Supplement of Nat. Hazards Earth Syst. Sci., 22, 1931–1954, 2022
https://doi.org/10.5194/nhess-22-1931-2022-supplement

Natural Hazards and Earth System Sciences
Open Access
EGU

Supplement of

Hidden-state modeling of a cross-section of geoelectric time series data can provide reliable intermediate-term probabilistic earthquake forecasting in Taiwan

Haoyu Wen et al.

Correspondence to: Haoyu Wen (s160049@e.ntu.edu.sg)

Contents

Section S1. Handling of Sudden Spikes in V TSs

When we first visualized the geoelectric TSs, we noticed that for many stations, there are sudden spikes in the geoelectric TS, such as the NS-direction of FENG station as shown in Fig. S1(a). We believe these outliers are due to unknown technical errors and were not correct recordings of the actual geoelectric fields. Such spikes lead to similar spikes in the V TS, as shown in Fig. S1(c); whereas C, S, K TSs (see Figs S1(b), (d), (e)) are not susceptible to this effect. To improve the data quality of the V TS, we applied a simple procedure of imposing an artificial upper bound on the V TS, which is the 99^{th} percentile value across the whole V TS, and thereafter reduce all V values exceeding this upper bound down to this upper bound value. As a result, the processed V TS in Fig. S1(f) no longer has spikes. We applied this procedure to the V TSs for all 20 stations, and obtained new V TSs for the subsequent k-means clustering and BWA.

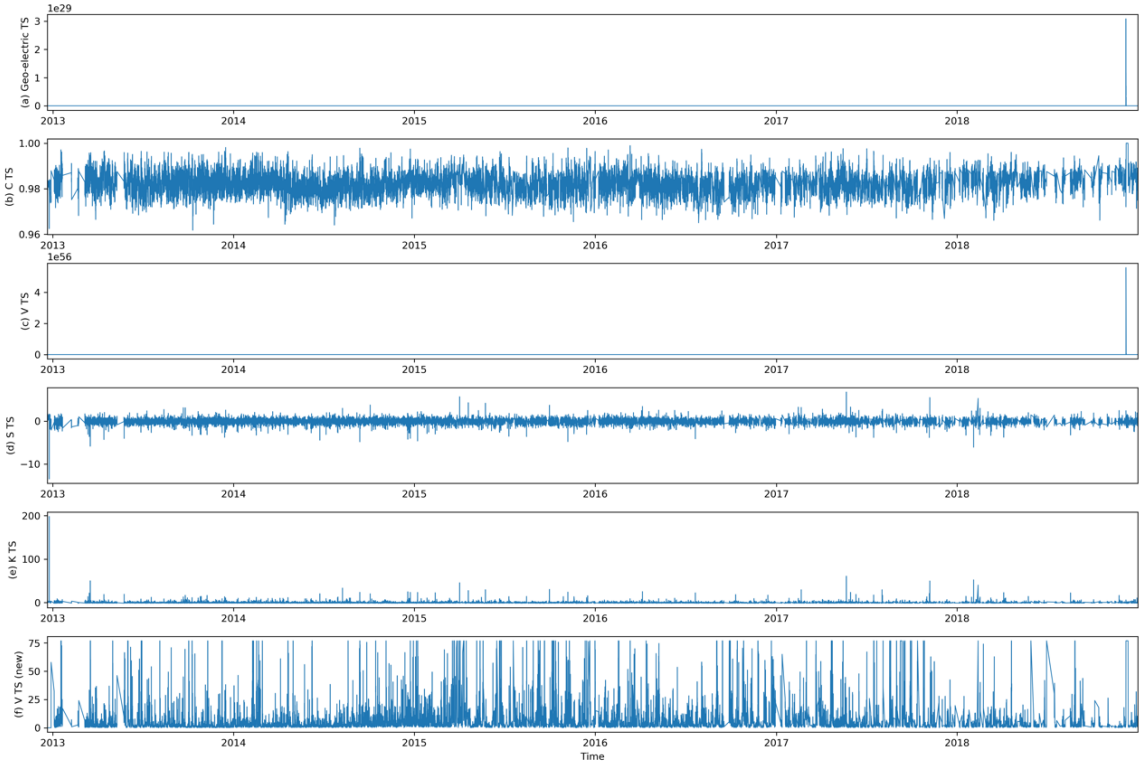

Figure S1. (a) The geoelectric TS of FENG station, NS direction; (b) the corresponding C TS; (c) the corresponding V TS; (d) the corresponding S TS; (e) the corresponding K TS; (f) the new V TS after outlier removal.

Section S2. Simulated Example of a Hidden Markov Model (HMM) and the Baum-Welch Algorithm (BWA)

Before illustrating the HMM, we start with an example of a Markov model, which is a toy model of local weather conditions described by states: $\{S_1, S_2\} = \{Sunny, Rainy\}$, and the transition matrix:

$$\mathbf{A} = \begin{array}{c} \\ Sunny \\ Rainy \end{array} \begin{array}{cc} Sunny\ tomorrow & Rainy\ tomorrow \\ 0.9 & 0.1 \\ 0.25 & 0.75 \end{array},$$

which tells us that there is a 90% chance of it being sunny tomorrow, if it is sunny today, etc. To estimate the transition probability $p_{j \leftarrow i}$, we observe the states of the system for a long time, and divide the total number of times we find the transition $j \leftarrow i$ by the number of transitions starting from i.

