Long-Term Forecasting of Strong Earthquakes in North America, South America, Japan, Southern China and Northern India With Machine Learning

Victor Manuel Velasco Herrera[1]*, Eduardo Antonio Rossello[2], Maria Julia Orgeira[2], Lucas Arioni[2], Willie Soon[3,4], Graciela Velasco[5], Laura Rosique-de la Cruz[6], Emmanuel Zúñiga[7] and Carlos Vera[1]

[1]Instituto De Geofísica, Universidad Nacional Autónoma De México, Mexico City, Mexico, [2]Universidad De Buenos Aires, Facultad De Ciencias Exactas y Naturales, IGEBA, Universidad De Buenos Aires-CONICET, Buenos Aires, Argentina, [3]Center for Environmental Research and Earth Sciences (CERES), Salem, MA, United States, [4]Institute of Earth Physics and Space Science (ELKH EPSS), Sopron, Hungary, [5]Instituto De Ciencias Aplicadas y Tecnología, Universidad Nacional Autónoma De México, Ciudad Universitaria, Mexico City, Mexico, [6]Comisión Nacional para El Conocimiento y Uso De La Biodiversidad, Mexico City, Mexico, [7]CONACYT, Instituto de Geografía, Universidad Nacional Autónoma De México, Mexico City, Mexico

Strong earthquakes (magnitude ≥7) occur worldwide affecting different cities and countries while causing great human, ecological and economic losses. The ability to forecast strong earthquakes on the long-term basis is essential to minimize the risks and vulnerabilities of people living in highly active seismic areas. We have studied seismic activities in North America, South America, Japan, Southern China and Northern India in search for patterns in strong earthquakes on each of these active seismic zones between 1900 and 2021 with the powerful mathematical tool of wavelet transform. We found that the primary seismic activity patterns for $M \geq 7$ earthquakes are 55, 3.7, 7.7, and 8.6 years, for seismic zones of the southwestern United States and northern Mexico, southwestern Mexico, South American, and Southern China-Northern India, respectively. In the case of Japan, the most important seismic pattern for earthquakes with magnitude $7 \leq M < 8$ is 4.1 years and for strong earthquakes with $M \geq 8$, it is 40 years. Every seismic pattern obtained clusters the earthquakes in historical intervals/episodes with and without strong earthquakes in the individually analyzed seismic zones. We want to clarify that the intervals where no strong earthquakes do not imply the total absence of seismic activity because earthquakes can occur with lesser magnitude within this same interval. From the information and pattern we obtained from the wavelet analyses, we created a probabilistic, long-term earthquake prediction model for each seismic zone using the Bayesian Machine Learning method. We propose that the periods of occurrence of earthquakes in each seismic zone analyzed could be interpreted as the period in which the stress builds up on different planes of a fault, until this energy releases through the rupture along faults and fractures near the plate tectonic boundaries. Then a series of earthquakes can occur along the fault until the stress subsides and a new cycle begins. Our machine learning models predict a new period of strong earthquakes between 2040 ± 5 and 2057 ± 5, 2024 ± 1 and 2026 ± 1, 2026 ± 2 and 2031 ± 2, 2024 ± 2 and 2029 ± 2, and 2022 ± 1 and 2028 ± 2 for the five active seismic zones of United States,

Mexico, South America, Japan, and Southern China and Northern India, respectively. In additon, our methodology can be applied in areas where moderate earthquakes occur, as for the case of the Parkfield section of the San Andreas fault (California, United States). Our methodology explains why a moderate earthquake could never occur in 1988 ± 5 as proposed and why the long-awaited Parkfield earthquake event occurred in 2004. Furthermore, our model predicts that possible seismic events may occur between 2019 and 2031, with a high probability of earthquake events at Parkfield around 2025 ± 2 years.

1 INTRODUCTION

Different natural phenomena like the fall of meteorites, tsunamis, volcanic eruptions, droughts, ice ages, the reversal of geomagnetic field, forest fires, droughts, earthquakes, and others can pose a significant danger and threat to human life and humanity's economic developments and resource managements (Murray, 2021).

Earthquakes are caused not only by natural seismic and tectonic processes but often time can also be induced by various anthropogenic activities such as nuclear bomb detonations, large dams, and subsurface exploitation of natural resources. The danger and risk posed by usually low intensity earthquakes induced by anthropogenic activities can be indeed mitigated by reducing or completely stopping the human activities that are responsible by these types of minor earthquakes. In a sharp contrast, especially earthquakes of great intensity that are caused by natural processes cannot be avoided but only forewarned with their often catastrophic and damaging impacts minimized.

Different sources and mechanisms have been suggested as triggers and modulators of earthquakes (see, for example Batakrushna et al., 2022, for a full review). For example, even the Sun's activity has been suggested as a significant agent causing earthquakes (Anagnostopoulos et al., 2021). Other proximate causes discussed in the literature include pole tide (Shen et al., 2005), pole wobble (Lambert and Sottili, 2019), surface ice and snow loading (Heki, 2019), glacial isostatic rebound (Hampel et al., 2007), heavy precipitation (Hainzl et al., 2006), atmospheric pressure (Liu et al., 2009), sediment unloading (Calais et al., 2016), seasonal groundwater change (Tiwari et al., 2021), seasonal hydrological loading (Panda et al., 2020). In addition, the Earth's rotation and tidal spinning have also been suggested as driver of plate tectonic activity.

The present geological paradigm about solid Earth is the plate tectonic theory which describes that the lithosphere is segmented into a series of plates that are in constant motions due to mantle mobility or convection. As a result of their interaction, a series of geological, mainly convergent and divergent, processes take place at their plate margins, ranging from seismicity, orogenic processes, and volcanism. The World Stress Map (WSM)[1]

compiles the orientation of maximum horizontal stress (σ_{Hmax}) where we delimited our study areas in **Figure 1** (Heidbach et al., 2016).

The dynamics of the plate tectonics provide a framework to understand the evolutive shape and dynamics of the earth's surface. Plate boundaries involve either divergence, like at oceanic spreading centers and continental rifts, or convergence, such as subduction (ocean to continent or ocean to ocean) and collision zones with different angles of displacement ranging from orthogonal towards subparallel one.
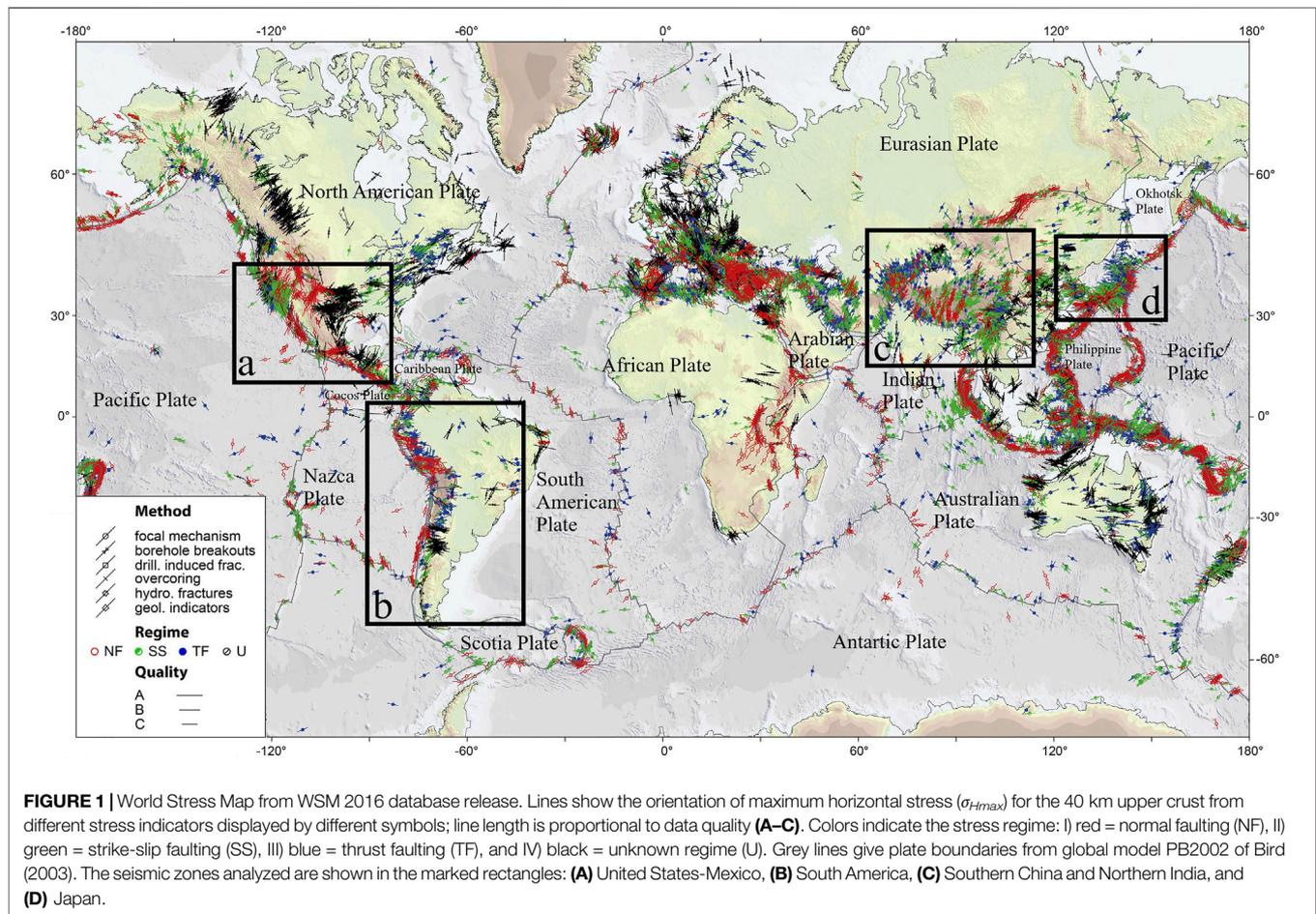
Only minor cases involve transform boundaries that facilitate plate kinematics on the global sphere. These boundaries accommodate plate-parallel relative displacement by strike-slip motion on vertical or steeply dipping faults. Due to these frictional contacts between the different types of plates, seismicity is triggered, producing a succession of earthquakes that progressively decrease in intensity in increasingly distant/remote areas away from the seismic center/zone.

The sliding between tectonic plates is quite varied. Some plates slide without any consequences on Earth's surface, while catastrophic failures punctuate others. Also, after a few hundred meters some earthquakes stop. Nevertheless, others continue to collapse even after thousands of kilometers (Kanamori and Brodsky, 2004).

The driving mechanisms of plate tectonics remain not well unknown or poorly understood. Are they due to internal factors or external astronomical forces? We are hoping that the analysis of seismic patterns could provide some clues and information about the sources and mechanisms that are responsible for both tectonic movements and earthquakes.

Earthquake forecasting is one of the most difficult areas of research even though it is clear that its early prognosis can save many lives (Jain et al., 2021). Deterministic prediction of the exact coordinates of the epicentre, its depth, magnitude and exact time of one earthquake at the moment remains difficult and possibly impossible (see, for example, Shcherbakov et al., 2019; Beroza et al., 2021). Ogata (1988) suggests that the seismic pattern and temporal variation are usually very complicated. Furthermore, temporal seismic clustering is complex and difficult to discern or anticipate in advance. Different models have been proposed to analyze space-time clusters of seismicity in a region. One example is the Epidemic Type Aftershock Sequence (ETAS) model. This model suggests that the earthquake of a particular magnitude (M) in a region during a period of time can be approximately

---

[1]http://www.world-stress-map.org/casmo/

FIGURE 1 | World Stress Map from WSM 2016 database release. Lines show the orientation of maximum horizontal stress (σ_{Hmax}) for the 40 km upper crust from different stress indicators displayed by different symbols; line length is proportional to data quality **(A–C)**. Colors indicate the stress regime: I) red = normal faulting (NF), II) green = strike-slip faulting (SS), III) blue = thrust faulting (TF), and IV) black = unknown regime (U). Grey lines give plate boundaries from global model PB2002 of Bird (2003). The seismic zones analyzed are shown in the marked rectangles: **(A)** United States-Mexico, **(B)** South America, **(C)** Southern China and Northern India, and **(D)** Japan.

considered as a Poisson process (Ogata, 1988). In addition, the method of the minimum area of alarm for earthquake magnitude prediction (Gitis and Derendyaev, 2020) and a method for earthquake predictions based on alarms (Zechar and Jordan, 2008) have all been suggested and evaluated.

