between Fig. 4A,C and between Fig. 4D,F.

The improvement can be also confi med by the sharpened distribution of the absolute errors between real and predicted magnitudes. For instance, Fig. S7 compares the absolute magnitude errors ($|\Delta M_w|$) from predictions without and with FT-based features in the best-so-far rules.

To some extent, it is an anticipated result due to two reasons. FT-based features' many LFs ($\mathcal{L}_{FT}$) and the associated free parameters can offer additional fitting power to prediction like a deep learning model with more neurons. Also, many $\mathcal{L}_{FT}$'s can be regarded as many higher terms in the Fourier series (here, up to 10 harmonics) which can contribute to smooth fitting strength.

To incorporate the FT-based features, the ML-identifi d rules should have additional Fourier bases like Eq. (6). When prediction rules use only CRS-based LFs (Eq. 4), the incorrect peaks and over-smoothing issues remain (see Fig. 4C,F). In contrast, the combination of smooth CRS-based LFs $\mathcal{L}_i, i = (E, P, \omega, L)$ and the Fourier series-based LFs $\mathcal{L}_{FT}$ in Eq. (5) can offer enhanced accuracy of reproducing large rare peaks without incorrect peaks and over-smoothing issues (see Fig. 4A,D). The Gibbs phenomenon (i.e., over/undershoot issues near a jump discontinuity) appears not in effect at the present prediction rules using the Fourier series-based LFs. Th s may be attributed to the fact that the large EQs in this framework are regarded as "point sources" not lines, thus not necessarily leading to sudden discontinuities of target distributions. In the future extensions, when EQs are regarded as 2D line sources, the Gibbs phenomenon may negatively affect the predictions, which shall be addressed later.

Also, in the future investigations, comprehensive validations of the ML-identifi d prediction rules should be done to confi m general applicability to a wide range of EQ sizes. For instance, Fig. S8 shows preliminary test predictions of "quiet" period without large EQs ($M_w < 5.5$ during the period) by using the best-so-far rules. Fig. S8A shows the prediction 33 days before the Ridgecrest EQ (2019/7/6; $M_w = 7.1$); Fig. S8B, 32 days before. No false alarms with spurious large EQ predictions are detected. Since the proposed approach trained and

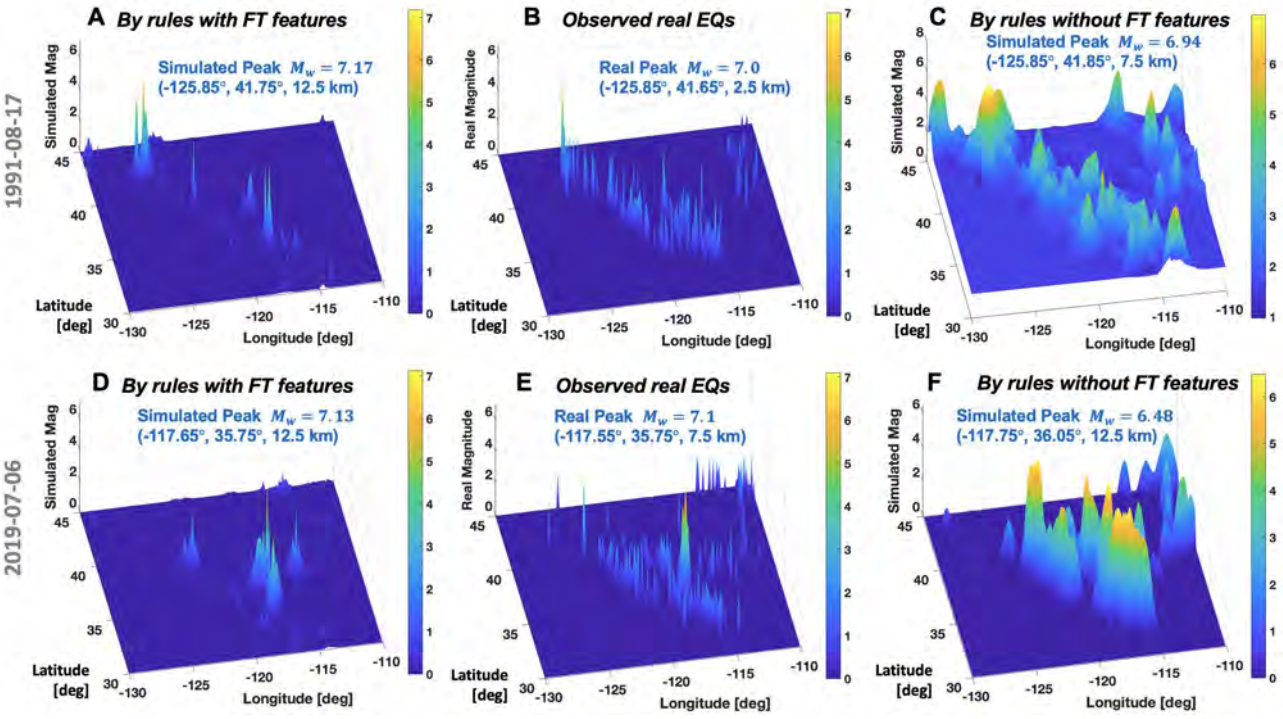

Figure 4. Positive role of FT-based new features in improving prediction accuracy of the best-so-far rules: (**A**–**C**) 1991/8/17 (1004246); (**D**–**F**) 2019/7/6 (Target date ID 1014431). (**A**) and (**D**) harness a combination of Fourier series-based LFs and CRS-based LFs by Eq. (5) whereas (**C**) and (**F**) utilize CRS-bases LFs of the four pseudo physics quantities by Eq. (4).

unraveled all the prediction rules with large EQs data ($M_w \geq 6.5$), this preliminary test result appears promising. But, to draw a concrete conclusion about the proposed approach and also to be practically meaningful tool (like[28,32]), future extension should conduct comprehensive tests over broad ranges of EQs.

The outcomes of this study add a new dimension to research for predicting individual large EQs. Gauss curvature-based unique signatures of large EQs may be remembered and distinguished by unsupervised ML methods while the data-driven prediction rules can be better customized for individual large EQs with new data. The overall processes can be managed and evolved by another global ML like reinforcement learning, thereby shedding light on purely data- and ML-driven predictions of large EQs. The handshake among ML methods, Fourier, and Gauss may help answer the long-standing enigma of seismogenesis.

Methods

Data preparation

Th s study collected and processed raw earthquake catalog data available in Ref.[18] from January 1980 through October 2019. Without any prejudice, all the recorded earthquakes within the past 40 years are included, and the total number of earthquakes is 1,895,190. According to the calendar-based date, all earthquakes within one day is stored in one data file. The day-based earthquake catalog data file is named as 1,000,000 for January 1st, 1980, 1,000,001 for January 2nd, 1980, and so on. Each file contains the number of data points in the file followed by longitude, latitude, depth, and magnitude of each earthquake. As illustrated in Fig. S1, one epoch is defined as 30-day time range. All earthquakes within the 30-days window are considered to belong to the same epoch. A frame of epochs consists of many consecutive epochs and serves as the training base for rule-learning glass-box machine learning algorithm. Within one frame of epochs, the last epoch is used as a target while all the previous epochs are used for training of hidden rules. As illustrated in Fig. S1A, the target epoch is completely disjointed from the frame of epochs used for training and rule-learning. As explained in Fig. S1, this study sets one-day interval between consecutive frames of epochs. By marching frames of epochs with one-day increment, this paper can dramatically increase the number of total frames of epochs to 14,600. For interested researchers, all the processed data sets of the refi ed epochs with one-day interval are publicly available at[40].

Th s paper focuses on prediction rule-learning about a large target EQ ($M_w \geq 6.5$) positioned at the last day of the target of epoch (Fig. S1A). Thus, all ML-identifi d rules of this paper are specifi ally trained to predict a future large event 30 days before the target EQ (i.e., D-30 case in Fig. S1B). In the future extension, shorter time-window predictions (e.g., a few days ahead) shall be possible by placing the target EQ at the earlier positions of the target epoch (D-2 or D-1 cases in Fig. S1B). In contrast, by defini g wider target epochs, longer time-widow predictions (e.g., months or years ahead) shall be also possible, which will be meaningful for complementing the existing long-term EQ forecasting methods.

