Pure Appl. Geophys.

https://doi.org/10.1007/s00024-021-02686-2

Pure and Applied Geophysics

Assessing Markovian Models for Seismic Hazard and Forecasting

Q. J. Gutierrez Peña,[1] F. A. Nava Pichardo,[1] E. Glowacka,[1] R. R. Castro Escamilla,[1] and V. H. Márquez Ramírez[2]

Abstract—We propose the use of statistical measures to quantify robustness, uncertainty, and significance for Markovian models of large magnitude seismic systems, and we also propose a method for choosing the best of different models by using the normalized measures in a discriminant function. We tested the proposed methods on earthquakes occurring in an area around Japan, divided into four regions; modeling the system as having four states, where each state corresponds to the region where the latest large earthquake, larger than a given threshold moment magnitude, has occurred. Our results show that for the 7.0–7.3 threshold magnitude range the seismicity of this region does occur according to a Markovian process, with optimum results for threshold magnitude 7.1, whereas for magnitudes outside this range seismicity is less Markovian.

Keywords: Markov chains, seismic hazard.

1. Introduction

Markovian models are a useful and widespread tool in many fields of research, including linguistics, communication, meteorology, economy, and others (e.g. Fritchman, 1967; Grant & Steven, 1991; Khmelev & Tweedie, 2001; Raible et al., 1999). In seismology, Vere-Jones (1966) used a Markovian model for aftershock occurrences, and Markovian models have been used for seismic hazard estimates for large earthquakes (Fujinawa, 1991; Hagiwara, 1975; Heng, 2002; Knopoff, 1971; Nava et al., 2005; Patwardhan et al., 1980; Ramin, 2012; Ünal & Celebioglu, 2011; Votsi et al., 2010; Votsi et al., 2013; Yildiz & Demir, 1999).

While in many applications, including some seismological ones, there is abundant data allowing reliable estimates of the transition probabilities between states, in seismological applications, for the case of large magnitude earthquakes the relative paucity of major and great earthquakes compared with the extension of most seismic catalogs limits the reliability of the Markovian seismic hazard estimates. Hence, it is of the utmost importance to determine ways to quantify the reliability of hazard estimates, and this is one of the themes we will touch in this paper.

Since usually there is not a unique way of modeling a given problem or system, another theme we will consider is how to quantify the reliability measures and the results from each model, in order to decide which model and/or set of parameters is optimal within the limits permitted by the data.

2. Background

We present here a very short review of some basic concepts about Markov chains that will be used in what follows.

2.1. Markov Process

A *finite Markov process* is a stochastic process, with a finite number N_s of states, $\{s_k;\, k = 1, \ldots, N_s\}$, $N_S > 1$, for which the probability of transition from the current state to a given state in the next trial (which may be an occurrence or a time interval) depends only on the current state, and not on any previous states. Let $s_{k(n)}$ be the state of the system at the n'th trial, then

[1] División de Ciencias de la Tierra, CICESE, Ensenada, Baja California, Mexico. E-mail: quiriatj@gmail.com
[2] Centro de Geociencias, UNAM, Juriquilla, Querétaro, Mexico.

Published online: 10 March 2021

$$\Pr\left[s_{k(n+1)}|s_{k(n)}, s_{k(n-1)}, \ldots, s_{k(0)}\right] = \Pr\left[s_{k(n+1)}|s_{k(n)}\right], \quad (1)$$

i.e., the system has no memory about the states that occurred before the present one. Let $s_{k(n)} = i$ and $s_{k(n+1)} = j$, then if the transition probabilities between these states are always the same, independently of the trial or time interval, the Markov chain is *homogeneous*, and we can write

$$\Pr\left[s_{k(n+1)} = j|s_{k(n)} = i\right] = \hat{p}_{ij}, \quad (2)$$

where the tilde denotes an exact or theoretical probability.

2.2. Markovian Multiple Step Probabilities

Let the state at step m be $s_{k(m)} = i$, then the probability that n steps later the system will be in state $s_{k(m+n)} = j$, the n-step transition probability, is given by

$$\Pr\left[s_{k(m+n)}|s_{k(m)}\right] = \hat{p}_{ij}^{(n)}, \quad (3)$$

where $\hat{p}_{ij}^{(n)}$ is an element of matrix $\hat{P}^{(n)}$ (Chapmann–Kolmogorov equation).

If numbers $\hat{\pi} = \left[\hat{\pi}_j; j = 1, \ldots, N_S\right]$ exist, such that

$$\lim_{n\to\infty} \hat{p}_{ij}^{(n)} = \hat{\pi}_j, \forall i, \quad \sum_{j=1}^{N_S} \hat{\pi}_j = 1, \quad (4)$$

so that the limit transition probabilities

$$\lim_{n\to\infty} \hat{P}^n = \hat{\Pi} = \begin{Bmatrix} \hat{\pi} \\ \vdots \\ \hat{\pi} \end{Bmatrix} \quad (5)$$

no longer depend on the current state, then the system represented by the transition matrix is *ergodic*, and the probabilities $\hat{\pi}_j$ are called *limiting* or *stationary probabilities*.

3. Empirical Estimation of Probabilities

When studying a real, supposedly Markovian system, for which the transition probabilities cannot be postulated or determined theoretically, the transition probabilities are estimated from the observed Markov chain, i.e. from the finite observed sequence of $N + 1$ states values taken by the random variable S

$$\left\{s_{k(n)}; n = 0, 1, 2, \ldots, N\right\}. \quad (6)$$

3.1. Transitions and Probabilities

The square transition matrix Θ is built by counting the observed transitions among successive states:

$$\Theta = \left[\theta_{ij}; i, j = 1, \ldots, N_S\right], \quad (7)$$

where θ_{ij} is the observed number of transitions from state i to state j.

Let ξ_i be the total number of transitions that originated from state i

$$\xi_i = \sum_{j=1}^{N_S} \theta_{ij}, \quad (8)$$

and from all total transitions we form the vector

$$\Xi = [\xi_i; i = 1, \ldots, N_S]. \quad (9)$$

From the Θ and Ξ matrices, the empirical transition probability matrix (TPM), $P = \left[p_{ij};\ i, j = 1, \ldots, N_S\right]$, is built as

$$p_{ij} = \frac{\theta_{ij}}{\xi_i}, \quad (10)$$

so that

$$\sum_{j=1}^{N_S} p_{ij} = 1,\ \forall i. \quad (11)$$

This TPM, P, is the seismic hazard evaluation resulting from the study, because if the system is currently at some state i, the probability for the next state being j is given by p_{ij}.

It should be remembered that Borel's law of large numbers tells us that p_{ij} given by (10) will tend to the "true" probability $\hat{p}_{ij}$ when $\xi_i \to \infty$; an impossible condition to attain when dealing with large magnitude earthquakes. In these applications it must be kept in mind that probability estimates are based on only a small number of realizations of a random process, so that probability estimates carry uncertainties that must be reckoned with when using seismic hazard estimates. How these uncertainties depend on the number of transitions will be discussed later.

3.2. Empirical Stationary Probabilities

The empirical, approximate, stationary probabilities, Π, are obtained from P^n after a finite number of steps, n, when convergence is achieved for a given precision (i.e., for a given number of decimals), i.e. for each column

$$\pi_j = p_{ij}^{(n)}, \ \forall i, \tag{12}$$

then

$$\Pi = \begin{bmatrix} \pi \\ \vdots \\ \pi \end{bmatrix}; \ \pi \equiv [\pi_1, \pi_2, \ldots, \pi_{N_S}] \tag{13}$$

The desired precision is mainly chosen according to how many decimals may be considered possibly significant in a probability estimate; a large number of decimals is also subject to numerical noise. In what follows we will consider precision to six decimal places.

4. Methods for Measuring Stability, Robustness, and Uncertainty of Markovian Probability Estimations

A simple determination of transition probabilities will not be very useful for seismic hazard evaluation if there is no estimate about how reliable they are; in what follows we will consider some aspects of this problem that depend largely on the size of the sample used for the probability estimation.