The studies of earthquake precursors such as observing crustal geochemical fluids and gases, ultra-low frequency magnetic signals, atmospheric effects including ionospheric total electron content measurements, and several recording seismicities in regions experiencing earthquakes in terms of atmospheric, geochemical, and historical information can all help to improve and refine earthquake prediction (see for example, Pulinets and Boyarchuk, 2005; Ouzounov et al., 2018; Pulinets and Ouzounov, 2018). Since 2007, the Collaboratory for the Study of Earthquake Predictability (CSEP) has actively conducted and rigorously evaluated earthquake forecasting experiments as well as the prospective evaluations of earthquake forecast models and prediction algorithms (see, for example, Schorlemmer et al., 2018). CSEP's main targets and focuses are to optimize earthquake forecasting, advance forecast model development, test model hypotheses, and improve seismic hazard assessments.

The medium-term prediction of the strongest earthquakes has been carried out by the *M8* algorithm, which is an algorithm for

evaluating times of increased probability (TIPs) for strong earthquakes (Keilis-Borok and Kossobokov, 1990) from intensity of an earthquake flow and rate differential on a specific seismic region of earthquake source concentration and clustering. Also, the prediction of extreme events such as earthquakes demonstrates the efficiency and potential of the algorithms based on a pattern recognition approach as example the *M8* algorithm (Kossobokov and Soloviev, 2008, 2018). In addition, the *M8* algorithm shows that the hypothesis that the largest earthquake events are mere random variations in seismically active regions can be confidently rejected (Kossobokov and Soloviev, 2021). Kossobokov et al. (2015) suggested that "forecasting earthquake information must be reliable, tested, confirmed by evidence, and not necessarily probabilistic". We disagree slightly with this opinion and interpretation of Kossobokov et al. (2015). Probabilistic forecasts in the last century have provided new results to understand natural phenomena (see, for example Wigner, 1967; Landau and Lifshitz, 1988b; Feynman et al., 2011b). In this work, we show the results of a Bayesian model of Machine Learning, which is a probabilistic model. We do agree with Kossobokov et al. (2015) that all forecasts which are either probabilistic or not probabilistic must indeed be confirmed by evidence. We think that only future events will show if our

probabilistic Machine Learning predictions are on the right track or not.

In recent years artificial intelligence (AI), deep learning (DL), machine learning (ML) (see, for example, Essama et al., 2021) have been applied to earthquake forecasting. In particular the use of ML in the study of earthquakes has been implemented in the detection, arrival time measurement, phase association, location and characterization (Beroza et al., 2021). In addition, the use of ML has focused on forecasting the exact magnitude of the next strong earthquake in different seismic zones (see, for example, Yousefzadeh et al., 2021).

In this paper, we propose a new method of analysis and algorithm for forecasting strong earthquakes (i.e., magnitude ≥ 7). We suggest that one promising progress to earthquake forecasting may consist in changing the prediction paradigms from an "exact" approach to probabilistic forecasting of future seismic activity cycles. This work aims to find the temporal seismic patterns of high and low seismicity in four major seismic zones: 1) the United States and Mexico, 2) South America, 3) Japan, and 4) Southern China and Northern India as sketched in **Figure 1**. We have made a probabilistic long-term earthquake prediction using a Bayesian ML model in these seismic zones based on the seismic patterns deduced from our wavelet analyses.

2 SEISMIC STUDY ZONES

We are analyzing the earthquake activity records in four major seismic zones in this work, and in turn we have made the probabilistic prediction for large earthquake (magnitudes $M \geq$ 7). The probability density function (PDF) of the spatial coordinates of all earthquake zones has been calculated for each seismic zone analysed. The PDFs of the longitudes and latitudes are shown at the top and left panels in **Figures 4**, **9**, **13**, **15**.

- Seismicity related to transform and subduction margins in North America (Southwestern regions of both United States and Mexico)

The scope of the tectonics setting related to the seismicity along Mexico's Pacific coast is divided into two regions: Northern and Southern (**Figure 4**).

In the northern Mexican subduction zone, the Gulf of California spreading center and the triple junction point around the Jalisco and the Michoacán Blocks represent the most active seismogenic belts inducing significant seismic hazard in the Jalisco-Colima-Michoacán region (Dañobeitia et al., 2016). The oblique to sub-parallel motions between the North American and Pacific plates at the latitude of the San Andreas fault produce a broad zone of large-magnitude earthquake activity mainly associated with dextral strike-slip faults extending more than 500 km into the continental interior. This seismic and tectonic activity patterns define the western limits of plate interaction as well as dominate the overall pattern of seismic strain release (Castro et al., 2017). Due to the

Rivera Transform Fault, this seismic source corresponds to the shallow seismicity (mean depth value of 16 km), showing strike-slip faulting mechanisms delimiting the southwestern border between the Rivera and the Pacific Plates (Sawires et al., 2021).

Most of the earthquakes in Southwestern Mexico are due to the subduction process between the Cocos and North American tectonic plates (Pardo and Suárez, 1995). The subduction zone extends 1,300 km along the coast of the Pacific Ocean from the Chiapas state to the Jalisco states showing an angle in the range of $12°–45°$ (Suárez et al., 1990; Singh and Pardo, 1993). The Rivera plate moves with a relative velocity between 2.5 and 5.0 cm/year (Kostoglodov and Bandy, 1995), and the relative velocity of the Cocos plate in the Pacific Coast is in the range of 5.0–7.5 cm/year (Singh and Pardo, 1993). The subduction earthquakes are originated on the Pacific Coast with reverse fault focal mechanism and depths in the range of 10–40 km. The rupture lengths of these earthquakes are between 50 and 250 km, and their widths vary between 75 and 150 km (García et al., 2005). The deeper, in-slab earthquakes are also related to the subduction process, and their epicentres are located inside the continent. The hypocenters of these events have occurred at depths between 40 and 150 km in Mexico's central and western zone, and they are produced by the rupture of the subducted lithosphere (Jara et al., 2015).

- Seismicity in South America (subduction oceanic lithospheric plate versus continental lithospheric plate, Andean case)

The South American plate is bounded for the subduction of the oceanic Nazca plate towards the west and the South Atlantic crustal oceanic section of the plate towards the east (**Figure 9**). This westerly subduction and the easterly spreading stresses due to the opening of the Oceanic Middle Ridge produces a compressional stress pattern on all the continents (**Figure 1**).

Earthquakes along the Andean cordillera show a progressively deep from the Chilean trench towards east associated with reverse (majority) or strike-slip faulting mechanisms with the principal significant compressional stress (σ_1) roughly in E-W direction. Along the Pacific coast, the deadliest earthquakes associated with tsunamis were registered during the last decades.

The seismicity in central Chile observed from 0 to 30 km depth beneath the western Andean thrust is due to the subduction of the Nazca plate. It shows essential seismicity beneath the Principal Cordillera located at a depth of 10 km, and deeper seismicity (~15 km) aligned with the main Andean thrust (Ammirati et al., 2019).

Rivas et al. (2019) determine in the foreland Andean region that has 44 seismic locations with focal depths mechanisms showing mainly reverse and in less proportion strike-slip solutions ranging between 10 and 30 km and magnitudes $1.2 \leq M \leq 4$. The intermediate principal stress (σ_2) is also compressional and more significant than the lithostatic pressure (σ_v). In the mid-plate South America, earthquakes seemed related to purely compressional stresses pattern (both σ_{Hmax} and σ_{hmin} larger than σ_v). Along the Atlantic margin, the regional stresses are affected by coastal effects due to transform

fault stresses as well as flexural effects from sediment load at the continental margin (Assumpcao et al., 2016).

This coastal effect tends to make σ_{Hmax} parallel to the coastline and σ_{hmin} (usually σ_3) perpendicular to the coastline. Few available breakout data and *in-situ* measurements are consistent with the pattern derived from the earthquake focal mechanisms (Heidbach et al., 2016). The Rio de la Plata craton and surrounding areas of *Argentina* and Uruguay located on the Atlantic margin of the South America plate have always been known as having deficient and shallow earthquakes activity related to preexisting faults (Rossello et al., 2020).

- Seismicity in Japan (subduction oceanic lithospheric plate versus oceanic lithospheric plate)

Japan's most densely populated area is subjected to intense crustal stress due to the convergence by subduction of two oceanic lithospheric plates (**Figure 12**). This compressional context produces consequently high seismic activity, among other geological processes associated with the subduction, such as volcanism and exhumation (Heidbach and et al., 2016). There are approximately 5,000 minor earthquakes recorded per year, mainly between 3.0 and 3.9 magnitudes and around 160 earthquakes with a magnitude of 5 or higher, which caused significant damages or casualties.

Geological analyses previously performed in the field, such as geomorphological markers, trench surveys and radiometric dating along the Nojima fault, associated with the 6.9 magnitude 1995 Kobe earthquake, revealed significant activity during the Pleistocene-Holocene times. Mainly, at least two significant earthquakes preceding the 1995 earthquake occurred in the last 1800 years (Lin, 2018). According to these authors, there would be no consistency between the recurrence intervals of seismic events proposed in previous contributions.

In this region, among many other areas of interest, the Hinagu-Futagawa fault zone (HFFZ) represents a study case, where the 7.1 magnitude Kumamoto earthquake occurred in 2016, which produced a surface rupture of ~40 km in length. Detailed geological studies in the field and radiometric dating (Lin et al., 2017) allowed inferring four events in the last 4,000–5,000 years on this fault, suggesting a mean late Holocene recurrence interval of 1,000 years for the associated morphogenic earthquakes. However, as expressed by the authors mentioned above, these results contradict previous studies estimating recurrence intervals of 3,600–11,000 and 8,000–26,000 years for the target segments of the Hinagu and Futagawa faults, respectively. Different methods have been carried out to predict seismic events in this region (Uyeda, 2013), albeit with numerous difficulties inherent to the complexity of the discipline and the non-linear nature of such complex chains of phenomena.

- Seismicity in Southern China-Northern India (collisional margin continental lithosphere plate versus continental lithosphere plate, Himalayan case)

The seismicity of this zone (**Figure 15**) corresponds to the sutured obducted margin formed by the collision of the Indian plate against the Asian plate, which is occurring for the last tens of millions of years.

As a result of this collision, the high Himalaya orogen was formed (Tapponnier et al., 1982). As the process is still active even today, intense crustal N-S oriented stress (Heidbach et al., 2016) associated with intense seismic activity and cortical deformation were produced (**Figure 1**).

Numerous contributions (Bilham, 2019) have been analyzed through different methodologies, rupture zones and rupture propagation directions, regular convergence rates, as well as evaluated the slip potential in different segments of the Himalaya and the occurrence of potential high magnitude earthquakes. There is increasing evidence that Himalayan seismicity may be bimodal: blind earthquakes (up to M ~ 7.8) tend to cluster in the lower part of the seismogenic zone, while infrequent large earthquakes (Mw 8+) propagate up the Himalayan frontal thrust (Dal Zilio et al., 2019).

Recently, Michel et al. (2021), considering many variables and uncertainties, suggest that earthquakes of magnitude greater than 8.7 (such as the one that occurred in 1950) are the most likely candidates to be the largest possible earthquakes in the Himalayas. However, they emphasize that, given the magnitude frequency distribution model used, the probability of a magnitude 8 + earthquake occurring in 100 years ranges from ~60–80%. The most likely associated recurrence time for such an event exceeds 1,000 years.