Data transformation from raw EQ catalog into spatio-temporal information index

Temporal convolution is carried out after spatial convolution is done as

$$\overline{\Pi}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l) = \int \omega(\tau; T_l)\overline{\Pi}_{S}^{(t_{past})}(\boldsymbol{\xi}_j; L_k)dt_{past} \tag{7}$$

where the one-dimensional (1D) Gaussian weight $\omega(\tau; T_l) = (T_l(2\pi)^{1/2})^{-1} \exp\left(-\frac{\tau^2}{2T_l^2}\right); \tau = |t - t_{past}|, t \geq t_{past}$, meaning the time gap between the current and the past time. And, the spatial convolution is done by

$$\overline{\Pi}_{S}^{(t)}(\boldsymbol{\xi}_j; L_k) = \int_{\mathrm{V}} \omega(\boldsymbol{\xi}_j, \mathbf{x}_i^{(t)}; L_k)M_i^{(t)}/10 \; d\mathbf{x} \tag{8}$$

where the 3D Gaussian weight $\omega(\boldsymbol{\xi}_j, \mathbf{x}_i^{(t)}; L_k) = (L_k(2\pi)^{1/2})^{-N} \exp\left(-\frac{|\mathbf{x}_i^{(t)} - \boldsymbol{\xi}_j|^2}{2L_k^2}\right)$; V means the entire lithosphere domain under consideration. Here, T_l $(l = 1, \ldots, n_T)$ and L_k $(k = 1, \ldots, n_L)$ are influence ranges in time and 3D space, respectively. Therefore, there can be at most $n_L \times n_T$ spatio-temporal IIs at one reference volume and a time. Following the preliminary investigations done in Ref.[17], this study adopts $n_L = 2$ with $L_1 = 10$ (km) and $L_2 = 25$ (km) while $n_T = 2$ with $T_1 = 3$ (month) and $T_2 = 6$ (month). This combination appears to lead to the best-so-far prediction performance since it embraces dual impacts of close and far EQs both in three-dimensional spatial domain and in one-dimensional temporal domain.

Flexible and expressive link functions

Pursuing the interpretability, this paper adopts an expressive link function (LF) using transparent, flex ble bases that can describe a mathematical expression between input features and the hidden rules as output. LF is denoted as $\mathcal{L}$ where $\boldsymbol{\theta}$ is a set of free parameters prescribing the LF. The cubic spline regression (CRS) curves consist of a few cubic polynomials connected at knots so that the curves are continuous up to the second derivatives[38]. For practical cubic spline bases[39] (denoted as b_i), LFs look like

$$\mathcal{L}(\overline{\Pi}_{ST}; \; \mathbf{a}, \mathbf{x}^*) = \sum_{i}^{p} a_i b_i(\overline{\Pi}_{ST}) \tag{9}$$

$$CRS-BasedLF:\mathcal{L}^{(k,l)}(\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j;L_k,T_l);\ \boldsymbol{\theta}^{(k,l)})=\sum_{i=1}^{p}a_i^{(k,l)}b_i^{(k,l)}(\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j;L_k,T_l)); \tag{10}$$

where $b_1(x)=1, b_2(x)=x$, and

$$b_{i+2}(x)=\frac{[(x_i^*-\frac{1}{2})^2-\frac{1}{12}][(x-\frac{1}{2})^2-\frac{1}{12}]}{4}-\frac{[(|x-x_i^*|-\frac{1}{2})^4-\frac{1}{2}(|x-x_i^*|-\frac{1}{2})^2+\frac{7}{240}]}{24}, \tag{11}$$

for $i=1\ldots p-2$. Here, x_i^* is i_{th} knot location. To fully describe one LF, we need to identify $p+(p-2)$ unknowns, i.e. $\mathbf{a}=\{a_1,\ldots,a_p\}$ and $\mathbf{x}^*=\{x_1^*,\ldots,x_{(p-2)}^*\}$. The CRS-based LF can accommodate simple monotonic rule to highly nonlinear rule. Importantly, the adopted CRS is not used for the direct "regression" but for the transparent expression searching.

Bayesian evolutionary algorithm

Th s paper adopts the Bayesian evolutionary algorithm to fi d the total free parameters of hidden rules and also to enable smooth evolution of the rules. In particular, the total free parameters include the parameters of the pseudo released energy rule in Eq. (1), CRS LFs of prediction rules in Eq. (4), and Fourier transform-based LFs of predition rules in Eq. (5). A combination of the fitness-proportionate probability (FPP) scheme of the genetic algorithm and the Bayesian update scheme is at the center of the method (see details in Ref.[17]). On the Bayesian evolutionary framework, in total 71,600 organisms (i.e., candidates for total free parameters of hidden rules) and 20 generations are used, 4 alleles per gene are used, and the variable-wise mutation with rate of 0.005 is used. Preliminary investigations narrow down the best-performing search ranges such that (i) exponential LFs' two parameters reside in $[0,3]\cup[0,10]$; (ii) CRS LFs' parameters of five bases all reside in $[-2,2]$ and three knots' loci are $[0,1/3]\cup[1/3,2/3]\cup[2/3,1]$; (iii) Fourier Transform-based LFs' parameters of Fourier bases all reside in $[-2,2]$. As done in Ref.[17], the three-fold error measure (i.e., fitness) is based on differences in magnitude, three-dimensional location, and the wrong peaks' count (false alarms) between real observations and best-so-far rule-driven predictions. The key steps of the Bayesian evolutionary algorithm are presented in Table S2 of Supplementary Information.

Computation cost

All computational simulations, training, and predictions of this paper were conducted on NOVA, a high-performance computing cluster of Iowa State University. NOVA cluster consists of compute nodes with two 18-Core Intel Skylake 6140, 1.5 TB of fast NVME local storage, and 192 GB of memory. All nodes and storage are connected via Mellanox EDR (100 Gbps) switch. Given one target EQ, the rule-learning simulation used 144 cores and finished within 12 h, which includes all steps: new feature generations, FFT, and rule-learning with Bayesian evolutionary algorithm. Once the best-so-far rule's free parameters are identifi d and stored, one separate prediction using top 10 best-so-far rules costs only 4 min with 16 cores.

Data availability

The processed 40-years data sets consisting of the month-based epochs and the refi ed day-based epochs are shared on a cloud storage[40]. Other supplementary data and parallel programs supporting other findings of this paper will be available upon request to the author.

Received: 17 February 2023; Accepted: 20 September 2023
Published online: 25 September 2023

References

1. Ross, Z. E., Cochran, E. S., Trugman, D. T. & Smith, J. D. 3D fault architecture controls the dynamism of earthquake swarms. *Science* **368**, 1357–1361 (2020).
2. Ross, Z. E., Men-Andrin, M., Egill, H. & Thomas, H. H. Generalized seismic phase detection with deep learning. *Bull. Seismol. Soc. Am.* **108**(5A), 2894–2901 (2018).
3. DeVries, P. M. R., Viegas, F., Wattenberg, M. & Meade, B. J. Deep learning of aftershock patterns following large earthquakes. *Nature* **560**, 632. https://doi.org/10.1038/s41586-018-0438-y (2018) (**pmid: 30158606**).
4. Feng, B. & Fox, G. C. Spatiotemporal Pattern Mining for Nowcasting Extreme Earthquakes in Southern California. arXiv:2012.14336 [physics.geo-ph] (2020).
5. Wang, Q., Guo, Y., Yu, L. & Li, P. Earthquake prediction based on spatio-temporal data mining: An LSTM network approach. *IEEE Trans. Emerg. Top. Comput.* **8**, 148–158. https://doi.org/10.1109/TETC.2017.2699169 (2017).
6. Tan, Y. J. *et al.* Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 Central Italy sequence. *Seismic Rec.* **1**, 11–19. https://doi.org/10.1785/0320210001 (2021).
7. Beroza, G. C., Segou, M. & Mousavi, M. Machine learning and earthquake forecasting-next steps. *Nat. Commun.* **12**, 4761. https://doi.org/10.1038/s41467-021-24952-6 (2021) (**pmid: 34362887**).
8. Mousavi, S. M., Zhu, W., Ellsworth, W. & Beroza, G. C. Unsupervised clustering of seismic signals using deep convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.* **16**(11), 1693–1697 (2019).
9. Zhu, S., Li, S., Peng, Z. & Xie, Y. Imitation Learning of Neural Spatio-Temporal Point Processes (2021). arXiv:1906.05467v4.
10. Mignan, A. & Broccardo, M. Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. *Seismol. Res. Lett.* **91**, 2330–2342 (2020).
11. Segou, M. The physics of earthquake forecasting. *Seismol. Res. Lett.* **91**, 1936–1939 (2020).
12. Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D. & Rhoades, D. A. Earthquake likelihood model testing. *Seismol. Res. Lett.* **78**, 17–29 (2007).
13. Zechar, J. D. *et al.* The collaboratory for the study of earthquake predictability perspectives on computational earth science. *Concurr. Comput. Pract. Exp.* **22**, 1836–1847 (2010).