4.1. Stability

An empirical Markov probability estimation is essentially a never-ending process, because any new datum will, in general, change the estimations (the only exception is when the new datum contributes to a previous $p_{ij} = 1$ probability, which is a rarity). What follows is adapted from Nava et al. (2005).

Suppose we have determined a Markovian TPM, and suppose the current state is state i, when a new transition to state k occurs, the estimation of p_{ij} changes from $p_{ij} = \theta_{ij}/\xi_i$ to one increased estimate

$$p_{ij}^{+} = \frac{\theta_{ij} + 1}{\xi_i + 1}; \ j = k, \tag{14}$$

And $N_S - 1$ diminished estimates

$$p_{ij}^{-} = \frac{\theta_{ij}}{\xi_i + 1}; \ j \neq k. \tag{15}$$

The corresponding changes are

$$\Delta p_{ij}^{+} = p_{ij}^{+} - p_{ij} = \frac{1 - p_{ij}}{\xi_i + 1}; \ j = k, \tag{16}$$

and

$$\Delta p_{ij}^{-} = p_{ij}^{-} - p_{ij} = \frac{-p_{ij}}{\xi_i + 1}; \ j \neq k; \tag{17}$$

the smaller the absolute values of these changes, the more stable the probability estimates are.

These possible changes can be used as lower and upper limits to the estimated probabilities, and these possible variations should be taken into account whenever considering a forecast. Thus, before event j occurs the probabilities can be expressed as

$$p_{ij}^{-} \leq p_{ij} \leq p_{ij}^{+}. \tag{18}$$

It is clear from (16) that the largest possible change caused by a new transition results when the transition is to the state with the smallest probability (and the fewest transitions) $p_{im} = \min\{p_{ij}; j = 1, \ldots, N_S\}$, so that this is the probability to be used for the worst-case possible change estimation. From (11), $\sum_{j \neq m} \Delta p_{ij}^{-} = -\Delta p_{im}^{+}$, so that the total largest possible absolute change for the whole row is

$$\Delta_i = 2\Delta p_{im}^{+} = 2\frac{1 - p_{im}}{\xi_i + 1}; \tag{19}$$

for the worst possible case, $p_{im} = 0$, Δ_i depends only on ξ_i and ranges from a largest value of 1, for $\xi_i = 1$, to 0 for $\xi_i \to \infty$.

However, $\xi_i = 1$ is an absurd proposition; a distribution of N_S states could hardly be estimated from a single transition; which leads to the question of what is a reasonable minimum value for ξ_i. Since we want a ξ_i that would guarantee that the worst possible change would not be too large, if we set the worst-case maximum allowable change to be less than the uniform probability, $\Delta_i^{\max} < 1/N_S$, then from (19), $\xi_i > 2N_S - 1$, so we will always require $\xi_i \geq 2N_S$.

4.2. Robustness

We say that a Markovian hazard estimation from state i is *robust* if the probabilities $\left[p_{ij};\, j = 1, \ldots, N_S\right]$ in the whole i'th row of matrix P are not substantially changed by a new observation.

An estimate will be robust if possible changes are small; hence, we define the robustness of row i as

$$\rho_i = 1 - \Delta_i = 1 - 2\frac{1 - p_{im}}{\xi_i + 1}. \tag{20}$$

Figure 1 illustrates how robustness increases rapidly with ξ_i for various p_{im}; the actual value of p_{im} makes a significant difference for $\xi_i < 50$; in all cases $\rho_i \geq 0.95$ for $\xi_i \geq 39$, and $\rho_i \geq 0.98$ for $\xi_i \geq 99$, so that robustness is a real concern only for small samples.

Low robustness means that there are not enough observations to make a reliable transition probability distribution estimate, so that only robust TPM rows should be used for forecasting. Let us mention here that many studies report only their P matrices, without even mentioning how many data were used to get it, let alone stating their Ξ values, so that the user has no idea about the reliability of the stated transition probabilities. Since for a given N the ξ_i values may vary widely from one row to another, it is extremely important that Ξ should always be reported.

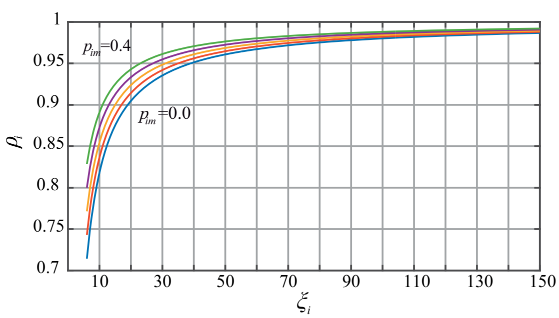

Figure 1
Row robustness as a function of ξ_i for $p_{im} = 0.0$, 0.1, 0.2, 0.3, 0.4

For a measure of robustness for the whole TPM, let us remember that the strength of a chain is that of its weakest link, and use as an estimate of the whole matrix robustness the value of the least robust row:

$$\rho = \min\{\rho_i;\, i = 1, \ldots, N_S\}. \tag{21}$$

Another, different, aspect of robustness, not treated here, is for estimates of states defined in time or space or some other kind of intervals, is how much the transition probabilities change if the starting time or interval length or some other interval parameter is slightly changed. If measured probabilities do no change very much when intervals are slightly modified, then the system is stable.

4.3. Uncertainty and Sample Size

For many seismological applications, particularly those dealing with large magnitude events, samples tend to be small, often times of the order of a few tens, while reliable estimation of the "true" probabilities requires large samples. How small can a sample be and still be useful, and how large are the uncertainties associated with small samples, are difficult questions that need to be considered in order to estimate how reliable are the seismic hazard estimates from any given study.

To illustrate this point, we can generate a TPM by a random numerical realization of an N transitions sample based on the "true" (postulated) transition probabilities shown below

$$P^T = \begin{bmatrix} 0.200000 & 0.600000 & 0.120000 & 0.080000 \\ 0.450000 & 0.250000 & 0.200000 & 0.100000 \\ 0.100000 & 0.150000 & 0.100000 & 0.650000 \\ 0.150000 & 0.200000 & 0.450000 & 0.200000 \end{bmatrix};$$

a comparison of these "true" probabilities, p_{ij}^T with those resulting from the "observed" realization, p_{ij}, of Sect. 3, shows at a glance that the "empirical" ones differ from the "true" ones; denoting the absolute difference (error) by

$$\varepsilon_{ij} = \left|p_{ij}^T - p_{ij}\right|. \tag{22}$$

To illustrate how estimation errors depend on sample size, we use these "true" probabilities to do a Monte Carlo estimation of the errors that may be expected from samples of different sizes. Figure 2 shows, on the left, the maximum and mean absolute errors averaged over $N_r = 10,000$ realizations for different sample sizes. For comparison, on the right of Fig. 2 are shown the same errors for a uniform probability matrix with $p_{ij} = 1/N_S \, \forall i, j$.

It is clear that small samples can lead to quite large deviations from the true transition probabilities but, unless some additional information is available, there is nothing to be done to improve the probability estimates, except increasing the sample size. Unfortunately, in many cases it is not possible to increase the sample size because of catalog limitations; in these cases, the possible differences from the true probabilities are an important factor to take into account when evaluating the trustworthiness of forecasts based on the observed transition probabilities.

For any given sample length, the magnitude of the errors depends on the characteristics of the particular matrix; they are larger for matrices close to uniform, and are less for "spiky" matrices (matrices having rows with one large probability and several small ones). Hence, to correctly assess the uncertainty (the possible errors) in our observed matrix P, let us suppose that it is representative of the "true" TPM that characterizes the Markovian system and according to which, our particular observed realization was randomly generated; then the errors in random chains resulting from P should not be very different to those for the true TPM.