3 DATA AND METHOD

3.1 Spatial Clustering

For this work, the orographic basemap comes from the ArcMap map library. The ocean bathymetry layer is obtained from the General Bathymetric Chart of the Oceans (GEBCO)[2]. The seismic records of the period 1902–2021 of the North America, South America, China and Japan regions taken from the U.S. Geological Survey (USGS)[3], are processed in a geographic information system (GIS) to create a vector layer of geo-referenced points. Continuous surface maps with seismic magnitude information are designed with the Kriging-type interpolation method within a GIS. This probabilistic method is widely used to generate seismic maps (Türker and Bayrak, 2018; Teves-Costa et al., 2019; Moradia et al., 2020) to its efficiency in predicting information from one variable to the other. Through the spatial structure of its discrete values. Eight classes with equal intervals (0.5) are defined to better represent and interpret the spatial results. The processing and plotting of seismic magnitude data on a map is done using GIS. Plate Boundary and Movement Information from USGS was added to the interpolated map to find out its geographic location and plate type, as well as its displacement. A spatial data filter is applied to the seismic record layer to extract

---

values ≥ 7 for inclusion in the interpolated map to locate the largest earthquakes. Using spatial analysis tools, density zones from the seismic records are defined for the filtered data.

3.2 Seismic Events

In order to analyze and search for any coherent seismic patterns, it is necessary to have an excellent catalogue of seismic activity. Seismic patterns should show periods when there is seismic activity as well as when there is no seismic activity at a certain magnitude if it exists. In this work, we have analyzed the seismic activity in four different and important seismic zones:

1) Seismicity in North America (**Figure 4**)
2) Seismicity in South America (**Figure 9**)
3) Seismicity in Japan (**Figure 12**)
4) Seismicity in Southern China-Northern India (**Figure 15**)

The public data on seismic activity have been obtained from USGS. These data contain the list of all registered seismic activity. The table also contains information on the date of each seismic event, its magnitude, depth, type of magnitude, longitude, and latitude. We will analyze seismic activity for strong earthquakes magnitudes ≥ 7 available from 1900 to 2021. Also, we only analyze the earthquakes that are delimited in the areas/zones shown and discussed in **Figures 4, 9, 12, 15**.

3.3 Time Series With Data Gaps

One of the biggest problems of analyzing incomplete time series (such as the seismic activity catalogues) is to extract the information of the phenomena. A solution used regularly to analyze this type series is to apply different interpolation methods Sturges (1983). However, any interpolation can lead to an underestimation of spectral power at both higher and lower frequencies. Another technique sometimes used is to extract the mean value of the data (Carroll et al., 1997). Though, this technique can fail for data records that have a trend.

In particular, gaps in geological, geophysical, geographic, and seismic databases exist because those records contain errors, or were not complete nor homogeneous, and were created with data from different epochs (Jopek and Kaňuchová, 2017; Soon et al., 2019). In this case, a Bayesian block in the geosciences record can be applied to suppress the inevitable corrupting observational errors (see, for example, Scargle et al., 2013). The statistical inference using a Bayesian approach is used for analyzing an incomplete database (Gelman and Meng, 2005). In addition, spectral analysis has also been used to study time series with missing data (Maoz et al., 1997; Ding et al., 1998; Ramírez-Rojas et al., 2019). For example, the Fourier Transform (FT) is applied for analyzing these data. Nevertheless, this method may not often be suitable for non-stationary and irregularly spaced time series (Velasco Herrera V. M. et al., 2022). Therefore a new and reliable method is required to study a time series with gaps such as the seismic events.

The classical wavelet technique (see, e.g., Torrence and Compo (1998); Velasco Herrera et al. (2017)) is used to analyzed non-stationary times series, but this technique can only be used for regularly spaced time series. This is why a modified wavelet

technique has been proposed and used to analyse incomplete time series. This technique has been named gapped wavelet (Frick et al., 1997; 1998; Soon et al., 2019). In our study, we have used the gapped wavelet spectral algorithm to analyze seismic events. Another variant for the analysis of irregularly spaced time series with wavelet has been proposed in Salcedo et al. (2012) recently.

3.4 Gapped Wavelet

In order to find seismic patterns in an area, it is necessary to analyze the periodicities that the seismic time series could have. Since earthquakes do not occur continuously, this implies frequent inactive or zero-activity "gaps" in the seismic catalogues. That is why we propose to analyze the seismic catalogues with the gapped wavelet transform. The gapped wavelet transform (W_g) of a time series with data gaps $f_g(t)$ is a matrix (Ω) and is defined by Frick et al. (1997) as:

$$W_g\{f_g(t)\} = \phi\,\Omega(t,a) \tag{1}$$

with

$$\phi = \sqrt{\frac{1}{a\,C(a,t)}} \tag{2}$$

$$\psi'(t',t,a) = \left[h\left(\frac{t'-t}{a}\right) - C(a,t)\right]\Phi\left(\frac{t'-t}{a}\right)G(t') \tag{3}$$

$$C(a,t) = \frac{\int_{-\infty}^{\infty} h\left(\frac{t'-t}{a}\right)\Phi\left(\frac{t'-t}{a}\right)G(t')\,dt'}{\int_{-\infty}^{\infty}\Phi\left(\frac{t'-t}{a}\right)G(t')\,dt'} \tag{4}$$

$$G(t) = \begin{cases} 1, & \text{if the signal is registered} \\ 0, & \text{lost data or no data reported} \end{cases} \tag{5}$$

$$\Omega(t,a) = \int_{-\infty}^{\infty}\psi'^{*}(t',t,a)f_g(t')\,dt' \tag{6}$$

where t is the time index and a is the wavelet scale, the superscript (*) indicates the complex conjugate, and ψ is the mother function. We applied the Morlet's mother function (ψ) to analyze the power spectral density (PSD) of seismic activity since this mother function does not only provide a higher periodicity resolution but also is a complex function that allows calculating the inverse wavelet transform (Torrence and Compo, 1998; Velasco Herrera et al., 2017; Soon et al., 2011, 2019). Then $\Phi(t) = e^{(-t^2/2)}$, $h(t) = e^{iw_o t}$, with $w_o = 6$.

The meaningful wavelet periodicities with a confidence level greater than 95% must be inside the cone of influence (COI), and thin black contours mark the interval of 95% confidence (Torrence and Compo, 1998). The global spectra show the power contribution of each periodicity inside the COI. Also, we established the significance levels in the global wavelet spectra with a simple red noise model that increases power with decreasing frequency (Gilman et al., 1963). The uncertainties of every peak position are obtained from the peak full width at half maximum (Mendoza et al., 2006).

3.5 Inverse Wavelet Spectral Analysis

The decomposition of f_g in channel (y_n) can be obtained from the inverse wavelet (Torrence and Compo, 1998) as:
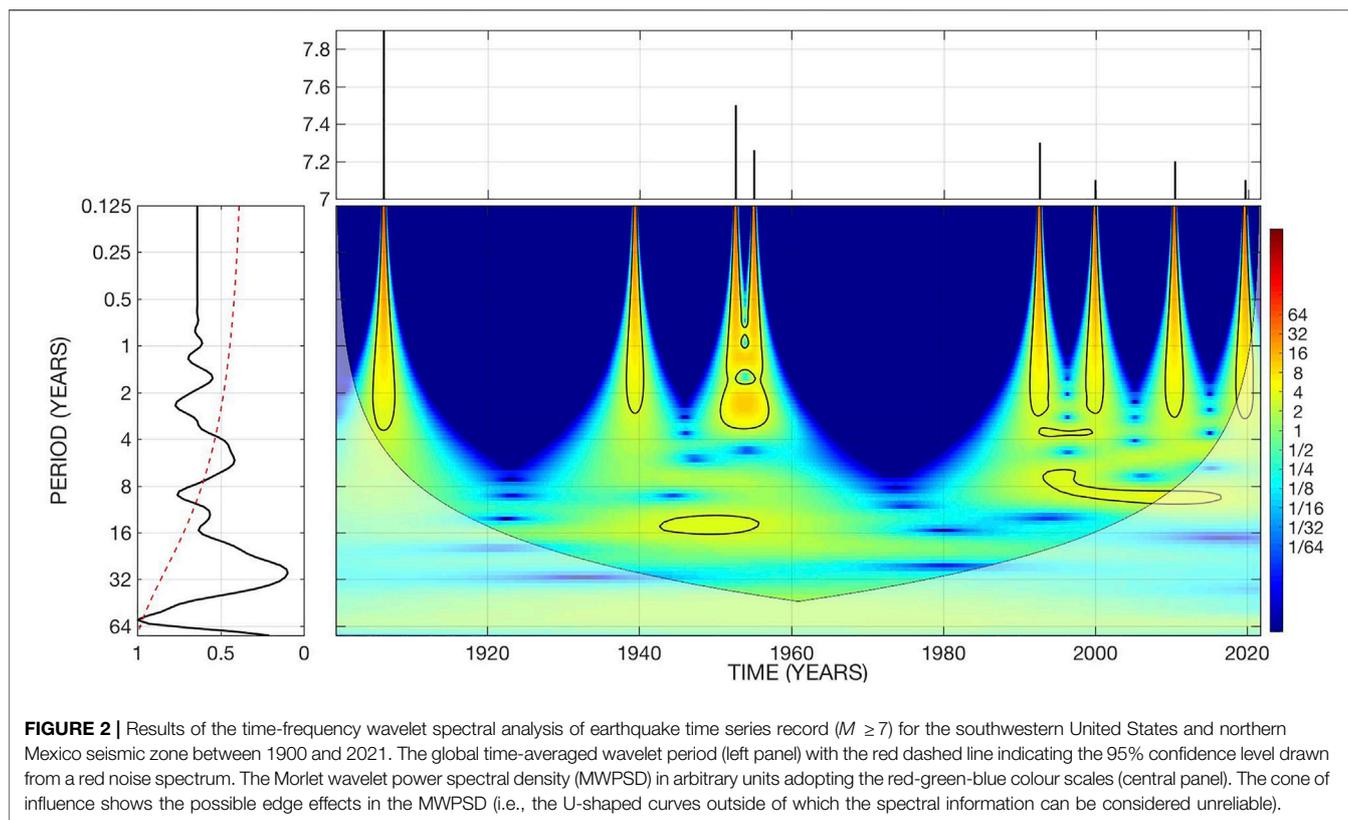
FIGURE 2 | Results of the time-frequency wavelet spectral analysis of earthquake time series record ($M \geq 7$) for the southwestern United States and northern Mexico seismic zone between 1900 and 2021. The global time-averaged wavelet period (left panel) with the red dashed line indicating the 95% confidence level drawn from a red noise spectrum. The Morlet wavelet power spectral density (MWPSD) in arbitrary units adopting the red-green-blue colour scales (central panel). The cone of influence shows the possible edge effects in the MWPSD (i.e., the U-shaped curves outside of which the spectral information can be considered unreliable).

$$y_n = \frac{\delta_j \delta t^{1/2}}{C_\delta \psi_o(0)} \sum_{j=j_1}^{j_2} \frac{Re\left(\mathbf{W_g}\left(\mathbf{s_j}\right)\right)}{s_j^{1/2}} \qquad (7)$$

where j_1 and j_2 define the scale range of the specified spectral bands, $\psi_o(0)$ is an energy normalization factor, C_δ is a reconstruction factor, and δ_j is a factor for scale averaging. For the Morlet wavelet, $\delta_j = 0.6$, $C_\delta = 0.776$, and $\psi_o(0) = \pi^{-1/4}$.

The input data in the $\mathbf{W}_g$ are the seismic catalogues between 1900 and 2021. The $\mathbf{W}_g$ has two main outputs as shown in **Figures 2, 5, 7, 10, 13**; the global spectrum, which shows the periodicities existing in the seismic record with the 95% confidence level above the red noise spectrum drawn as red dashed line (left panel) and the wavelet power spectral density (PSD) that shows the evolution over time of these periodicities (central panel).