To illustrate how an HMM works using the same example, suppose we are observing a remote friend's daily activity, without knowing the local weather. Suppose each day, he/she only chooses between going out and staying home, therefore the observable set $\{O_1, O_2\}$ is $\{Out, Stay\}$. We call the conditional probability $P(o_t = O_q | s_t = S_l)$ of observing O_q given the HS S_l at any time t as the emission probability. The set of all emission probabilities can be organized into an emission matrix $\mathbf{B}$, where $\mathbf{B}(l, k) = P(o_t = O_q | s_t = S_l)$. For this example, we can write the emission matrix $\mathbf{B}$ as:

$$\mathbf{B} = \begin{array}{c} \\ Sunny \\ Rainy \end{array} \begin{array}{cc} Out & Stay \\ 0.8 & 0.2 \\ 0.1 & 0.9 \end{array},$$

which tells us that if the local weather is sunny, he/she goes out with an 80% chance, etc. Additionally, if we know that there is a 20% probability that the first day is rainy, we can write the initial state distributions as $\boldsymbol{\pi_0} = (0.8, 0.2)$.

With this toy model, we can run a simple simulation of 100 days. This is illustrated in Fig. S2, where we show both the HS TS and the observation TS. Visually, we can indeed notice that our friend staying home more on rainy days, and going out more on sunny days.

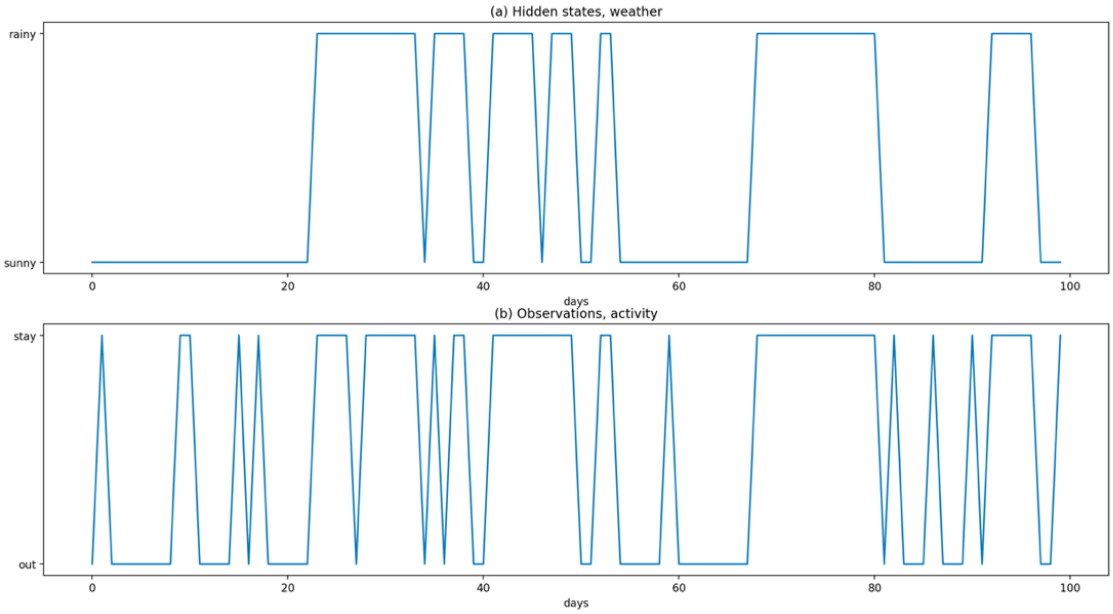

Figure S2. A 100-day simulation of the toy weather HMM, showing (a) the HS TS of weather, and (b) the observation TS of activity.

To illustrate the BWA, we applied it on the weather-activity model. Instead of only having 100 days, this time we simulated the model for 10,000 days to accumulate more data so that BWA can work properly. We then ran the BWA for 100 iterations. In the end, the estimated transition matrix $\widetilde{\mathbf{A}}$ is

$$\widetilde{\mathbf{A}} = \begin{array}{c} \\ Sunny \\ Rainy \end{array} \begin{array}{cc} Sunny\ tomorrow & Rainy\ tomorrow \\ 0.9172 & 0.0828 \\ 0.2532 & 0.7468 \end{array},$$

and the estimated emission matrix $\widetilde{\mathbf{B}}$ is

$$\widetilde{\mathbf{B}} = \begin{array}{c} \\ Sunny \\ Rainy \end{array} \begin{array}{cc} Out & Stay \\ 0.7963 & 0.2038 \\ 0.0852 & 0.9148 \end{array}.$$

This result agrees well with the actual model parameters. The algorithm also yielded highly-accurate posterior probabilities $P(rainy\ at\ t|\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\boldsymbol{\pi}}_{\mathbf{0}})$ for $t = 1, 2, \ldots, 100$, as shown in Fig. S3. The posterior probability is accurate if it is close to 1 for rainy days, and close to 0 for sunny days. We see that this is true during those rainy periods that are widely separated by the sunny state, and less so during time periods where there are rapid transitions between the sunny and rainy states. From this example, we get a feel of the power of the BWA given sufficiently many observations.