To estimate the possible errors we use the Monte Carlo approach: based on P, generate N_r realizations of random chains N elements long; for each chain k obtain the synthetic TPM $\hat{P}$ and get the N_S^2 absolute errors

$$\varepsilon_{k,ij} = \left|\hat{p}_{ij} - p_{ij}\right|, \; i,j = 1, \ldots, N_S; \tag{23}$$

finally, get the mean and the standard deviation of the $N_\varepsilon = N_r N_S^2$ absolute errors. The mean absolute expected estimation error due to sample size

$$\varepsilon = \frac{1}{N_\varepsilon} \sum_{k=1}^{N_r} \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \varepsilon_{k,ij}. \tag{24}$$

4.4. Other Uncertainty Sources

Besides this basic uncertainty, other uncertainties in probability estimations are associated to uncertainties in the data. For instance, if transitions between earthquake magnitudes or magnitude classes are being studied, the fact that magnitudes are usually rounded to one decimal place should be taken into account; likewise, location uncertainties should be considered when studying transitions in space, etc. When using a magnitude threshold, errors in magnitude determination may cause either "gaps" in the data or the inclusion of events that should not participate, both of which result in erroneous apparent transitions.

The effects of uncertainties in the data on the transition probabilities may be hard, or impossible, to estimate in closed form, but estimates can be made

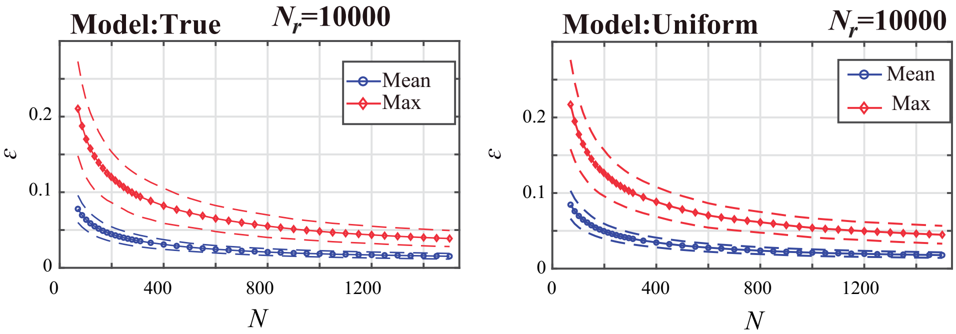

Figure 2
Mean (circles) and maximum (diamonds) absolute errors; the dashed line below and above each curve are the corresponding ± one standard deviations. On the left, we use true probabilities P^T, and on the right we use a uniform probability matrix

using the Monte Carlo approach, where random (hopefully realistic) variations are applied to the data, and likely variations to the probabilities are estimated from the results of a very large number of realizations.

5. Methods for Measuring Markovianity

The results of a Markovian hazard estimation will be significant only if the studied system, as seen through the chain observed for the model employed, turns out to be indeed Markovian, which means that the estimated transition probabilities do depend on the current state.

Hence, the Markovian results should be checked versus those of the null hypothesis, H0, which is that the observed transitions and the resulting transition probabilities, are due to chance and do not depend on the initial state. The probability transition matrix that does not depend on the initial state and comes immediately to mind is the uniform TPM, U, where each row is the uniform distribution, i.e. $u_{ij} = 1/N_S$, $\forall i, j$. However, in order to be more strict, we will use the more realistic H0 TPM of stationary state occurrence probabilities Π (13).

The differences in constitution and properties between the Markovian, P, and the H0, Π, matrices will be called the *Markovianity* of the model and of the system it represents.

We will propose five measures of Markovianity based on easily evaluated differences; they are examples of possible measures and the readers may wish to establish their own measures. In order to determine how significant a difference is, it is necessary to know the range of each measure; some limits can be ascertained theoretically while others can be estimated from the H0 distribution. Some of the theoretical limits are not practical because of data limitations and of the properties of Markovian TPMs, in which case we will use observed limits.

5.1. Markovianity from Convergence to Stationary State Probabilities

A simple and straightforward way of quantifying the Markovianity of a TPM is by measuring how many times has a TPM to be multiplied by itself to achieve the stationary probabilities with a given precision (the probabilities in each column being equal to a given number of decimals, for all columns). We will denote this measure by M_p, where p specifies the number of decimal places for which agreement is required. As mentioned above, we consider six decimals a reasonable precision, so that we will work with M_6.

This is a very important measure because it measures how important the influence of an initial state is, based on how many transitions, are necessary to lose the memory of an initial state, in contrast to stationary probabilities where multiplication of Π by itself yields Π again.

5.2. Markovianity from Absolute Differences with $\boldsymbol{\Pi}$

Let the absolute difference between an element of P and the corresponding element of Π be

$$\delta_{ij} = \left| p_{ij} - \pi_j \right|, \tag{25}$$

then, the total mean absolute difference for a given matrix is

$$\delta = \frac{1}{N_S^2} \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \delta_{ij}, \tag{26}$$

and will have a minimum value of zero when the row probabilities are equal to the stationary ones.

5.3. Markovianity from Bhattacharyya Overlap of $\boldsymbol{P}$ and $\boldsymbol{\Pi}$

The Bhattacharyya coefficient (Battachatyya, 1943) measures the overlap between two statistical probability distributions; the overlap between the Markovian transition probabilities and the stationary ones for row i of the TPM is

$$b_i = \sum_{j=1}^{N_S} \sqrt{p_{ij} \pi_j};\ i = 1, \ldots, N_S; \tag{27}$$

and $0 \le b_i \le 1$; it equals one for $p_{ij} = \pi_j, \forall j$, and would equal zero when non-zero probabilities in one distribution correspond to zero probabilities in the other distribution, which cannot happen because there are no zero probabilities in π.

Since we are interested in differences rather than overlaps, we will define the non-overlap for row i as

$$\beta_i = 1 - b_i = 1 - \sum_{j=1}^{N_S} \sqrt{p_{ij}\pi_j};\ i = 1, ..., N_S, \quad (28)$$

and the Bhattacharyya mean non-overlap for the whole matrix will be

$$\beta = \frac{1}{N_S}\sum_{i=1}^{N_S} \beta_i; \quad (29)$$

$\beta = 0$ for $P = \Pi$ (total overlap), and will have largest values (always < 1) when non-overlap between P and Π is high.

5.4. *Markovianity from Matrix Entropy*

Since a Markovian TPM should be less homogeneous, i,e, less disordered, than the TPM for uniform probability or than one having the same probability independently of the current state, an H0 system, we can use a measure of orderliness, like the entropy, to measure the difference between a markovian system and the H0 one.

The Shannon, or information, entropy of the i'th row of the P matrix is defined as (Shannon, 1948)

$$S_i^P = -h\sum_{j=1}^{N_S} p_{ij}\log_2 p_{ij}, \quad (30)$$

where h is a conventional positive constant that relates entropies calculated using logarithms with different bases; we will choose $h = 1$, so that the entropy will be expressed in bits. According to Shannon's (1948) definition of the ("surprise") self-information of a probability estimate, the information of the transition from state i to state j is

$$I_{ij} = -\log_2 p_{ij}, \quad (31)$$

so that S_i^P is the expected value of the information in row i, and we will consider the total entropy of the TPM, P, to be the sum

$$S^P = \sum_{i=1}^{N_S} S_i^P. \quad (32)$$

When p_{ij} tends to zero or to one, the corresponding term in (30) tends to zero, so that very low or very high probabilities contribute little to the entropy; actually, the entropy attains its maximum value, equal to $N_S \log_2 N_S$, for the uniform probability distribution with $p_{ij} = 1/N_S$, $\forall i, j$, and its minimum value would be zero for P constituted of ones and zeros.

While it is possible to take the uniform probability entropy as a reference, it is more realistic to take as reference entropy the entropy of the state stationary probabilities, which are the same for any row so that the total reference entropy is

$$S^\pi = -N_S\sum_{j=1}^{N_S} \pi_j \log_2 \pi_j. \quad (33)$$

The difference between the observed entropy (32) and the null hypothesis reference entropy (33),

$$S = S^P - S^\pi, \quad (34)$$

will be our measure of the orderliness of a Markovian system compared with that of a system that does not depend on the current state.

