https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences Discussions
Open Access

Hidden-State Modelling of a Cross-section of Geoelectric Time Series Data Can Provide Reliable Intermediate-term Probabilistic Earthquake Forecasting in Taiwan

Haoyu Wen,[1] Hong-Jia Chen,[2] Chien-Chih Chen,[2,3] Massimo Pica Ciamarra,[1] Siew Ann Cheong[1]

[1]Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371
[2]Department of Earth Sciences, National Central University, Taoyuan 32001, Taiwan
[3]Earthquake-Disaster & Risk Evaluation and Management Center, National Central University, Taoyuan 32001, Taiwan

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

Abstract. Geoelectric time series (TS) has long been studied for its potential for probabilistic earthquake forecasting, and a recent model (GEMSTIP) directly used the skewness and kurtosis of geoelectric TS to provide Time of Increased Probabilities (TIPs) for earthquakes in several months in future. We followed up on this work by applying the Hidden Markov Model (HMM) on the correlation, variance, skewness, and kurtosis TSs to identify two Hidden States (HSs) with different distributions of these statistical indexes. More importantly, we tested whether these HSs could separate time periods into times of higher/lower earthquake probabilities. Using 0.5-Hz geoelectric TS data from 20 stations across Taiwan over 7 years, we first computed the statistical index TSs, and then applied the Baum-Welch Algorithm with multiple random initializations to obtain a well-converged HMM and its HS TS for each station. We then divided the map of Taiwan into a 16-by-16 grid map and quantified the forecasting skill, i.e., how well the HS TS could separate times of higher/lower earthquake probabilities in each cell in terms of a *discrimination power* measure that we defined. Next, we compare the *discrimination power* of empirical HS TSs against those of 400 simulated HS TSs, then organized the statistical significance values from these cellular-level hypothesis testing of the forecasting skill obtained into grid maps of *discrimination reliability*. Having found such significance values to be high for many grid cells for all stations, we proceeded with a statistical hypothesis test of the forecasting skill at the global level, to find high statistical significance across large parts of the hyperparameter spaces of most stations. We therefore concluded that geoelectric TSs indeed contain earthquake-related information, and the HMM approach to be capable at extracting this information for earthquake forecasting.

Keywords. Electric properties; statistical methods; time-series analysis; earthquake dynamics; earthquake early warning; earthquake interaction, forecasting, and prediction.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

1 Introduction

Earthquakes (EQs) are one of the most destructive natural hazards that can befall us, with the potential to take many human lives and deal serious damages to economies and environments. On 26 December 2004, an M_w-9.1 (M_w is the moment magnitude scale) earthquake struck near Sumatra, Indonesia, and the subsequent tsunami waves of up to 30 m high resulted in 227,898 people dead or missing. In addition, 1,740,000 people lost their homes to the tsunami in 14 countries (Survey, 2004). This will be remembered as one of the deadliest earthquakes in recorded human history. On 12 May 2008, an M_w-7.9 earthquake in Sichuan, China killed over 69,000 people, and injured 374,176. As of July 2008, another 18,222 were reported missing (Sina-News, 2008). More recently on 12 January 2010, an M_w-7.0 earthquake shook Haiti. By 24 January, this earthquake, along with over 52 aftershocks with magnitude greater than 4.5, had caused the death of 160,000 people (Kolbe et al., 2010), and severe damage to or the collapse of 280,000 buildings (Renois, 2010). This disaster brought the country to bankruptcy, and its people experienced a humanitarian crisis never before encountered. Until we learn how to build earthquake-proof buildings and cities, it is imperative for us to work towards better forecasting/prediction capabilities against EQs, to inform pre-EQ evacuation, post-EQ relief, as well as expediting critical reinforcement works for selected buildings and infrastructures. To achieve this goal, the scientific community has done much work discovering precursors and models that are useful for the forecasting/prediction of EQs.

First, let us clarify that in the seismological community, the terms “prediction” and “forecast” are often used interchangeably (Kagan, 1997; Ismail-Zadeh, 2013). When they are distinguished, the term “prediction” emphasizes the issuing of an *alarm* with high accuracy and reliability indicating the time, location, and magnitude of the next large EQ (Geller et al., 1997), whereas the term “forecast” is a statement about the probability of EQ(s) within the specified spatial-temporal window (Ismail-Zadeh, 2013). Till this day, it is extremely difficult to make accurate and specific EQ predictions (Geller et al., 1997). However, the forecasting of EQs is a far more tractable task: a method that performs better than random guesses (the null hypothesis) is recognized as having *predictive power* or *predictive skill* (“prediction” and “forecast” used as synonym here) (Kagan, 1997). In this paper, we will also use the two terms interchangeably.

If we categorize EQ forecasting methods according to their time scales, we can organize them into three categories: long-term (decades ahead), intermediate-term (a few years ahead), and short-term (days or a few months ahead) (Peresan et al., 2005; Kanamori, 2003). EQ forecasting at different time scales serve different purposes. For a region of interest, a long-term EQ forecasting aims to estimate the probabilities of large EQs in the next decades or more. In most past studies, the primary input data was the historical EQ catalog, which allowed statistical modellings of the occurrence times of large and medium sized EQs (Kagan and Jackson, 1994; Sykes, 1996; Papazachos et al., 1987; Papadimitriou, 1993; Papazachos et al., 1997), assuming that EQs’ occurrences in the same spatial area follow a Poisson process of relatively constant rate. One such example is the probabilistic seismic hazard assessment (PSHA) first established by Cornell in 1968 (Cornell, 1968). This became a popular method for long-term seismic hazard assessment implemented in many countries (Tavakoli and Ghafory-Ashtiany, 1999; Petersen, 1996; Meletti et al., 2008; Vilanova and Fonseca, 2007; Nath and Thingbaijam, 2012; Wang et al., 2016). In this method, we take into account both historical EQ catalog information, as well as ground motion characteristics for the modelling of energy attenuation over spatial distances, thus providing a map of seismic hazard rates that varies across location for the next 50 years. Long-term EQ forecasting such as PSHA can be valuable for location-specific seismic risk evaluation, thereby providing guidelines or criteria for local construction projects. For example, a building that is expected to last 100 years must be able to withstand 10 large EQs of the magnitude that occurs once every 10 years locally. What long-term EQ forecasting does not do, would be to tell people how to do things differently at any time.

For intermediate-term EQ forecasting, the aim is to detect deviations of EQ rates from their long-time values, to assess increased probabilities of EQs within the next one to ten years. For example, if a region usually has a magnitude-6 EQ every 10 years, and 15 years have passed without one, the region would be in a state of increased probability. A famous example for the intermediate-term EQ forecasting is the M8 algorithm (Kossobokov et al., 2002; Peresan et al., 2005; Keilis-Borok, 1996), developed by Healy et al. (1992). The M8 algorithm used the EQ catalog as input, and returned as output the Time of Increased Probability (TIP) for EQs of magnitude 7.5 and above for the next one year. Another example is the CN algorithm (Peresan et al., 2005; Keilis-Borok, 1996) developed by Keilis-Borok and Rotwain (1990), that also took the EQ catalog as input to produce as output TIP for strong EQs (defined specifically for different regions) within the next half to a few years. In the literature, we also found the self-organizing spinodal (SOS) model (Chen, 2003; Rundle et al., 2000), which used the increased activity of medium-sized EQs as precursors to large EQs that could occur within the next several years or decades. Finally, one of the more successful methods at this time scale is pattern informatics (Nanjo et al., 2006), which was demonstrated to be effective at predicting $M \geq 5$ EQs in Japan between 2000 and 2009. Intermediate-term EQ forecasting can, for example, help local authorities prioritize inspections and reinforcements of old buildings over the construction of new ones.

Short-term EQ forecasting use a variety of methods to forecast the time, place, and magnitude of a specific large EQ. Here we commonly find methods using the EQ catalog as input data, and apply machine learning approaches (Asim et al., 2017; Reyes et al., 2013), as well as Hidden Markov Model (HMM) approaches (Yip et al., 2018; Chambers et al., 2012). For example, in (Chambers et al., 2012) an HMM was trained to track the waiting time between EQs with magnitudes above 4 in southern California and western Nevada (Yip et al., 2018), giving EQ forecasts for up to ten days in the future. Apart from using EQ catalog data, there is an increased variety of methods using other data inputs, such as the widely used Seismic Electric Signals (SESs) (Uyeda et al., 2000; Varotsos et al., 2013; Varotsos et al., 2002; Varotsos et al., 2017; Varotsos and Lazaridou, 1991; Varotsos et al., 1993), to look for EQ precursors in the form of abnormal changes to the geoelectric potential. In addition to looking for specific SES-type precursors, we also found papers using methods such as artificial neural networks (ANNs) (Moustra et al., 2011), Fisher Information (Telesca et al., 2005a; Telesca et al., 2009), and multi-fractal analysis (Telesca et al., 2005b) directly on geoelectric time series (TSs) data to make short-term EQ forecasting. Other data that can be used include the combination of geoelectric and magnetic data (Kamiyama et al., 2016; Sarlis, 2018), GPS crustal movements (Kamiyama et al., 2016; Wang and Bebbington, 2013), electromagnetics of the atmosphere (Hayakawa and Hobara, 2010), and lithosphere dynamics (Shebalin et al., 2006). Short-term EQ forecasting can guide emergency responses such as evacuations and pre-emptive relief efforts, although they are usually not reliable enough based on our current level of understandings.

Among all these precursors, our recent research interest was on the potential use of geoelectric TSs for

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

EQ forecasting (Chen and Chen, 2016; Chen et al., 2020; Jiang et al., 2020; Telesca et al., 2014; Chen et al., 2017). In 2016 and 2017, Chen and his colleagues (Chen and Chen, 2016; Chen et al., 2017) analyzed the data of 20 geoelectric stations in Taiwan (Fig. 1) and studied the association between skewness and kurtosis of the geoelectric data and $M_L \geq 5$ EQs, where M_L is the Richter magnitude scale. Through statistical analyses, they found significant correlations between geoelectric anomalies and these large EQs. They then developed an EQ forecasting algorithm named GEMSTIP to extract TIPs for future EQs. TIPs were identified through differences in the distributions of skewness and kurtosis with those found during normal periods. Moreover, Jiang et al. (2020) investigated the geoelectric signals before, during, and after EQs by the shifting correlation method, and found that the lateral and vertical electrical resistivity variation and subsurface conductors might amplify SESs, which agreed with the findings by Sarlis (Sarlis et al., 1999) and Huang (Huang and Lin, 2010).

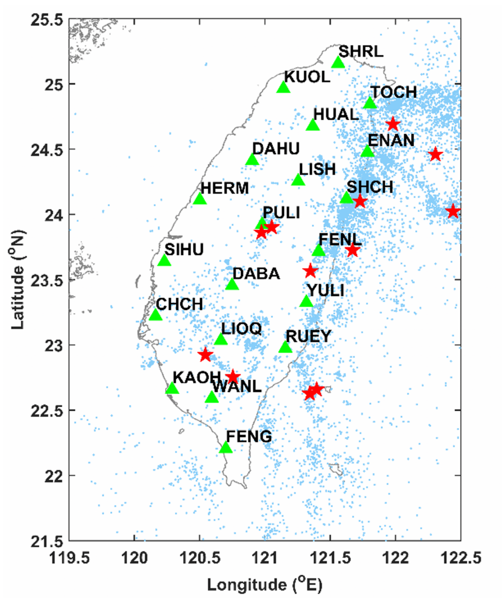

Figure 1: Map of the spatial distributions of seismicity and geoelectric stations (green triangles) in Taiwan. In this figure, past EQs with $M_L \geq 3$ are shown as light blue dots while past EQs with $M_L \geq 6$ are shown as red stars.

Inspired by these findings, in this paper we wanted to take a closer look at the relationship between the EQ times and statistical indexes of geoelectric TSs, namely correlation (C), variance (V), skewness (S), and kurtosis (K). During initial explorations, we computed the TSs of these indexes (see Sect. 2.2 for computation details) on geoelectric TSs given by the 20 stations over the 7-year period of Jan. 2012 – Dec. 2018 (see Sect. 2.1 for data details). We then aggregated the distribution of the indexes' values within different times-to-failure (TTFs, i.e., time remaining to the next EQ) intervals. In Fig. 2, we show the normalized frequency distributions of C, V, S, K computed from KAOH station at different TTFs (using 0.9-day intervals) for $M_L \geq 4$ EQs within 2 degrees longitude-latitude of KAOH station. In this figure, we see bands of darker-colored pixels across the TTFs. Specifically, for C, V, and S, there are sudden shifts in the average position of the bands, suggesting that there are two regimes (short TTFs and

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

long TTFs) where the geoelectric fields show qualitatively different behaviors. For all statistical indexes, we find the darkest pixels concentrated in the long-TTF regime, whereas in the short-TTF regime, the pixels show a lower variability in their intensities. We suspect that this second phenomenon is the result of fewer samples at longer TTFs.

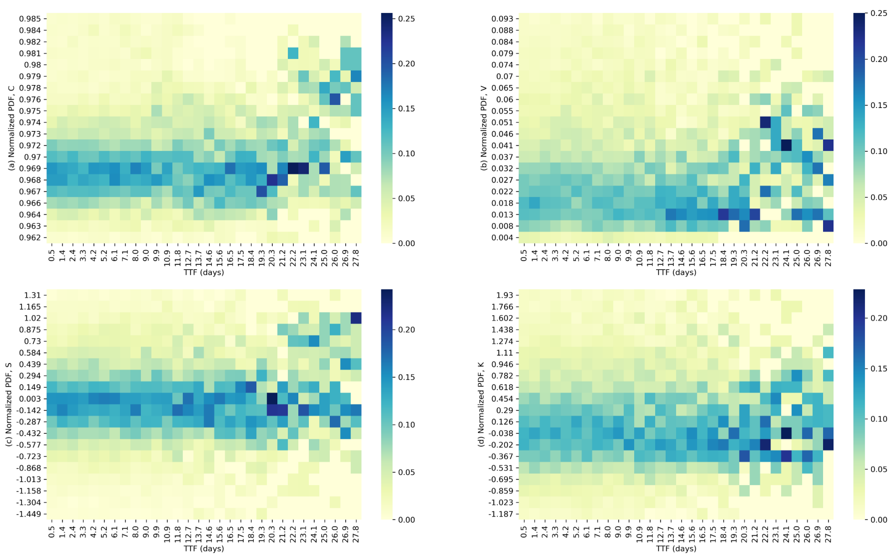

Figure 2: Heatmaps of normalized probability density functions of C, V, S, K at different times-to-failure (TTFs), for the east-west component of the geoelectric TS. The TTFs are computed using $M_L \geq 4$ EQs within 2 degrees longitude-latitude from station KAOH.