14. Pardo, E., Garfias, C. & Malpica, N. Seismic phase picking using convolutional networks. *IEEE Trans. Geosci. Remote Sens.* **57**(9), 7086–7092 (2019).
15. 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).*
16. Rouet-Leduc, B. *et al.* Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.*https://doi.org/10.1002/2017GL074677 *(2017).*
17. Cho, I. Gauss curvature-based unique signatures of individual large earthquakes and its implications for customized data-driven prediction. *Nat. Sci. Rep.* **12**, 8669 (2022).
18. United States Geological Survey (USGS), Earthquake Catalog. *USGS* (2020). https://earthquake.usgs.gov/earthquakes/search/ (accessed Nov 2020).
19. Gray, A. *Modern Differential Geometry of Curves and Surfaces* 2nd edn. (CRC Press, 1997).
20. *www.fftw.org*.
21. Frigo, M. & Johnson, S. G. The design and implementation of FFTW3, proceedings of the IEEE: Special issue on program generation, optimization, and platform. *Adaptation* **93**(2), 216–231 (2005).
22. Rundle, J. B. *et al.* Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. *Rev. Geophys.* **41**(4), (2003).
23. Keilis-Borok, V. I. & Soloviev, A. A. *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction* (Springer, 2003).
24. Tiampo, K. F. *et al.* Eigenpatterns in southern California seismicity. *J. Geophys. Res.* **107**(B12), 2354 (2002).
25. Sotolongo-Costa, O. & Posadas, A. Fragment-asperity interaction model for earthquakes. *Phys. Rev. Lett.* **92**(4), 048501 (2004).
26. Toda, S. & Stein, R. S. Long- and short-term stress interaction of the 2019 ridgecrest sequence and Coulomb-based earthquake forecasts. *Bull. Seismol. Soc. Am.* **110**(4), 1765–1780 (2020).
27. Field, E. H. *et al.* Long-term time-dependent probabilities for the third Uniform California Earthquake Rupture Forecast (UCERF3). *Bull. Seismol. Soc. Am.* **105**(2A), 511–543 (2015).
28. Field, E. H. *et al.* A spatiotemporal clustering model for the third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bull. Seismol. Soc. Am.* **107**(3), 1049–1081 (2017).
29. Shcherbakov, R., Zhuang, J., Zöller, G. & Ogata, Y. Forecasting the magnitude of the largest expected earthquake. *Nat. Commun.* **10**, 4051 (2019).
30. Gulia, L. & Wieber, S. Real-time discrimination of earthquake foreshocks and aftershocks. *Nature* **574**, 193 (2019).
31. Gerstenberger, M. C., Wieber, S., Jones, L. M. & Reasenberg, P. A. Real-time forecasts of tomorrow's earthquakes in California. *Nature* **435**, 328 (2005).
32. Milner, K. R. *et al.* Operational earthquake forecasting during the 2019 Ridgecrest, California, earthquake sequence with the UCERF3-ETAS model. *Seismol. Res. Lett.* **91**, 1567–1578 (2020).
33. Page, M. T. *et al.* The UCERF3 grand inversion: Solving for the long-term rate of ruptures in a fault system. *Bull. Seismol. Soc. Am.* **104**(3), 1184–1204 (2014).
34. Cho, I. *et al.* A framework for glass-box physics rule learner and its application to nano-scale phenomena. *Nat. Commun. Phys.* **3**(78) (2020).
35. Cho, I. *et al.* Unraveling hidden rules behind the wet-to-dry transition of bubble array by glass-box physics rule learner. *Nat. Sci. Rep.* **12**, 3191 (2022).
36. Bazroun, M. *et al.* Flexible and interpretable generalization of self-evolving computational materials framework. *Comput. Struct.* **260**, 106706 (2021).
37. Cho, I. A framework for self-evolving computational material models inspired by deep learning. *Int. J. Numer. Methods Eng.* **120**(10), 1202–1226 (2019).
38. Wood, S. *Generalized Additive Models: An Introduction with R* (CRC Press, 2006).
39. Gu, C. *Smoothing Spline ANOVA Models* (Springer Science and Business Media, 2013).
40. https://iastate.box.com/s/noffbpufy9va3zel3wthbdibnvera2je.

Acknowledgements

The author is grateful for the various research supports. Th s work was supported in part by the National Science Foundation under grants CSSI-1931380 and CMMI-2129796. High-performance computing (HPC) for this study is partially supported by the HPC@ISU equipment at Iowa State University, some of which has been purchased through funding provided by the NSF under MRI grant CNS-1229081 and CRI grant 1205413.

Author contributions

I.H.C. is responsible for all algorithms and programs presented as well as writing of the manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-43181-z.

Correspondence and requests for materials should be addressed to I.H.C.

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

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

Sharpen Data-Driven Prediction Rules of Individual Large Earthquakes with Aid of Fourier and Gauss

In Ho Cho[1]

[1]**CCEE Department, Iowa State University, Ames, IA 50011, USA**

1 Refined frames of epochs with day-based increment

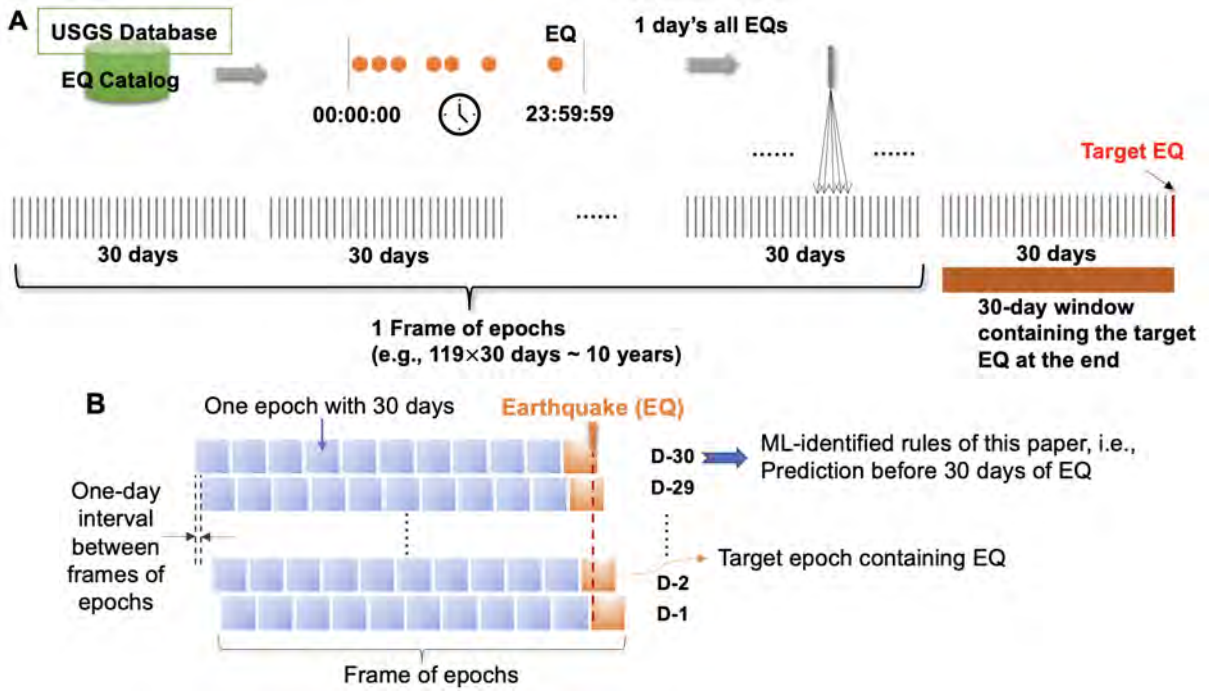

Fig. S1. (A) Detailed procedure of generating a frame of epochs from USGS earthquake catalog database; **(B) Refined frames of epochs with one-day interval.** 10-years data before the target earthquake (EQ) are used for prediction rule-learning. This paper focuses on prediction of target EQ 30 days before.