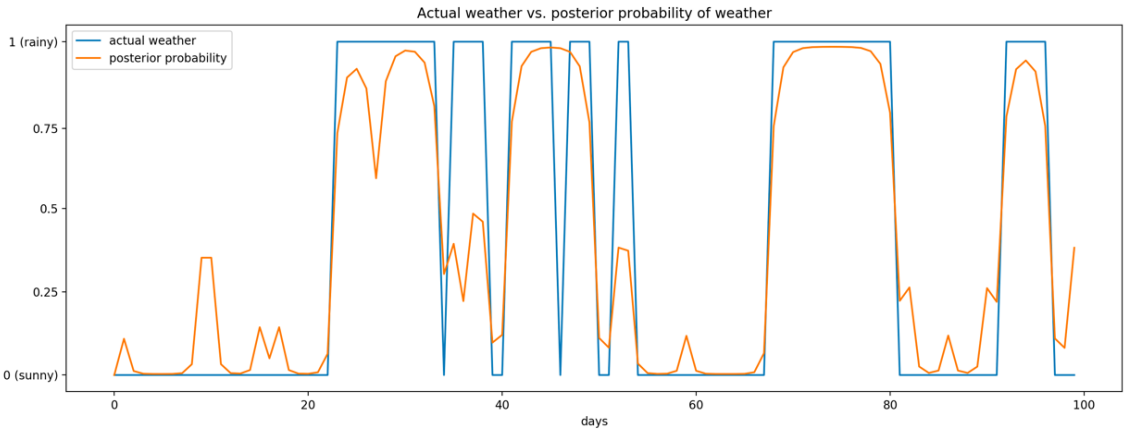

Figure S3. The rainy-state posterior probability (orange curve) estimated by the BWA for the first 100 days of a 10,000-day simulation. In this figure, we also show the actual weather conditions in the simulation (blue curve).

Section S3. Mathematical Details of K-means Clustering

In the k-means clustering of the set of N continuous-valued vectors $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \dots, \boldsymbol{W}_N\}$, we start by choosing $Q \leq N$ clusters: $\boldsymbol{G} = \left\{G_1^{(0)}, G_2^{(0)}, \dots, G_Q^{(0)}\right\}$, where cluster $G_q^{(0)}$ is initialized with a random center $\boldsymbol{\mu}_q^{(0)}$, and Q is the number of total clusters we choose for the k-means clustering. We then assign each vector $\boldsymbol{W}_n$ to a cluster $G_q^{(0)}$: $G_q^{(0)} \equiv G_q^{(0)} \cup \boldsymbol{W}_n$, such that the squared Euclidean norm $\left\|\boldsymbol{W}_n - \boldsymbol{\mu}_q^{(0)}\right\|^2$ is minimized for $\boldsymbol{W}_n$. After assigning all vectors in $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \dots, \boldsymbol{W}_N\}$ this way, $G_q^{(0)} = \left\{\boldsymbol{W}_{q,1}, \boldsymbol{W}_{q,2}, \dots, \boldsymbol{W}_{q,n_q}\right\}$ would contain n_q feature vectors. We can improve on this initial clustering by updating the position of the centers by:

$$\boldsymbol{\mu}_q^{(1)} = \frac{1}{n_q} \sum_{i=1}^{n_q} \boldsymbol{W}_{q,i},$$

and re-assigning the N continuous-valued vectors $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \dots, \boldsymbol{W}_N\}$ to these new clusters $\boldsymbol{G} = \left\{G_1^{(1)}, G_2^{(1)}, \dots G_Q^{(1)}\right\}$ with updated centers. After repeating this procedure for enough times, the clusters will converge to $\boldsymbol{G}^* = \left\{G_1^*, G_2^*, \dots G_Q^*\right\}$, where $G_q^* = \left\{\boldsymbol{W}_{q,1}^*, \boldsymbol{W}_{q,2}^*, \dots, \boldsymbol{W}_{q,n_q^*}^*\right\}$. Ultimately, the k-means clustering algorithm ensures that the sum of Q within-cluster sum of squares (WCSS) for each cluster is minimized, which can be written as:

$$\underset{\boldsymbol{G}^*}{argmin} \sum_{q=1}^{Q} \sum_{\boldsymbol{W}_{q,i}^* \in G_q^*} \left\|\boldsymbol{W}_{q,i}^* - \boldsymbol{\mu}_q^*\right\|^2.$$

Section S4. Mathematical Details of Baum Welch Algorithm

At each iteration i, the forward procedure computes the probability $\alpha_{l,t}^{i} = P\big(o_1, o_2, \dots, o_t, s_t = S_l | (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})\big)$ that the observations up to time t are $o_1, o_2, \dots, o_t$, and the HS s_t at time t takes on the value S_l, given the model parameters $(\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})$. This is done by setting $\alpha_{l,0}^{i} = \boldsymbol{\pi_l}\mathbf{B_i}(l, o_0)$, and computing $\alpha_{l,t+1}^{i} = \mathbf{B_i}(l, o_{t+1}) \sum_{m=1}^{2} \alpha_{m,t}^{i} \mathbf{A_i}(m, l)$ for all l and t. The backward procedure computes the probability $\beta_{l,t}^{i} = P(o_{t+1}, \dots, o_T | s_t = S_l, (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i}))$ that the rest of the observations are $o_{t+1}, \dots, o_T$ given that $s_t = S_l$ and model parameters $(\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})$. This is done by setting $\beta_{l,T}^{i} = 1$, and computing $\beta_{l,t}^{i} = \sum_{m=1}^{2} \beta_{m,t+1}^{i} \mathbf{A_i}(l, m) \mathbf{B_i}(m, O_{t+1})$ for all l and t.