3.6 Machine Learning Algorithms for Forecasting Seismic Activity

3.6.1 Non-Linear Autoregressive eXogenous Model

The system's state can describe the dynamics of seismic activity from its input (V)-output (Y) behavior, which describes its evolution over time. Different models can approximate the state of the system. In particular, we use the Non-linear Autoregressive eXogenous (NARX) (Suykens et al., 2005) model in order to create forecasting models of seismic activity variation ($\hat{\mathbf{Y}}$) that is defined as:

$$\hat{Y}[k+1] = \Xi\left(\begin{matrix} V[k,P] \\ Y[k,Q] \end{matrix}\right) \qquad (8)$$

where Ξ is a transfer function of the state of the system at the moment "k" to the moment "$k+1$", that depends intrinsically on the input and output data (V and Y, respectively), p and Q are the delay times, and $\hat{\mathbf{Y}}$ is the estimated seismic activity at a time "$k+1$".

We used the Least-Squares Support-Vector Machines (LS-SVM) algorithms to estimate the transfer function (Ξ), a non-linear function (Suykens et al., 2005) as:

$$\Xi = \sum_{k=1}^{n} \omega_k \mathbf{D}^k + \mathbf{B} \qquad (9)$$

where D is training data, in our case the seismic records. Also D^k denotes the input data, *i.e.,* the seismic records at time "k" (discrete time index from $k = 1, \ldots, n$), ω_k is the weighting factor which in turn has functional dependence on V_k and $\mathbf{B}$ is the bias term.

We use the Bayesian inference ML model (Suykens et al., 2005) obtained from the seismic records to provide a probabilistic earthquake prediction of the variation in the seismic activity. Bayes's theorem (Bayes, 1763) is the basis of our ML model and can be expressed as follows:

$$p(\Xi|D) = \frac{p(D|\Xi)}{p(D)} p(\Xi) \qquad (10)$$

where Ξ is the Least-Squares Support-Vector Machines (**Eq. 9**) regression model.

Furthermore, Bayes's theorem is used to deduce the optimal parameters in the LS-SVM model (**Eq. 9**). In addition, we use the radial basis function (RBF) kernel in this LS-SVM method. In this work, we have applied and modified the LS-SVM algorithms and toolbox by Suykens et al. (2005).

3.6.2 Algorithms for the Estimation of the Next High or Active Phase of Seismic Activity

We apply the following iterative steps to forecast the next high seismic activity season:

1) Use wavelet transform (**Eq. 1**) to find the periodicities (seismic patterns) in each seismic zone analyzed. The results are shown in **Figures 2**, **5**, **7**, **10**, **13**.
2) The decomposition of the seismic record in time series called "channels" with the periodicities obtained in step (1) can next be obtained using the inverse wavelet (**Eq. 7**).
3) Selection of the model lags P and Q for each Bayesian inference model that has been analyzed in each seismic zone.
4) Use the Radial Basis Function (RBF) kernel.
5) For training, validation, testing and deduction of the hyper-parameters of the model. Use the K-fold cross-validation.
6) Set aside $1/K$ of data. Train the model with the remaining $(K - 1)/K$ data. Measure the accuracy obtained on the $1/K$ data that we had set aside. K independent training is therefore acquired. The final accuracy will be the average of the previous K accuracies. Note that we are hiding or withholding a $1/K$ part of the training set during each iteration. This is applied at the time of training. After these K iterations, we obtain K accuracies that should be similar to each other; this would be an indicator whether the model is working well or not. In this work, $K = 10$ is adopted, but, it is possible to vary K between 5 and 10.
7) Determination of the weight and bias.
8) Estimation of next high cycle of earthquake activity using **Eq. 9**. Before forecasting the following period of strong earthquake activity, it is necessary to quantify the ability of the Machine Learning to "predict" the recent clustered earthquakes. We use 80% of the Bayesian clustered model (that is, data from 1900 to 1996) as input data to "forecast" the remaining 20% of the Bayesian clustered model (i.e., 1997–2021) in each seismic zone analyzed. The Bayesian clustered model of the historical earthquakes shows that all the historically strong earthquakes were manifested during the positive phase of the Bayesian clustered model; this fact indicates that our model has no overtraining nor undertraining. Furthermore, the wavelet analysis shows that the high and low seasons of strong earthquakes have a multiannual and multidecadal variations, so the Bayesian model we deduced is not overly complex, which implies that the validation is simple. We do not show the validation figures but instead choose to concentrate on the forecasting result.
9) Computation of a cost function.

10) Test of the accuracy of the estimate next high cycle of earthquake activity.
11) Test of the cost function: if this function was small enough, we stopped. Otherwise, we change one of the parameters and repeat from step (2) onwards.

We have used and modified the LS-SVM algorithms and toolbox by Suykens et al. (2005) for this goal.

We want to add that the accuracy of any forecast of seismic activity is limited by an uncertainty principle (Velasco Herrera et al., 2015). Great precision in the spatial location forecast implies a significant uncertainty in the temporal forecast. This is why in this work, we focus on the problem of temporal forecasting by proposing a new Bayesian Machine Learning composite method. In our case, the possible zones where the following strong earthquakes in each seismic zone analyzed could occur have been essentially clustered and pre-determined according to the methodology described in **Section 3.1**.

4 RESULTS

In this section, we show the time-frequency seismic patterns of strong earthquakes ($M \geq 7$) from 1900 to 2021 in the following four major seismic zones: 1) the United States and Mexico, 2) South America, 3) Japan, and 4) Southern China and Northern India, using the wavelet transform. After finding seismic patterns in each of these seismic zones, the oscillation with a given periodicity that groups the historical earthquakes ($M \geq 7$) into high and low seismicity will be obtained using the inverse wavelet transform. Each of these oscillations is used to create a probabilistic long-term earthquake prediction model for each seismic zone analyzed using the Bayesian Machine Learning method.

4.1 Southwestern United States and Mexico

Figure 2 shows the wavelet analysis of the strong earthquake records ($M \geq 7$) for the southwestern United States and northern Mexico (see **Figure 4**) between 1900 and 2021. The top panel shows that only seven strong earthquakes have been recorded in this century-long interval and they are heterogeneously distributed between 1900 and 2021. The global wavelet spectrum (left panel) shows periodicities (seismic patterns) at 1.2 ± 0.5, 2.4 ± 0.9, 9.2 ± 2, 15.7 ± 4, and 55 ± 10 years. The time evolution of the power spectral density (PSD) for these periodicities is illustrated in the central panel.

Each of the periodicities shown in the global wavelet spectrum (seismic patterns) obtained with the wavelet transform implicitly provides information about the intrinsic properties of the tectonic plates of the southwestern United States, the interaction between these plates, the sources both internal and external that modulate the tectonic movement as well as the dynamics of strong earthquakes. In particular, we are selecting seismic pattern of 55 years that indicates the period of recurrence of strong earthquakes. The other periodicities shown in the global wavelet spectrum will be discussed in other future analysis

FIGURE 3 | **(A)** Probabilistic earthquake prediction for the southwestern United States and northern Mexico. Bayesian inference of LS-SVM model (blue line and shade) compared with the historical earthquakes clustered in 3 groups (I-III) and shown with vertical blue bars from 1900 to 2021. In addition, the probabilistic earthquake prediction is shown for the two future periods (cluster IV and V). The blue shaded area represents the 95% confidence intervals of the Bayesian model. Furthermore, we show and contrast the earthquake activity events at Parkfield, California based on **(B)** the model proposed by Bakun and Lindh (1985) and **(C)** the results from our Machine Learning algorithm with the adopted periodicity of 35-years. Please see **Section 4.1.1** for the discussion of the earthquake activity and event at Parkfield, California.

FIGURE 4 | Earthquakess in the Southwestern United States and Mexico. **(A)** Seismic activity between 1900 and 2021 is shown as yellow triangles for 5 ≤ earthquakes magnitudes < 7. Earthquakes magnitudes ≥7 are marked with red triangles. Plate motion vectors (rotated based on the direction of plate movement), plate limits and earthquakes data were taken from the USGS. The PDF (probability density distribution) of the longitude and latitude shows that the seismic zones have a bimodal and trimodal distribution, respectively (see main text for more details and discussion). **(B)** Distribution of seismic hazards in the Southwestern United States and Mexico. The probabilistic spatial clustering occurrence for 5 ≤ earthquakes magnitudes < 6 is shown in shades of green; for 6 ≤ earthquakes magnitudes < 7 is shown in shades of orange; and for earthquakes magnitudes ≥7 is shown in shades of red.

because in this work the objective is to make a forecast for strong earthquakes.

Figure 3A shows the probabilistic earthquake prediction model (blue line/shade). This is a model with a 55-year periodicity, and it can be seen that the historical seismic events of the southwestern United States and northern Mexico (vertical blue bars) could be grouped into three groups (I-III). The fact that historical earthquakes can be grouped would indicate that the activity and event are not mere random process but that they are the result of a complex interaction between the tectonic plates and the internal and external factors that modulate the tectonic movement and trigger a series of earthquakes that occur only in the positive phase of the 55-year periodicity. This pattern holds for our all new results for other seismic zones discussed below.

We want to note that there are historical records of strong earthquakes before 1900 and that these earthquakes also occur in the positive phase of the 55-years oscillation. However, in this paper, we do not show these older historical earthquakes. According to the Bayesian ML model, the next active seismic period ($M \geq 7$) could probably start in 2040 ± 5 and end in 2057 ± 5 (cluster IV), and no earthquake would be expected in this seismic zone from 2058 ± 5 to 2077 ± 5. Then another new active

period of $M \geq 7$ earthquakes would begin again around 2078 (cluster V).

A cursory study of **Figure 4A** shows that earthquakes could apparently occur anywhere. The PDF of the longitude and latitude shows that the seismic zone has a bimodal and trimodal distribution with maxima at −117° and −93°; and 15°, 25°, and 37°, respectively. However, **Figure 4B** shows the seismic activity's grouping into different classes (see Methodology). In **Figure 4B** earthquakes with magnitudes between 5 and 6 are shown in shades of green. Earthquakes with magnitudes between 6 and 7 are shown in all yellow-orange shades. Earthquakes greater than seven are shown in shades of red-brown. If strong earthquakes (red triangles) have a random distribution, then no more than one should occur in the same area. In addition, it can be seen the grouping of strong earthquakes that are in the areas delimited by a black curve, which are practically around the geological faults. This result can be interpreted in terms of two scenarios for the successive strong earthquakes.

The first scenario is that the next high seismic season, including the "Big One" could occur according to our model between 2040 ± 5 and 2057 ± 5 (see **Figure 3**), around any of the areas outlined with a black line (**Figure 4B**). Also, it is possible that occur around the yellow-orange areas, which are the areas

where historically, earthquakes of categories 6 and 7 have occurred.

The second scenario is that successive strong earthquakes can occur arbitrarily and sporadically in any zone. However, the fact that strong earthquakes are clustered temporally and spatially may indicate that there are both temporal and spatial seismic patterns that had not previously been considered for their study and forecast. So from the ML point of view, this second scenario is less likely.

4.1.1 Forecasting Earthquakes at Parkfield, California

The Parkfield section of the San Andreas fault (California, United States) was officially recognized by the United States government as a seismic physics laboratory for developing earthquake forecasts due to its apparent regularity in six moderate earthquakes (magnitudes between 5 and 7) since 1857 (Bakun and Lindh, 1985). The interval between these seismic events at Parkfield on 9 January 1857; 2 February 1881; 3 March 1901; 10 March 1922; 8 June 1934; and 28 June 1966 is, on average, 21–22 years (Bakun and McEvilly, 1984; Bakun and Lindh, 1985; Bakun et al., 2005). Based on the average recurrence time of 22 years, the Parkfield recurrence model by Bakun and Lindh (1985) forecasted that the next following characteristic Parkfield earthquake could have occurred on 1988.0 ± 5.2 years, i.e., the possible next Parkfield earthquake should occur between 1983 and 1993. However, this expected Parkfield earthquake never occurred within the anticipated interval until 2004 (Bakun et al., 2005).