2 Summary of equations and formulas of salient steps

Table S1. Key Equations and Formulas

[Data Transformation-I] Spatio-Temporal Information Index

From USGS (*18*), prepare raw data matrix of all past EQs, $\{\lambda, \phi, -h, M\}_i^{(t)}$

- $II_{local} \in \mathbb{R}[0,1]$, ${II_{local}}^{(t)}(\mathbf{x}_i^{(t)}) = M_i^{(t)}/10$ where $\mathbf{x}_i^{(t)} = (x,y,z)_i^{(t)}$
- $\overline{II}_S^{(t)}(\boldsymbol{\xi}_j; L_k) = \int_{\mathrm{V}} \omega(\boldsymbol{\xi}_j, \mathbf{x}_i^{(t)}; L_k) II_{local}^{(t)}(\mathbf{x}_i^{(t)}) d\mathbf{x}$
  where $\omega(\boldsymbol{\xi}_j, \mathbf{x}_i^{(t)}; L_k) = (L_k(2\pi)^{1/2})^{-N} \exp\left(-\frac{|\mathbf{x}_i^{(t)} - \boldsymbol{\xi}_j|^2}{2L_k^2}\right) = \mathcal{N}(\mathbf{x}_{(i)}^{(t)}, L_k^2)$
- $\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l) = \int \omega(\tau; T_l) \overline{II}_S^{(t_{past})}(\boldsymbol{\xi}_j; L_k) dt_{past}$
  where $\omega(\tau; T_l) = (T_l(2\pi)^{1/2})^{-1} \exp\left(-\frac{\tau^2}{2T_l^2}\right) = \mathcal{N}(t, T_l^2); \tau = |t - t_{past}|, t \geq t_{past}$

[Data Transformation-II] Pseudo Physics Features $\mathcal{U}$

- Pseudo Energy: $E_r^{*(t)}(\boldsymbol{\xi}_j) = \max\left[\sum_{k=1}^{n_L=2} \sum_{l=1}^{n_T=2} \mathcal{L}^{(k,l)}(\overline{II}_{ST}^{(t)}(\boldsymbol{\xi}_j; L_k, T_l);\ \boldsymbol{\theta}^{(k,l)}), 0.0\right]$
- Pseudo Power: $E_r' = \frac{\partial E_r^{(t)}(\boldsymbol{\xi}_j)}{\partial t}$
- Pseudo Vorticity: $\boldsymbol{\omega} := \nabla_g \times \left(\nabla_g \frac{\partial E_r^{(t)}(\boldsymbol{\xi}_j)}{\partial t}\right)$
  $= \left(\frac{\partial}{\partial \phi}\frac{\partial E_r'}{\partial h} - \frac{\partial}{\partial h}\frac{\partial E_r'}{\partial \phi},\ \frac{\partial}{\partial h}\frac{\partial E_r'}{\partial \lambda} - \frac{\partial}{\partial \lambda}\frac{\partial E_r'}{\partial h},\ \frac{\partial}{\partial \lambda}\frac{\partial E_r'}{\partial \phi} - \frac{\partial}{\partial \phi}\frac{\partial E_r'}{\partial \lambda}\right)$
  where $\nabla_g \mathrm{f}(\boldsymbol{\xi}_j) = \mathbf{J}\nabla \mathrm{f}(\boldsymbol{\xi}_j)$ (see full details in (*17*))
- Pseudo Laplacian: $\nabla_g^2 E_r^{(t)}(\boldsymbol{\xi}_j) = \frac{\partial^2 E_r'}{\partial \lambda^2} + \frac{\partial^2 E_r'}{\partial \phi^2} + \frac{\partial^2 E_r'}{\partial h^2}$

[Data Transformation-III] Gauss Curvature-Based Features $\mathbb{K}$

- $\kappa_1 = H + C;\ \kappa_2 = H - C$
  where $H = \frac{GL - 2FM + EN}{2(EG - F^2)}$ and $C = \sqrt{A^2 + B^2}$ (see full details in (*17*))

[Data Transformation-IV] Fourier Transform-Based Features $\mathbb{F}$

- Full details are presented in Table 1 of this paper

3 Link functions of Fourier transform-based new features

$$\mathcal{L}_{FT}({\kappa_2}^{(E)}) = \sum_{i=1}^{10}(c_i \cos(2\pi \overline{f}_i^{(2)} \kappa_2^{(E)}) + d_i \sin(2\pi \overline{f}_i^{(2)} \kappa_2^{(E)})) \tag{12}$$

$$\mathcal{L}_{FT}({\kappa_1}^{(P)}) = \sum_{i=1}^{10}(e_i \cos(2\pi \overline{f}_i^{(3)} \kappa_1^{(P)}) + f_i \sin(2\pi \overline{f}_i^{(3)} \kappa_1^{(P)})) \tag{13}$$

$$\mathcal{L}_{FT}({\kappa_2}^{(P)}) = \sum_{i=1}^{10}(g_i \cos(2\pi \overline{f}_i^{(4)} \kappa_2^{(P)}) + h_i \sin(2\pi \overline{f}_i^{(4)} \kappa_2^{(P)})) \tag{14}$$

$$\mathcal{L}_{FT}({\kappa_1}^{(V)}) = \sum_{i=1}^{10}(o_i \cos(2\pi \overline{f}_i^{(5)} \kappa_1^{(V)}) + p_i \sin(2\pi \overline{f}_i^{(5)} \kappa_1^{(V)})) \tag{15}$$

$$\mathcal{L}_{FT}({\kappa_2}^{(V)}) = \sum_{i=1}^{10}(q_i \cos(2\pi \overline{f}_i^{(6)} \kappa_2^{(V)}) + r_i \sin(2\pi \overline{f}_i^{(6)} \kappa_2^{(V)})) \tag{16}$$

$$\mathcal{L}_{FT}({\kappa_1}^{(L)}) = \sum_{i=1}^{10}(s_i \cos(2\pi \overline{f}_i^{(7)} \kappa_1^{(L)}) + t_i \sin(2\pi \overline{f}_i^{(7)} \kappa_1^{(L)})) \tag{17}$$

$$\mathcal{L}_{FT}({\kappa_2}^{(L)}) = \sum_{i=1}^{10}(u_i \cos(2\pi \overline{f}_i^{(8)} \kappa_2^{(L)}) + v_i \sin(2\pi \overline{f}_i^{(8)} \kappa_2^{(L)})) \tag{18}$$

Table S2. Recap of the Bayesian Evolutionary Algorithm

Three-Fold Error Function ($\mathcal{J}$) in terms of Magnitude, Location, and False Alarms

- $\mathcal{J} = (1 - a_{cnt}) \sum_{\tilde{k} \in Top} \omega_{MD}^{(\tilde{k})} E_{MD}^{(\tilde{k})} / n(Top) + a_{cnt} E_{cnt}$

  where definitions of terms are summarized below.

  Magnitude & distance error $E_{MD}^{(\tilde{k})} := a_M \text{erf}\left(\frac{|M_{obs}^{(t+1)}(\boldsymbol{\xi}_{\tilde{k}}) - M_{pred}^{(t+1)}(\boldsymbol{\xi}_{k^*})|}{M_{obs}^{(t+1)}(\boldsymbol{\xi}_{\tilde{k}})} \right) + (1 - a_M) \text{erf}\left(\frac{||\boldsymbol{\xi}_{\tilde{k}} - \boldsymbol{\xi}_{k^*}||_2}{r_{max}} \right)$

  False alarm error $E_{cnt} := \frac{1}{2} \text{erf}\left(\frac{|n(Top) - n(Top_{pred})|}{n(Top)} \right) + \frac{1}{2} \text{erf}\left(\sum_{\forall \tilde{k} \in Top_{pred}^{-1}} \frac{|M_{thr} - M_{pred}^{(t+1)}(\boldsymbol{\xi}_{\tilde{k}})| / M_{thr}}{n(Top_{pred}^{-1})} \right)$

  $Top :=$ a set of indices of reference volumes that contains the sorted real magnitudes $> M_{thr}$

  $Top_{pred} :=$ a set of indices of reference volumes of sorted predicted magnitudes $> M_{thr}$

  Top_{pred}^{-1} is the complement of the set Top_{pred}

  $k^*(\tilde{k}) \in Top_{pred}$ is the closest reference volume to $\tilde{k} \in Top$; $k^*(\tilde{k}) := \text{argmin}_{\forall k \in Top_{pred}} ||\boldsymbol{\xi}_{\tilde{k}} - \boldsymbol{\xi}_k||_2$