We then reach the update procedure. We start by calculating the probability $\gamma_{l,t}^{i} = P(s_t = S_l | o_1, o_2, \dots, o_T, (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i}))$, which is the conditional probability of $s_t = S_l$ given the full observation TS and the model parameters $(\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})$. This is computed by:

$$\gamma_{l,t}^{i} = \frac{P\big(o_1, o_2, \dots, o_T, s_t = S_l \big| (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})\big)}{P\big(o_1, o_2, \dots, o_T \big| (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})\big)} = \frac{\alpha_{l,t}^{i} \beta_{l,t}^{i}}{\sum_{m=1}^{2} \alpha_{m,t}^{i} \beta_{m,t}^{i}}.$$

Next, we calculate the probability $\xi_{l,m,t}^{i} = P(s_t = S_l, s_{t+1} = S_m | o_1, o_2, \dots, o_T, (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i}))$, which is the probability of the HS making a transition from S_l to S_m going from time t to $t+1$, given the full observation TS and the model parameters $(\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})$. This is computed by:

$$\xi_{l,m,t}^{i} = \frac{P\big(o_1, o_2, \dots, o_T, s_t = S_l, s_{t+1} = S_m \big| (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})\big)}{P\big(o_1, o_2, \dots, o_T \big| (\mathbf{A_i}, \mathbf{B_i}, \boldsymbol{\pi_i})\big)}$$

$$= \frac{\alpha_{l,t}^{i} \mathbf{A_i}(l, m) \beta_{m,t+1}^{i} \mathbf{B_i}(m, O_{t+1})}{\sum_{r=1}^{2} \sum_{p=1}^{2} \alpha_{r,t}^{i} \mathbf{A_i}(r, p) \beta_{p,t+1}^{i} \mathbf{B_i}(p, O_{t+1})}.$$

Now, we can update the new model parameters as:

1) $$\pi_{i+1} = \gamma_{t=1}^{i}$$

2) $$\mathbf{A_{i+1}}(l, m) = \frac{\sum_{t=1}^{T-1} \xi_{l,m,t}^{i}}{\sum_{t=1}^{T-1} \gamma_{l,t}^{i}}$$

3) $$\mathbf{B_{i+1}}\big(l, O_q\big) = \frac{\sum_{t=1}^{T} 1_{o_t = O_q} \gamma_{l,t}^{i}}{\sum_{t=1}^{T} \gamma_{l,t}^{i}},$$

where $1_{o_t = O_q} = \begin{cases} 1 \;\; if \; o_t = O_q \\ 0 \;\; otherwise \end{cases}$.

Section S5. Visualizing HMMs' Posterior Probability TSs of 15 Random BWA Initializations

In this section, we provide visualizations of the outputs of the BWA for 15 random initializations for each of the 20 stations in Fig. S4. This information may be useful for readers who want to see the complete outputs of the BWA, as well as the posterior probabilities for all 20 stations (the first row for each sub-plot). We noticed that it is not uncommon for the BWA to fail with random initializations. These are shown as the noisy bottom rows in all stations. However, once the BWA manages to converge, it is very likely to yield consistent results. This is supported by the observation of almost identical HSs for well-converged models represented by the first several rows for each station. Additionally, Fig. S4 also provides a clear visualization of how the frequencies of HS transitions differ from station to station. For example, the HSs for LIOQ make rapid transitions and never stay the same for more than half a year, while the HSs for DAHU can remain unchanged for almost three years.

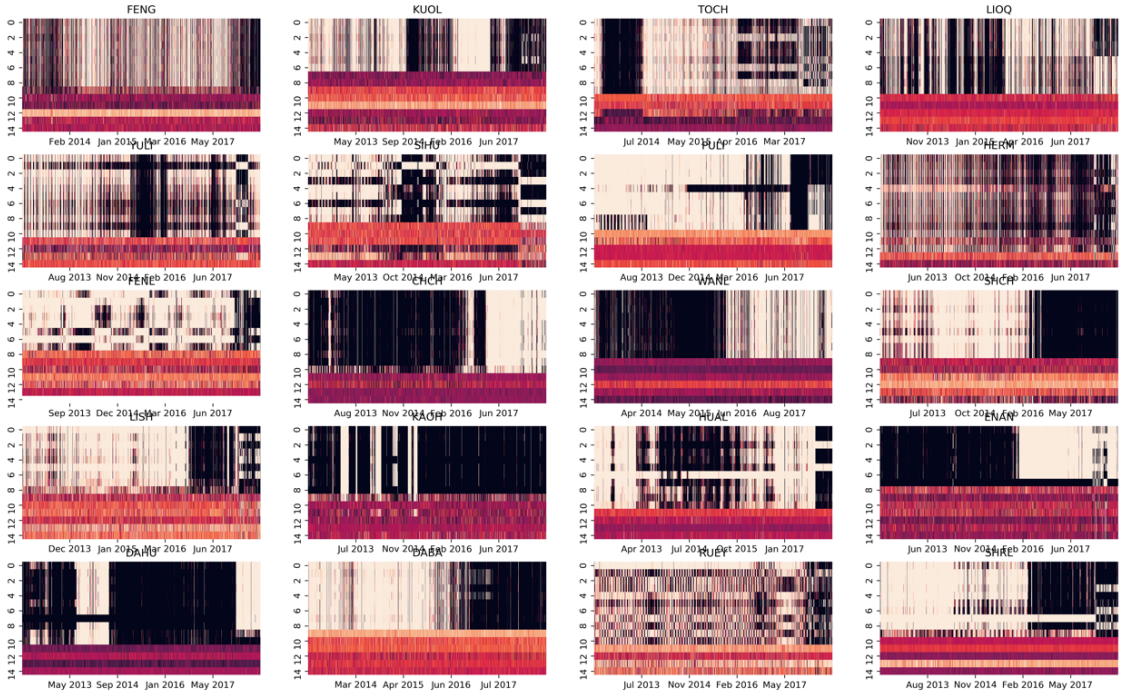

Figure S4. The heatmaps of the 15 HMM models' posterior probability TSs for S_1 for all 20 stations (bright for 0, dark for 1). For each station, TSs are sorted by model score from highest to lowest, with the most well-converged model at index 0 (first row), and those failing to converge as the bottom rows. (Note: the labelling of S_1 and S_2 are not synchronized across 20 stations, and all subplots are obtained with optimal HMM hyper-parameters as indicated in Fig. 12 of the main text.)