To overcome this problem, which is created by superimposing the index TSs of different lengths between EQs, we decided to discover such regimes directly from the geoelectric TSs by using HMMs. The HMM is well known for being data-driven, enabling us to search and use more general statistical features beyond limited templates that we currently know (Beyreuther and Wassermann, 2008). Additionally, its explicit incorporation of the time dimension into the model is a distinct advantage for providing holistic and time-sensitive representations, especially in the application of EQ forecasting (Beyreuther and Wassermann, 2008). In our HMM, we defined two hidden states (HSs) as the high-level representations of geoelectricity, featuring unique distributions of C, V, S, K. Here we chose to use only two, instead of more HSs, because 2-state HMM have already been successfully applied to model regimes with different EQ frequencies using EQ catalogs as the only inputs (Yip et al., 2018; Chambers et al., 2012). Thereafter, for each monitoring station, we obtained the TS of posterior HS probability, or HS TS, using the TSs of C, V, S, K and the Baum-Welch Algorithm (BWA). We then partitioned the time periods under study according to the HS TSs, and investigated whether these HS TSs that are obtained purely from geoelectric data can separate time periods of high versus low EQ ($M_L \geq 3$) probabilities, with high statistical confidence.

The goal of this investigation is to decide whether the HMM-modelling of geoelectric TS could provide features (i.e., HS TSs) of true forecasting skill for intermediate-term EQ forecasting. Therefore, we are

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

more concerned with statistical significance, than with evaluating the exact forecasting accuracy, or the forecasting of specific EQs. In this regard, we also note that the same HMM approach described in this paper can be applied to many other geophysical high-frequency time series data, such as geomagnetic or GPS ground movement data, even though we only used geoelectric data as the input of the HMM, to show that the underlying seismic dynamics is indeed clearly separable into distinct regimes of higher versus lower seismic activities (as supported by (Yip et al., 2018; Chambers et al., 2012)).

For the sake of our readers, we organize our Data and Methods in Sect. 2, Results and Discussions in Sect. 3, and Conclusions in Sect. 4. In Sect. 2, we provide information on the EQ catalog, the geoelectric TSs, how we pre-processed the latter, and subsequently computed the index TSs of C, V, S, K from them. We then explain how an HMM and the Baum-Welch Algorithm works, before applying them to our problem. We also explain why we did not estimate individual HMMs from the index TSs of C, V, S, K, but one HMM for each station from an observation TS aggregating C, V, S, K through k-means clustering. At the end of this section, we present our procedures to quantify how informative the HSs are against EQ activities, by defining and analyzing EQ Grid Maps, EQ Frequencies, and EQ Frequency Ratios (R_F). In Sect. 3, we first used the R_F grid map of one of the 20 stations to illustrate how we can compare a Discrimination Power (D) grid map against 400 simulated grid maps of D, to obtain the Discrimination Reliability (R_D) grid map, which are cellular-level statistical significances that the HSs are useful for EQ forecasting. We then performed significance tests to verify that the HSs' forecasting power are also significant at the global level, using a metric of Global Confidence Level (GCL) that we defined. To end Sect. 3, we explored how robust the GCL values are across the hyperparameter space and clarified how we chose the optimal hyperparameters for each station. Finally, we conclude in Sect. 4.

2 Data and Methods

2.1 Data Description

The 1-Hz geoelectric TSs data used in this paper was provided by the 20 monitoring stations located across Taiwan (see Fig. 1), which are collectively named Geoelectric Monitoring System (GEMS). The spacings among stations are generally 50 km. The geoelectric data here is the self-potential data, which is the natural electric potential differences in the earth, measured by dipoles placed 1–4 km apart within each station. Each station can output two sets of high-frequency geoelectric TSs, measuring alone the NS direction and the EW direction. Depending on the spatial constraints of some stations, the azimuths of the dipoles might deviate from the exact NS or EW directions by 10–40°. For the purpose of this study, we used the geoelectric TSs provided by the GEMS with the same time span as the EQ catalog data, which is from January 2012 to December 2018. We down sampled the data to 0.5 Hz, and used these in subsequent analyses.

The HMMs that we will show in Sect. 3 partitioned the 20 geoelectric TSs into two HSs, distinguished by the local statistics of their geoelectric fields. We believe these HSs can also exhibit different seismicities within their time durations. To check this, we used EQ catalog data compiled by the Central Weather Bureau (CWB), in charge of monitoring EQs in the region of Taiwan (Shin et al., 2013). The

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

CWB seismic network is highly dense and provides an abundant set of waveform data. Due to considerable EQs recorded, the seismotectonics of Taiwan is well depicted, showing the complicated subduction between the Philippine Sea plate and Eurasian plate (Kuo-Chen et al., 2012; Yi-Ben, 1986). Despite the dense seismic network, the EQ catalog was shown to be incomplete at small magnitudes due to the detection threshold of seismic instruments and the coverage of networks (Fischer and Bachura, 2014; Nanjo et al., 2010; Rydelek and Sacks, 1989). The completeness magnitude (M_c), defined as the lowest magnitude above which all EQs are reliably detected, in Taiwan is approximately between 2 and 3 (Chen et al., 2012; Mignan et al., 2011). Chen et al. (2012) showed the temporal variation of M_c, while Mignan et al. (2011) provided the spatial information of that. In this study, for the conservative estimate, we took the completeness magnitude of 3 and analyzed EQs with $M_L \geq 3$, during the period from January 2012 to December 2018 in the area of 119.5–122.5° E and 21.5–25.5° N, as shown in Fig. 1, in which the locations of strong events with $M_L \geq 6$ are marked. Some of these events were destructive. For instance, at 03:57 on 6 February 2016 (UTC+8), an M_L-6.6 EQ occurred in the southern part of Taiwan (120.54° E, 22.92° N). This event struck at a depth of around 14.6 km (Chen et al., 2017; Lee et al., 2016; Pan et al., 2019). Such a comparatively shallow depth caused more intensities on the surface, and resulted in wide-spread damage which included 117 deaths and over 500 wounded.

In the latest update of the GEMSTIP model, Chen et al. (2021) found out that by applying a specific bandpass filter on the geoelectric TS, the model became better at anticipating EQs using the skewness and kurtosis TSs. The filter they used is the order-3 Butterworth bandpass filter with lower and higher cut-off frequencies of $f_1 = 10^{-4.0}$ Hz and $f_2 = 10^{-1.75}$ Hz respectively. These lower and upper cut-off frequencies were determined to give the optimal signal-to-noise ratio by Chen et al. (2021).

Similar to the GEMSTIP model, our HMM modelling also searched for EQ-related information in skewness and kurtosis TSs computed from the geoelectric TS, we conveniently utilized the insight from Chen et al. (2021), and applied the same Butterworth filter on our geoelectric TS data before computing the index TSs. This filter was applied using *scipy.signal (v1.4.1)* package in *Python (v3.6.5)*, with instructions from (Scipy-Cookbook, 2012), which also demonstrated a clear working example of the Butterworth bandpass filter that readers can refer to.

2.2 Computation of Index TSs of $\boldsymbol{C}, \boldsymbol{V}, \boldsymbol{S}, \boldsymbol{K}$

For each station, there are two geoelectric TSs (NS and EW) of frequency 0.5-Hz. Each geoelectric TS will produce four statistical index TSs (C, V, S, K). For each station, we therefore obtained up to 8 index TSs, 4 for each direction (NS and EW). Starting from the 0.5-Hz geoelectric TS, we computed one index point for every non-overlapping time window of length L_w geoelectric TS data points. Later in Sect. 3.5, we will discuss in detail how we chose the optimal L_w individually for each station in parameter space that we tested: $[0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.25]$ (days). As can be noticed from Fig. 12, 11 out of 20 stations' optimal choice was $L_w = 0.02 \text{ or } 0.03$ days, which we suppose can be a good compromise between timely monitoring of state shifts and updating at a comfortable frequency for the human decision makers. Potential decisions that such an update frequency may enable includes the forward deployment of relief materials such as back-up generators, portable water treatment units, tents, medical supplies, refresher training of emergency response teams, as well as administrative prioritizing of re-certification works for buildings and structures in regions where more EQs are expected soon.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

Next, we present the definitions for each index. Within each time window, let us write the geoelectric field as $\{X_n\}_{n=1,\ldots,L_W}$. The correlation C that we used in this paper is the lag-1 Pearson autocorrelation of $\{D_n = X_{n+1} - X_n\}_{n=1,\ldots,L_W-1}$, which is the *difference sequence* of $\{X_n\}_{n=1,\ldots,L_W}$. Mathematically,

$$C(\{X_n\}) = AC1(\{D_n\}) = \frac{\mathbb{E}[(D_n - \mu_D)(D_{n+1} - \mu_D)]}{{\sigma_D}^2}, \tag{1}$$

where $\mathbb{E}$ is the expectation, μ_D is the mean of $\{D_n\}_{n=1,\ldots,L_W-1}$ and σ_D is the standard deviation of $\{D_n\}_{n=1,\ldots,L_W-1}$. The range of C is $[-1, 1]$, and C measures how fast the TS relaxes back to the equilibrium. If C is close to 1, X would tend to increase or decrease persistently; if C is around 0, X would be equivalent to random walks; and if C is close to -1, every increase in X would tend to be followed by a similar decrease.

The variance V of $\{X_n\}_{n=1,\ldots,L_W}$ is the sequence's second standard central moment. It is a positive number that measures how drastically the values in the sequence are different from each other, with higher values indicating higher difference. It is defined as:

$$V(\{X_n\}) = \mathbb{E}[(X_n - \mu_X)^2], \tag{2}$$

where μ_X is the mean of $\{X_n\}_{n=1,\ldots,L_W}$. Additionally, we observed astronomically extreme values in the V TSs for most stations, which were caused by unknown technical errors, and we therefore considered them outliers that have to be removed for consistent data quality. We discuss how we removed them in detail in Supporting Information Sect. A. From here onwards, the V TSs will always refer to those after the outlier-removal process.

The skewness S of $\{X_n\}_{n=1,\ldots,L_W}$, or the sequence's third standard central moment, is defined as:

$$S(\{X_n\}) = \mathbb{E}\left[\left(\frac{X_n - \mu_X}{\sigma_X}\right)^3\right], \tag{3}$$

where σ_X is the standard deviation of $\{X_n\}_{n=1,\ldots,L_W}$. It is a real number measuring how asymmetric the distribution of $\{X_n\}_{n=1,\ldots,L_W}$ is about the mean. For a perfectly symmetric distribution such as the normal distribution, the skewness is 0. A positive skewness means the distribution has a longer tail to the right, and a negative skewness means the distribution has a longer tail to the left.

The kurtosis K of $\{X_n\}_{n=1,\ldots,L_W}$, or the sequence's fourth standard central moment, is defined as:

$$K(\{X_n\}) = \mathbb{E}\left[\left(\frac{X_n - \mu_X}{\sigma_X}\right)^4\right]. \tag{4}$$

It is a real number measuring how frequently extreme values (values very far from the mean) appear in the distribution. The higher the number, the more frequently extreme values can be found. As a reference, the kurtosis of the normal distribution is $K = 3$. If $K > 3$, we say that the distribution is *leptokurtic*, meaning the distribution has fatter tails and more frequent extreme values compared to the normal distribution. If $K < 3$, the distribution is said to be *platykurtic*, meaning the distribution has thinner tails, and extreme values appear less frequently compared to the normal distribution.

2.3 Estimation of HMM Using the Baum-Welch Algorithm

A Markov model is a stochastic model that can be used to describe a system whose future state s_{t+1} is drawn from a set of L states $\{S_l\}_{l=1,\ldots,L}$ with probabilities $p_{j\leftarrow i} = P(s_{t+1} = S_j | s_t = S_i)$ conditioned

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

by its current state s_t. The probabilities $p_{j\leftarrow i}$ can be organized into a transition matrix $\boldsymbol{A}$, where $\boldsymbol{A}(i,j) = p_{j\leftarrow i}$. The HMM is an extension of the Markov model, with the additional property that the system state s_t is not explicitly known, hence the word "hidden" in the name. Instead, what can be observed from an HMM at any time t is an observable o_t drawn from a size-Q observable set $\{O_q\}_{q=1,\ldots,Q}$. Just as in a Markov model, the future state s_{t+1} of an HMM is drawn from the set $\{S_l\}_{l=1,\ldots,L}$ with probabilities $p_{j\leftarrow i}$ (similarly conditioned by the current state s_t) taken from the transition matrix $\boldsymbol{A}$. At time t, the observable o_t is emitted with a probability $P(o_t = O_q|s_t = S_l)$ that depends on which HS $s_t = S_l$ the system is in. These probabilities can be organized into an $L \times Q$ emission matrix $\boldsymbol{B}$, where $\boldsymbol{B}(l,k) = P(o_t = O_q|s_t = S_l)$. Additionally, we call the HS probability distributions at the initial time as $\boldsymbol{\pi_0} = \{P(S_1), P(S_2), \ldots, P(S_L)\}$. With this, we have fully specified the HMM: the sets of HSs $\{S_l\}_{l=1,\ldots,L}$ and observations $\{O_q\}_{q=1,\ldots,Q}$ as well as the model parameters that are collectively called $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi_0})$.

In common real-world applications of HMM, the question is to estimate the probability distributions of the HS TS given the observation TS and the model parameter, namely $P(s_t = S_l|\{o_t\}_{t=1,\ldots,T}, \lambda)$. More often than not, the model parameter λ is unknown and has to be simultaneously estimated as well. One of the most common ways to do this is the Baum-Welch Algorithm (BWA) (Zhang et al., 2014; Oudelha and Ainon, 2010; Yang et al., 1995; Bilmes, 1998), which belongs to the family of Expectation Maximization methods (Bilmes, 1998). Starting from randomly initialized model parameters λ, the algorithm runs recursively to maximize the likelihood of the model given the observation TS. When the algorithm converges, we will obtain a set of estimated model parameters $\tilde{\lambda} = (\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}}_{\boldsymbol{0}})$, as well as a posterior probability $P(s_t = S_l|\{o_t\}_{t=1,\ldots,T}, \tilde{\lambda})$ TS. We include more details on the BWA in Sect. 2.5. Additionally, for readers who want an intuitive demonstration of how HMM and BWA works, we attached a simulation of a simple HMM and its BWA application in Supporting Information Sect. B.