  $\text{erf}(z) := \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ where $z \in \mathbb{C}$ and $\text{erf}(z) \in [-1, 1]$

  Magnitude-dependent scale-up factor $\omega_{MD}^{(\tilde{k})} := \exp(M_{obs}^{(t+1)}(\boldsymbol{\xi}_{\tilde{k}})/10.0);\ \tilde{k} \in Top$

  Other coefficients and hyper-parameters: $a_M = 0.5, a_{cnt} = 0.1, r_{max} = 200$ km,

  $M_{thr} = 6.8$ for $M_w \geq 7.0$ group and $M_{thr} = 6.3$ for $M_w \in [6.5, 7.0)$ group

(for full explanations and details see (*17*))

Rule-Learning by Bayesian Evolutionary Algorithm

Learning $M_{FT\&CRS}^{(t+1)}(\boldsymbol{\xi}_j)$ in Eq. (5) means to learn all free parameters Θ of all LFs, $\Theta = (\theta_E, \theta_P, \theta_\omega, \theta_L, \theta_{FT})$ via repetition of solution space exploration, random mutation, fitness-proportionate spawning, and Bayesian inheritance.

- Fitness: $\mathcal{F}(s) = \frac{(1 + \mathcal{J}(s))^{-1}}{\sum_{\forall s \in S} [(1 + \mathcal{J}(s))^{-1}]}$

  where s denotes an individual in the entire generation S

- Bayesian Fitness Score: $\mathcal{F}_B(s) = \frac{1}{\kappa} \frac{\mathcal{F}(s; S^*) \mathcal{F}^*(s)}{\sum_{\forall s \in S^*} \mathcal{F}^*(s)}$

  where $\kappa = \sum_{\forall s \in S^*} \frac{\mathcal{F}(s; S^*) \mathcal{F}^*(s)}{\sum_{\forall s \in S^*} \mathcal{F}^*(s)}$

- Probability for selection: $\text{Prob}(\text{parent}_k | s) \propto \mathcal{F}_B(s),\ (k = 1, 2)$

4 Summary of Fourier transform-based new feature generation

As explained in main text, the coordinate vector K consists of the principal Gauss curvatures (κ_1, κ_2) of four pseudo physics quantities at time t at a reference volume $\boldsymbol{\xi}$ as

$$K(t; \boldsymbol{\xi}_j) := ((\kappa_1, \kappa_2)_E, (\kappa_1, \kappa_2)_P, (\kappa_1, \kappa_2)_V, (\kappa_1, \kappa_2)_L)) \tag{19}$$

where E, P, V and L stand for the pseudo released energy, the pseudo power, the pseudo vorticity's first term, and the pseudo Laplacian's first term, respectively, all calculated at time t and a reference volume $\boldsymbol{\xi}_j$. The set of Gauss curvature-based coordinates at $\boldsymbol{\xi}_j$ up to the present time $t_n = n \times \Delta t$ is given by

$$\mathbb{K}(n; \boldsymbol{\xi}_j) := \{K(t; \xi_j) \in \mathbb{R}^8 |\ t = 1, ..., t_n\} \in \mathbb{R}^{(n \times 8)} \tag{20}$$

where Δt is the sampling interval, one day in this paper. Regarding $\mathbb{K}(n; \boldsymbol{\xi}_j) \in \mathbb{R}^{(n \times 8)}$ as a matrix, the m_{th} column, denoted as $\mathbf{k}^{(m)}(n; \boldsymbol{\xi}_j)$ $(m = 1, ..., 8)$, corresponds to the time series of a principal Gauss curvature of a pseudo physics quantity. For instance, $\mathbf{k}^{(1)}(n; \boldsymbol{\xi}_j)$ the 1st column of $\mathbb{K}(n; \boldsymbol{\xi}_j)$, means the time series of κ_1 of the pseudo released energy up to this time t_n at a reference volume $\boldsymbol{\xi}_j$ whereas the 8th column $\mathbf{k}^{(8)}(n; \boldsymbol{\xi}_j)$ means the time series of κ_2 of the pseudo Laplacian's first term.

To generate the Fourier transform-based new features, this study performed the column-wise fast Fourier transform (FFT), i.e., each column of $\mathbb{K}(n; \boldsymbol{\xi}_j)$. The FFT generates the resultant set F_{PSD} consisting of the power spectral densities $\mathbf{p}_{PSD} \in \mathbb{R}^n$ and the associated frequencies $\mathbf{f} \in \mathbb{R}^n$. In short,

$$\mathbb{K}(n; \boldsymbol{\xi}_j) \xrightarrow[column-wise]{FFT} F(n; \boldsymbol{\xi}_j) := \{\mathbf{p}_{PSD}^{(1)}, \mathbf{f}^{(1)}, ..., \mathbf{p}_{PSD}^{(8)}, \mathbf{f}^{(8)}\}. \tag{21}$$

Then, for each column, we can remove the DC component and sort the column vectors in descending order with respect to the magnitude of PSD. The resulting sorted set is denoted as

$\overline{F}$ given by

$$F(n;\boldsymbol{\xi}_j) \xrightarrow[by\ PSD]{Sort} \overline{F}(n;\boldsymbol{\xi}_j) := \{\overline{\mathbf{p}}_{PSD}^{(1)}, \overline{\mathbf{f}}^{(1)}, ..., \overline{\mathbf{p}}_{PSD}^{(8)}, \overline{\mathbf{f}}^{(8)} |\ \overline{p}_{i-1}^{(m)} \geq \overline{p}_{i}^{(m)}, \forall i \in [1,n]\} \tag{22}$$

where $\overline{\mathbf{p}}_{PSD}^{(m)} = (\overline{p}_1^{(m)}, ..., \overline{p}_n^{(m)})^{\mathrm{T}}$. Thus, $\overline{p}_i^{(m)}$ is the i_{th} entity of the sorted column vector $\overline{\mathbf{p}}_{PSD}^{(m)}$ in descending order. $\overline{\mathbf{f}}^{(m)}$ is the sorted frequency vector according to the $\overline{\mathbf{p}}_{PSD}^{(m)}$. In general, $\overline{\mathbf{f}}^{(m)} \neq \overline{\mathbf{f}}^{(m')}$ and $\overline{\mathbf{f}}^{(m)} \neq \mathbf{f}^{(m)}$ for all $m \neq m'$ since the sorting took place column-wise with respect to individual column's PSD.

To generate practically meaningful features, amongst many peaks in the power spectra, this paper extracted the top 10 magnitudes and the associated frequencies. FT-based new feature set is denoted as $\mathbb{F}(n;\boldsymbol{\xi}_i) \in \mathbb{R}^{10\times 16}$,

$$\overline{F}(n;\boldsymbol{\xi}_j) \xrightarrow[Top10]{Extract} \mathbb{F}(n;\boldsymbol{\xi}_i) := \{\overline{\mathbf{p}}_{top}^{(1)}, \overline{\mathbf{f}}_{top}^{(1)}, ..., \overline{\mathbf{p}}_{top}^{(8)}, \overline{\mathbf{f}}_{top}^{(8)} |\ \overline{\mathbf{p}}_{top}^{(m)} \subset \overline{\mathbf{p}}_{PSD}^{(m)}, \overline{\mathbf{f}}_{top}^{(m)} \subset \overline{\mathbf{f}}^{(m)}\} \tag{23}$$

In particular, $\overline{\mathbf{f}}_{top}^{(1)}$ and $\overline{\mathbf{f}}_{top}^{(2)}$ corresponds to the pseudo released energy; $\overline{\mathbf{f}}_{top}^{(3)}$ and $\overline{\mathbf{f}}_{top}^{(4)}$, to the pseudo power; $\overline{\mathbf{f}}_{top}^{(5)}$ and $\overline{\mathbf{f}}_{top}^{(6)}$, to the pseudo vorticity; $\overline{\mathbf{f}}_{top}^{(7)}$ and $\overline{\mathbf{f}}_{top}^{(8)}$, to the pseudo Laplacian.