Section S6. Optimal HMMs' Distribution of (C, V, S, K) Indexes in S_1 and S_2

In Figs S5-S9, we show the distributions of the 4 indexes in HSs S_1 and S_2 for all 20 stations and both directions (NS and EW), for the HMM obtained with optimal hyper-parameters (available as Fig. 12 in the main text). We note that C and V consistently showing different distributions for S_1 and S_2, while the distributions of S and K in S_1 and S_2 can occasionally be not so different in some stations, such as FENG and CHCH. Nevertheless, from these figures, we can see that all four indexes can be useful for the HMM by showing different distributions for S_1 and S_2.

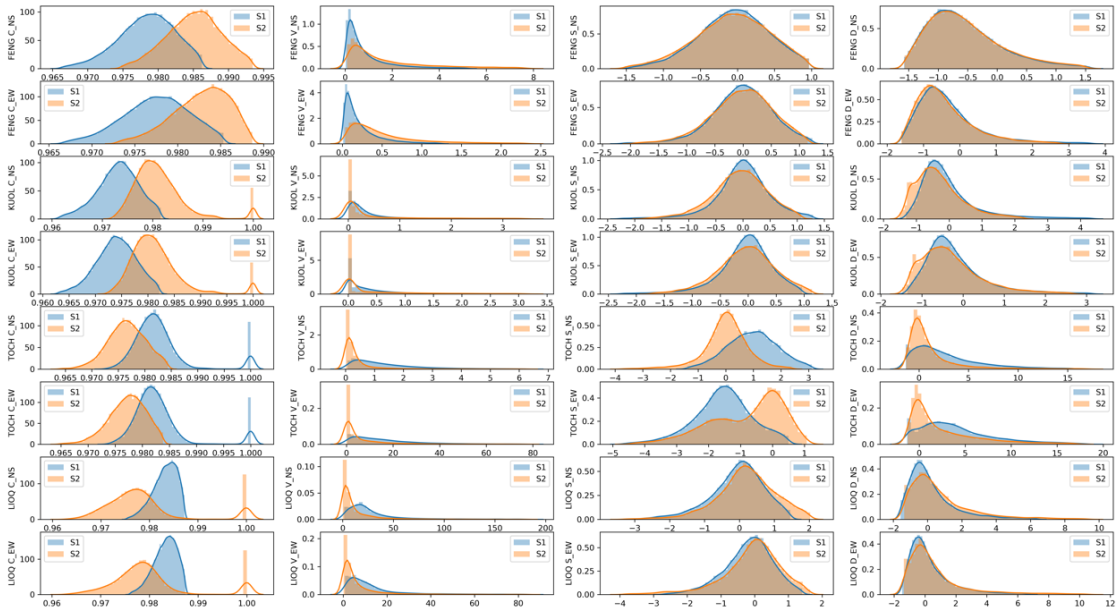

Figure S5. Distributions of indexes (C, V, S, K) in S_1 and S_2, for HMMs obtained with optimal hyper-parameters using geoelectric TSs (both NS and EW) for stations FENG, KUOL, TOCH, and LIOQ.

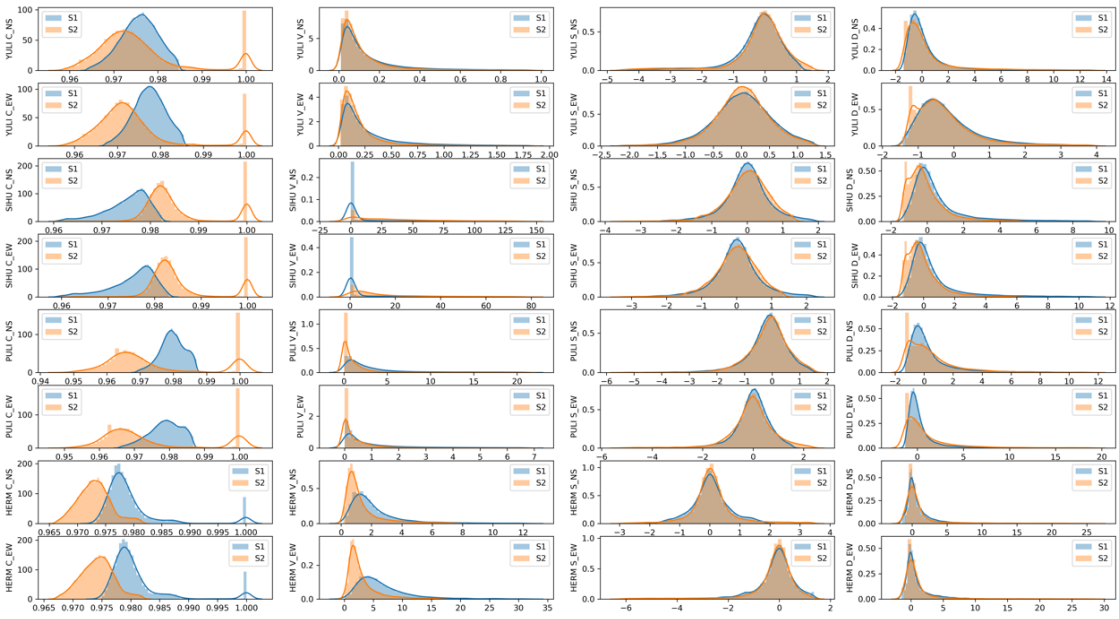

Figure S6. Distributions of indexes (C, V, S, K) in S_1 and S_2, for HMMs obtained with optimal hyper-parameters using geoelectric TSs (both NS and EW) for stations YULI, SIHU, PULI, and HERM.