HMMs are traditionally applied in fields such as speech recognition (Palaz et al., 2019; Novoa et al., 2018; Chavan and Sable, 2013; Abdel-Hamid and Jiang, 2013), bioinformatics, and anomaly detection (Qiao et al., 2002; Joshi and Phoha, 2005; Cho and Park, 2003). It has also been used for short-term EQ forecasting, using observations from EQ catalogs (Yip et al., 2018; Chambers et al., 2012; Ebel et al., 2007), as well as GPS measurements of ground deformations (Wang and Bebbington, 2013). To the best of our knowledge, there is no past HMM study on geoelectric TSs for EQ forecasting. In this paper, we argue that the HMM is an objective tool, because the HSs were estimated only from the geoelectric TSs, and thereafter validated against the EQ catalog. We believe this statistical procedure limits the bias that we could introduce into our prediction model when we optimized the model. This will be even clearer by the end of Sect. 2.5 where we summarize the entire procedure.

2.4 HMM Modelling and Inputs to the BWA

In the context of this study, we assume for simplicity two seismicity states of the earth crust beneath each station. These are our HSs $\{S_1, S_2\}$, since they cannot be directly observed. What we can observe directly are the geoelectric TSs for each station. Our goal is to reconstruct the HS TSs so that the distributions of indexes (C, V, S, K) of the geoelectric TSs in S_1 and S_2 are as different as possible. To do this, we

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

computed 4 index TSs each for NS and EW geoelectric fields using the procedure described in Sect. 2.2, and organized them into a TS of 8-dimensional feature vectors $\boldsymbol{F}_t = [C_{NS,t}, V_{NS,t}, S_{NS,t}, K_{NS,t}, C_{EW,t}, V_{EW,t}, S_{EW,t}, K_{EW,t}]$. The values of each of the indexes are continuously distributed, but the standard BWA requires discrete observations $\{O_q\}_{q=1,\ldots,Q}$ as input. In this section, we discuss possible ways to convert $\boldsymbol{F}_t$ into discrete observations for the BWA, and why we chose one particular method for implementation.

One way to do so would be to model each component of $\boldsymbol{F}_t$ as samples drawn from known distributions, such as a normal distribution or a gamma distribution. Unfortunately, as we can see from Fig. 3 (introduced in the next paragraph), none of the known distributions fit the empirical data well. Alternatively, we can discretize the components of $\boldsymbol{F}_t$ by binning them. In other words, we represent the distribution of each component with a histogram, with a specific choice of the number of bins (50 for example). This will effectively convert the continuous values of each component of $\boldsymbol{F}_t$ into discrete values, such as integer labels from 1 to 50 if we use 50 bins. Let us write the discretized $\boldsymbol{F}_t$ as $\bar{\bar{\boldsymbol{F}}}_t = [\bar{\bar{C}}_{NS,t}, \bar{\bar{V}}_{NS,t}, \bar{\bar{S}}_{NS,t}, \bar{\bar{K}}_{NS,t}, \bar{\bar{C}}_{EW,t}, \bar{\bar{V}}_{EW,t}, \bar{\bar{S}}_{EW,t}, \bar{\bar{K}}_{EW,t}]$.

If we do this for the TSs of individual components, such as the TS of $\bar{\bar{C}}_{NS,t}$, and use them as inputs for the BWA, we will obtain one HS TSs for each of the 8 components. In Fig. 3, we show (A) the estimated emission matrix $\widetilde{\boldsymbol{B}}$ in Figs 3(a), (c), (e), (g), and (B) the posterior probability TSs in Figs 3(b), (d), (f), (h) for 4 components: $\bar{\bar{C}}_{NS,t}, \bar{\bar{V}}_{NS,t}, \bar{\bar{S}}_{NS,t}, \bar{\bar{K}}_{NS,t}$ of KAOH station. These posterior probability TSs are different, which is not what we desire. Therefore, instead of this, we would like to use all 8 components in $\bar{\bar{\boldsymbol{F}}}_t$ as a single input to the BWA, to obtain a single HS TS for each station.

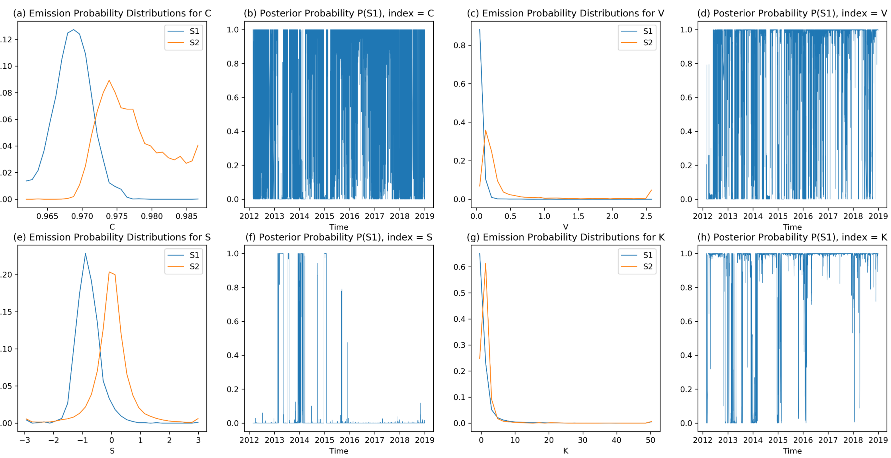

Figure 3: The output of BWA: the emission probability, or the probability mass functions, as well as their posterior HS probability TSs, for $\bar{\bar{C}}_{NS,t}$ (a, b), $\bar{\bar{V}}_{NS,t}$ (c, d), $\bar{\bar{S}}_{NS,t}$ (e, f), and $\bar{\bar{K}}_{NS,t}$ (g, h), respectively, using station KAOH's geoelectric TS data, with 50 bins.

The BWA has no problem dealing with high-dimensional problems, provided the inputs are discrete. However, this method would work well only if the overall number of possible observations is small. If

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

we use 50 bins for each of the eight indexes, there would be $D = 50^8 \approx 3.91 \times 10^{13}$ possible observations, meaning the emission matrix would be of dimension 3.91×10^{13}-by-2. Reducing the number of bins to just 10 for each index, we still have $D = 10^8$ possible observations. This latter space is still too large for the BWA to search through exhaustively in a reasonable amount of time, even though we feel 10 bins for each index may already be too coarse and likely to miss subtle details. Furthermore, with so many possible observations, we expect the emission probabilities to be significantly different from 0 only for a very small subset of the D possible observations.

We do not know *a priori* what the elements of this very small subset are. They may occur as isolated points in the search space, or they may occur in groups of closely spaced points. In the continuous feature space, each of these groups of observations represents a cluster of similar feature vectors. To determine the number of such clusters, and where they occur in the 8-dimensional continuous feature space, we mapped similar feature vectors to the same label using the k-means clustering algorithm (Gupta et al., 2010; Wen et al., 2006; Dash et al., 2011), which is commonly used for discretizing continuous vectors such as $\boldsymbol{F}_t$. We chose to use the k-means clustering for discretizing $\boldsymbol{F}_t$ because of its low computational cost as well as its reliability in grouping similar feature vectors in the feature space. In so doing, we created a discrete feature space with reasonable size, as high-level labels of different geoelectric dynamics.

In the k-means clustering of the set of N continuous-valued vectors $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_N\}$, we start by choosing $Q \le N$ clusters: $\boldsymbol{G} = \left\{G_1^{(0)}, G_2^{(0)}, \ldots, G_Q^{(0)}\right\}$, 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 $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_N\}$ this way, $G_q^{(0)} = \left\{\boldsymbol{W}_{q,1}, \boldsymbol{W}_{q,2}, \ldots, \boldsymbol{W}_{q,n_q}\right\}$ would contain n_q feature vectors. We can improve on this initial clustering by updating the position of the centers by:

$$\boldsymbol{\mu}_q^{(1)} = \frac{1}{n_q} \sum_{i=1}^{n_q} \boldsymbol{W}_{q,i}\,, \tag{5}$$

and re-assigning the N continuous-valued vectors $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_N\}$ to these new clusters $\boldsymbol{G} = \left\{G_1^{(1)}, G_2^{(1)}, \ldots G_Q^{(1)}\right\}$ with updated centers. After repeating this procedure for enough times, the clusters will converge to $\boldsymbol{G}^* = \left\{G_1^*, G_2^*, \ldots G_Q^*\right\}$, where $G_q^* = \left\{\boldsymbol{W}_{q,1}^*, \boldsymbol{W}_{q,2}^*, \ldots, \boldsymbol{W}_{q,n_q^*}^*\right\}$. 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:

$$\underset{\boldsymbol{G}^*}{argmin} \sum_{q=1}^{Q} \sum_{\boldsymbol{W}_{q,i}^* \in G_q^*} \left\|\boldsymbol{W}_{q,i}^* - \boldsymbol{\mu}_q^*\right\|^2 . \tag{6}$$

The indexes $C_{NS,t}$, $V_{NS,t}$, $S_{NS,t}$, $K_{NS,t}$ have highly disparate dynamic ranges, and should not be directly combined into a feature vector. Therefore, before the clustering, we first standardized our indexes by dividing them by their respective standard deviations. The purpose of this step is to ensure the weights associated with each index during the k-means clustering are equal, so as not to bias our search for features with high dynamic range. Mathematically, the feature vector of standardized indexes at time t, $\boldsymbol{F'}_t$ can be written as:

$$\boldsymbol{F'}_t = \left[\frac{C_{NS,t}}{\sigma(C_{NS,t})}, \frac{V_{NS,t}}{\sigma(V_{NS,t})}, \frac{S_{NS,t}}{\sigma(S_{NS,t})}, \frac{K_{NS,t}}{\sigma(K_{NS,t})}, \frac{C_{EW,t}}{\sigma(C_{EW,t})}, \frac{V_{EW,t}}{\sigma(V_{EW,t})}, \frac{S_{EW,t}}{\sigma(S_{EW,t})}, \frac{K_{EW,t}}{\sigma(K_{EW,t})}\right]. \quad (7)$$

We then implemented k-means clustering using the *Scikit-learn* package (v0.23.1) in *Python* (v3.6.5), on the sequence of feature vectors $\boldsymbol{F'}_t$ covering the time period from January 2012 to December 2018. The choice of the number of clusters Q was determined as part of the hyperparameter optimization, described in Sect. 3.5. In this way, we matched each $\boldsymbol{F'}_t$ to a discrete label $o_t \to O_q$ (where q is an integer from 1 to Q), to obtain the TS of discrete observations $\{o_1, o_2, \dots, o_t, \dots, o_T\}$ for each station as its input to the BWA.

2.5 Implementation of BWA

In this section, we describe how we implemented the BWA to obtain one HS TS for each station. We start by describing how we initialized and iterated the BWA, as well as how we delt with local optima in the BWA results by using multiple initializations.

The first step of the BWA is to initialize the HMM model parameters $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$. Since we had no prior knowledge on the model parameters, we initialized parameters $(\boldsymbol{A_0}, \boldsymbol{B_0}, \boldsymbol{\pi_0})$ randomly. After this, we iterated BWA's expectation maximization steps 30 times, starting with iteration index $i = 1$. Each iteration comprises of the forward procedure, the backward procedure, and the update.

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

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

$$\gamma_{l,t}^i = \frac{P\big(o_1, o_2, \dots, o_T, s_t = S_l \big| (\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{\pi_i})\big)}{P\big(o_1, o_2, \dots, o_T \big| (\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{\pi_i})\big)} = \frac{\alpha_{l,t}^i \beta_{l,t}^i}{\sum_{m=1}^{2} \alpha_{m,t}^i \beta_{m,t}^i}. \quad (8)$$

Next, we calculate the probability $\xi_{l,m,t}^i = P(s_t = S_l, s_{t+1} = S_m | o_1, o_2, \dots, o_T, (\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{\pi_i}))$, 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 $(\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{\pi_i})$. This is computed by:

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

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

$$= \frac{\alpha^i_{l,t}\boldsymbol{A_i}(l,m)\beta^i_{m,t+1}\boldsymbol{B_i}(m,O_{t+1})}{\sum_{r=1}^{2}\sum_{p=1}^{2}\alpha^i_{r,t}\boldsymbol{A_i}(r,p)\beta^i_{p,t+1}\boldsymbol{B_i}(p,O_{t+1})}. \tag{9}$$

Now, we can update the new model parameters as:

1) $$\pi_{i+1} = \gamma^i_{t=1} \tag{10}$$

2) $$\boldsymbol{A_{i+1}}(l,m) = \frac{\sum_{t=1}^{T-1}\xi^i_{l,m,t}}{\sum_{t=1}^{T-1}\gamma^i_{l,t}} \tag{11}$$

3) $$\boldsymbol{B_{i+1}}(l,O_q) = \frac{\sum_{t=1}^{T}1_{o_t=O_q}\gamma^i_{l,t}}{\sum_{t=1}^{T}\gamma^i_{l,t}}, \tag{12}$$

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

As the iteration goes, the BWA improves the likelihood of observing the input observation TS $o_1, o_2, \ldots, o_T$ given the model parameters $(\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{\pi_i})$, which converges when the improvements on the posterior probability $P(o_1, o_2, \ldots, o_T|(\boldsymbol{A_i}, \boldsymbol{B_i}, \boldsymbol{\pi_i}))$ become minimal. In practice, we found that 30 iterations were long enough for most models to converge. We therefore obtained the estimated model parameters $(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}}) = (\boldsymbol{A_{30}}, \boldsymbol{B_{30}}, \boldsymbol{\pi_{30}})$, as well as the posterior probability TS of $P(s_t = S_l|o_1, o_2, \ldots, o_T, \widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}})$ for both HSs and all t, which we write in short form as: $\boldsymbol{P_1} = (P(s_1 = S_1), P(s_2 = S_1), \ldots, P(s_T = S_1))$ and $\boldsymbol{P_2} = (P(s_1 = S_2), P(s_2 = S_2), \ldots, P(s_T = S_2))$. Here, we noted that BWA assigns the indexing of HSs randomly; therefore, the S_1 of one station is not guaranteed to be equivalent to the S_1 of another station.

We cannot simply do the above BWA estimation once to get $(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}})$, because the BWA converges to local optima instead of the global optimum in the model parameter space (Bilmes, 1998; Yang et al., 2017; Larue et al., 2011). Also, the initial parameters have a significant influence on the local optimum where the BWA converges. In order to obtain a global optimum result within a reasonable computation time, we ran 15 BWA estimations in parallel for each station, with different random initial parameters. For each station, we then chose the model with the highest model score given by $P\left(o_1, o_2, \ldots, o_T|(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}})\right)$ for subsequent analysis. Later in Fig. 4(a), we also show all 15 HMMs to demonstrate how consistent the converged models are. We can write the posterior probability TS of this model as $\widetilde{\boldsymbol{P}}_{\boldsymbol{1}} = \left(P(s_1 = S_1), P(s_2 = S_1), \ldots, P(s_T = S_1)|o_1, o_2, \ldots, o_T, (\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}})\right)$.