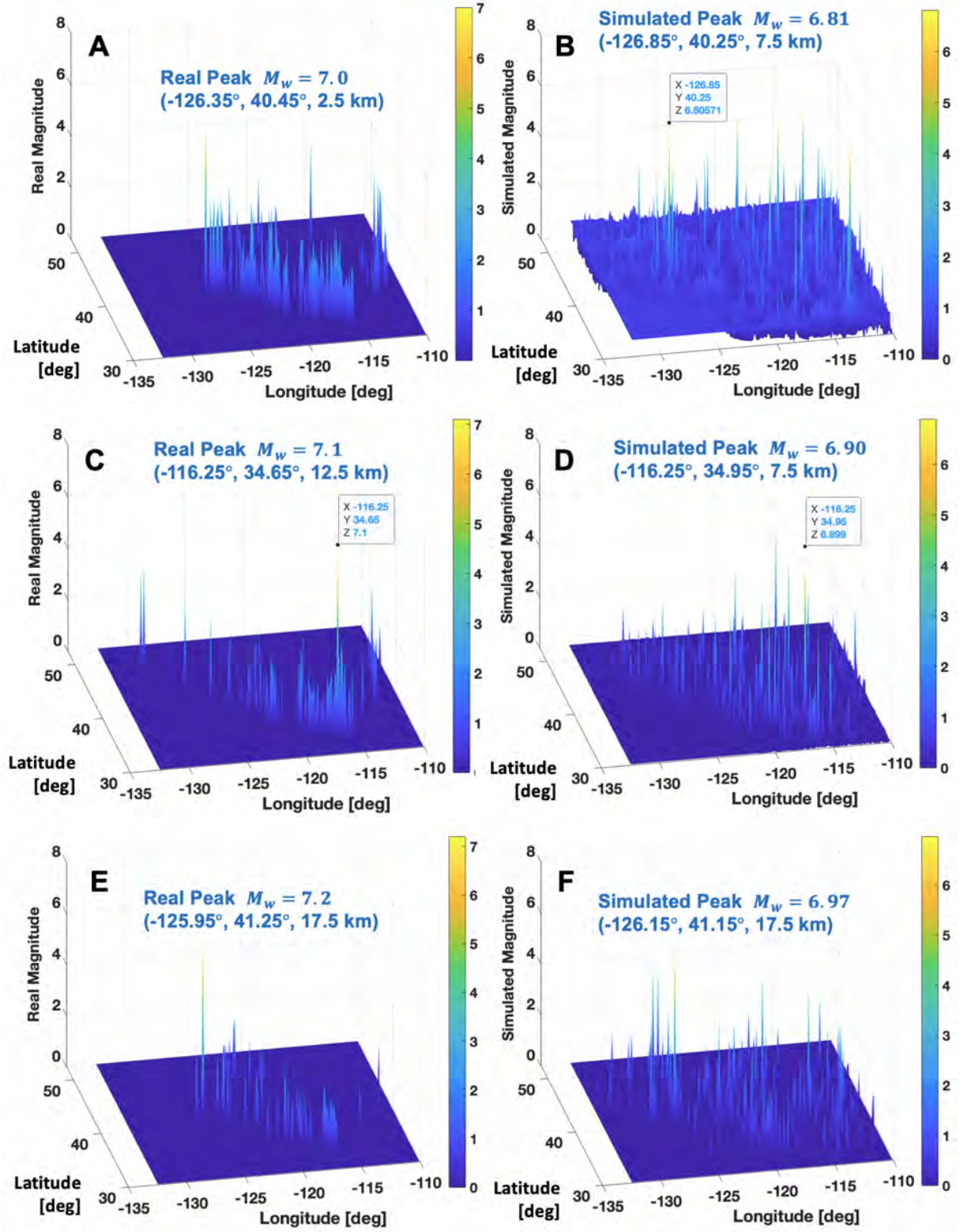

Fig. S2. Reproduction of large magnitude events $M_w > 7.0$ by using the customized ML-identified data-driven prediction rules with FT-based new features: (A-B) Observed real and simulated earthquake events on 1994/9/1 (target day ID 1005357); **(C-D)** 1999/10/16 (1007228); **(E-F)** 2005/6/15 (1009297).

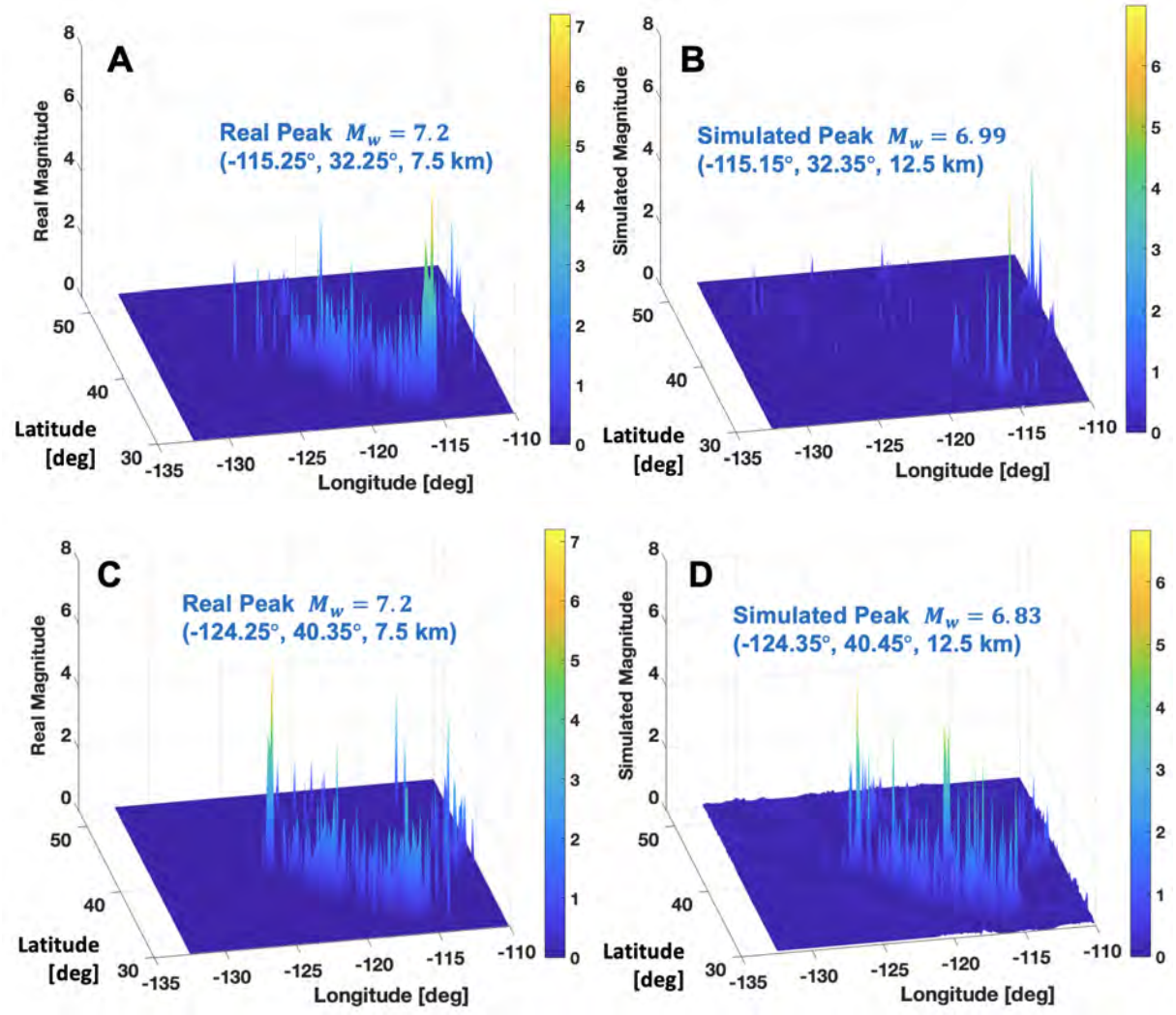

Fig. S3. Reproduction of large magnitude events $M_w > 7.0$ by using the customized ML-identified data-driven prediction rules with FT-based new features: (A-B) Observed real and simulated earthquake events on 2010/4/4 (target day ID 1011051); **(C-D)** 1992/4/25 (1004498).