5.5. *Markovianity from Kullback–Leibler Distance*

The Relative Entropy or Kullback–Leibler Distance (Kullback & Leibler, 1951), between the distribution $p_i = [p_{i1}, p_{i2}, ..., p_{iN_S}]$ and the reference probability distribution $\pi = [\pi_1, \pi_2, ..., \pi_{N_S}]$, is defined as

$$\kappa_i = \sum_{j=1}^{N_S} p_{ij}\log_2\left(\frac{p_{ij}}{\pi_j}\right), \quad (35)$$

and the mean K–L distance for the whole P matrix is

$$\kappa = \frac{1}{N_S}\sum_{i=1}^{N_S}\kappa_i = \frac{1}{N_S}\sum_{i=1}^{N_S}\sum_{j=1}^{N_S} p_{ij}\log_2\left(\frac{p_{ij}}{\pi_j}\right), \quad (36)$$

which is always well-defined because π has no null elements. $\kappa = 0$ when P and Π are identical.

6. *Method for Evaluating a Model Through Measure Normalization and Discriminant*

As mentioned before, the various measures we use to qualify a model, robustness, uncertainty and

Markovianity, have different units, magnitudes, limits, and behaviors, so we will now propose ways of normalizing them so as to make the information each one of them provides compatible with the information from the others. The normalizations we propose, are adapted to the particular system we are studying; other researchers considering similar or different systems may wish to make their own normalization schemes.

As mentioned above, we will normalize all measures to the range from zero (worst case, undesirable) to one (perfect, desirable). Since all measures are different, and have different value ranges, normalizations are accordingly different, and are designed so as to differentiate between good and bad values in the actual observed ranges, although taking into account theoretical limits whenever possible.

Of course there is not a unique way to do the normalizations, so that the ones we present here should be considered as suggestions; other normalization ways and/or functions can be chosen by each researcher.

6.1. *Robustness*

As an indicator of the robustness of the whole TPM (Sect. 4.2), we use the minimum robustness (21). Robustness ρ tends to one as ξ_i increases and, since the minimum acceptable ξ_i is $2N_S$, its minimum value is

$$\rho_0 = 1 - \frac{2}{2N_S + 1}. \tag{37}$$

Hence, the normalized robustness is

$$\hat{\rho} = \frac{\rho - \rho_0}{1 - \rho_0}. \tag{38}$$

6.2. *Uncertainty and Sample Size*

The minimum uncertainty (Sect. 4.3) is, of course, cero, but as was mentioned in Sect. 4.3 the necessary number of samples is not attainable for the kind of applications we are considering here; hence, we will use as minimum error ε_0 the smallest ε for the whole group of models. Let us use as largest allowable error above which there is no confidence in the measured probabilities, $\varepsilon_x = 1/2N_S$, and compute the normalized error ε as

$$\hat{\varepsilon} = \begin{cases} 1 - \frac{\varepsilon - \varepsilon_0}{\varepsilon_x - \varepsilon_0};\ \varepsilon_0 \le \varepsilon \le \varepsilon_x \\ 0; \varepsilon > \varepsilon_x \end{cases}. \tag{39}$$

6.3. *Markovianity from Stationary State Probabilities Convergence*

We define the measure M_p (Sect. 5.1) as the number of times that a TPM has to be multiplied by itself in order to achieve the stationary probabilities with a given precision M_p, where p specifies the number of decimal places. Here, we will work with M_6.

In order to normalize this measure, we note that its minimum value is zero, and it has no a-priori upper value; hence, we will use as upper value, M_{6x}, the maximum observed value for the whole model set, and normalize as

$$\hat{M}_6 = \frac{M_6}{M_{6x}}. \tag{40}$$

6.4. *Markovianity from Absolute Differences Between* $\boldsymbol{P}$ *and* $\boldsymbol{\Pi}$

The minimum value that the mean absolute difference δ (26) (Sect. 5.2) can take is obviously zero when $P = \Pi$; we will use as maximum value, δ^x, the highest observed value for the whole model set, and normalize as

$$\hat{\delta} = \frac{\delta}{\delta^x}. \tag{41}$$

6.5. *Markovianity from Bhattacharyya Non-overlap of* $\boldsymbol{P}$ *and* $\boldsymbol{\Pi}$

In Sect. 5.3 the Bhattacharyya mean non-overlap of the P and Π row distributions, β, was defined by (29); this measure equals zero for $p_{ij} = \pi_j$, $\forall j$, but the theoretical maximum of one is not realistic, because π has no null elements. Hence we will use as maximum, β^x, the highest observed value, and normalize as

$$\hat{\beta} = \frac{\beta}{\beta^x}. \quad (42)$$

6.6. Markovianity from Matrix Entropy

The entropy difference S (Sect. 5.4) has a maximum value of zero for $S^P = S^\pi$, but the minimum value of $-S^\pi$ is not realistic because S^P is never zero; hence, we will use as minimum value S^m the lowest observed entropy value for the whole model set, and normalize as

$$\hat{S} = 1 - \frac{S - S^m}{-S^m}. \quad (43)$$

6.7. Markovianity from Kullback–Leibler Distance

(From Sect. 5.5) The relative entropy κ equals cero for $p_{ij} = \pi_j$, $\forall j$, but its maximum value is not realistic; we will use the maximum observed value for the whole model set as κ^x, and normalize as

$$\hat{\kappa} = \frac{\kappa}{\kappa^x}. \quad (44)$$

6.8. Discriminant

We have proposed 7 normalized measures that rate different aspects of a Markovian study, all ranging from 0 (bad) to 1 (good). A sum of all measures, divided by seven, can be an estimate of the overall goodness of the study, but, since not all measures are equally important, we will assign to each a weight that rates its relative importance, to define a discriminant

$$Q = w_\rho \hat{\rho} + w_\varepsilon \hat{\varepsilon} + w_{M_6} \hat{M}_6 + w_\delta \hat{\delta} + w_\beta \hat{\beta} + w_S \hat{S} + w_\kappa \hat{\kappa}, \quad (45)$$

where

$$\sum w_{(\bullet)} = 1. \quad (46)$$

Actually, we subjectively assign relative importance values $q_{(\bullet)}$ from which weights are derived as $w_{(\bullet)} = q_{(\bullet)} / \sum q_{(\bullet)}$.

7. Application

We now proceed to apply our proposed measures and discriminant-based scoring to models of a real seismic system, in order to show how they work and to show how they can extract information about the applicability of Markovian seismic hazard studies.

7.1. System and States

We will consider a system consisting of a seismically active geographical area, divided into N_S regions corresponding to major, seismically active, tectonic features. Each state of the system corresponds to the region where the latest large earthquake (above a given threshold magnitude M^T) has occurred. For instance, if the latest large earthquake occurred in region number one, the system is currently in state one; if the next large earthquake occurs in region three, the system will now be in state three, and one transition from state one to three has occurred.

The rough physical idea behind our model is that large and great earthquakes have rupture areas and seismic slips large enough to locally or regionally influence the ongoing plate motions and, hence, the future seismicity. Whether or not this hypothesis is correct will be shown by the Markovianity or non-Markovianity of our results.

7.2. Study Area and Regions

We chose as an example of application of our evaluation method a study area that includes Japan and the main subduction zones around it, because this area combines features such as high seismicity [on average there is one $M_W = 7$ earthquake per year (Matsu'ura, (2017)], several seismic networks which together add up to more than 2000 instruments (seismometers, accelerometers, and GPS) (Okada et al., 2004), and a seismic instrumental catalog with more than 100 years of recording. Also, although tectonic processes in Japan are complex, they are relatively simpler than those of other areas in the western Pacific plate having similar seismicity rates.

Japan is located in a four-plate convergence zone, where two oceanic plates interact with two continental plates. The oceanic Pacific Plate subducts under the Amur, Okhotsk, and Philippine Sea plates, with relative speeds of about 9 cm/year, 9 cm/year, and 5 cm/year, respectively (Ismail-Zadeh et al., 2013); whereas the Philippine Sea plate subducts under the Amur plate with relative speed of about 5 cm/year (Fig. 3a).