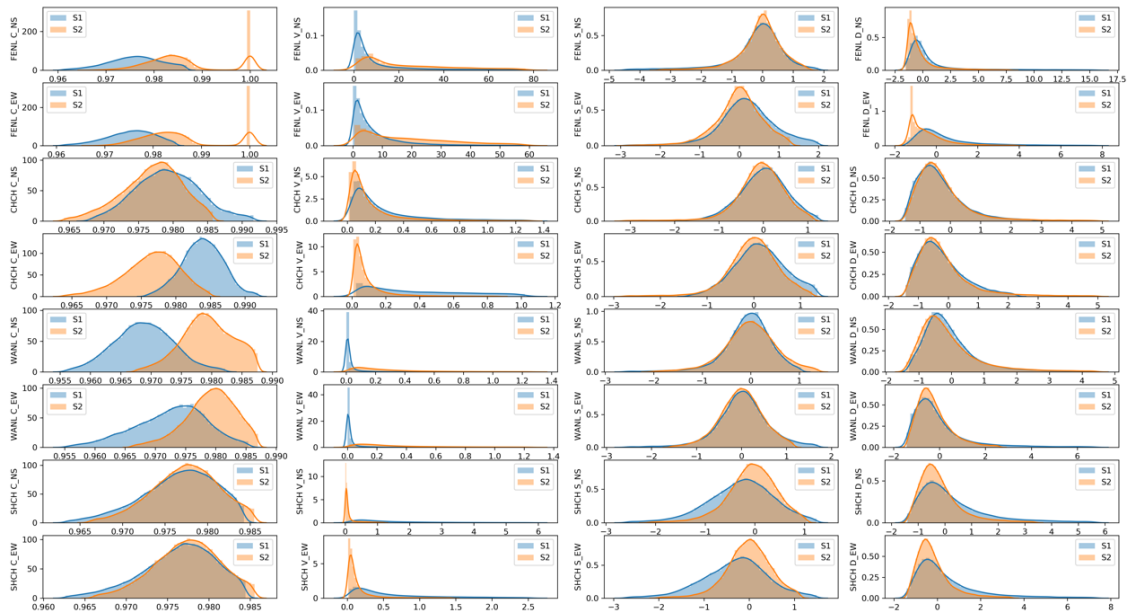

Figure S7. Distributions of indexes (C, V, S, K) in S_1 and S_2, for HMMs obtained with optimal hyper-parameters using geoelectric TSs (both NS and EW) for stations FENL, CHCH, WANL, and SHCH.

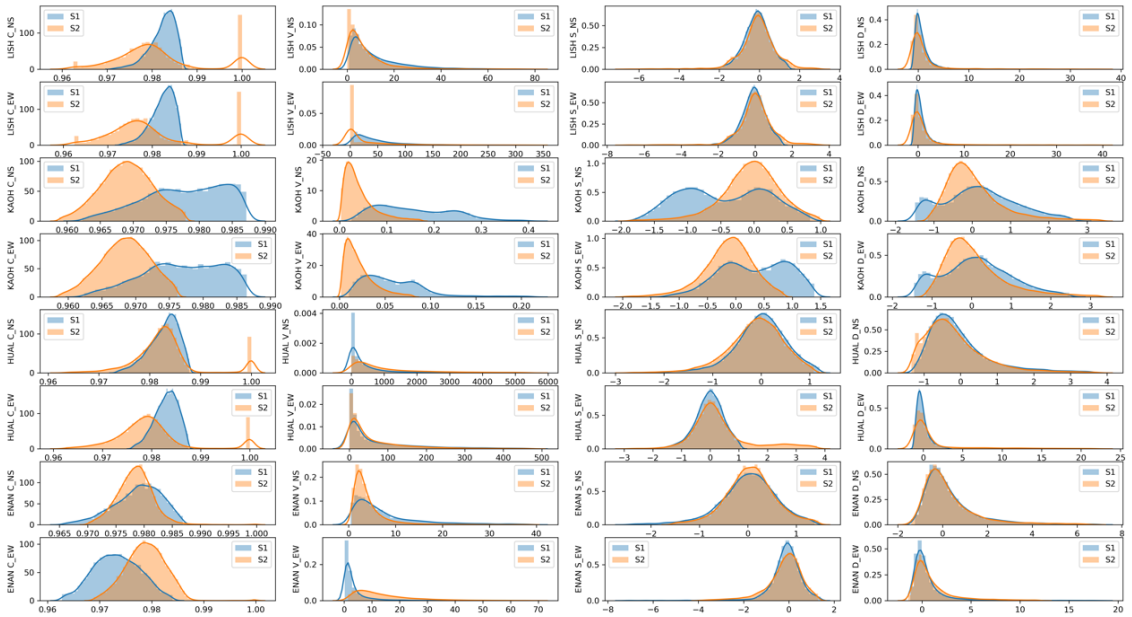

Figure S8. Distributions of indexes (C, V, S, K) in S_1 and S_2, for HMMs obtained with optimal hyper-parameters using geoelectric TSs (both NS and EW) for stations LISH, KAOH, HUAL, and ENAN.