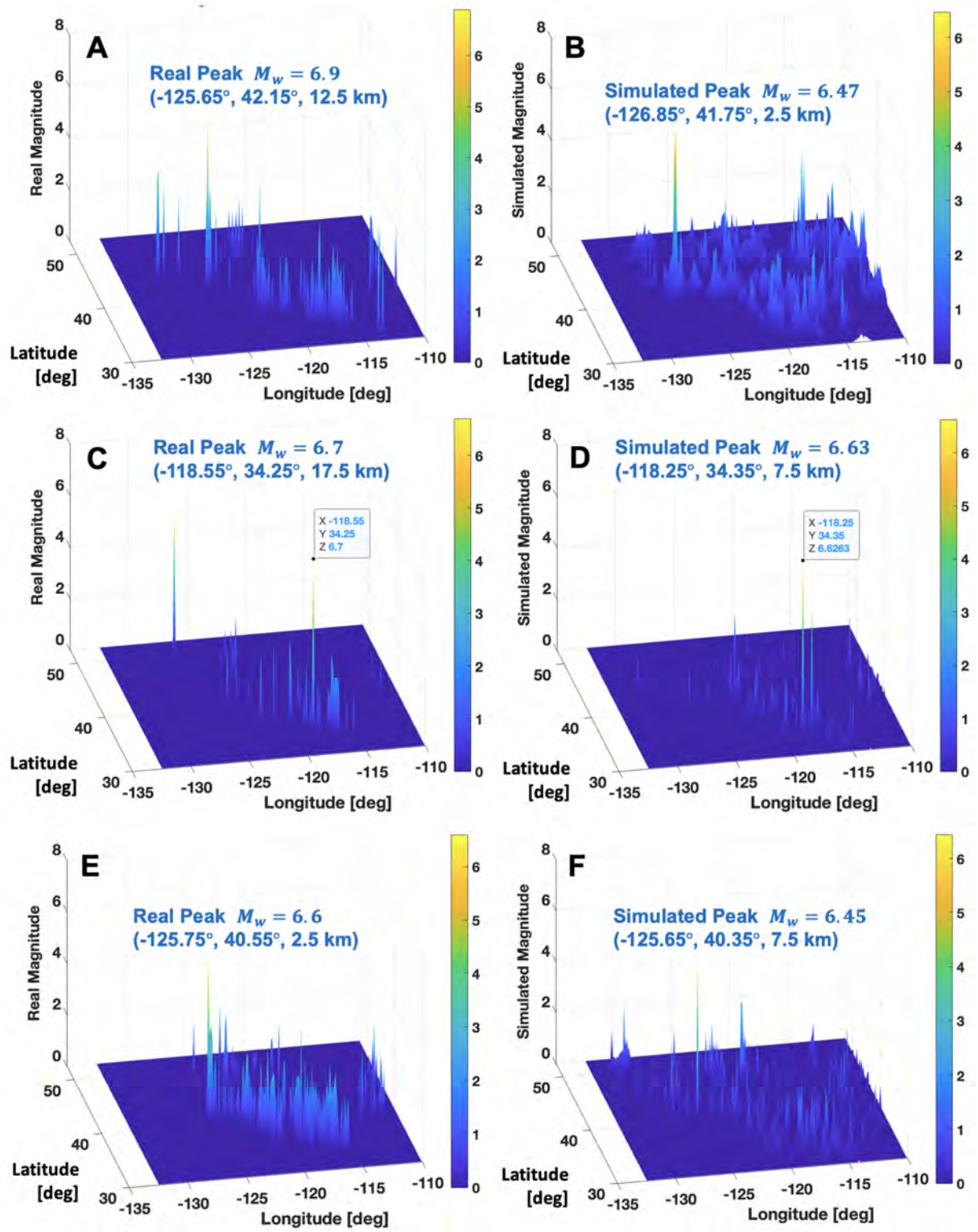

Fig. S4. Reproduction of large magnitude events $M_w \in [6.5, 7.0)$ by using the customized ML-identified data-driven prediction rules with FT-based new features: (A-B) Observed real and simulated earthquake events on 1994/7/13 (target day ID 1004211); **(C-D)** 1994/1/17 (1005130); **(E-F)** 1995/2/19 (1005528).

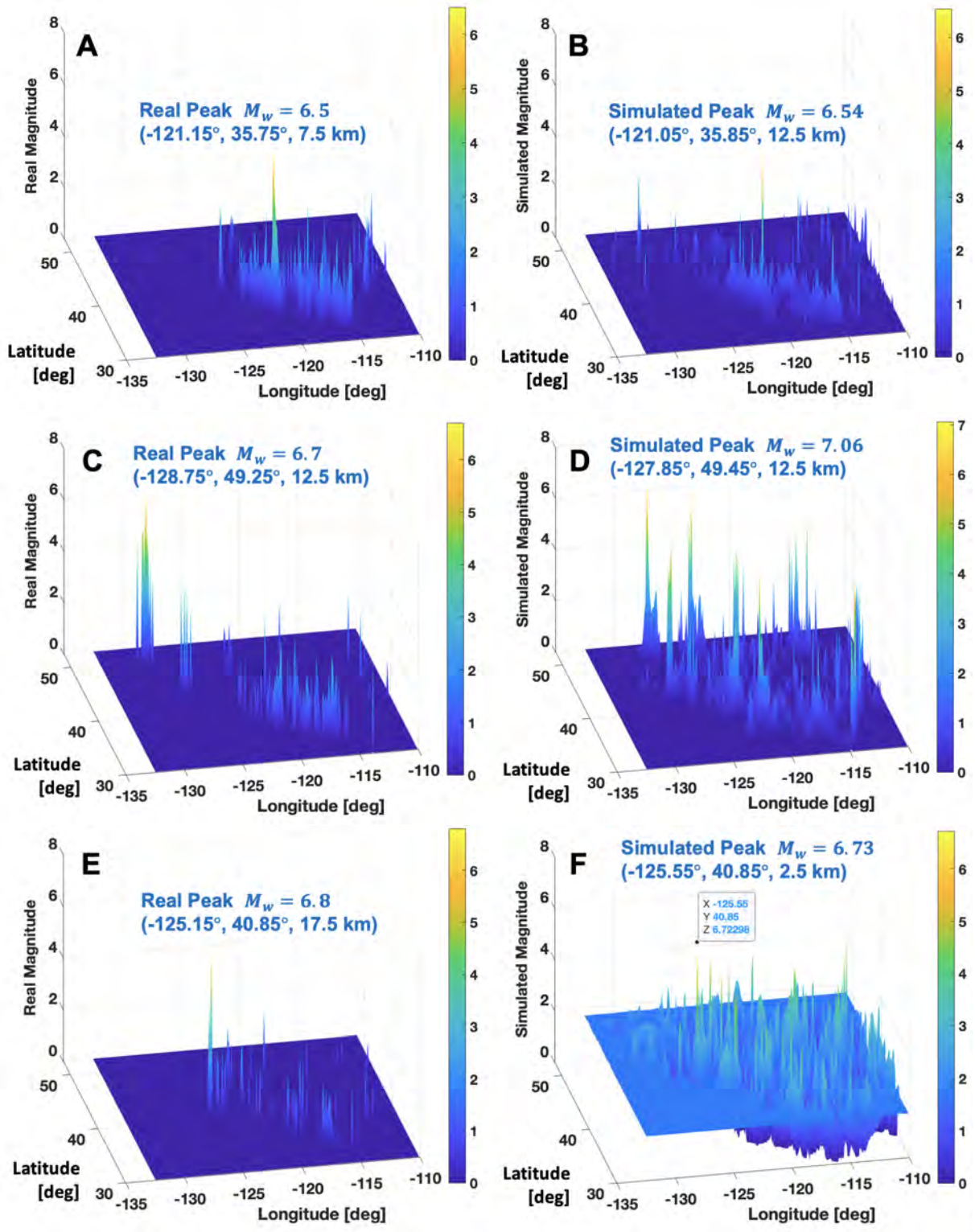

Fig. S5. Reproduction of large magnitude events $M_w \in [6.5, 7.0)$ by using the customized ML-identified data-driven prediction rules with FT-based new features: (A-B) Observed real and simulated earthquake events on 2003/12/22 (target day ID 1008756); **(C-D)** 2004/11/2 (1009072); **(E-F)** 2014/3/10 (1012487).