For each initial condition, the BWA randomly assigns one HS to be S_1, and the other to be S_2. To show all 15 HMMs simultaneously in Fig. 4(a), we need to standardize S_1 and S_2 across all HMMs. For this purpose, we set $\widetilde{\boldsymbol{P}}_{\boldsymbol{1}}$ as the "standard". For the remaining 14 posterior probabilities $\{\boldsymbol{P^i_1}\}_{i=2,\ldots,15}$, we checked their Expected Absolute Difference $EAD = mean(|\widetilde{\boldsymbol{P}}_{\boldsymbol{1}} - \boldsymbol{P^i_1}\})$ from $\widetilde{\boldsymbol{P}}_{\boldsymbol{1}}$, whose value ranges from 0 and 1. If $EAD > 0.5$, $\boldsymbol{P^i_1}$ is more similar to $\widetilde{\boldsymbol{P}}_{\boldsymbol{2}}$ than to $\widetilde{\boldsymbol{P}}_{\boldsymbol{1}}$, and we proceed to swap the HS indexing for the i^{th} HMM by assigning $\boldsymbol{P^i_1}(\boldsymbol{new}) \equiv \boldsymbol{P^i_2}$ and $\boldsymbol{P^i_2}(\boldsymbol{new}) \equiv \boldsymbol{P^i_1}$. Otherwise, $\boldsymbol{P^i_1}$ corresponds to the same HS as $\widetilde{\boldsymbol{P}}_{\boldsymbol{1}}$, and we leave its HS indexing unchanged. In this way, we standardized all 15 models so that their $\boldsymbol{P_1}$ can be visualized together in Fig. 4(a), with the $\widetilde{\mathbf{P}}_{\mathbf{1}}$ TSs sorted by their model scores $P\left(o_1, o_2, \ldots, o_T|(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}})\right)$, and the optimal model at the first row. In Fig. 4(b), we show

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

the actual posterior probability TS of this optimal model. The figures of 15 HMMs for all 20 stations are included in the Supporting Information Sect. C.

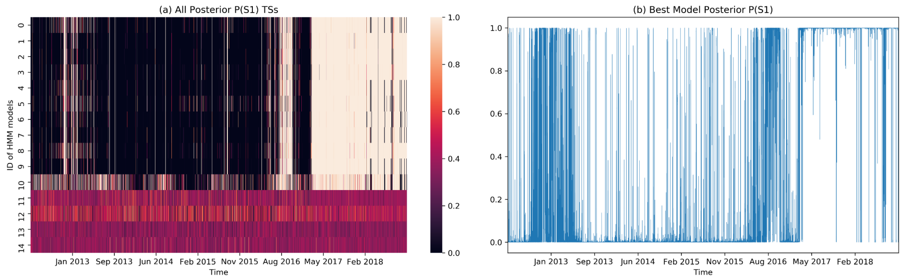

Figure 4: The step-by-step data visualization for station CHCH, showing (a) a heatmap of the 15 HMM models' posterior probability TSs for S_1, sorted by model score from highest to lowest. The posterior probabilities for the last 4 HMMs are messy, because the BWA estimations do not converge; (b) the optimal model's posterior HS probability TS for S_1, $\tilde{P}_1$ (obtained using optimal hyperparameters: $[L_w, Q] = [0.02\ (\text{day}), 30]$).

We summarize the procedures used to obtain $\tilde{P}_1$, starting from a pair of geoelectric TSs for each GEMS station in the form of a flow chart in Fig. 5. It is noteworthy that the full procedure contains essentially only 2 hyperparameters: Q and L_w. The figures shown in the results section will use the optimal hyperparameters, whose identification procedure will be discussed in detail later in Sect. 3.5. Additionally, for each station's optimal HMM, we plotted the distribution of indexes (C, V, S, K) at both HSs in Supporting Information Sect. D.

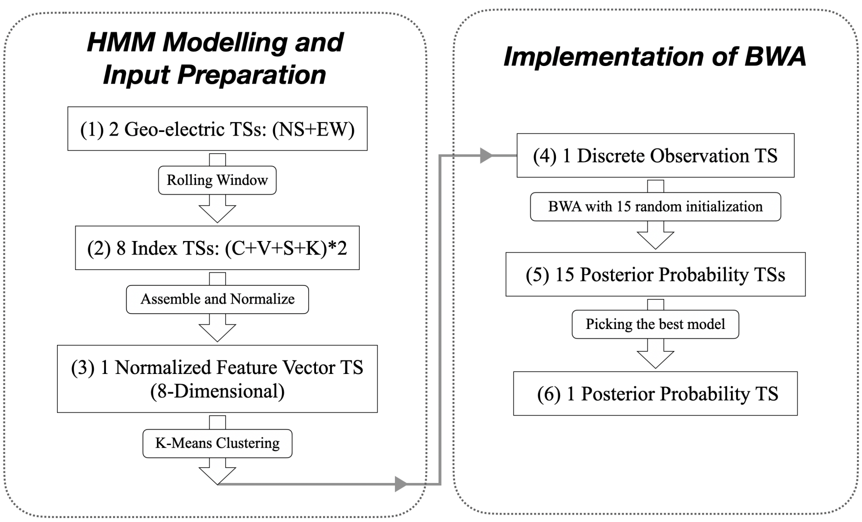

Figure 5: Flow chart summarizing the procedures of obtaining the optimal posterior probability TS $\tilde{P}_1$ from the data of one GEMS station.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

2.6 EQ Grid Map, EQ Frequency, and EQ Frequency Ratio

Up to this point, we did not incorporate any EQ catalog information into $\widetilde{\boldsymbol{P}}_{\mathbf{1}}$ for each station. Unlike many past EQ studies looking for specific precursory features within the geoelectric data, we made no specific assumptions regarding what these EQ precursors might look like. Instead, we let the BWA search for specific precursory features within the 8-dimensional feature space.

After the HMM modeling, we then checked locally whether S_1 and S_2 would effectively partition time periods with significantly lower EQ probabilities from those with significantly higher EQ probabilities. We think of one HS as a *passive* state (with significantly lower EQ probabilities) and the other HS as an *active* state (with significantly higher EQ probabilities), but we cannot call the former S_1 and the latter S_2 because we have not yet standardized these HS labels across the 20 stations. To do so, we need to match the HS TS of each station to the EQ catalog to determine the EQ frequencies of S_1 and S_2 for this station, and use S_1 and S_2 as the HS labels of the active and passive states respectively (relabeling when necessary). In the remainder of this section, we describe in detail how this is done.

For each GEMS station we started from $\widetilde{\boldsymbol{P}}_{\mathbf{1}}$, and classified time periods across the 7 years as belonging to two sets T_1 and T_2. The time point t_i was assigned to T_1 if $\tilde{P}_{t_i}(S_1) > 0.5$ and T_2 if $\tilde{P}_{t_i}(S_2) > 0.5$. After this is done, we checked how EQs are distributed between T_1 and T_2 for different regions across Taiwan. For this task, we first made a 16-by-16 grid map of Taiwan, so that EQs within the same grid cell (ix, iy), for ix and iy in $[0,1,\ldots,15]$, are grouped together (see Fig. 6).

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions
Open Access

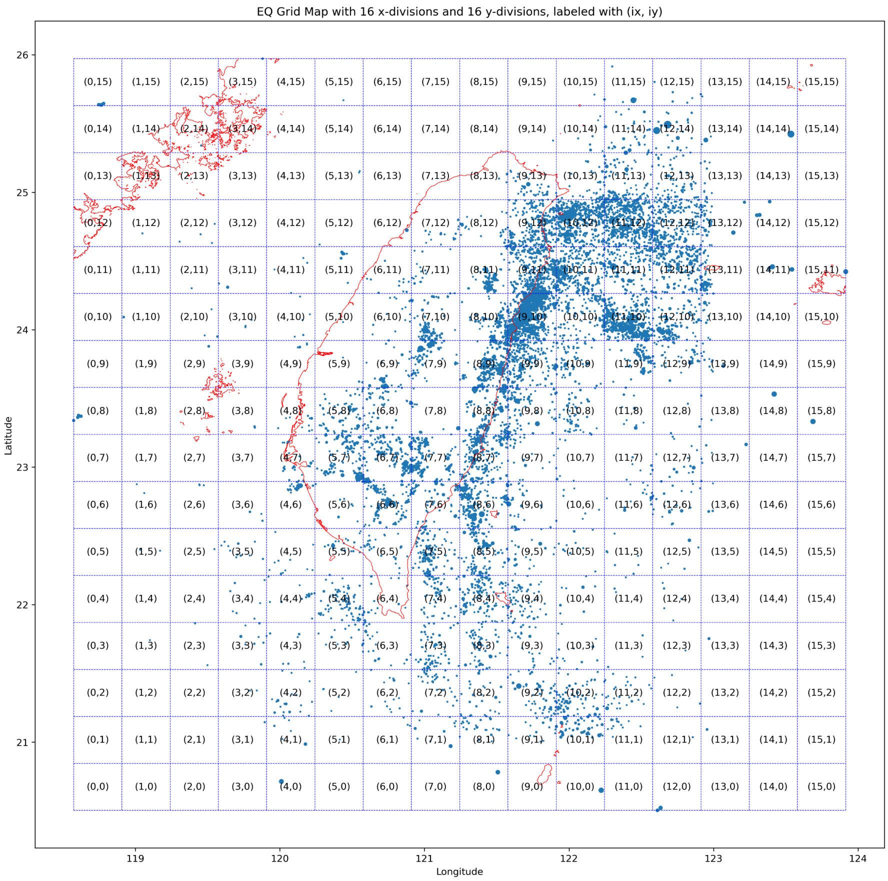

Figure 6: A sample EQ Grid Map with 16-by-16 divisions, in which each cell measures $0.3330°$(longitude)-by-$0.3418°$(latitude). All EQs of $M_L \geq 3$ are labeled with blue circles, with the radius of each circle being proportional to the natural exponential of EQ's magnitude.

For each grid cell (ix, iy), we defined the *EQ Frequencies* for HSs S_1 and S_2 as:

$$F_{EQ,1} = \frac{N_1}{|T_1|}, F_{EQ,2} = \frac{N_2}{|T_2|}, \tag{13}$$

where N_1 is the number of EQs occurring within T_1, N_2 is the number of EQs occurring within T_2, $|T_1|$ is the total duration of T_1 time periods, and $|T_2|$ is the total duration of T_2 time periods. From Fig. 6, we see that the spatial distribution of EQs is highly heterogeneous, so we may find a grid cell with about 10 EQs but also another grid cell with about 1000 EQs. This tells us that we should not directly compare the EQ frequencies, but should instead compare the *EQ Frequency's Ratio*, defined as:

$$R_F = \frac{F_{EQ,1}}{F_{EQ,1} + F_{EQ,2}}. \tag{14}$$

For any cell containing at least one EQ, the range of its R_F is $[0,1]$. Intuitively, any cell with $R_F < 0.5$ is a region having lower EQ frequency in S_1 compared to S_2; and any cell with $R_F > 0.5$ is a region having a higher EQ frequency in S_1 compared to S_2. For example, for a cell with $R_F = 0.2$, $F_{EQ,1}$ is

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

only $1/4$ of $F_{EQ,2}$. The R_F value quantifies how one HS has a higher or lower EQ frequency than the other. In Sect. 3, we will present how we deep dived into the spatial-temporal correlations between HS TSs ($\widetilde{\boldsymbol{P}}_{\mathbf{1}}$) and EQ activities for all 20 stations, starting from 20 grid maps of R_F values.

3 Results and Discussions

In this section, we present the results obtained for all 20 stations, as well as additional treatments that we felt are necessary to investigate whether the HS TSs have significant forecasting power for EQs.

3.1 EQ Frequency's Ratio ($\boldsymbol{R_F}$) Grid Maps

Once we obtained the $\widetilde{\boldsymbol{P}}_{\mathbf{1}}$ TS for each station, the natural first step of our analysis was to examine the R_F values for all cells in the 16-by-16 grid map. We show this procedure for CHCH station in Fig. 7, where we visualize the grid maps for N_1 and N_2 in Figs 7(a) and (b) respectively, to clearly show how many EQs occurred during T_1 and T_2. The resulting R_F grid map is shown in Fig. 7(c), where there are cells with values close to 0.5 (white-color cells) and cells with values far from 0.5 (red for close to 0; green for close to 1). White-color cells are regions whose EQ activities are weakly correlated with the HSs, since the time periods of S_1 and S_2 are not very different in terms of EQ frequency; whereas red/green cells are regions with significantly lower/higher EQ frequencies in S_1.

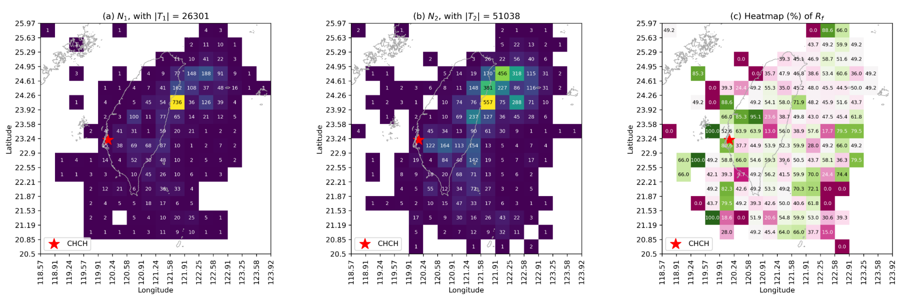

Figure 7: The step-by-step data visualization for station CHCH, showing (a) the grid map showing the number of $\boldsymbol{M_L \geq 3}$ EQs during $\boldsymbol{S_1}$'s time periods, $\boldsymbol{N_1}$; (b) the grid map showing the number of $\boldsymbol{M_L \geq 3}$ EQs during $\boldsymbol{S_2}$'s time periods, $\boldsymbol{N_2}$; and (c) the grid map showing the EQ Frequency Ratio, $\boldsymbol{R_F}$ $(\times \mathbf{0.01})$. Results were obtained using optimal hyperparameters: $[\boldsymbol{L_w}, \boldsymbol{Q}] = [\mathbf{0.02}\ (\mathbf{day}), \mathbf{30}]$.