Our methodology proposed for analyzing strong earthquakes can also be applied to study moderate earthquake activity and events at Parkfield. We seek to analyze and explain why the seventh earthquake could never occur in 1988.0 ± 5.2 years as predicted, but instead occurred between 1993 and 2007. In addition, we made the forecast for the eighth earthquake in Parkfield. One of the differences between the methodology proposed by Bakun and Lindh (1985) and the methodology in this work is that while Bakun and Lindh (1985) had chosen an average value between the events, we use the periodicities (seismic patterns) of the Parkfield earthquakes. A mean value is a global characteristic of a phenomenon or event, while a periodicity is an intrinsic property of the phenomenon or events (see, for example, Velasco Herrera et al., 2021; Velasco Herrera et al., 2022a; Velasco Herrera et al., 2022b). Furthermore, a periodicity is intrinsically related to spectral and temporal power, which is a fundamental concept of physics (Landau and Lifshitz, 1988a; Feynman et al., 2011a). In another deeper sense, a periodicity is also related to the underlying symmetry of the physical phenomenon (Wigner, 1967).
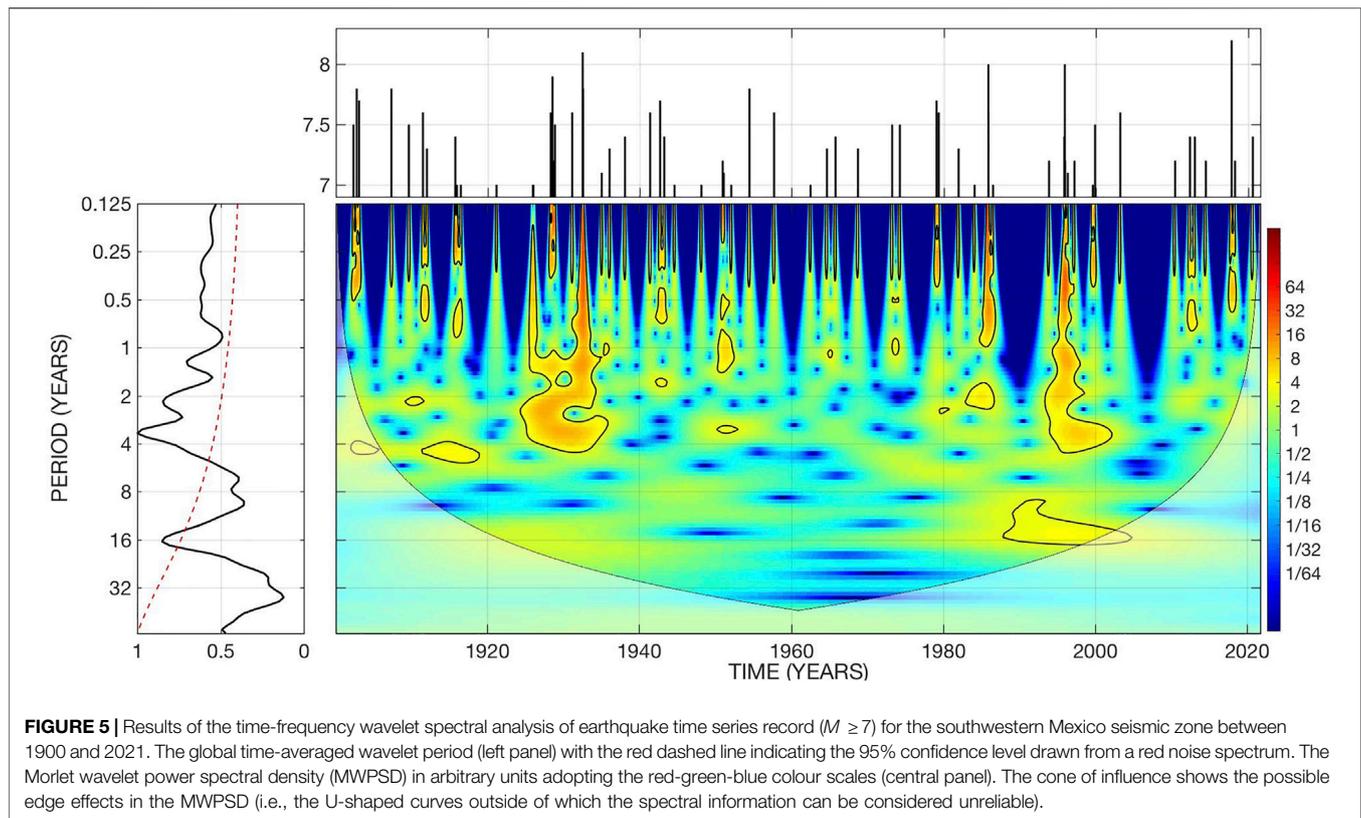
The mean value has the characteristic that all events associated with this value oscillate around it. In fact, the Parkfield seismic events vary between 12 and 38 years, so the mean value of these seismic events is 21 years. Therefore, using the mean value of seismic events to forecast earthquakes from the point of view of signal theory, signal processing, and machine learning is not the most appropriate. While, the periodicities in the wavelet spectra of earthquakes allow us to identify the intrinsic properties of the earthquakes, determine the characteristics of the earthquake's

source, and ultimately determine the interaction between the earthquake and its source and/or the interaction between the earthquake and the faults involved (see for example, Ramírez-Rojas et al., 2019; Soon et al., 2019). In addition, the periodicities cluster the events in high and null seasons (Velasco Herrera V. M. et al., 2022), which allows for constructing models for the forecasts of events with Machine Learning (Velasco Herrera et al., 2021; Velasco Herrera et al., 2022a; Velasco Herrera et al., 2022b).

Figure 3B shows the grouping of earthquakes in Parkfield with the period of 22 years proposed by Bakun and Lindh (1985) for the first six seismic events between 1857 and 1966. For the first five events, the group of historical earthquakes are in the positive phase of the 22-years oscillation grouped by the clusters I to V. Also, these five events occur practically during the five maxima of this periodicity. However, in cluster VI, there was no Parkfield earthquake, but the sixth characteristic earthquake in Parkfield occurred in the positive phase of Cluster VII, which was the 1966 event.

With these six seismic events in Parkfield, we can offer a prognosis for the next characteristic earthquakes of Parkfield with the 22-year period proposed by Bakun and Lindh (1985). The result of this forecast is clustering labeled VII to X. According to the periodicity of 22 years, the seventh characteristic earthquake must have occurred in the positive phase of cluster VII which was between 1979 and 1990. But, the seventh event occurred in 2004 (this seventh characteristic event seismic is not shown in or directly indicated on **Figure 3B** because it was never used for the forecast) i.e., during the positive phase of cluster IX. Therefore, earthquakes did not occur in Parkfield in two clusters (VI and VIII). Based on these results, the empirical evidence suggests that the periodicity of 22 is not a characteristic seismic pattern of earthquakes around the Parkfield seismic zone. This is because, according to the grouping of the 22-years oscillation, the forecast either can be fulfilled or cannot be fulfilled. The fact that an earthquake did not occurred (as predicted) could mean that there were no human and economic losses for the population that lives near these active seismic areas. However, for the development of the science of earthquake forecasting, the fact that a predicted earthquake or earthquake event did not occurred means that the proposed model does not have all the necessary information. In turn, such failure simply means that it is necessary for a more serious and careful re-analysis of the data and a deeper understanding of the particular seismic zone.

Almost 2 decades after the 2004 Parkfield earthquake and nearly 4 decades after the pioneering Bakun and Lindh (1985) forecast, it is necessary to explain why physically and geologically, an earthquake could not occur in 1988.0 ± 5.2 years as proposed. In addition, it is necessary to correct the pioneering/original model proposed by Bakun and Lindh (1985). Wavelet analysis (figure not shown) of the earthquake activity around the Parkfield seismic section show periodicities of 6.2, 11.1, 20.8 and 35.1 years. We are not going to focus on the periodicities of 6.2 and 11.1 years in this paper. We note that the periodicity of 20.8 ± 5 years, with its associated uncertainty, is practically the 22-years oscillation

FIGURE 5 | Results of the time-frequency wavelet spectral analysis of earthquake time series record ($M \geq 7$) for the southwestern Mexico seismic zone between 1900 and 2021. The global time-averaged wavelet period (left panel) with the red dashed line indicating the 95% confidence level drawn from a red noise spectrum. The Morlet wavelet power spectral density (MWPSD) in arbitrary units adopting the red-green-blue colour scales (central panel). The cone of influence shows the possible edge effects in the MWPSD (i.e., the U-shaped curves outside of which the spectral information can be considered unreliable).

proposed by Bakun and Lindh (1985), which was presented in **Figure 3B**.

Next we will focus on the result of the 35.1± 7-year periodicity from our wavelet analysis. **Figure 3C** shows the groupings of historical earthquakes with a periodicity of 35.1 ± 7 years. It can be observed that all, absolutely all, Parkfield earthquakes are occuring during the positive phase of the six clusters identified. This is the first contrast of our results with the model proposed by Bakun and Lindh (1985). Furthermore, it is recognized that more than one event can occur within a single cluster. Such is the case of cluster IV, where two seismic events occurred.

According to the Bayesian Machine Learning model with the chosen period of 35 years, the characteristic earthquakes in Parkfield can *never* occur in the negative phase, especially between 1978 and 1992, as proposed by Bakun and Lindh (1985). Our model suggests that during cluster VI (which is between 1994 and 2007), the seventh seismic event would occur. Again, in this forecast, we do not use the 2004 earthquake to make the actual forecast; which is the reason why we did not plotted the 2004 event in **Figure 3C**. Indeed, we wish to highlight in this paper that the seventh event at Parkfield is correctly "forecasted" within the cluster VI. In addition, we want to note that the next season in which at least one characteristic Parkfield earthquake can occur would be around cluster VII starting in 2019 and ending in 2032. Except for cluster IV, Parkfield seismic events occur around the maximum of each cluster. So with a high probability, the eighth earthquake in Parkfield seismic zone/ section can be reasonably expected to occur at around 2025 ±

2 years. This date will be coming soon, so it will be possible to follow in great detail the seismic events in this well-recognized recognized seismic laboratory at Parkfield, California. We are cautiously hopeful that this is indeed a significant area for the development of the science of earthquake forecasts in order to proffer effective and reliable early warnings with the worthy objective of minimizing human and economic losses.

4.1.2 Mexican Earthquake

Figure 5 shows the wavelet analysis of the earthquake records ($M \geq 7$) for southwestern Mexico (see **Figure 4**) between the years 1900 and 2021. A more significant number of earthquake is observed here when compared to the southwestern United States and northern Mexico. This shows a completely different seismic activity and pattern between Northern and Southern Mexico. Also, this could indicate that in southwestern Mexico there is less viscosity between the tectonic plates when compared to the southwestern United States and northern Mexico. The global wavelet spectrum (left panel) shows periodicities at 1.3 ± 0.4, 2.2 ± 0.5, 3.8 ± 0.6, 7.7 ± 1, 16.4 ± 3, 30.1 ± 4, and 56 ± 5 years. The time evolution of the power spectral density (PSD) for these periodicities is shown in the central panel. We have selected the periodicity of 3.7 years to group the strong earthquakes in southwestern Mexico.