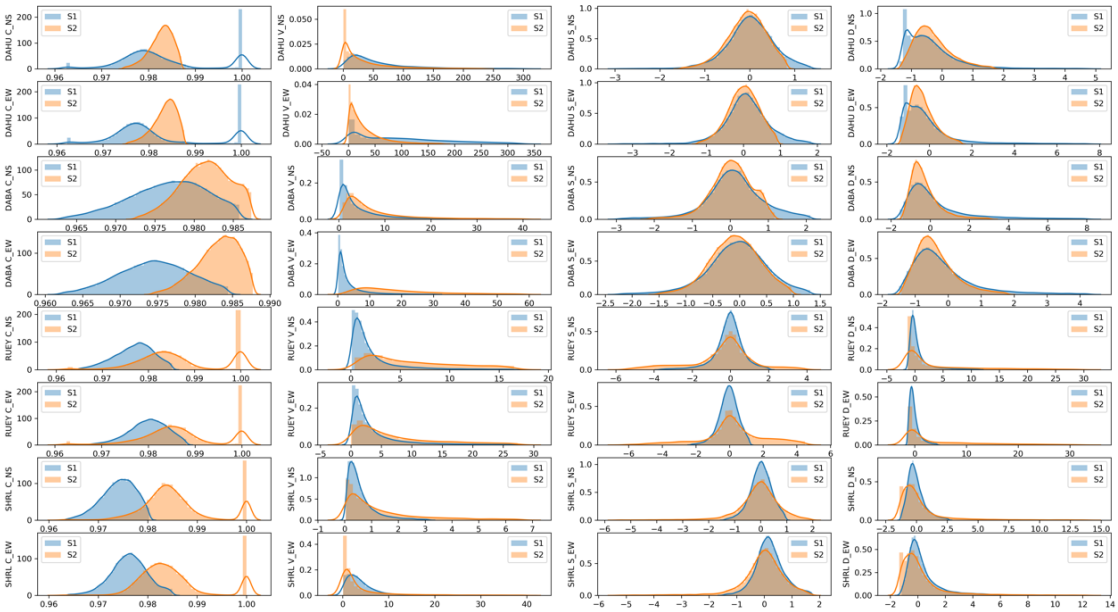

Figure S9. Distributions of indexes (C, V, S, K) in S_1 and S_2, for HMMs obtained with optimal hyper-parameters using geoelectric TSs (both NS and EW) for stations DAHU, DABA, RUEY, and SHRL.

Section S7. EQ Frequency Distributions Across Different Magnitudes for Each HS

In Figs S10-S12, we show the log-scale frequency distributions of EQs across different magnitudes in both HSs. We do this for all 20 stations, focusing on 3 cells with the most EQs on the EQ grid map (Fig. 6 in the main text). For each station and each cell, we call the HS with more EQs the 'active state' and the one with less EQs as the 'passive state'.

When the number of EQs in the active state is significantly higher than that in the passive state (such as Fig. S10: KAOH and ENAN), the active state would also cover most of the larger EQs. ($M > 4.5$ for KAOH and $M > 5$ for ENAN in Fig. S10). On the other hand, there are also cases (PULI, HERM, RUEY in Fig. S10) where the active state failed to cover most of the larger EQs. For PULI and RUEY, it is mostly because the active state does not have significantly more EQs than the passive state. The only intriguing case is HERM, because the active state has many smaller EQs ($M < 5.1$) but fail to cover larger EQs. Nevertheless, for most of cases in Figs S10-S12, the active states would generally cover more larger EQs. This is a favorable attribute for real-world forecasting applications, since a highly passive state would indeed have less EQs across different magnitudes.

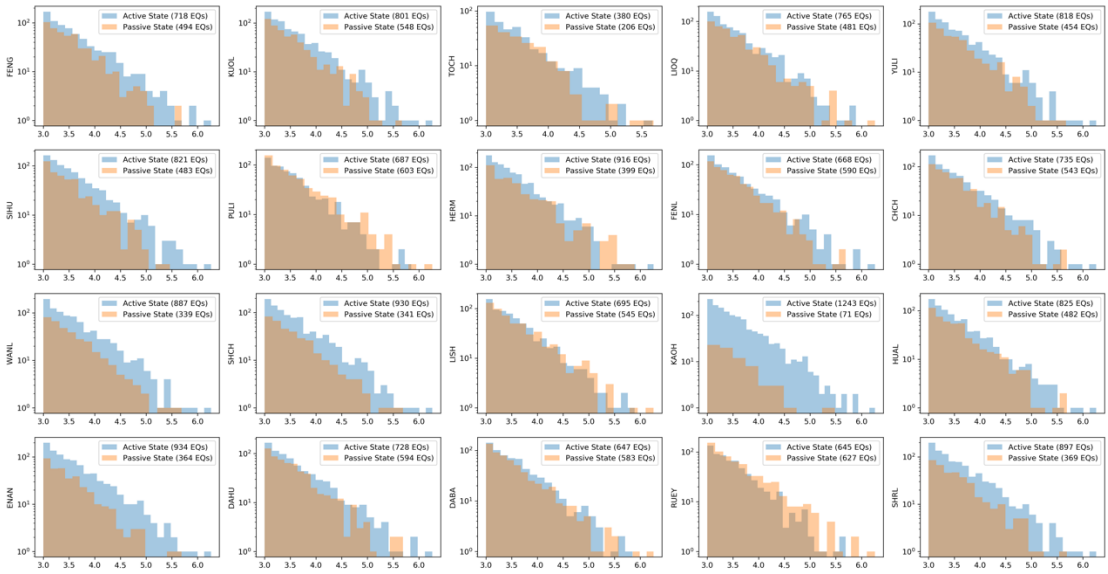

Figure S10. Distribution of log-scale EQ frequency (y-axis) at different magnitudes (x-axis) in both HSs for 20 stations, within a cell at longitude 121.58°E-121.91°E, latitude 23.92°N-24.26°N, containing 1361 EQs.