As can be seen in Fig. 7(c), for different regions the HS with higher EQ activities can be either S_1 or S_2. This is true not only for CHCH station, but also for all 20 stations, whose R_F grid maps are shown in Fig. 8. Although there is no consistent pattern of any state corresponding to higher EQ activities globally, we see in Fig. 8 that there are regions whose R_F values are far from 0.5 across many stations. This means that statistically speaking, one of the HSs has higher EQ activities than the other. In fact, if the active HS has a lot more EQs than the passive HS, it is also likely that the active HS cover most of the larger EQs (e.g., $M > 5$), which is a good attribute for potential EQ forecasting applications. This phenomenon is shown in Supporting Information Sect. E, where we visualized the EQ frequency distributions across different magnitudes for both HSs, for three selected cells with most EQ events.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

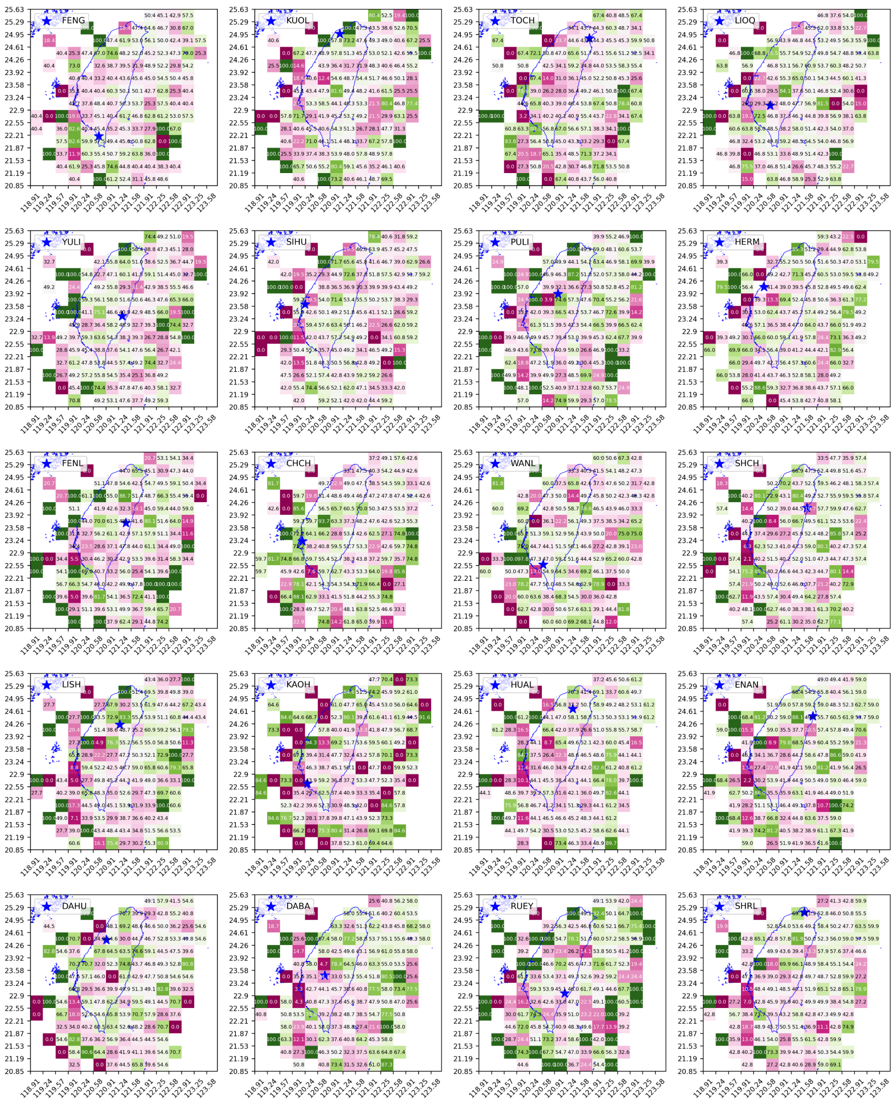

Figure 8: The grid maps of EQ Frequency's Ratio R_F $(\times 0.01)$ for 20 stations (obtained using optimal hyperparameters individually specified for each station in Fig. 12).

All in all, the findings in this section is important, but we cannot directly decide S_1 or S_2 to be the proxy for increased EQ probabilities, because they cannot be associated consistently with the active or the passive state. Instead, we should understand S_1 and S_2 as two high-level, fuzzy labels for tectonic dynamics related to EQ activities in different regions. There can be elements such as rock and soil formations, the underground water system, and fault lines, forming a complex dynamical system that influences where and when EQs become active. A concrete mapping between EQ activities and specific elements of the complex dynamical system would be very difficult, as this will involve high-resolution

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

subterrain surveys. Nevertheless, we can still measure how well S_1 and S_2 can partition the time periods so that one HS can have significantly more EQs than the other. To show this more clearly, we created grid maps of *discrimination power* D and present them in the next section.

3.2 Discrimination Power (*D*) Grid Maps

We defined the discrimination power D for each cell as:

$$D = |R_F - 0.5|. \tag{15}$$

The value of D ranges from 0 to 0.5, with 0.5 being the most discriminative since all EQs are found in one HS, and 0 being the least discriminative since EQ frequencies are identical between the two HSs. We show the grid maps of D for 20 stations in Fig. 9, which are easier to interpret compared to the grid maps in Fig. 8 where we had to use two different colors. Intuitively, for a region with $D = 0.25$ (not uncommon), one of its HSs would have an EQ frequency three times that of the other HS. It can be noted that cells around the edge of the map tend to have very high D values, because there are very few EQ events in these cells. This is not a problem, as we will take the number of EQs into account later in Sect. 3.3.

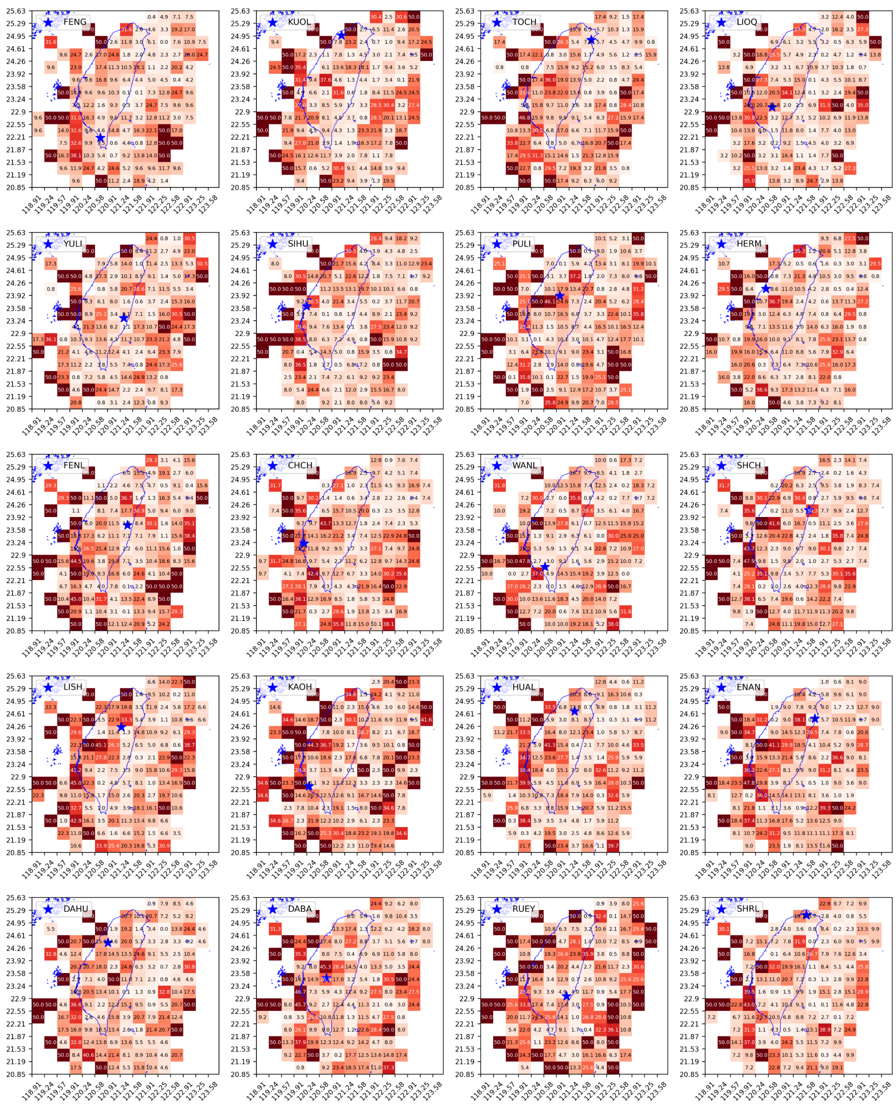

Figure 9: The grid map of discrimination power $D\,(\times 0.01)$ for 20 stations (obtained using optimal hyperparameters individually specified for each station in Fig. 12).

In some cells, we find D values close to 0.5, which seems to suggest that the seismicity associated with S_1 is very different from that associated with S_2. However, looking at Fig. 9, we see large variations in D values across the cells, and more importantly among some neighboring cells. We therefore wonder whether regions with high D values are statistically significant, or the products of random temporal clustering of EQs (Dieterich, 1994; Frohlich, 1987; Holbrook et al., 2006; Batac and Kantz, 2014). For example, if all EQs in a cell occurred within a single day in the 7-year period, any random assignment of HSs would produce the highest D value of 0.5. To address this concern, we investigated the significance of the grid maps of D through statistical tests in the next section.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences Discussions

3.3 Cellular-level Significance Tests of the Forecasting Power

Since we had the optimal HMMs for the 20 stations, we can test cellular statistical significance levels that their HSs can indeed separate time periods of higher/lower EQ probabilities, using D grid maps shown in Fig. 9. Specifically, for each grid cell and an empirical HS TS we carried out a statistical hypothesis testing using the following null hypothesis: any random HS TS would achieve the same or higher performance (in terms of D value). To create random HS TSs for the hypothesis testing, we chose to directly simulate the HMM using the same model parameters $(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\pi}})$ as the empirical HMM of the corresponding station. For each hypothesis testing of an empirical HS TS (actual HS TS obtained for each station), we created 400 simulated HS TSs, which were then used to create 400 grid maps of the Discrimination Power D. In Fig. 10, we show the empirical HS TSs alongside a random sample of 10 simulated HS TSs for YULI, SHRL, CHCH, and SIHU to illustrate the simulated counterparts. After this, in each cell, we had one empirical value of D that we can compare against a distribution of 400 simulated values of D. This allows us to compute for each cell the probability that its empirical D value is higher than its simulated counterparts. We named this quantity the *Discrimination Reliability* R_D, defined for each cell in the grid map as:

$$R_D = \frac{\#\,(simulated\ D < empirical\ D)}{400}. \qquad (16)$$

In the language of statistical hypothesis testing, the p-value for the test is given by $p = 1 - R_D$. The value of R_D ranges from 0 to 1. If R_D is close to 1, we are confident that the discrimination power of the empirical HS TS is statistically significantly high; otherwise, we have no such confidence.

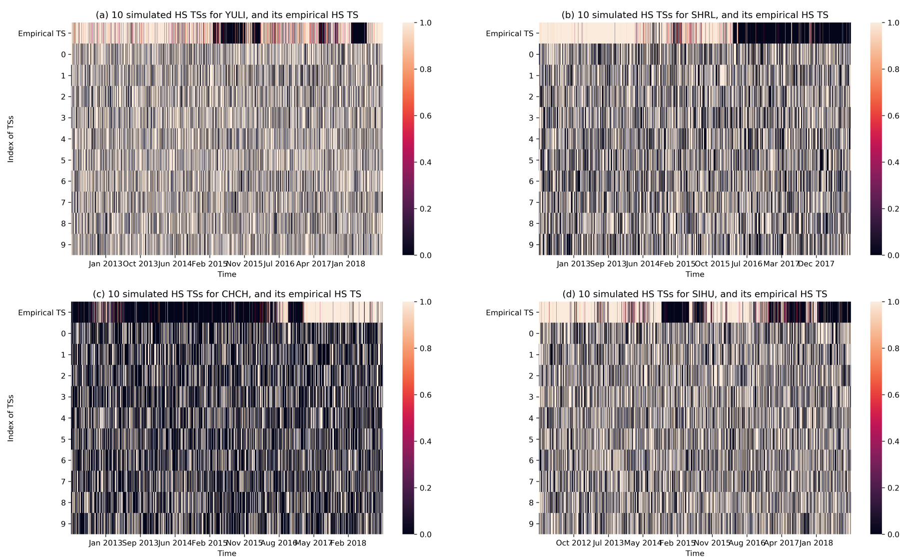

Figure 10: The empirical HS TS and 10 simulated HS TSs, for stations (a) YULI, (b) SHRL, (c) CHCH, and (d) SIHU. The simulated HS TSs have HS transition frequencies and HS total durations similar to the empirical HS TS, but have none of the temporal correlations in the empirical HS TS. Results are obtained using optimal hyperparameters individually specified for each station in Fig. 12.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

In Fig. 11, we show the grid maps of R_D values (in percentage) for all 20 stations. Dark-red cells are regions with R_D close to 1, and white-pink cells are regions with R_D close to 0. From these grid maps, we can better appreciate the utility of HS TSs across the grid map, since the R_D value is a statistical significance measure of the HS-EQ correlation, unlike the discrimination power D. To explain this, let us take the example of station LIOQ (upper left of Fig. 11), whose physical location is marked by the blue star, within a dark-red grid cell of $R_D = 0.992$. This means that the empirical HS TS performs better than random guesses (i.e., simulated HS TSs) at separating time periods of low/high EQ frequencies, with a statistical significance of $p = 0.008$. This means that it is improbable for a simulated HS TS to have such a high D, and therefore the empirical HS TS is unlikely to be a product of random chance. This is a very strong demonstration of the mutual information between the HS TS obtained from geoelectric TS, and the EQ catalog that was not used to train the HMM.