Figure 6 shows the probabilistic earthquake forecast ($M \geq 7$) for southwestern Mexico. This model has adopted a nominal recurrent periodicity of 3.7 years and it can be seen that this model groups the historical earthquakes (black bars) into
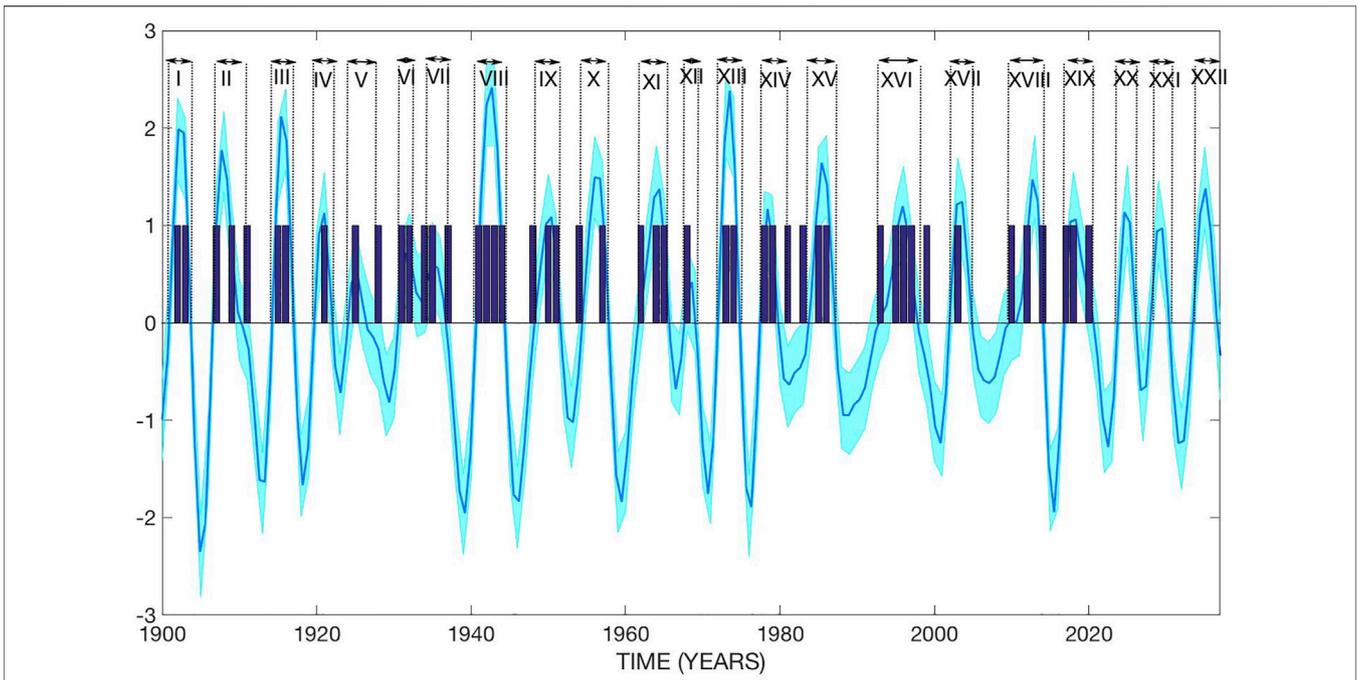
FIGURE 6 | Probabilistic earthquake prediction($M \geq 7$) for southwestern Mexico. Bayesian inference of LS-SVM model (blue line and shade) compared with the historical earthquakes clustered in nineteen clusters (I-XIX) and shown with vertical blue bars from 1900 to 2021. In addition, the probabilistic earthquake prediction is shown for the following three periods (clusters XX-XXII). The blue shaded area represents the 95% confidence intervals of the Bayesian model.
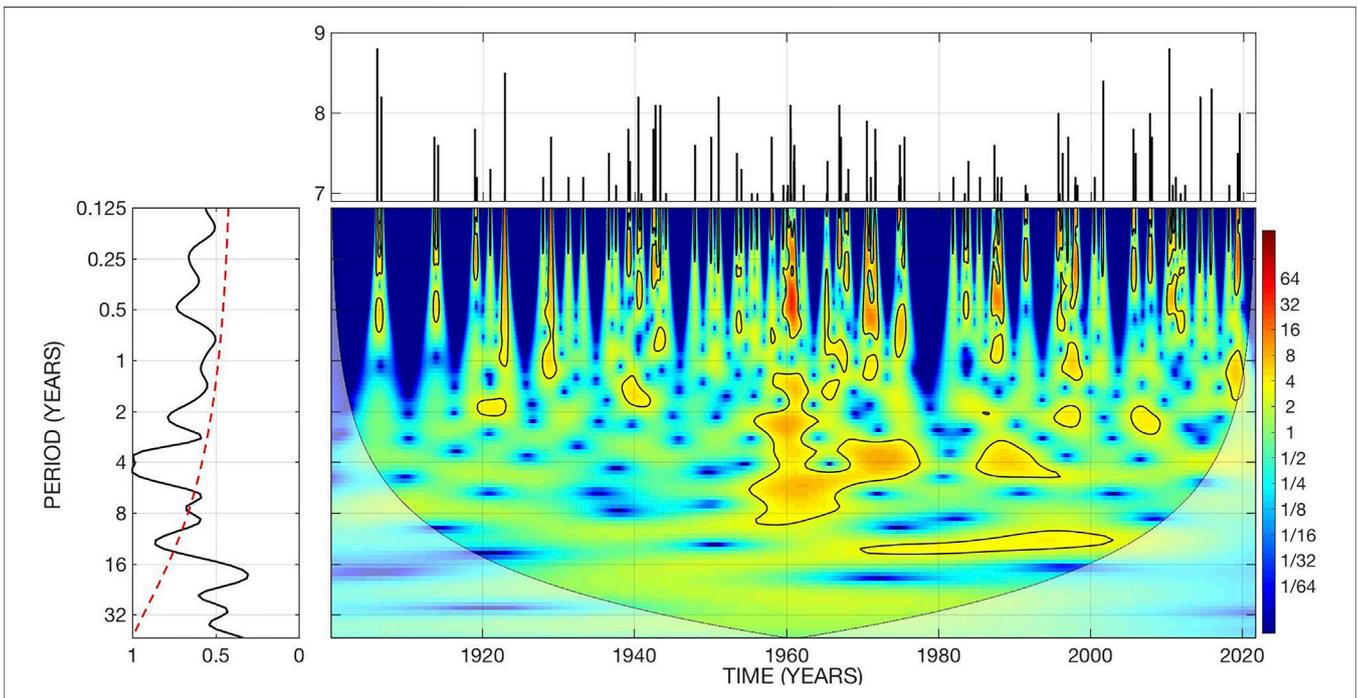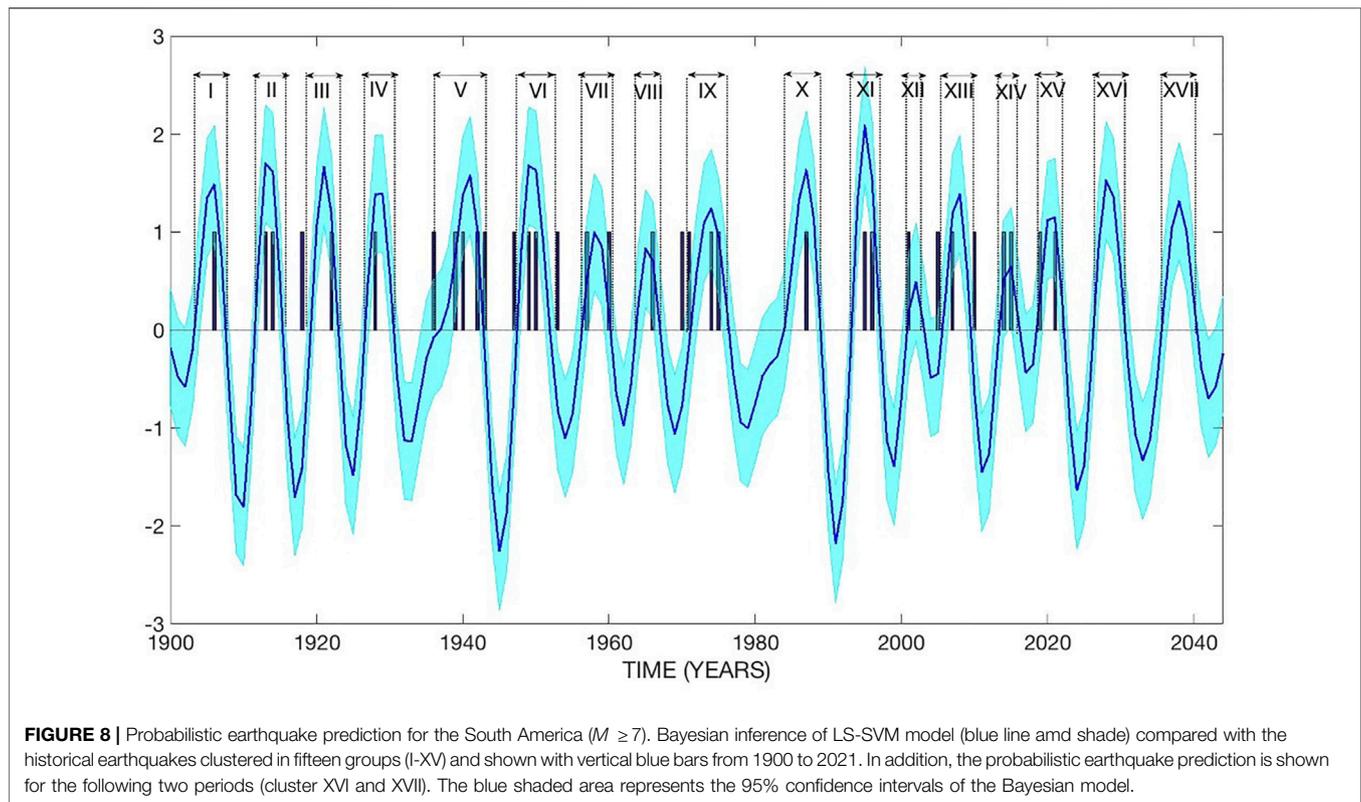

FIGURE 7 | Results of the time-frequency wavelet spectral analysis of earthquake time series record ($M \geq 7$) for the South American seismic zone between 1900 and 2021. The global time-averaged wavelet period (left panel) with the red dashed line indicating the 95% confidence level drawn from a red noise spectrum. The Morlet wavelet power spectral density (MWPSD) in arbitrary units adopting the red-green-blue colour scales (central panel). The cone of influence shows the possible edge effects in the MWPSD (i.e., the U-shaped curves outside of which the spectral information can be considered unreliable).

FIGURE 8 | Probabilistic earthquake prediction for the South America ($M \geq 7$). Bayesian inference of LS-SVM model (blue line amd shade) compared with the historical earthquakes clustered in fifteen groups (I-XV) and shown with vertical blue bars from 1900 to 2021. In addition, the probabilistic earthquake prediction is shown for the following two periods (cluster XVI and XVII). The blue shaded area represents the 95% confidence intervals of the Bayesian model.

nineteen clusters. Furthermore, it can be seen that these seismic events also occur in the positive phase of the 3.7-years oscillation. This characteristic shows that strong earthquakes in southwestern Mexico are not temporally random. According to this model, the next period of greater earthquakes ($M \geq 7$) would start in 2024 ± 1 and last until 2026 ± 1. This is clearly a testable scientific proposition.

In **Figure 4B**, the cluster of strong earthquakes in southwestern Mexico can be studied. It can be seen that these earthquakes occur preferentially in the subduction zone. In addition, the following strong earthquakes may occur in the cluster zones and in the yellow-orange areas where magnitude 6 to 7 earthquakes have historically occurred. In addition to the interplate earthquakes in southwestern Mexico, there are also intraplate earthquakes in central Mexico. If successive earthquakes occur in this area, it could cause significant damages in the cities of central Mexico with serious human and economic losses (e.g., Novelo-Casanova et al., 2013).

4.2 South America
Figure 7 shows the wavelet analysis for earthquakes ($M \geq 7$) in the South American seismic zone. The global wavelet spectrum shows the periodicities of 1.10.3, 2, 2 ± 0.5, 3.6 ± 0.7, 4.5 ± 0.7, 7.7 ± 1.7, 12.1 ± 2.5, 24.6 ± 5.5, and 46.8 ± 8.4 years.

In order to group strong earthquakes, we use the periodicity of 7.7 years. This nominal choice of longer recurrent periodicity may indicate or imply that there is greater viscosity between the tectonic plates in South America than in southwestern Mexico. **Figure 8** shows the probabilistic earthquake forecast (blue line/

shade) model of 7.7-years that grouped historical earthquakes into fifteen clusters. According to this model, the next seismically active period would begin in 2026 ± 2 and end in 2031 ± 2 (cluster XVI).