Based on the tectonics and seismicity of Japan, we chose four regions, each one corresponding to a major subduction process, shown in Fig. 3b; the largest number of regions appropriate to the number of large earthquakes. Regions one and two correspond to the subduction of the Pacific plate beneath the Okhost and the Amur plates, respectively (Ishibashi, 2004). In order to pinpoint the boundary between these two regions, we used hypocenters from earthquakes with magnitudes $M_W > 0$ from the USGS catalog to make four seismic profiles parallel to the trench to observe the subduction of the Pacific plate (Fig. 3c). The profiles show different dips for the southwestern and the northeastern parts of the Pacific plate, and the change of dip coincides with the change of azimuth in the trench.

Region three contains the seismicity and tectonic structures corresponding to the interaction between the Pacific and Philippine Sea plates (Lallemand, 2016). Region four contains the seismicity and tectonic structures of the Philippines and Amur plates interaction (Lee & Kim, 2016).

7.3. Magnitude Thresholds and Results

To illustrate the measuring, scoring, and discriminant scheme, we will build different models using the same regions and different threshold magnitudes, and will choose the threshold magnitude that results in the best compromise between Markovianity and number of data.

We will use data from the ISC seismic catalog (1904–2016), which reports magnitudes $M_W \geq 5.5$ so that we can be certain that magnitudes above $M_W = 6.5$ will be complete. In what follow we will use threshold magnitudes, M^T, ranging from $M_W = 6.5$, below which (as will be seen later) events are too small to significantly affect plate motion, to $M_W = 7.4$, which is the largest magnitude for which we have enough data to meet the $\xi_i \geq 2N_S$ requirement.

Table 1 shows the total number of transitions N, and the Θ, Ξ, π and P matrices for all considered threshold magnitudes. Note that, although P embodies the final results of the study, Θ and Ξ are essential for evaluating the results.

Now, let us see how results differ for different threshold magnitudes, and whether there are results that would be useful hazard estimates.

For the first row, transitions from state one, let us remember that region one is close only to region two, so that we can expect these two regions to have strong interaction. As expected, probabilities p_{11} and p_{12} are consistently larger than the rest, with p_{11} larger for $M^T < 6.7$ and p_{12} larger for $M^T \geq 6.7$.

For the second row, p_{22} has the highest probability for $M^T < 7.0$, and the highest probability for larger M^T is p_{21}.

For row three $p_{32} > p_{33}$ for $M^T < 7.0$, after which p_{33} is higher (except for $M^T \geq 7.3$). Curiously, for $M^T \geq 6.7 p_{31} > p_{34}$ although region 4 is closer to region 3 than region 1.

For the fourth row, p_{42} is always the largest and p_{43} the smallest. It appears as if regions three and four interact through region 2.

Something to note is the low probability of p_{34} and p_{43}, despite that regions 3 and 4 are geographically close together yet do not interact with each other as regions 1 and 2 do.

To decide which is the best model, i.e., which is the optimal threshold magnitude, we will proceed to evaluate the stability, accuracy, and Markovianity measures.

Figure 4 shows the number of transitions (top left) for reference, and the values of the seven different measures for each threshold magnitude, and clearly shows the problem of judging which model is best, because measures have different magnitudes, units and characteristics, i.e., some measures are good when their values are high (ρ, M_6, δ, β and κ) and others when their values are low (ε and S).

Figure 5 shows the normalized measures; now, for all of them high values are desirable and low ones are undesirable. Measures that depend more or less directly on the number of samples decrease with magnitude, while Markovianity measures present

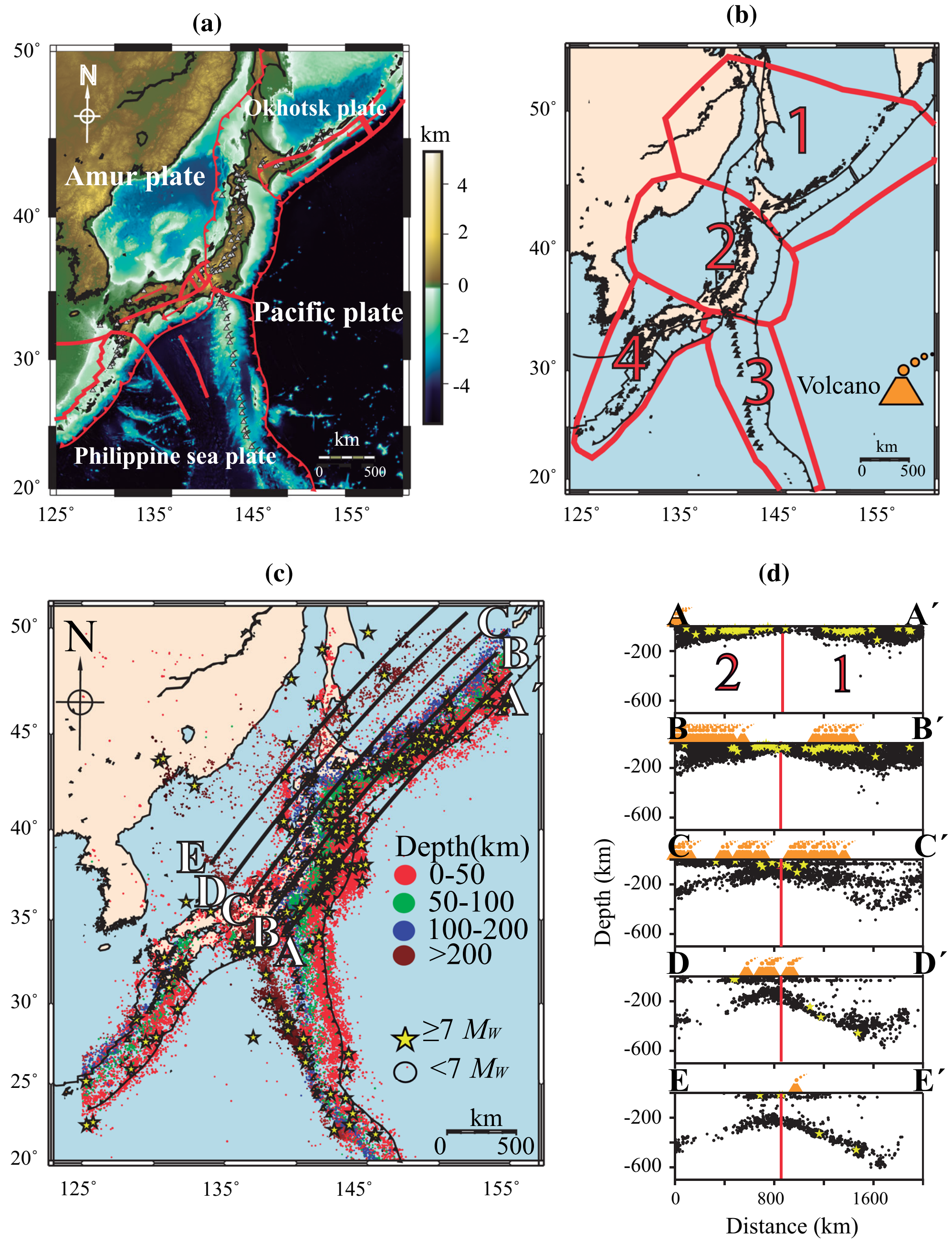

Figure 3
Seismicity and tectonic setting in the area around Japan. **a** Tectonic setting, red lines indicate tectonic boundaries. **b** The four regions, brown triangles indicate volcanoes. **c** Seismicity (USGS catalog, 1963–2017), with magnitudes > 0; black lines are seismic sections. **d** Seismic sections, red lines indicate the limits between regions 1 and 2

high values for magnitudes 7.1 to 7.3. Since the normalizations are linear, the measures do not change shape on normalization; their units change, and some measures with negative values are now positive.