In the proximity of station LIOQ located within $22.55–23.58° N$, we can see a clear pattern of cells with $R_D \geq 0.9$ (dark-red color), while $R_D \geq 0.9$ occasionally for most cells outside this general region. This pattern suggests the geoelectric information from station LIOQ is approximately local. This is consistent with the logical requirement for direct/indirect structural relation between station LIOQ and region X, such as being close to the same subterrain fault line, for the information at station LIOQ to be useful for region X. As a corollary, information given by station LIOQ is less likely to be useful for far away regions, as they are less likely to have such structural relations with station LIOQ. In application scenarios, this means that the state of EQ probabilities of region X can be estimated using stations closer to the region. Last but not least, it is also worth mentioning that most cells at the edge of the map seldom have high R_D values. This is consistent with the fact that these cells typically have very few EQ events to provide high statistical significance.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions

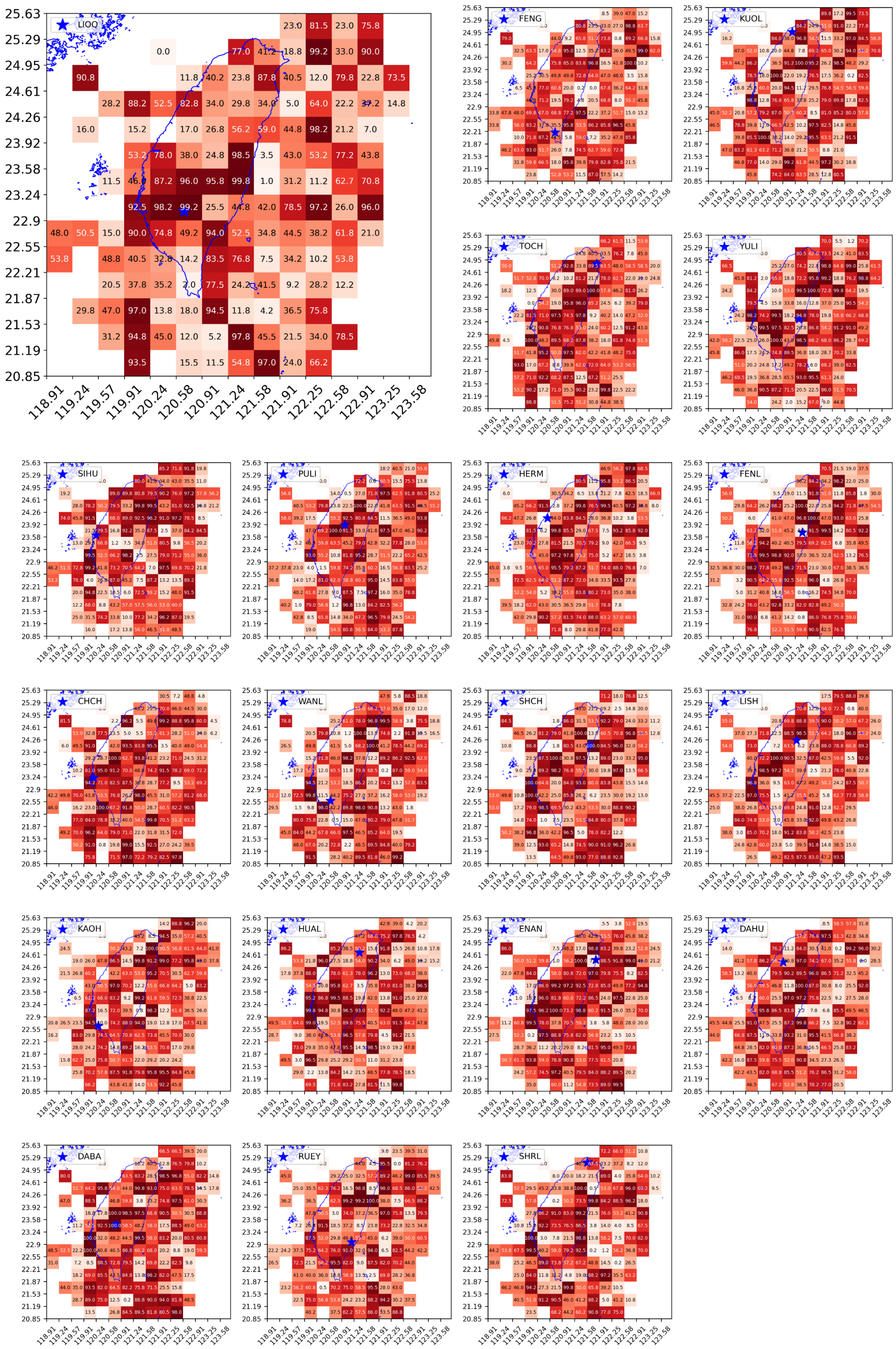

Figure 11: The grid map of Discrimination Reliability R_D ($\times 0.01$) for 20 stations (obtained using optimal hyperparameters individually specified for each station in Fig. 12).

Based on our discoveries on the HS-EQ correlations so far, it is clear that the HS TSs can provide usable EQ forecasts for real-world applications. For example, in a high-performance grid cell such as the one where LIOQ is situated, the corresponding HS TS can tell us confidently ($p = 0.008$) whether the current time is within the "active state" featuring more frequent EQs or the "passive state" featuring less frequent EQs. Let us note that the above statement is about how the EQ frequency deviates from its long-term value. Since the regime switches after every few months to every few years (e.g., CHCH in Fig. 10), what we have is therefore an intermediate-term EQ forecasting method. For grid cells with high R_D, the corresponding HS TS alone is sufficient to make such intermediate-term EQ forecasts. However, we also have grid cells where none of the 20 stations provide sufficiently high R_D value for intermediate-term EQ forecasting on their own. These could still be useful if we combine all 20 HS TSs as input features, for higher-level forecasting algorithms trained individually for each grid cell. For example, for any region (grid cell), if we want to decide whether it currently belongs to the active regime or the passive regime, an algorithm use the input from all 20 stations to decide the "local" HS for the given grid cell. This high-level algorithm can for example be weight-based model averaging (Marzocchi et al., 2012) or decision trees (Asim et al., 2016). Additionally, the value of R_D can be helpful for the algorithm to decide how to weigh the information given by all 20 stations. For example, we can consider only stations with $R_D \geq R_{D\_min}$ at the given grid cell. The user-defined threshold R_{D_min} can take on constant values (e.g., 0.9) across the grid map, or be location specific, such as being lower (e.g., 0.8) for grid cells where few of the 20 stations have $R_D \geq 0.9$. We hope to explore this in future works.

Due to the nature of our HSs, we cannot use them to forecast specific EQs or issue evacuation alarms. What the HSs can do, however, is to provide information with forecasting skill to decision makers, in regions where the HS switched from the passive state to the active state convincingly (i.e., the observed active state is persistent and not a temporary fluctuation), to take courses of action that can lower the potential damage with minimal costs. For example, in the passive state, the building inspection authority can prioritize the inspection and issuing safety permits to new projects over re-inspections of old buildings. With the arrival of an active state that might last a few months to a few years, local authorities would have the incentive to clear up pending re-inspection works, so that fewer old buildings are exposed to potential EQ damage. Other than the re-inspection of old buildings, local authorities could also increase the frequency of safety education and drills to vulnerable groups such as students and construction workers, to reduce potential injuries or fatalities due to panic or lack of understanding. Additionally, disaster relief services may use the HS's information to re-deploy the stockpile of relief materials such as food, clothing, tents, and first-aid kits, whenever necessary. In doing so, the stockpile of relief materials can be brought closer to high-risk regions within a convincing active state, to be distributed to victims more cost-effectively after a major EQ.

3.4 Global-level Significance Tests of the Forecasting Power

From Fig. 11 alone, we have demonstrated the HS TSs are able to separate time periods of low/high EQ probabilities for regions (cells in the grid map) with high R_D values. While the forecasting power of HS TSs in each of these cells is statistically significant, the more critical among us may wonder whether some of these cells can be significant purely by chance, even though there is in reality no persistent correlation between EQs and HSs. For example, any simulated HS TS in Fig. 10 would have at least a few cells with high R_D values. Therefore, in this next section, we will answer the question of "whether

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

these HS TSs indeed contain useful information about EQs, or the number of 'significant' cells can be explained by a random null model where the EQs and HSs are mutually uninformative, because we test a large number of cells assuming that they are statistically independent."

In order to answer this question, we need to define a performance metric that can quantify the performance of each station with a single value, instead of a grid map of R_D values. We start by assuming that all stations have zero forecasting skill, but as a result of our statistical test, some cells may still end up with high R_D by chance. A truly informative station should have significantly more cells with high R_D than random guesses. Taking the number of EQs into the consideration, we further propose that a truly informative station should have significantly higher EQs counts located in high-performing cells. On the grid map, let us define cells with $R_D \geq R_{D\_min}$ as *satisfactory* cells, and the rest as *unsatisfactory* cells, where R_{D_min} is the user-defined threshold that determine how high the R_D should be in order to be considered "high-performing". As mentioned earlier, it is possible to work out schemes that allow for regionally acceptable R_{D_min}. Here for simplicity let us consider a scheme with a uniform R_{D_min} across all cells in the grid map. With this setting we can proceed to define the single-value performance metric for each station, as *the ratio of EQs in satisfactory cells*, or R_{EQS} as:

$$R_{EQS} = \frac{\sum_{satisfactory\ cells} N_{EQ}}{\sum_{all\ cells} N_{EQ}}, \tag{17}$$

where N_{EQ} is the number of EQs in each cell. This *ratio of EQs in satisfactory cells* takes on values $0 \leq R_{EQS} \leq 1$. Intuitively, if $R_{EQS} = 0.4$, it means that given the R_{D_min} value, 40% of all EQs are located within satisfactory cells, and are therefore "forecasted" by the station to the level required by the user (i.e., R_{D_min}). Therefore, to show that a station has more forecasting power than random guesses, we proceed to test a given station against the null hypothesis is that a random guess (simulated HS TS) can have the same or higher R_{EQS} than the empirical HS TS.

We carried out this hypothesis test station by station, by first computing the R_{EQS} values of its empirical HS TS as well as for 400 HS TSs simulated using the HMM parameters for the given station. We then defined the *global confidence level* as:

$$GCL = \frac{\#\ (simulated\ R_{EQS} < empirical\ R_{EQS})}{400}. \tag{18}$$

Similar to the p-value for the cellular-level hypothesis test, the p-value for this global-level hypothesis test is given by $p = 1 - GCL$, where GCL ranges from $[0,1]$, and gives the probability that the empirical HS TSs having higher R_{EQS} than its simulated counterparts. For example, if a station has $GCL = 0.99$, we can say that given the specified R_{D_min}, we are 99% confident that the empirical HS TS yields higher R_{EQS} than its simulated counterparts.

In Fig. 12, we show the results of our global-level significance tests, for a choice of $R_{D\_min} = 0.95$, in the form of histograms of the 400 simulated R_{EQS} values, compared against the empirical R_{EQS} values. Except for LIOQ and LISH stations, we can see from Fig. 12 that all the other stations have GCL values close to 1. This tells us that the empirical R_{EQS} values of the 18 stations are statistically significant. We also observed that for $R_{D\_min} = 0.95$, the simulated R_{EQS} values are mostly around (or below) 0.05, meaning that only 5% of EQs are located in satisfactory cells by chance. In contrast, the empirical R_{EQS} values are mostly above 0.2, except for TOCH, LIOQ, PULI, HERM, and LISH. This is an important finding, as it shows the HS TSs' EQ forecasting utility to be significant at the global level, through the

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

use of R_{EQS} and its significance level GCL.

Last but not least, the histograms for each station in Fig. 12 are created with individually optimized hyperparameters, namely L_w (length of time window to compute indices C, V, S, and K, in days) and Q (number of clusters for the k-means clustering). The optimal hyperparameter values for each station are indicated in the titles for each station. Let us discuss the details of this optimization process in the next section.

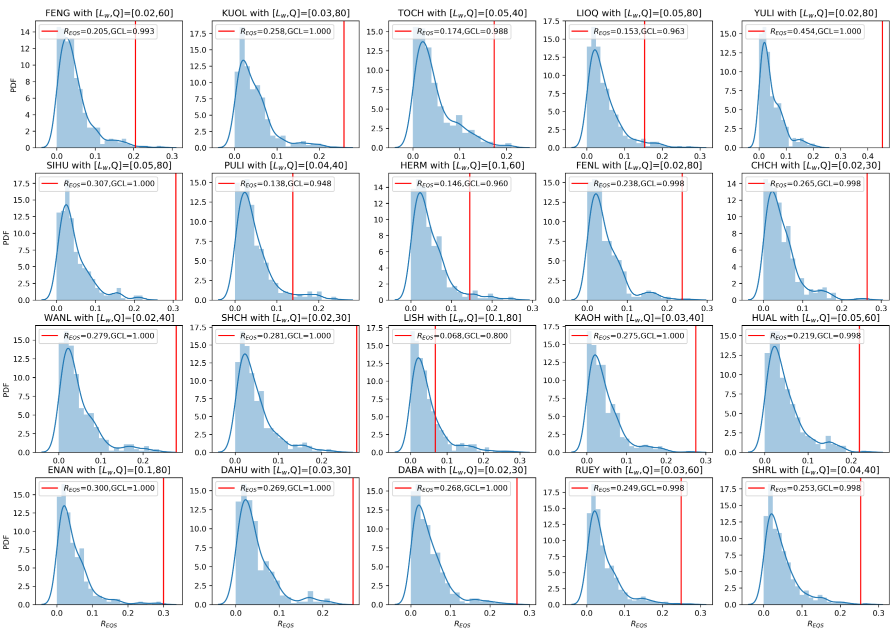

Figure 12: Histograms (blue) of 400 simulated R_{EQS} values compared against the empirical R_{EQS} (red vertical line) for 20 stations and $R_{D\_min} = 0.95$, with the GCL values in the legends. The hyperparameters of $[L_w, Q]$ optimized for each station are shown at each subplot's titles.

3.5 Significance Levels Across the Hyperparameter Space

Typically, a forecasting model's performance may be sensitive to our choice of hyperparameters. If possible, we would like to choose hyperparameters that make the model the most predictive. If there are too many hyperparameters, this optimization would be challenging in the high-dimensional search space. Fortunately, there are only two hyperparameters needed to obtain the HS TS: $[L_w, Q]$. In this section, we show how the model performance (GCL) will vary across the tested hyperparameter space, as well as how we chose the hyperparameters $[L_w, Q]$, for each station. Due to the high computational cost to test each combination of $[L_w, Q]$ (about 40 mins per station on a desktop with 4-GHz quad-core i7 processors, 16-GB of RAM, running macOS Mojave 10.14.6), we performed a coarse grid search over 28 points in the parameter space, consisting of 7 different L_w values: $[0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.25]$ days (or $[28.8, 43.2, 57.6, 72, 144, 288, 360]$ mins) and 4 different Q values: $[30, 40, 60, 80]$. We decided on this search space based on our experience during the model

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

development stage. For real-world applications, where more computational resources can be invested, this hyperparameter optimization can be done over a larger and finer grid, in which case better results can be expected.