In **Figure 9A** shown the seismic activity in South America between 1900 and 2021. Also, the PDF of the longitude and latitude shows that the seismic zone has a trimodal and quadrimodal distribution with maxima at −81°, −71°, and −66°; and −44°, −32°, −21°, and −2°, respectively.

In **Figure 9B**, the spatial clustering of strong earthquakes in South America is shown. It can be seen that these earthquakes occur preferentially in the subduction zone. In addition, the following strong earthquakes may occur in the cluster zones and yellow-orange areas where magnitude 6 to 7 earthquakes have historically occurred.

4.3 Japan
Owing to the significant and relatively more frequent seismic activity in Japan, we have divided earthquakes in the Japanese zone into two groups. The first group consists of earthquakes $7 \leq M < 8$. After the 11 March 2011s "Great East Japan Earthquake", offshore of the Tohoku region (see, for example, Davis et al., 2012, for a full discussion about the missed opportunity for disaster preparedness), there is a new fundamental question in developing earthquake forecasts in Japan: When could another similar or equal magnitude earthquake occur? To offer a possible answer, we analyze earthquakes in Japan for magnitudes equal to or greater than 8. Therefore our study of the second group is focused on the strongest earthquakes $M \geq 8$.
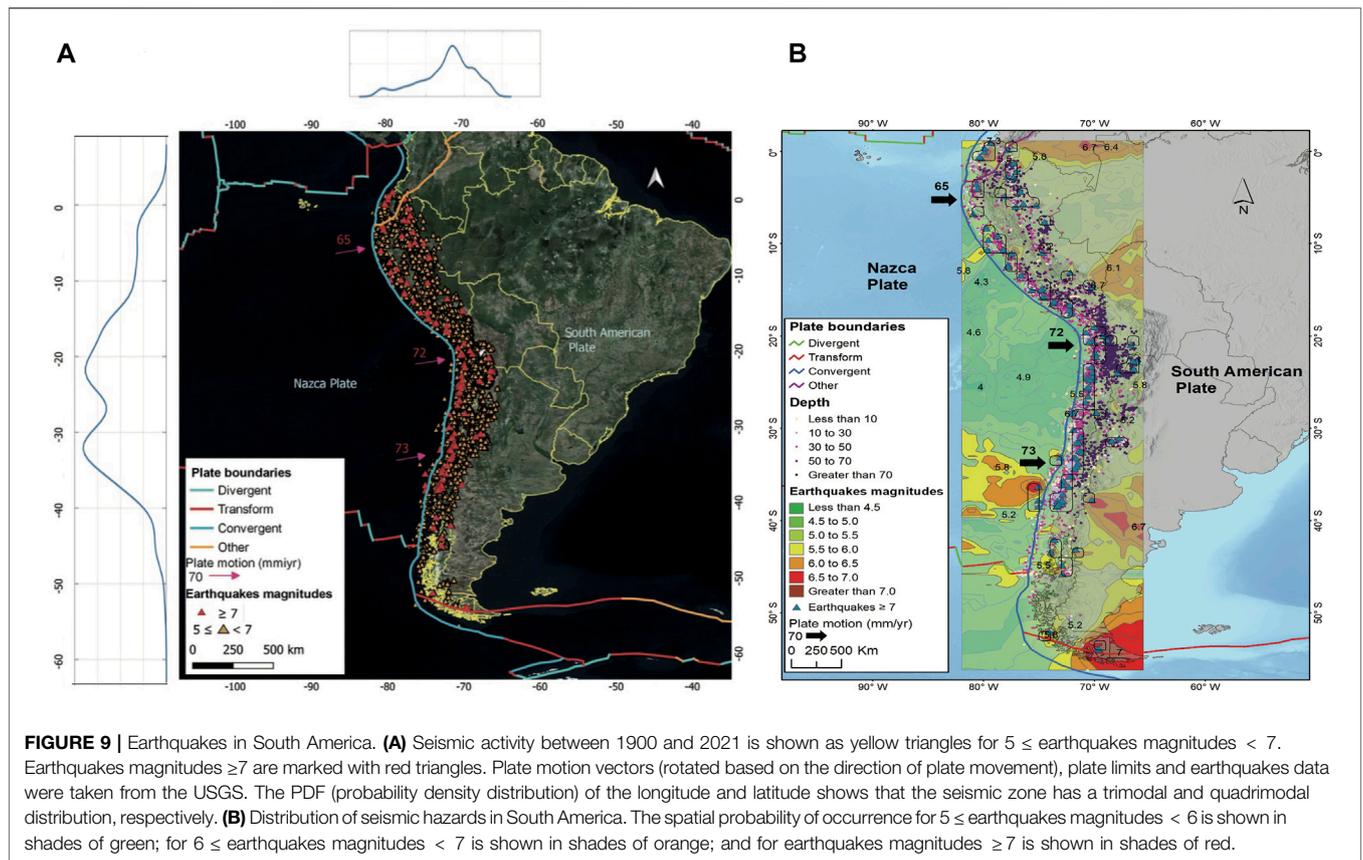
FIGURE 9 | Earthquakes in South America. **(A)** Seismic activity between 1900 and 2021 is shown as yellow triangles for 5 ≤ earthquakes magnitudes < 7. Earthquakes magnitudes ≥7 are marked with red triangles. Plate motion vectors (rotated based on the direction of plate movement), plate limits and earthquakes data were taken from the USGS. The PDF (probability density distribution) of the longitude and latitude shows that the seismic zone has a trimodal and quadrimodal distribution, respectively. **(B)** Distribution of seismic hazards in South America. The spatial probability of occurrence for 5 ≤ earthquakes magnitudes < 6 is shown in shades of green; for 6 ≤ earthquakes magnitudes < 7 is shown in shades of orange; and for earthquakes magnitudes ≥7 is shown in shades of red.

Figure 10A shows the wavelet analysis for the Japanese earthquakes of the first group. The global wavelet spectrum shows the periodicities of 1.2 ± 0.3, 2.4 ± 0.5, 4.1 ± 0.7, 10.9 ± 1.5, and 36.8 ± 5 years. **Figure 10B** shows the wavelet analysis for the second group of Japanese earthquakes $M \geq 8$. The global wavelet spectrum shows the periodicities of 1.1 ± 0.3, 1.7 ± 0.5, 5.1 ± 0.7, 23.1 ± 4.5, and 40 ± 7 years.

Figure 11A shows the probabilistic earthquake forecast (7 ≤ M < 8) model (blue line/shade) of 4.1-year that grouped historical earthquakes into twenty-nine clusters. We want to highlight a high seismic activity between clusters IX and X, XIV and XV, and XVI and XVII. Furthermore, there was no seismic activity in cluster XXIV. Seismic cluster number XXIX began in 2019 ± 1 and will end in 2022 ± 1. Therefore, it is possible that strong earthquakes may occur during 2022. During the preparation of this work, on 16 March 2022 an earthquake of magnitude 7.3 (37.702°N, 141.587°E) was recorded in Japan. So this strong earthquake verifies the accuracy of the Bayesian machine learning event classification and forecast for Japan in real time. According to this model, the next active seismic period would begin in 2024 ± 2 and end in 2029 ± 2 (cluster XXX). Here is another imminently testable forecast contributed by our Bayesian ML algorithm and analyses.

Figure 11B shows the probabilistic forecast of earthquakes $M \geq 8$. This model has a 40-years oscillation and groups historical earthquakes into three clusters. It is observed that

earthquakes occur in the positive phase and that the next period of high seismic activity would begin in 2035 ± 5 and end in 2050 ± 5 (cluster IV).

In **Figure 12A** shown the seismic activity in Japan between 1900 and 2021. The PDF of the longitude and latitude shows that the seismic zone has a trimodal and bimodal distribution with maxima at 131°, 142°, and 147°; and 37° and 43°, respectively. In **Figure 12B**, the spatial clustering of strong earthquakes in Japan is shown. It can be seen that these earthquakes occur preferentially in the subduction zone. In addition, the following strong earthquakes may occur in the cluster zones and yellow-orange areas where magnitude 6 to 7 earthquakes have historically occurred.

We would like to highlight that the earthquake of 16 March 2022 that was recorded in Japan with a magnitude of 7.3 (with epicenter at 37.702°N, 141.587°E) which was recorded off the coast of Japan's Fukushima prefecture. It is within the areas where, according to our spatial model shown in **Figures 12A,B** strong earthquake would be expected.