Table 1

M_6 and Θ, Ξ, π and P matrices for all considered threshold magnitudes

M^T	N	Θ				Ξ	P				π				M_6
		45	38	21	16	120	0.375000	0.316667	0.175000	0.133333					
6.5	449	42	48	25	37	152	0.276316	0.315789	0.164474	0.243421					
		14	28	25	18	085	0.164706	0.329412	0.294118	0.211765	0.267261	0.338530	0.189310	0.204900	08
		19	38	14	21	092	0.206522	0.413043	0.152174	0.228261					
		34	33	17	12	096	0.354167	0.343750	0.177083	0.125000					
6.6	375	32	38	25	31	126	0.253968	0.301587	0.198413	0.246032					
		12	26	22	14	074	0.162162	0.351351	0.297297	0.189189	0.256000	0.336000	0.197333	0.210667	08
		18	29	10	22	079	0.227848	0.367089	0.126582	0.278481					
		26	32	13	13	084	0.309524	0.380952	0.154762	0.154762					
6.7	317	29	34	17	28	108	0.268519	0.314815	0.157407	0.259259					
		13	17	17	09	056	0.232143	0.303571	0.303571	0.160714	0.264984	0.340694	0.176656	0.217666	08
		16	25	09	19	069	0.231884	0.362319	0.130435	0.275362					
		21	25	11	11	068	0.308824	0.367647	0.161765	0.161765					
6.8	253	27	32	12	20	091	0.296703	0.351648	0.131868	0.219780					
		12	12	11	08	043	0.279070	0.279070	0.255814	0.186047	0.268775	0.359684	0.169960	0.201581	07
		08	22	09	12	051	0.156863	0.431373	0.176471	0.235294					
		15	22	09	11	057	0.263158	0.385965	0.157895	0.192982					
6.9	212	22	23	12	17	074	0.297297	0.310811	0.162162	0.229730					
		09	12	11	05	037	0.243243	0.324324	0.297297	0.135135	0.268814	0.348733	0.180139	0.202314	07
		11	17	06	10	044	0.250000	0.386364	0.136364	0.227273					
		10	23	07	08	048	0.208333	0.479167	0.145833	0.166667					
7	173	22	16	11	11	060	0.366667	0.266667	0.183333	0.183333					
		09	06	10	05	030	0.300000	0.200000	0.333333	0.166667	0.277081	0.340489	0.180242	0.202188	09
		07	14	03	11	035	0.200000	0.400000	0.085714	0.314286					
		09	17	06	06	038	0.236842	0.447368	0.157895	0.157895					
7.1	134	14	14	09	10	047	0.297872	0.297872	0.191489	0.212766					
		08	06	11	01	026	0.307692	0.230769	0.423077	0.038462	0.283664	0.342448	0.204094	0.169794	12
		07	09	01	06	023	0.304348	0.391304	0.043478	0.260870					
		09	11	04	06	030	0.300000	0.366667	0.133333	0.200000					
7.2	105	12	08	06	09	035	0.342857	0.228571	0.171429	0.257143					
		03	05	08	02	018	0.166667	0.277778	0.444444	0.111111	0.283550	0.323770	0.184954	0.207726	12
		06	10	01	05	022	0.272727	0.454545	0.045455	0.227273					
		08	09	03	06	026	0.307692	0.346154	0.115385	0.230769					
7.3	082	12	06	04	06	028	0.428571	0.214286	0.142857	0.214286					
		03	05	04	00	012	0.250000	0.416667	0.333333	0.000000	0.315385	0.331361	0.161538	0.191716	10
		03	07	02	04	016	0.187500	0.437500	0.125000	0.250000					
		07	08	03	05	023	0.304348	0.347826	0.130435	0.217391					
7.4	068	10	04	03	05	022	0.454545	0.181818	0.136364	0.227273					
		03	03	01	01	008	0.375000	0.375000	0.125000	0.125000	0.337570	0.311187	0.132196	0.219046	08
		03	06	02	04	015	0.200000	0.400000	0.133333	0.266667					

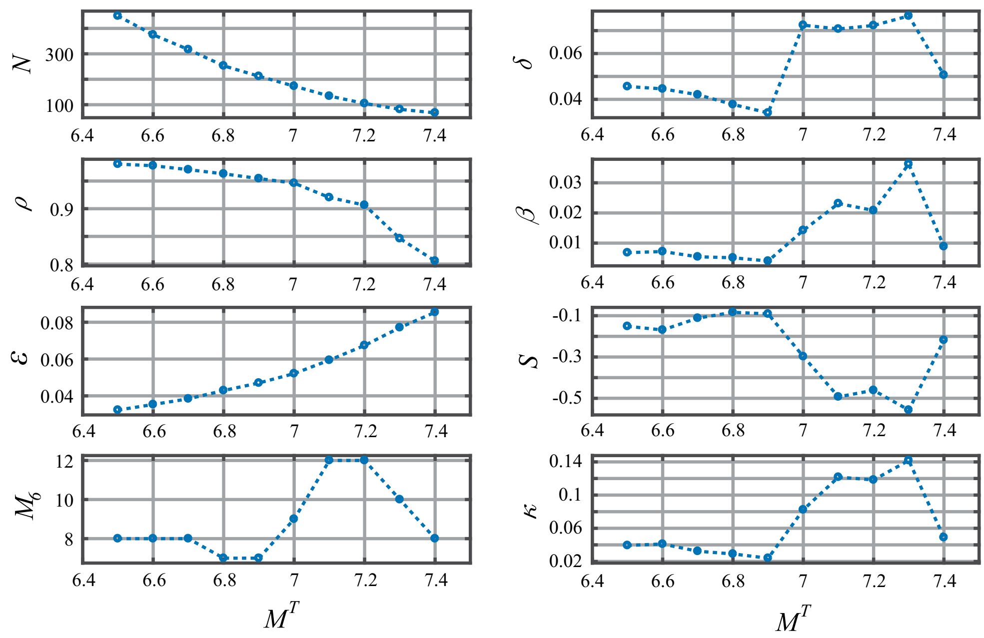

Figure 4
Markovianity measures. Top left shows the number of transitions for reference

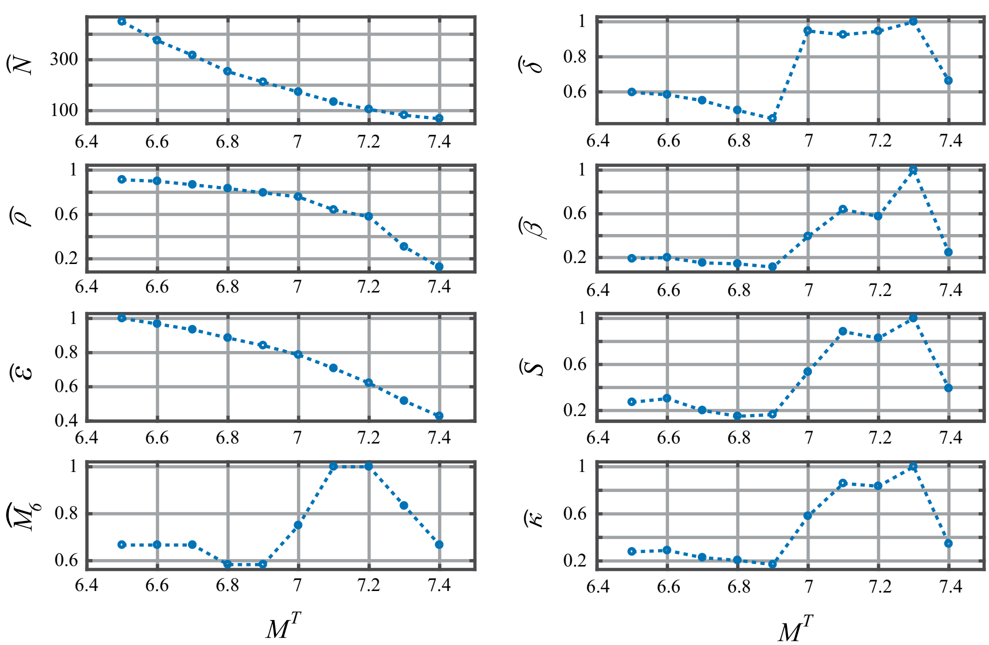

Figure 5
Normalized Markovianity measures. Top left shows the number of transitions for reference

The next step is to apply the discriminant (45) using the normalized measures for different threshold magnitudes, in order to choose the optimal threshold magnitude. We will show the discriminant behavior for different tentative weights sets.