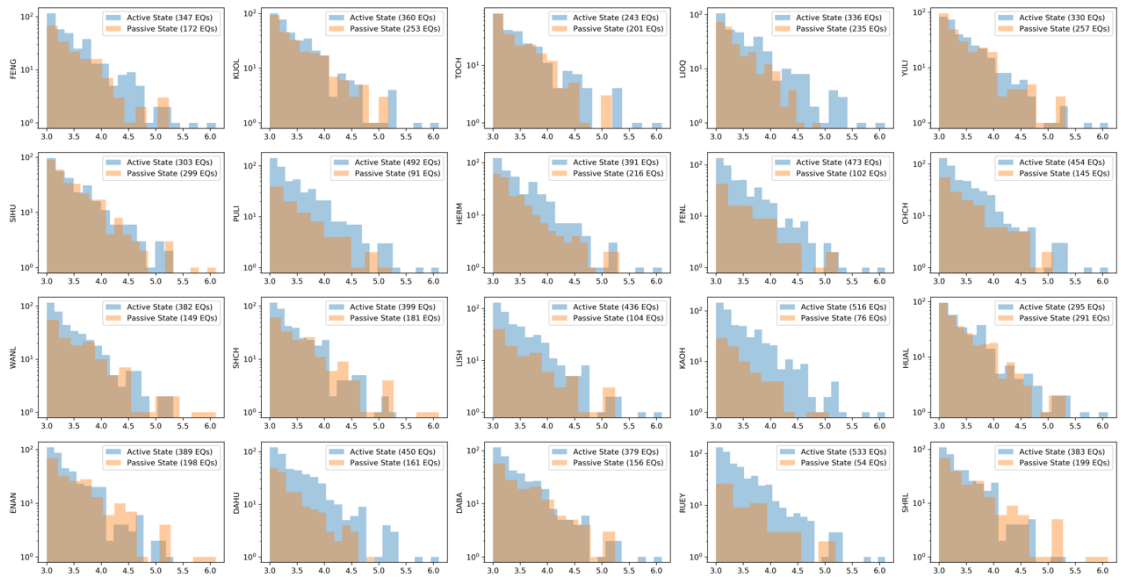

Figure S11. Distribution of log-scale EQ frequency (y-axis) at different magnitudes (x-axis) in both HSs for 20 stations, within a cell at longitude 121.91°E-122.25°E, latitude 24.61°N-24.95°N, containing 622 EQs.

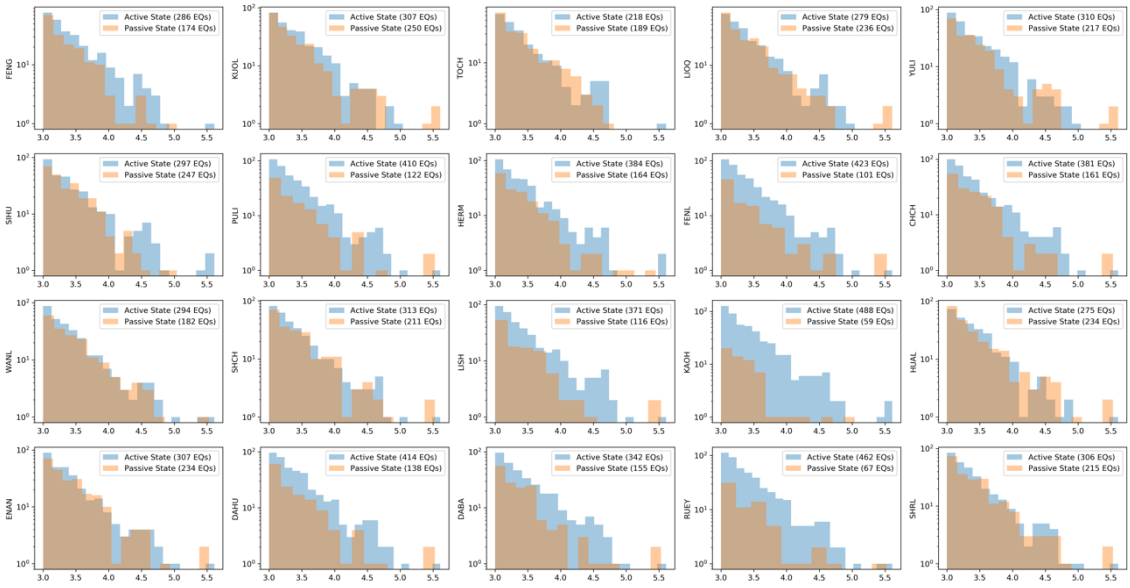

Figure S12. Distribution of log-scale EQ frequency (y-axis) at different magnitudes (x-axis) in both HSs for 20 stations, within a cell at longitude 121.58°E-121.91°E, latitude 24.26°N-24.61°N, containing 561 EQs.