For each choice of station and hyperparameter, we followed the same procedure of computing 1+400 R_{EQS} values, as well as the resulting GCL value. In Figs 13 and 14, we show the 20 heatmaps of R_{EQS} and GCL across the hyperparameter space, respectively for $R_{D\_min} = 0.95$. The results shown in Fig. 14 are more intuitive, where we found that for many stations, the GCL values approach 1 across broad regions of the hyperparameter space. This can for example be the full hyperparameter space for YULI station, or a patch within the hyperparameter space for KUOL station. There is just one station (LISH) with poor GCL values everywhere in the hyperparameter space, indicating that there might be exclusive factors that severely limit station LISH's forecasting power. For all other 19 stations, the GCL values are close to 1 either across a large area of the parameter space, or almost the entire parameter space (e.g., YULI, WANL, ENAN, DABA). This result is compelling, and is exactly what we needed for our goal: *to demonstrate the forecasting skill of the HS TS, which does not depend on highly optimized hyperparameters, but is valid over a broad range of hyperparameters.*

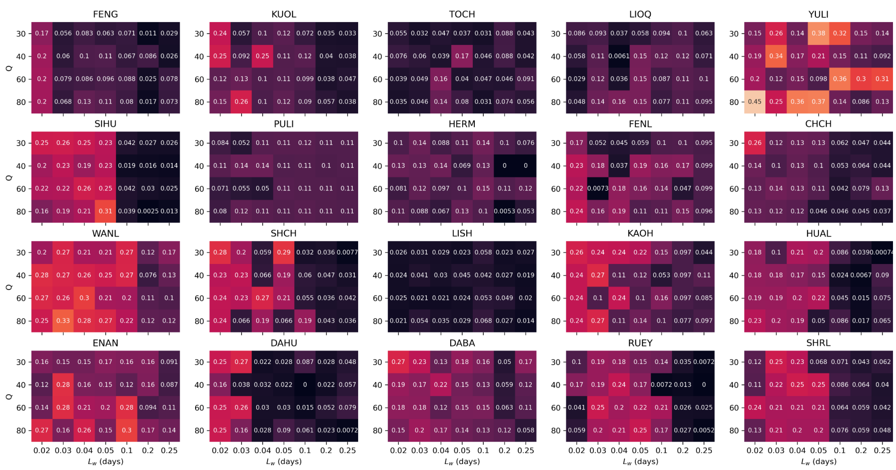

Figure 13: Heatmaps of R_{EQS} values for all 20 stations across tested hyperparameter space, given $R_{D\_min} = 0.95$.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Discussions
Open Access

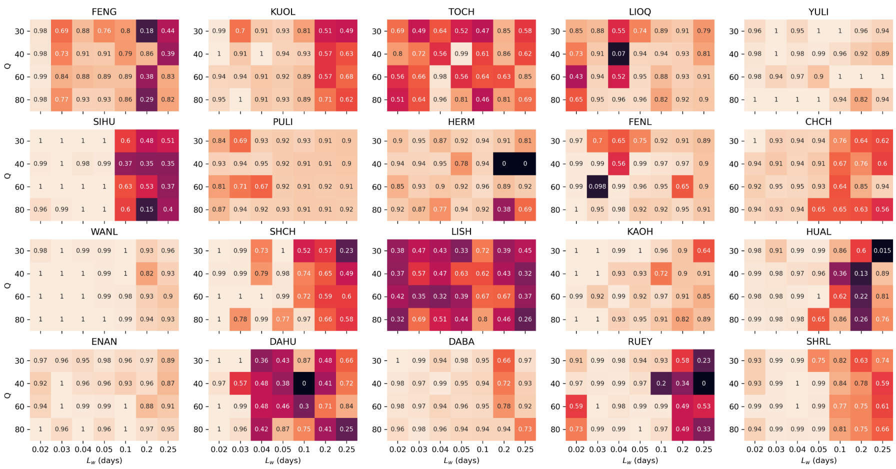

Figure 14: Heatmaps of GCL values for all 20 stations across tested hyperparameter space, given $R_{D\_min} = 0.95$.

To wrap up this section, let us describe how to select the optimal hyperparameter for each station. We did this in two steps: first, we selected the hyperparameters with highest GCL values (1 for many stations); next, in case of ties, we chose the hyperparameter with the highest R_{EQS} as the winner. For example, for WANL station in Fig. 14, there are many cells with $GCL = 1$. We therefore proceeded to check the heatmap for station WANL in Fig. 13, and identified the hyperparameter combination $L_w = 0.03$ and $Q = 80$ as optimal, since it has the highest R_{EQS} value. Using this selection procedure, we identified the optimal hyperparameter for each station, and used these individually optimal hyperparameters to create Figs 7 to 12. This selection procedure could also be adapted for real-world applications, when more historical data and computational power are available, to provide even better model performances.

4 Conclusions

EQ forecasting is an important research topic, because of the potential devastation EQs can cause. As pointed out by many past studies, there is a correlation between features within geoelectric TSs and individual large EQs. In those studies, different features of geoelectric TSs were explored for their use of EQ forecasting, among which the GEMSTIP model was the first one to directly use statistic index TSs of geoelectric TSs to produce TIPs for EQ forecasting. Inspired by this, we took a second look at the relationship between these statistic indexes and the timing of EQs, and found out that there is an abrupt shift of the indexes' distribution along the TTF axis. This suggested that there are at least two distinct geoelectric regimes, which can be modeled and identified using a 2-state HMM. This motivation is further backed by the knowledge that there can be drastic tectonic configuration changes before and after a large EQ, one important aspect of which being the telluric changes identified in the region around the epicenter of the EQ (Sornette and Sornette, 1990; Tong-En et al., 1999; Orihara et al., 2012; Kinoshita et al., 1989; Nomikos et al., 1997). Therefore, should there be two higher-level tectonic regimes featuring higher/lower EQ frequencies, we would expect to also find two matching geoelectric regimes with

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

contrasting statistical properties, which can be of good utility for EQ forecasting.

Specifically, we modeled the earth crust system as having two HSs identifiable with distinctive geoelectric features encoded by 8 index TSs from each station. To obtain the HMM for each station, we needed to run the BWA, which is most convenient to use with a discrete observation TS input. Therefore, we used the k-means clustering to convert the continuous TS of 8-dimensional index vectors into a discrete observation TS, and subsequently obtained a converged HMM for each station. We then investigated whether these HS TSs provide informative partitions of EQs, i.e., one of the HS can be interpreted as a passive state with less frequent EQs, while the other one as an active state with more frequent EQs. For this task, we defined the *EQ frequency's ratio* (R_F), which is the frequency of EQs in one of the HSs divided by the total frequency of the EQs. Using R_F we further defined the *discrimination power* (D), to measure how differently one HS is from the other HS in terms of the EQ frequency. We then plotted 16-by-16 grid maps of R_F and D for all 20 stations, and tested the statistical significance of D in each cell, by comparing the empirical value against the distribution of D from 400 simulated HS TSs, to end up with the grid maps of *discrimination reliability* (R_D) for all 20 stations. To further investigate the statistical significance level at the global scale, we defined R_{EQS} to measure the percentage of total EQs located within satisfactory cells, i.e., cells having $R_D \geq R_{D\_min}$ for a user-specified R_{D_min} value. This R_{D_min} value can be easily customized for different cells, but in this paper, we used a constant R_{D_min} value across the grid map for demonstration. By comparing the R_{EQS} value of the empirical model against those of 400 simulated models, we obtained one global significance value for each station, namely the *global confidence level* (GCL). This tells us how confident we can be that information contained in the empirical HS TSs can be used for EQ forecasting.

Finally, we showed how we optimized the GCL values through a grid-search in the 2-dimensional hyperparameter space and obtained the optimal combination of $[L_w, Q]$ individually for each station. As a result, among the 20 stations with optimized hyperparameters, there are 19 stations with $GCL > 0.95$, 15 of which having $GCL > 0.99$. Additionally, the confidence levels are also robust across the hyperparameter space for most stations. Based on these positive results, the Hidden Markov Modelling of the index TSs computed from geoelectric TSs is indeed a viable way to extract information that can be useful for EQ forecasting. As discussed in greater detail in Sect. 3.3, in real-world scenarios, the HS TSs can be useful for intermediate-term EQ forecasting either directly (for high R_D cells) or as input features for higher-level algorithms that take information from all 20 stations (for low R_D cells). Beyond our demonstration of extracting EQ-related information from geoelectric TSs, the HMM approach described in this paper can also be explored on other high-frequency geophysical data, such as those from geomagnetic, geochemical, hydrological, and GPS measurements, for EQ forecasting.

At this point, we would like to address the issue of out-of-sample testing (or cross-validation) to support the validity of our model. There are two ways to do this: (1) split a long time series into a training data set to calibrate the model, and a testing data set to validate the model; and (2) use whatever time series data is available to calibrate the model, before collecting more data to validate the model. If the model is statistically stationary (its parameters do not change with time), both approaches are acceptable. However, many would agree that an out-of-sample test with freshly collected data (approach (2)) is more impressive, especially if it is done in real time. We would certainly like to try this, and are writing a grant application to fund such a validation study. For this paper, however, we were not even able to use approach (1), because our geoelectric time series are not long enough. This is especially so if we require **(A)** the validation data is always temporally *after* the training data; and **(B)** the validation data is also intermediate-term for intermediate-term EQ forecasting. These two requirements cannot be fulfilled on our limited 7-year data, if we want to have a significant number (e.g., 10 times) of validations to produce confident claims. Therefore, in this paper, we limited our scope to demonstrating that our model has *forecasting skill*, without quantifying its *exact forecasting accuracy*. We argue that we have indeed achieved this, without the use of out-of-sample testing, because in Sect. 3.5, we showed *the forecasting skill is statistically significant regardless of the choice of the hyperparameters*, for 19 out of the 20 stations that we tested. Furthermore, the statistical hypothesis test has the advantage of giving rigorous p-value with moderate computation cost, through simulating the HMM for multiple null-hypothesis tests.

Code Availability

The python codes that we used to produce the results in this paper can be downloaded at GitHub: https://github.com/wenhy1111/HMM_Geoelectric_EQ.

Data Availability

The dataset of the index TSs for 20 stations computed using various time window of lengths (L_w) is available in a repository and can be accessed via a DOI link: https://doi.org/10.21979/N9/JSUTCD. For the 0.5-Hz geoelectric TS data for 20 stations, the data is available on request by contacting Hong-Jia Chen (redhouse6341@gmail.com) or Chien-Chih Chen (chienchih.chen@g.ncu.edu.tw). The EQ catalogue data is owned by a third party, the Central Weather Bureau in Taiwan.

Author Contribution

Author Contribution Statement: SAC and CCC came up with the research motivation; HJC and HW processed the data; SAC and HW analyzed the results; SAC, HW, and HJC drafted the manuscript; all co-authors read the manuscript and suggested revisions.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgments

No available acknowledgements.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

References

Abdel-Hamid, O. and Jiang, H.: Fast speaker adaptation of hybrid NN/HMM model for speech recognition based on discriminative learning of speaker code, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 7942-7946,

Asim, K., Martínez-Álvarez, F., Basit, A., and Iqbal, T.: Earthquake magnitude prediction in Hindukush region using machine learning techniques, Natural Hazards, 85, 471-486, 2017.

Asim, K. M., Idris, A., Martínez-Álvarez, F., and Iqbal, T.: Short term earthquake prediction in Hindukush region using tree based ensemble learning, 2016 International conference on frontiers of information technology (FIT), 365-370,

Batac, R. and Kantz, H.: Observing spatio-temporal clustering and separation using interevent distributions of regional earthquakes, Nonlinear Processes in Geophysics, 21, 735-744, 2014.

Beyreuther, M. and Wassermann, J.: Continuous earthquake detection and classification using discrete Hidden Markov Models, Geophysical Journal International, 175, 1055-1066, 2008.

Bilmes, J. A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, International Computer Science Institute, 4, 126, 1998.

Chambers, D. W., Baglivo, J. A., Ebel, J. E., and Kafka, A. L.: Earthquake forecasting using hidden Markov models, Pure and applied geophysics, 169, 625-639, 2012.

Chavan, R. S. and Sable, G. S.: An overview of speech recognition using HMM, International Journal of Computer Science and Mobile Computing, 2, 233-238, 2013.

Chen, C.-c.: Accelerating seismicity of moderate-size earthquakes before the 1999 Chi-Chi, Taiwan, earthquake: Testing time-prediction of the self-organizing spinodal model of earthquakes, Geophysical Journal International, 155, F1-F5, 2003.

Chen, H.-J. and Chen, C.-C.: Testing the correlations between anomalies of statistical indexes of the geoelectric system and earthquakes, Natural Hazards, 84, 877-895, 2016.

Chen, H.-J., Chen, C.-C., Ouillon, G., and Sornette, D.: Using skewness and kurtosis of geoelectric fields to forecast the 2016/2/6, ML6. 6 Meinong, Taiwan Earthquake, Terrestrial, Atmospheric and Oceanic Sciences, 28, 745-761, 2017.

Chen, H.-J., Chen, C.-C., Ouillon, G., and Sornette, D.: A paradigm for developing earthquake probability forecasts based on geoelectric data, The European Physical Journal Special Topics, 230, 381-407, 2021.

Chen, H.-J., Chen, C.-C., Tseng, C.-Y., and Wang, J.-H.: Effect of tidal triggering on seismicity in Taiwan revealed by the empirical mode decomposition method, Natural Hazards and Earth System Sciences, 12, 2193-2202, 2012.

Chen, H. J., Ye, Z. K., Chiu, C. Y., Telesca, L., Chen, C. C., and Chang, W. L.: Self-potential ambient noise and spectral relationship with urbanization, seismicity, and strain rate revealed via the Taiwan Geoelectric Monitoring Network, Journal of Geophysical Research: Solid Earth, 125, e2019JB018196, 2020.

Cho, S.-B. and Park, H.-J.: Efficient anomaly detection by modeling privilege flows using hidden Markov model, computers & security, 22, 45-55, 2003.

Cornell, C. A.: Engineering seismic risk analysis, Bulletin of the seismological society of America, 58, 1583-1606, 1968.

Dash, R., Paramguru, R. L., and Dash, R.: Comparative analysis of supervised and unsupervised discretization techniques, International Journal of Advances in Science and Technology, 2, 29-37, 2011.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Dieterich, J.: A constitutive law for rate of earthquake production and its application to earthquake clustering, Journal of Geophysical Research: Solid Earth, 99, 2601-2618, 1994.

Ebel, J. E., Chambers, D. W., Kafka, A. L., and Baglivo, J. A.: Non-Poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California, Seismological Research Letters, 78, 57-65, 2007.

Fischer, T. and Bachura, M.: Detection capability of seismic network based on noise analysis and magnitude of completeness, Journal of seismology, 18, 137-150, 2014.

Frohlich, C.: Aftershocks and temporal clustering of deep earthquakes, Journal of Geophysical Research: Solid Earth, 92, 13944-13956, 1987.

Geller, R. J., Jackson, D. D., Kagan, Y. Y., and Mulargia, F.: Earthquakes cannot be predicted, Science, 275, 1616-1616, 1997.