Figure 6 shows the discriminant values for equal weights for all measures. The peak discriminant value is for $M^T = 7.3$, with $M^T = 7.1$ a very close second.

Considering ε to be the most important measure, because it measures the expected uncertainty associated with sample size, and large uncertainties make results unreliable, we assign to this measure thrice the importance of the other measures, which results in the discriminant values shown in Fig. 7; $M^T = 7.1$ is the optimal threshold magnitude for this weights set.

Now, considering that M_6, although not as important as ε, is also a very important measure (for reasons mentioned above), we assign it twice the importance of the remaining measures. The resulting discriminant values are shown in Fig. 8, where $M^T = 7.1$ is again the optimal magnitude.

Finally, we present a last set obtained from the previous one by giving to δ and S one and half times the importance of β and κ. It could be argued that they measure similar things but the first two are more direct, as shown in Fig. 9 where the peak discriminant value is for $M^T = 7.1$.

Since weights are assigned subjectively, it is important to assign them carefully, sparingly, and according to explainable, and defensible, reasons; particularly if a change in weights leads to a change

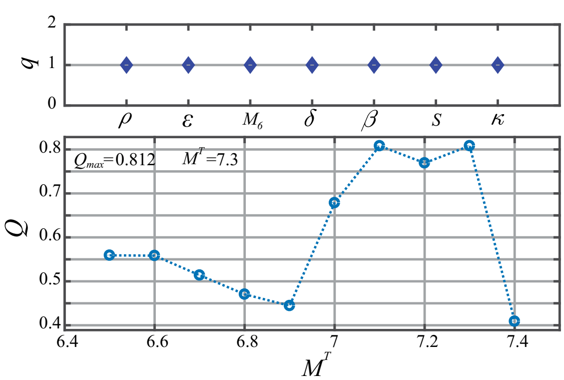

Figure 6
Discriminant. Top: assigned importance values for the measures ($q_\rho = 1$, $q_\varepsilon = 1$, $q_{M_6} = 1$, $q_\delta = 1$, $q_\beta = 1$, $q_S = 1$, $q_\kappa = 1,$). Bottom: discriminant value for each magnitude threshold

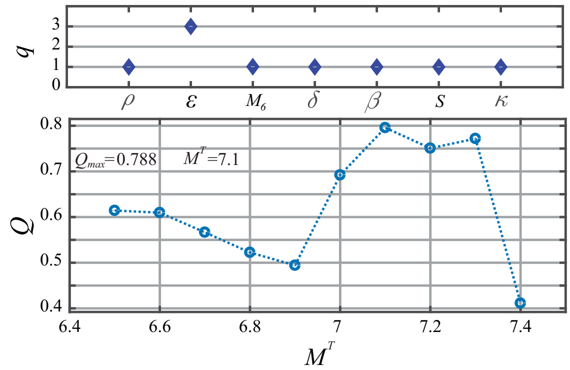

Figure 7
Discriminant. Top: assigned importance values for the measures ($q_\rho = 1$, $q_\varepsilon = 3$, $q_{M_6} = 1$, $q_\delta = 1$, $q_\beta = 1$, $q_S = 1$, $q_\kappa = 1,$). Bottom: discriminant value for each magnitude threshold

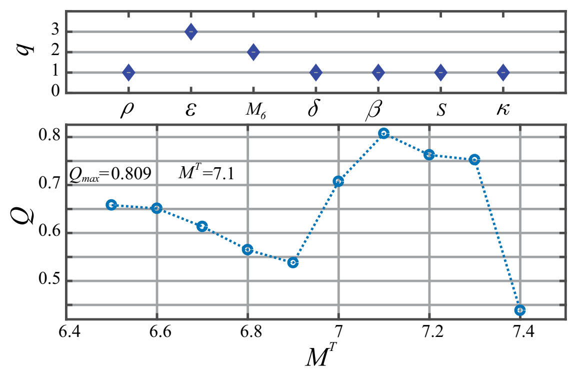

Figure 8
Discriminant. Top: assigned importance values for the measures ($q_\rho = 1$, $q_\varepsilon = 3$, $q_{M_6} = 2$, $q_\delta = 1$, $q_\beta = 1$, $q_S = 1$, $q_\kappa = 1,$). Bottom: discriminant value for each magnitude threshold

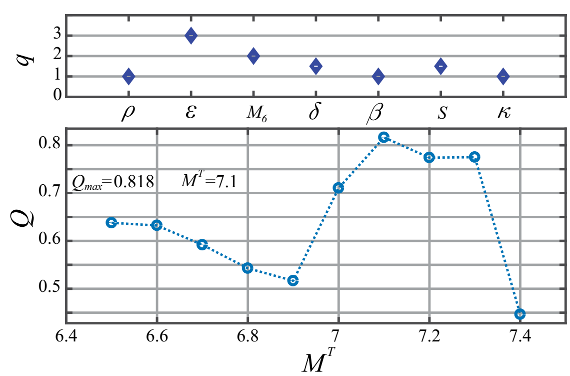

Figure 9
Discriminant. Top: assigned importance values for the measures ($q_\rho = 1$, $q_\varepsilon = 3$, $q_{M_6} = 2$, $q_\delta = 1.5$, $q_\beta = 1$, $q_S = 1.5$, $q_\kappa = 1,$). Bottom: discriminant value for each magnitude threshold

in the choice of the best model. According to these criteria, and seeing that all weight sets, excepting the equal weights one, result in discriminants with maxima for $M^T = 7.1$, and that other reasonable changes to these weights sets do not lead to a different preferable threshold magnitude, we choose the weight set in Fig. 8 as the simplest one that takes into account the most important features of the various measures, and the model with $M^T = 7.1$ as the most reliable one.

Thus, with respect to seismic hazard in the study area, we can say from the TPM for $M^T = 7.1$ that: after a large earthquake in region one, the next one will most probably occur in region two; an event in region 2 will most probably either repeat in region 2 or be followed by an event in region one; an event in region 3 will most probably repeat or, with a smaller probability, be followed by one in region one, but a following event in region 4 is quite improbable; finally, an event in region 4 will probably be followed by one in region two or region one, but has very small probability of being followed by one in region three.

As mentioned in previous paragraphs, for the optimal model there is a strong interaction between region one (Kuril trench) and two (Japan trench); both regions have in common that most earthquakes are shallow and do not exceed 150 km in depth (Fig. 3d), which may be due to the fact that these trenches are of the Chilean type, so there is strong coupling of the tectonic plates in this region (Uyeda & Kanamori, 1979). We think that the strong interaction between region one and two may be due to the proximity of both regions and to the shallow seismicity. We mentioned that region three (Izu-Bonin trench) has strong interaction with itself; in this region the shallow seismicity is low and most of the seismicity is located in a range of 200–500 km depth, this may be caused by the weak coupling between the Pacific and Philippine plates, since the Izu-Bonin trench is of the Mariana type (Uyeda & Kanamori, 1979) and the great depth of the earthquakes in this region may be a factor for the strong interaction with itself and for the weak interaction with the other regions. The largest interaction of region four is with region two, the seismicity in region four is shallow (does not exceed 150 km) and there is a gap in the seismicity where the fossil ridge subducts below the Amuria plate; in this region the trench is of the Chilean type; the large interaction between regions two and four may again be due to their proximity and shallow seismicity.