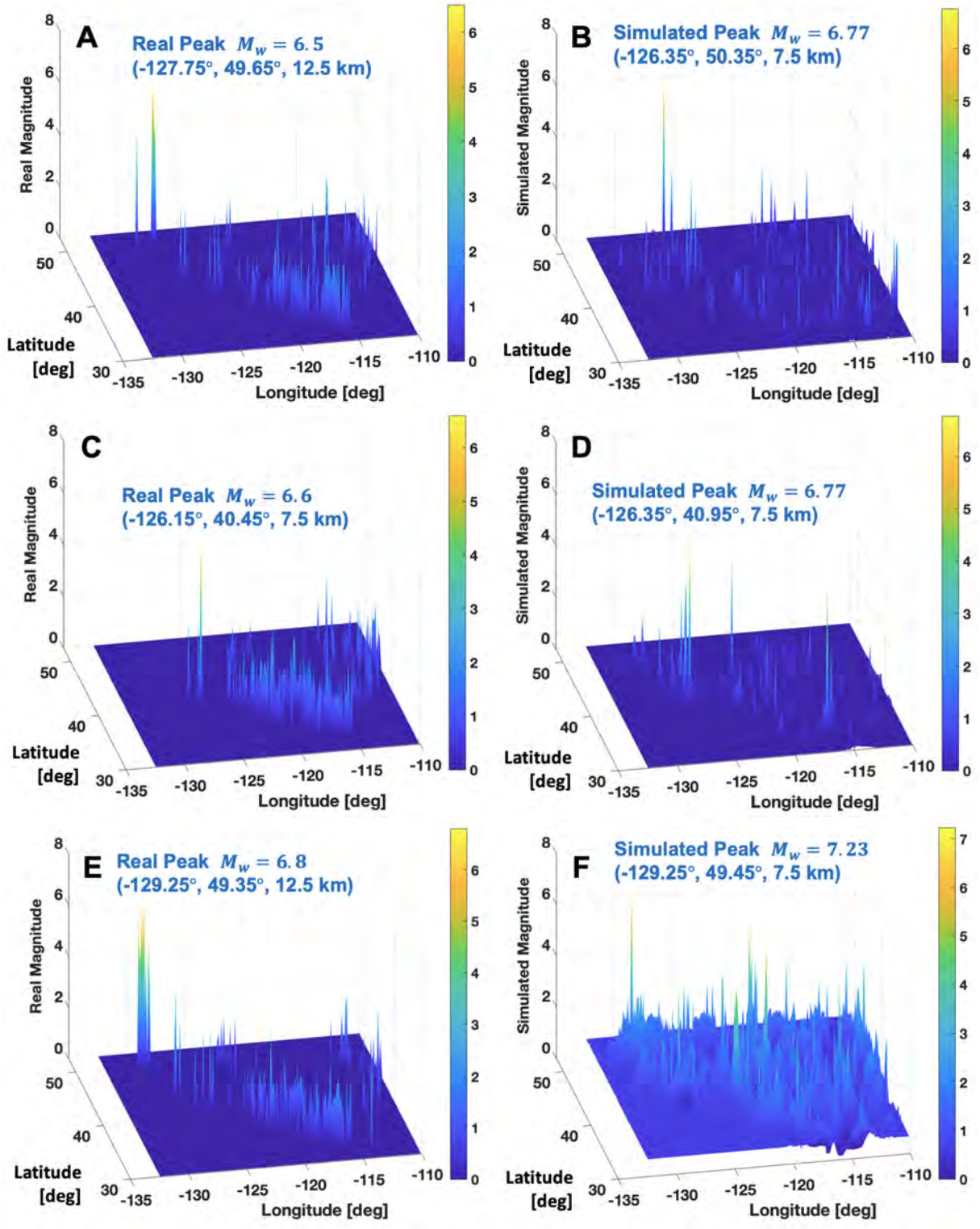

Fig. S6. Reproduction of large magnitude events $M_w \in [6.5, 7.0)$ by using the customized ML-identified data-driven prediction rules with FT-based new features: (A-B) Observed real and simulated earthquake events on 2014/4/24 (target day ID 1012532); **(C-D)** 2016/12/8 (1013491); **(E-F)** 2018/10/22 (1014174).

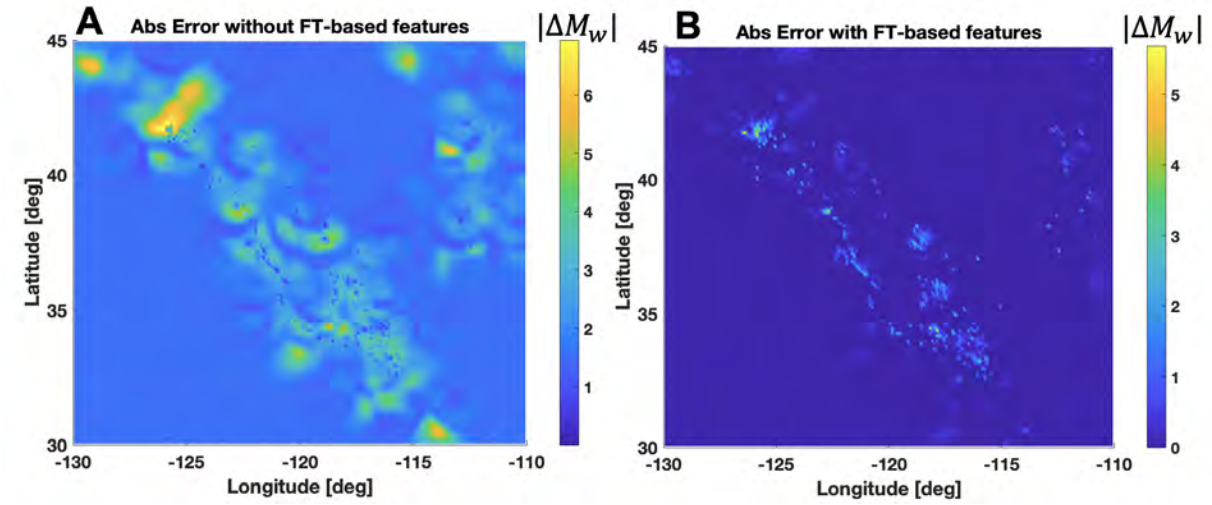

Fig. S7. Absolute errors between real and predicted magnitudes of 1991/8/17 EQ (target day ID 1004246): (A) By using the best-so-far rules without FT-based features; **(B)** By the rules with FT-based features.

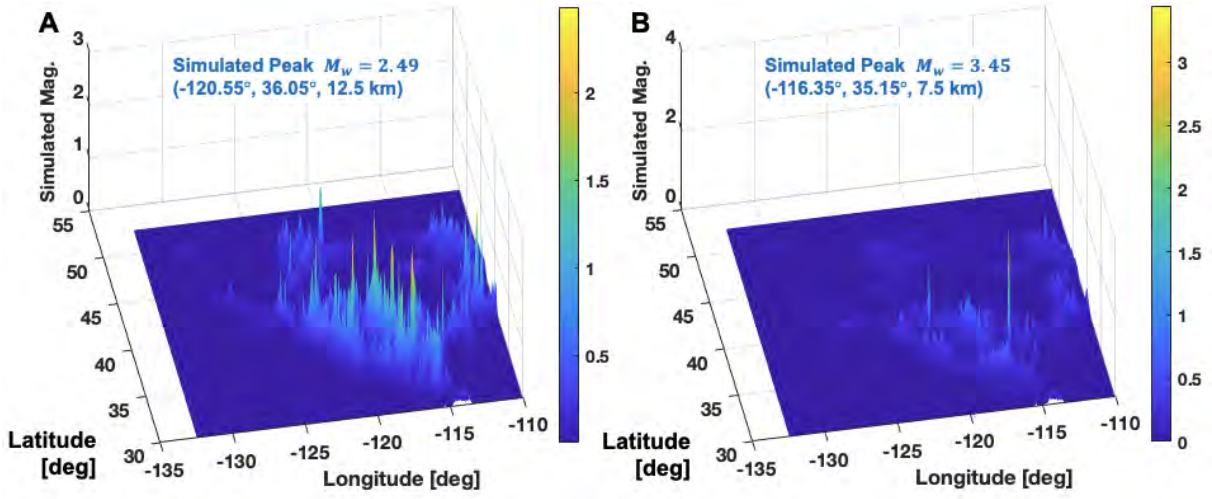

Fig. S8. Preliminary test predictions of small EQs ($M_w < 5.5$) **period by using the best-so-far rules: (A)** Predictions 33 days before and **(B)** 32 days before the Ridgecrest EQ (2019/7/6; $M_w = 7.1$). No false alarms with spurious large EQ predictions are detected, but a concrete conclusion requires comprehensive test over all size EQs in the future extension.