Gupta, A., Mehrotra, K. G., and Mohan, C.: A clustering-based discretization for supervised learning, Statistics & probability letters, 80, 816-824, 2010.

Hayakawa, M. and Hobara, Y.: Current status of seismo-electromagnetics for short-term earthquake prediction, Geomatics, Natural Hazards and Risk, 1, 115-155, 2010.

Healy, J. H., Kossobokov, V. G., and Dewey, J. W.: A test to evaluate the earthquake prediction algorithm, M8, 401, US Geological Survey1992.

Holbrook, J., Autin, W. J., Rittenour, T. M., Marshak, S., and Goble, R. J.: Stratigraphic evidence for millennial-scale temporal clustering of earthquakes on a continental-interior fault: Holocene Mississippi River floodplain deposits, New Madrid seismic zone, USA, Tectonophysics, 420, 431-454, 2006.

Huang, Q. and Lin, Y.: Selectivity of seismic electric signal (SES) of the 2000 Izu earthquake swarm: a 3D FEM numerical simulation model, Proceedings of the Japan Academy, Series B, 86, 257-264, 2010.

Ismail-Zadeh, A. T.: Earthquake Prediction and Forecasting, in: Encyclopedia of Natural Hazards, edited by: Bobrowsky, P. T., Springer Netherlands, Dordrecht, 225-231, 10.1007/978-1-4020-4399-4_106, 2013.

Jiang, F., Chen, X., Chen, C.-C., and Chen, H.-j.: Relationship between seismic electric signals and tectonics derived from dense geoelectric observations in Taiwan, Pure and Applied Geophysics, 177, 441-454, 2020.

Joshi, S. S. and Phoha, V. V.: Investigating hidden Markov models capabilities in anomaly detection, Proceedings of the 43rd annual Southeast regional conference-Volume 1, 98-103,

Kagan, Y. and Jackson, D.: Long-term probabilistic forecasting of earthquakes, Journal of Geophysical Research: Solid Earth, 99, 13685-13700, 1994.

Kagan, Y. Y.: Are earthquakes predictable?, Geophysical Journal International, 131, 505-525, 1997.

Kamiyama, M., Sugito, M., Kuse, M., Schekotov, A., and Hayakawa, M.: On the precursors to the 2011 Tohoku earthquake: crustal movements and electromagnetic signatures, Geomatics, Natural Hazards and Risk, 7, 471-492, 2016.

Kanamori, H.: Earthquake prediction: An overview, 2003.

Keilis-Borok, V. I.: Intermediate-term earthquake prediction, Proceedings of the National Academy of Sciences, 93, 3748-3755, 1996.

Keilis-Borok, V. I. and Rotwain, I.: Diagnosis of time of increased probability of strong earthquakes in different regions of the world: algorithm CN, Physics of the Earth and Planetary Interiors, 61, 57-72, 1990.

Kinoshita, M., Uyeshima, M., and Uyeda, S.: Earthquake Prediction Research by Means of Telluric Potential Monitoring: Progress Report No. 1: Installation of Monitoring Network, 東京大學地震研究

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

所彙報= Bulletin of the Earthquake Research Institute, University of Tokyo, 64, 255-311, 1989.

Kolbe, A. R., Hutson, R. A., Shannon, H., Trzcinski, E., Miles, B., Levitz, N., Puccio, M., James, L., Noel, J. R., and Muggah, R.: Mortality, crime and access to basic needs before and after the Haiti earthquake: a random survey of Port-au-Prince households, Medicine, conflict and survival, 26, 281-297, 2010.

Kossobokov, V., Romashkova, L., Panza, G., and Peresan, A.: Stabilizing intermediate-term medium-range earthquake predictions, Journal of Seismology and Earthquake Engineering, 4, 11-19, 2002.

Kuo-Chen, H., Wu, F. T., and Roecker, S. W.: Three-dimensional P velocity structures of the lithosphere beneath Taiwan from the analysis of TAIGER and related seismic data sets, Journal of Geophysical Research: Solid Earth, 117, 2012.

Larue, P., Jallon, P., and Rivet, B.: Modified k-mean clustering method of HMM states for initialization of Baum-Welch training algorithm, 2011 19th European Signal Processing Conference, 951-955,

Marzocchi, W., Zechar, J. D., and Jordan, T. H.: Bayesian forecast evaluation and ensemble earthquake forecasting, Bulletin of the Seismological Society of America, 102, 2574-2584, 2012.

Meletti, C., Galadini, F., Valensise, G., Stucchi, M., Basili, R., Barba, S., Vannucci, G., and Boschi, E.: A seismic source zone model for the seismic hazard assessment of the Italian territory, Tectonophysics, 450, 85-108, 2008.

Mignan, A., Werner, M., Wiemer, S., Chen, C.-C., and Wu, Y.-M.: Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs, Bulletin of the Seismological Society of America, 101, 1371-1385, 2011.

Moustra, M., Avraamides, M., and Christodoulou, C.: Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals, Expert systems with applications, 38, 15032-15039, 2011.

Nanjo, K., Rundle, J., Holliday, J., and Turcotte, D.: Pattern informatics and its application for optimal forecasting of large earthquakes in Japan, in: Computational Earthquake Physics: Simulations, Analysis and Infrastructure, Part II, Springer, 2417-2432, 2006.

Nanjo, K., Ishibe, T., Tsuruoka, H., Schorlemmer, D., Ishigaki, Y., and Hirata, N.: Analysis of the completeness magnitude and seismic network coverage of Japan, Bulletin of the Seismological Society of America, 100, 3261-3268, 2010.

Nath, S. and Thingbaijam, K.: Probabilistic seismic hazard assessment of India, Seismological Research Letters, 83, 135-149, 2012.

Nomikos, K., Vallianatos, F., Kaliakatsos, I., Sideris, E., and Bakatsakis, M.: The latest aspects of telluric and electromagnetic variations associated with shallow and intermediate depth earthquakes in the South Aegean, Annals of Geophysics, 40, 1997.

Novoa, J., Wuth, J., Escudero, J. P., Fredes, J., Mahu, R., and Yoma, N. B.: DNN-HMM based automatic speech recognition for HRI scenarios, Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, 150-159,

Orihara, Y., Kamogawa, M., Nagao, T., and Uyeda, S.: Preseismic anomalous telluric current signals observed in Kozu-shima Island, Japan, Proceedings of the National Academy of Sciences, 109, 19125-19128, 2012.

Oudelha, M. and Ainon, R. N.: HMM parameters estimation using hybrid Baum-Welch genetic algorithm, 2010 International Symposium on Information Technology, 542-545,

Palaz, D., Magimai-Doss, M., and Collobert, R.: End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition, Speech Communication, 108, 15-32,

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

2019.

Papadimitriou, E. E.: Long-term earthquake prediction along the western coast of South and Central America based on a time predictable model, in: Shallow Subduction Zones: Seismicity, Mechanics and Seismic Potential Part 1, Springer, 301-316, 1993.

Papazachos, B., Papadimitriou, E., Karakaisis, G., and Panagiotopoulos, D.: Long-term earthquake prediction in the circum-Pacific convergent belt, pure and applied geophysics, 149, 173-217, 1997.

Papazachos, B., Papadimitriou, E., Kiratzi, A., Papaioannou, C. A., and Karakaisis, G.: Probabilities of occurrence of large earthquakes in the Aegean and surrounding area during the period 1986–2006, pure and applied geophysics, 125, 597-612, 1987.

Peresan, A., Kossobokov, V., Romashkova, L., and Panza, G.: Intermediate-term middle-range earthquake predictions in Italy: a review, Earth-Science Reviews, 69, 97-132, 2005.

Petersen, M. D.: Probabilistic seismic hazard assessment for the state of California, 706, California Department of Conservation Division of Mines and Geology1996.

Qiao, Y., Xin, X., Bin, Y., and Ge, S.: Anomaly intrusion detection method based on HMM, Electronics letters, 38, 663-664, 2002.

Haitians angry over slow aid: https://www.smh.com.au/world/haitians-angry-over-slow-aid-20100204-ng2g.html, last access: 23 April.

Reyes, J., Morales-Esteban, A., and Martínez-Álvarez, F.: Neural networks to predict earthquakes in Chile, Applied Soft Computing, 13, 1314-1328, 2013.

Rundle, J. B., Klein, W., Turcotte, D. L., and Malamud, B. D.: Precursory seismic activation and critical-point phenomena, in: Microscopic and Macroscopic Simulation: Towards Predictive Modelling of the Earthquake Process, Springer, 2165-2182, 2000.

Rydelek, P. A. and Sacks, I. S.: Testing the completeness of earthquake catalogues and the hypothesis of self-similarity, Nature, 337, 251-253, 1989.

Sarlis, N., Lazaridou, M., Kapiris, P., and Varotsos, P.: Numerical model of the selectivity effect and the ΔV/L criterion, Geophysical research letters, 26, 3245-3248, 1999.

Sarlis, N. V.: Statistical Significance of Earth's Electric and Magnetic Field Variations Preceding Earthquakes in Greece and Japan Revisited, Entropy, 20, 561, 2018.

Butterworth Bandpass: https://scipy-cookbook.readthedocs.io/items/ButterworthBandpass.html, last access: 23 April.

Shebalin, P., Keilis-Borok, V., Gabrielov, A., Zaliapin, I., and Turcotte, D.: Short-term earthquake prediction by reverse analysis of lithosphere dynamics, Tectonophysics, 413, 63-75, 2006.

Shin, T.-C., Chang, C.-H., Pu, H.-C., Hsiao-Wei, L., and Leu, P.-L.: The geophysical database management system in Taiwan, TAO: Terrestrial, Atmospheric and Oceanic Sciences, 24, 11, 2013.

伤亡汇总_四川汶川强烈地震_新闻中心_新浪网 [Casualties of the Wenchuan Earthquake] (in Chinese): http://news.sina.com.cn/pc/2008-05-13/326/651.html, last access: 23 April.

Sornette, A. and Sornette, D.: Earthquake rupture as a critical point: Consequences for telluric precursors, Tectonophysics, 179, 327-334, 1990.

M 9.1 - 2004 Sumatra - Andaman Islands Earthquake: https://earthquake.usgs.gov/earthquakes/eventpage/official20041226005853450_30/executive, last access: 23 April.

Sykes, L. R.: Intermediate-and long-term earthquake prediction, Proceedings of the National Academy of Sciences, 93, 3732-3739, 1996.

Tavakoli, B. and Ghafory-Ashtiany, M.: Seismic hazard assessment of Iran, 1999.

https://doi.org/10.5194/nhess-2021-378
Preprint. Discussion started: 4 January 2022

Natural Hazards and Earth System Sciences
Open Access
Discussions

Telesca, L., Lapenna, V., and Lovallo, M.: Fisher Information Analysis of earthquake-related geoelectrical signals, Natural Hazards and Earth System Sciences, 5, 561-564, 2005a.

Telesca, L., Lapenna, V., and Macchiato, M.: Multifractal fluctuations in earthquake-related geoelectrical signals, New Journal of Physics, 7, 214, 2005b.

Telesca, L., Lovallo, M., Ramirez-Rojas, A., and Angulo-Brown, F.: A nonlinear strategy to reveal seismic precursory signatures in earthquake-related self-potential signals, Physica A: Statistical Mechanics and its Applications, 388, 2036-2040, 2009.

Telesca, L., Lovallo, M., Romano, G., Konstantinou, K. I., Hsu, H.-L., and Chen, C.-c.: Using the informational Fisher–Shannon method to investigate the influence of long-term deformation processes on geoelectrical signals: An example from the Taiwan orogeny, Physica A: Statistical Mechanics and its Applications, 414, 340-351, 2014.

TONG-EN, M., JI-LOU, X., YAN-QIONG, W., and SHU-ZHI, Y.: The variation characteristics of the telluric field in the process of earthquake [J], Chinese Journal of Geophysics, 4, 1999.

Uyeda, S., Nagao, T., Orihara, Y., Yamaguchi, T., and Takahashi, I.: Geoelectric potential changes: Possible precursors to earthquakes in Japan, Proceedings of the National Academy of Sciences, 97, 4561-4566, 2000.

Varotsos, P. and Lazaridou, M.: Latest aspects of earthquake prediction in Greece based on seismic electric signals, Tectonophysics, 188, 321-347, 1991.

Varotsos, P., Alexopoulos, K., and Lazaridou, M.: Latest aspects of earthquake prediction in Greece based on seismic electric signals, II, Tectonophysics, 224, 1-37, 1993.

Varotsos, P., Sarlis, N., and Skordas, E.: Long-range correlations in the electric signals that precede rupture, Physical Review E, 66, 011902, 2002.

Varotsos, P., Sarlis, N., Skordas, E., and Lazaridou, M.: Seismic Electric Signals: An additional fact showing their physical interconnection with seismicity, Tectonophysics, 589, 116-125, 2013.

Varotsos, P. A., Sarlis, N. V., and Skordas, E. S.: Identifying the occurrence time of an impending major earthquake: a review, Earthquake science, 30, 209-218, 2017.

Vilanova, S. P. and Fonseca, J. F.: Probabilistic seismic-hazard assessment for Portugal, Bulletin of the Seismological Society of America, 97, 1702-1717, 2007.

Wang, T. and Bebbington, M.: Identifying anomalous signals in GPS data using HMMs: An increased likelihood of earthquakes?, Computational Statistics & Data Analysis, 58, 27-44, 2013.

Wang, Y.-J., Chan, C.-H., Lee, Y.-T., Ma, K.-F., Shyu, J. B. H., Rau, R.-J., and Cheng, C.-T.: Probabilistic seismic hazard assessment for Taiwan, Terr. Atmos. Ocean. Sci., 27, 325-340, 2016.

Wen, F., Chen, Z.-h., Zhuo, R., and Zhou, G.-m.: Reinforcement learning method of continuous state adaptively discretized based on K-means clustering, Control and Decision, 21, 143, 2006.

Yang, F., Balakrishnan, S., and Wainwright, M. J.: Statistical and computational guarantees for the Baum-Welch algorithm, The Journal of Machine Learning Research, 18, 4528-4580, 2017.

Yang, L., Widjaja, B., and Prasad, R.: Application of hidden Markov models for signature verification, Pattern recognition, 28, 161-170, 1995.

Yi-Ben, T.: Seismotectonics of Taiwan, Tectonophysics, 125, 17-37, 1986.

Yip, C. F., Ng, W. L., and Yau, C. Y.: A hidden Markov model for earthquake prediction, Stochastic Environmental Research and Risk Assessment, 32, 1415-1434, 2018.

Zhang, Y., Zhao, D., and Liu, J.: The application of baum-welch algorithm in multistep attack, The Scientific World Journal, 2014, 2014.