But a seismic hazard estimation is not the only result from this application; our measures and discriminant tell us something about the seismic process itself and about the applicability of a Markovian model to the study region. All Markovianity measures have low values for $M^T < 7.0$ and for $M^T > 7.3$, which can be explained by remembering that each threshold magnitude results in a different Markov chain. Depending on where they occur and on the prevalent stress conditions in the hypocentral region, events with $M_W < 7.0$ may not be large enough to significantly influence the plate motions, so that some links in the observed chains may not be actually related to the Markovian interaction. At the other extreme, while very large earthquakes are certain to influence plate motion, by considering too large a threshold magnitude we are discarding events that should be reckoned with, i.e., there are missing links in the observed chain. Hence, it seems that a Markovian study is applicable to the region for the 7.1 to 7.3 magnitude range, and the M_6 measure tells us that a Markovian model applies best to the study region for threshold magnitudes 7.1 and 7.2, that give the best tradeoff between extraneous and missing links in the Markov chains.

8. *Conclusion*

It is not possible to determine whether results from a Markovian seismic hazard study are useful and reliable if only the transition probabilities matrix, P, is reported; measures quantifying uncertainty, robustness, and Markovianity of the transition probabilities or, at least, either the transitions matrix Θ or the Ξ matrix together with the total number of transitions N, should also be reported.

We propose a set of measures for Markovian models that, based on the above mentioned information, quantify robustness, uncertainty, and Markovianity. Also, since different measures have different units, magnitudes, and behaviors, we propose a way to normalize them in order to make them comparable.

Usually there is not an unique way to model a given system, models can be very different from each other, but even models having the same characteristics but different sets of parameters can result in widely varying hazard estimations; hence, it is very important to be able to decide which among the proposed models is the best one. We propose a way to choose among different models by applying the above mentioned normalized measures through use of a discriminant.

As an example, we applied our measures and decision method to an area around Japan; the system was modeled as having four states corresponding to which of four regions has the most recent earthquake above a given threshold magnitude, and we explored models with threshold magnitudes from $M_W = 6.5$ to $M_W = 7.4$. Our results show optimum tradeoff between reliability and Markovianity for $M^T = 7.1$. The expected behavior of the system is discussed in the previous section.

Our results show that nearby regions with shallow earthquakes will have strong interaction between them, on the other hand, regions with deep earthquakes will have strong interaction with themselves and weak interaction with other regions.

Our measures also shed some light on the system itself; our results indicate that for magnitudes smaller than 7.0 or larger than 7.3 the system is less Markovian than for the 7.0–7.3 magnitude range, which can be explained by spurious or missing links in the observed transition chains, respectively.

Of course, the measures, normalizations, and discriminant we propose are not unique, and the readers can implement their own to evaluate the results of Markovian seismic hazard analysis. Whatever the measures used, we exhort all Markovian seismic hazard studies to report their transition matrices Ξ, so that others may evaluate their results.

Acknowlegements

This research was partially funded by CONACYT scholarship 633456 and CONACyT grant 222795. We thank ISC (2013) and USGS for use of their seismic catalogs, and gratefully acknowledge the use of the Generic Mapping Tool (GMT).

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

References

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society., 35,* 99–109.

Fritchman, B. (1967). A binary channel characterization using partitioned Markov chains. *IEEE Transactions on Information Theory., 13*(2), 221–227.

Fujinawa, Y. (1991). A method for estimating earthquake occurrence probability using first and multiple-order Markov chain models. *Natural Hazards, 4,* 7–22. https://doi.org/10.1007/BF00126556

Grant, M., & Steven, T. (1991). Are stock returns predictable? A test using Markov Chains. *The Journal of Finance, 46*(1), 239–263. https://doi.org/10.1111/j.1540-6261.1991.tb03751.x

Hagiwara, Y. (1975). A stochastic model of earthquake occurrence and the accompanying horizontal land deformation. *Tectonophysics, 26*(1–2), 91–101. https://doi.org/10.1016/0040-1951(75)90115-8

Heng, T. (2002). Estimates of earthquake recurrences in the Chiayi-Tainan area, Taiwan. *Elsevier, 63*(1–2), 157–168. https://doi.org/10.1016/S0013-7952(01)00079-5

International Seismological Centre. (2013). On-line Bulletin. http://www.isc.ac.uk, Internatl. Seismol. Cent., Thatcham, United Kingdom.

Ishibashi, K. (2004). Status of historical seismology in Japan. *Annals of Geophysics, 47,* 2–3. https://doi.org/10.4401/ag-3305

Ismail-Zadeh, A., Honda, S., & Tsepelev, I. (2013). Linking mantle upwelling with the lithosphere descent and the Japan Sea evolution: a hypothesis. *Scientific Reports*. https://doi.org/10.1038/srep01137

Khmelev, D. V., & Tweedie, F. J. (2001). Using Markov Chains for Identification of Writer. *Literary and Linguistic Computing, 16*(3), 299–307. https://doi.org/10.1093/llc/16.3.299

Knopoff, L. (1971). A stochastic model for the occurrence of main sequence earthquakes. *Journal of Geophysical Research, 75,* 5745–5756.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. https://projecteuclid.org/euclid.aoms/1177729694

Lallemand, S. (2016). Philippine Sea Plate inception, evolution, and consumption with special emphasis on the early stages of Izu-Bonin-Mariana subduction. *Progress in Earth and Planetary Science, 3,* 15. https://doi.org/10.1186/s40645-016-0085-6

Lee G. H., & Kim H. J. (2016). Crustal Structure and Tectonic Evolution of the East Sea. In: K. I. Chang et al. (Eds.), *Oceanography of the East Sea (Japan Sea)*. Springer https://doi.org/10.1007/978-3-319-22720-7_17

Matsu'ura, R. S. (2017). A short history of Japanese historical seismology: past and the present. *Geoscience Letters, 4,* 3. https://doi.org/10.1186/s40562-017-0069-4

Nava, F., Herrera, C., Frez, J., & Glowacka, E. (2005). Seismic hazard evaluation using Markov chains; Application to the Japan area. *PAGEOPH, 162,* 1347–1366.

Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, F., & Yamamoto, A. (2004). Recent progress of seismic observation networks in Japan—Hi-net, F-net K-NET and KiK-net. *Earth Planet, 56,* 15–28. https://doi.org/10.1186/BF03353076

Patwardhan, A., Kulkarni, R., & Tocher, D. (1980). A semi Markov model for characterizing recurrence of great earthquakes. *Bulletin of the Seismological Society of America, 70,* 323–347.

Raible, C., Bischof, G., Fraedrich, K., & Kirk, E. (1999). Statistical single-station short-term forecasting of temperature and probability of precipitation: area interpolation and NWP combination. *American Meteorological Society, 14,* 203–214. https://doi.org/10.1175/1520-0434(1999)014%3c0203:SSSSTF%3e2.0.CO;2

Ramin, S. (2012). Forecasting time and place of earthquakes using a Semi-Markov model (with case study in Tehran province). *Journal of Industrial Engineering International*, 8(1). http://www.jiei-tsb.com/content/8/1/20.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(379–423), 623–656.

Ünal, S., & Celebioglu, S. (2011). A Markov chain modelling of the earthquakes occurring in Turkey. *Gazi University Journal of Science, 24*(2), 263–274.

Uyeda, S., & Kanamori, H. (1979). Back-Arc opening and the mode of subduction. *Journal of Geophysical Research., 84,* B3. https://doi.org/10.1029/JB084iB03p01049

Vere-Jones, D. (1966). A Markov model for aftershock occurrence. *Pure and Applied Geophysics, 64,* 31–42.

Votsi, I., Limnios, N., Tsaklidis, G., & Papadimitriou, E. (2010). Semi-markov models for seismic hazard assessment in certain areas of Greece. *Bulletin of the Geological Society of Greece, 43,* 2200–2209.

Votsi, I., Tsaklidis, G., Limnios, N., Papadimitriou, E., & Vallianatos, F. (2013). A Markov model for seismic hazard analysis along the Hellenic subduction Zone (Greece). *Bulletin of the Geological Society of Greece, 47*(3), 1376–1385. https://doi.org/10.12681/bgsg.10934

Yildiz, A., & Demir, K. (1999). An application of the semi-Markov model for earthquake occurrences in North Anatolia, Turkey. *Journal of the Balkan Geophysical Society, 2*(4), 90–99.

(Received May 25, 2020, revised February 10, 2021, accepted February 22, 2021)