4.4 Southern China-Northern India

Figure 13 shows the result of wavelet analysis for strong earthquakes with $M \geq 7$ around the Southern China-Northern India zone. The global wavelet spectrum shows the periodicities of 1 ± 0.3, 1.8 ± 0.5, 3.4 ± 0.7, 6.9 ± 1.2, 8.6 ± 1.3, 13 ± 2, and 20.7 ± 5 years.
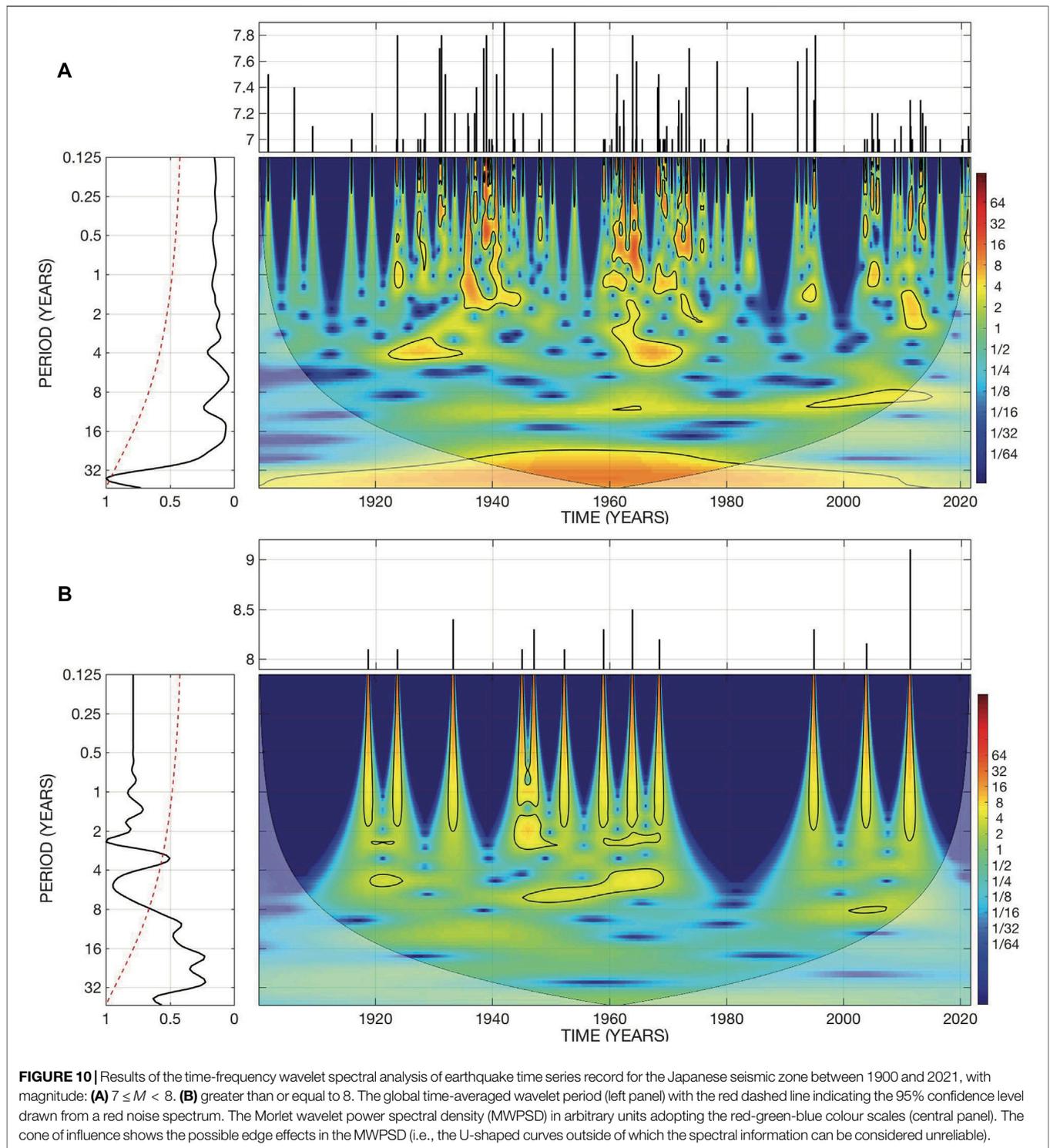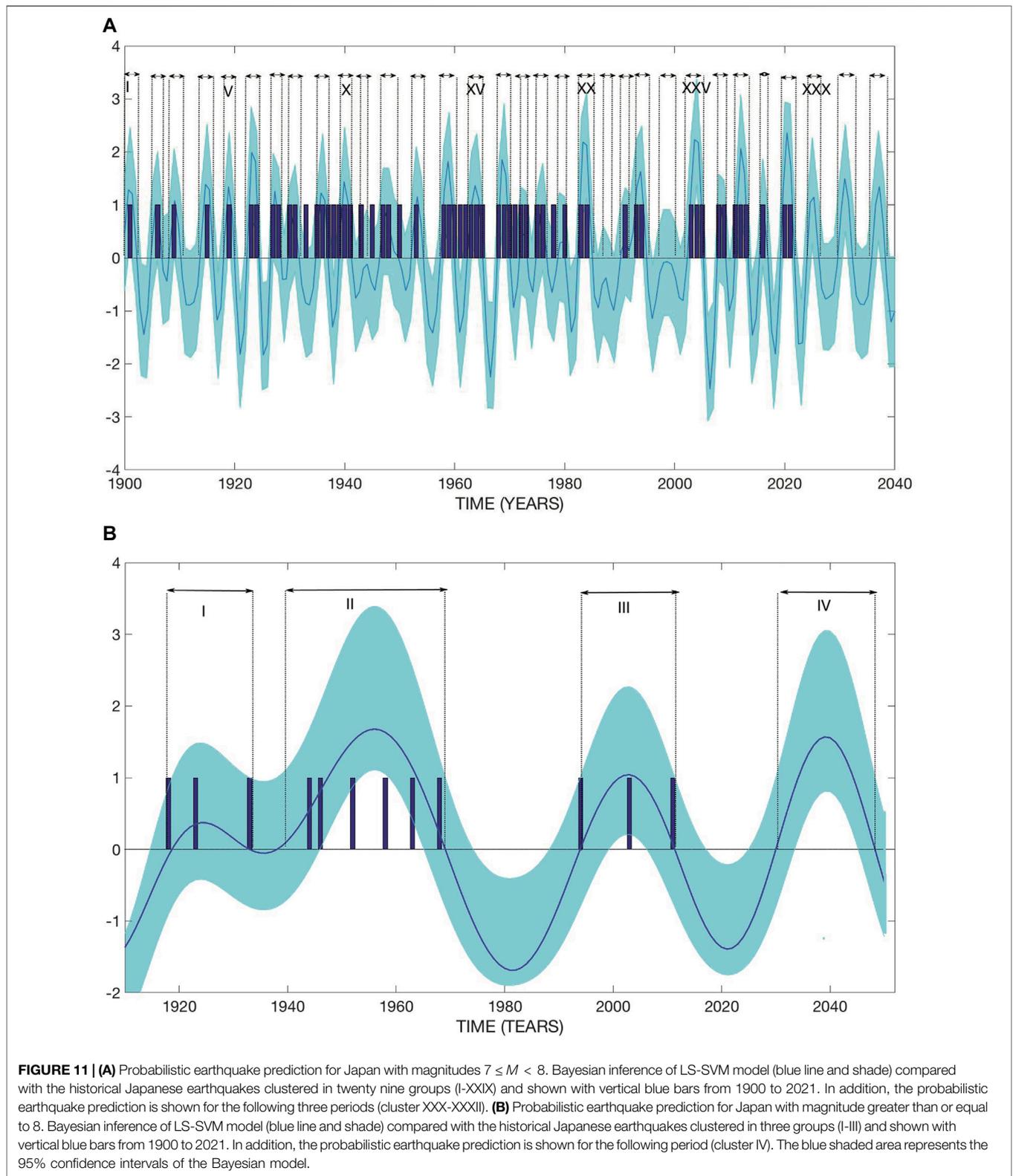
FIGURE 10 | Results of the time-frequency wavelet spectral analysis of earthquake time series record for the Japanese seismic zone between 1900 and 2021, with magnitude: **(A)** $7 \leq M < 8$. **(B)** greater than or equal to 8. The global time-averaged wavelet period (left panel) with the red dashed line indicating the 95% confidence level drawn from a red noise spectrum. The Morlet wavelet power spectral density (MWPSD) in arbitrary units adopting the red-green-blue colour scales (central panel). The cone of influence shows the possible edge effects in the MWPSD (i.e., the U-shaped curves outside of which the spectral information can be considered unreliable).
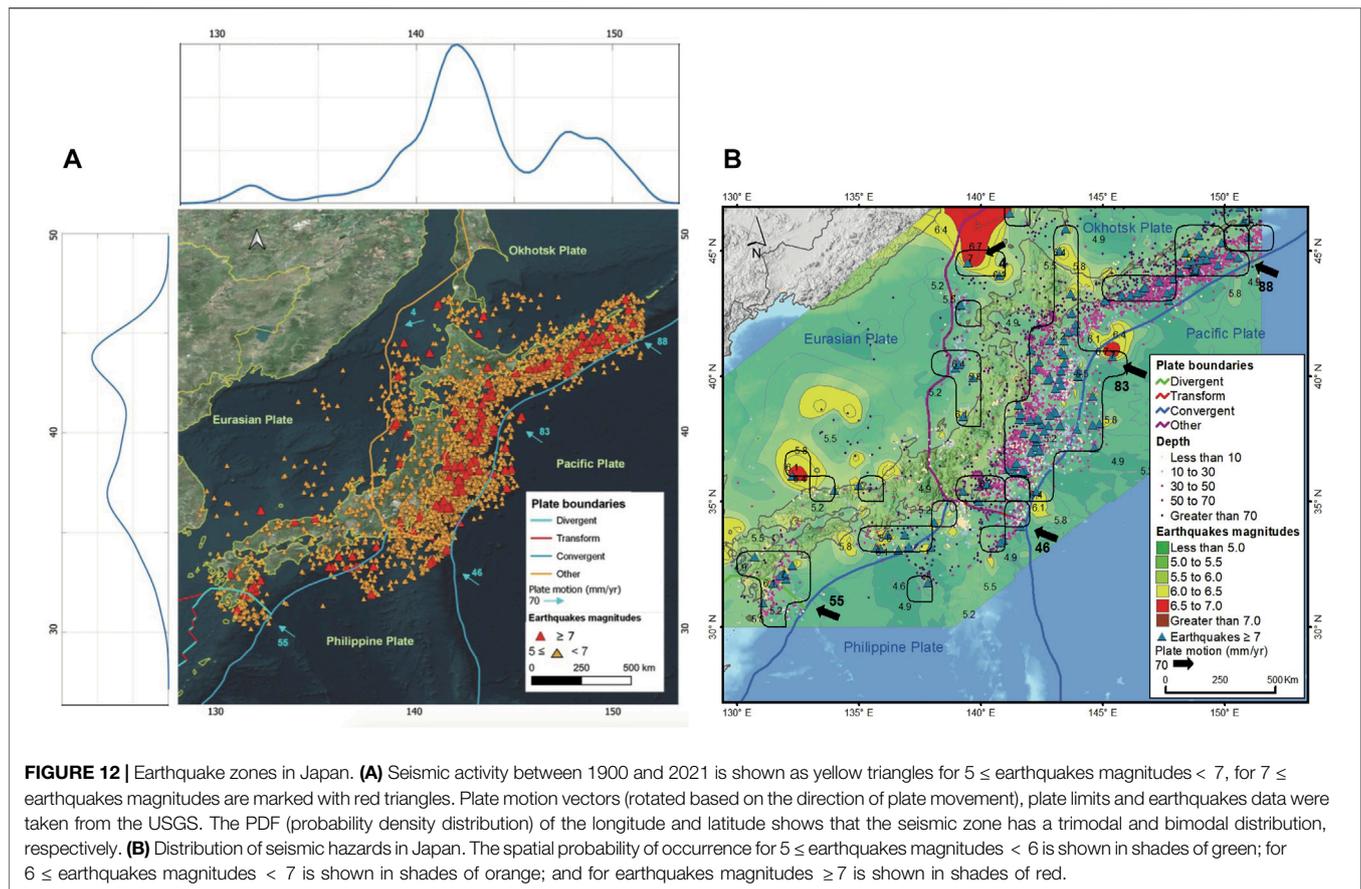
Figure 14 shows the probabilistic forecast of the Southern China-Northern India earthquakes with $M \geq 7$. This model has a 8.6-years oscillation and groups historical earthquakes into twelve clusters. It is observed that earthquakes occur in the positive phase and that the next period of high seismic activity would begin in 2022 ± 1 and end in 2028 ± 2 (cluster XIV).

In **Figure 15A** shown the seismic activity in Southern China-Northern India between 1900 and 2021. The PDF of the seismic activity longitude and latitude shows that the seismic zone has bimodal distributions with maxima at 77° and 94°; and 30° and 40°, respectively. In **Figure 15B**, the spatial clustering of strong earthquakes in Southern China-Northern India is shown. It can

FIGURE 11 | (A) Probabilistic earthquake prediction for Japan with magnitudes $7 \leq M < 8$. Bayesian inference of LS-SVM model (blue line and shade) compared with the historical Japanese earthquakes clustered in twenty nine groups (I-XXIX) and shown with vertical blue bars from 1900 to 2021. In addition, the probabilistic earthquake prediction is shown for the following three periods (cluster XXX-XXXII). **(B)** Probabilistic earthquake prediction for Japan with magnitude greater than or equal to 8. Bayesian inference of LS-SVM model (blue line and shade) compared with the historical Japanese earthquakes clustered in three groups (I-III) and shown with vertical blue bars from 1900 to 2021. In addition, the probabilistic earthquake prediction is shown for the following period (cluster IV). The blue shaded area represents the 95% confidence intervals of the Bayesian model.

be seen that these are intraplate earthquakes. In addition, the following strong earthquakes may occur in the cluster zones and yellow-orange areas where magnitude 6 to 7 earthquakes have

historically occurred and we would like to highlight that strong seismic activities are preferentially distributed around fault lines.

FIGURE 12 | Earthquake zones in Japan. **(A)** Seismic activity between 1900 and 2021 is shown as yellow triangles for 5 ≤ earthquakes magnitudes < 7, for 7 ≤ earthquakes magnitudes are marked with red triangles. Plate motion vectors (rotated based on the direction of plate movement), plate limits and earthquakes data were taken from the USGS. The PDF (probability density distribution) of the longitude and latitude shows that the seismic zone has a trimodal and bimodal distribution, respectively. **(B)** Distribution of seismic hazards in Japan. The spatial probability of occurrence for 5 ≤ earthquakes magnitudes < 6 is shown in shades of green; for 6 ≤ earthquakes magnitudes < 7 is shown in shades of orange; and for earthquakes magnitudes ≥ 7 is shown in shades of red.

5 DISCUSSION AND CONCLUDING REMARKS

Mechanisms conducting the plate motions and the Earth's geodynamics related to the triggering, the persistency and the potency of earthquakes are still not entirely clarified (Kanamori and Brodsky, 2004; Doglioni and Panza, 2015; Senapati et al., 2022, and several references cited in these papers) Although it is not the subject of diagnosis in the present work, the episodic elastic accumulation and release of energy produced by earthquakes depend on the multiple prevailing characteristics when the shear stress exceeds the internal friction between the sliding plates.
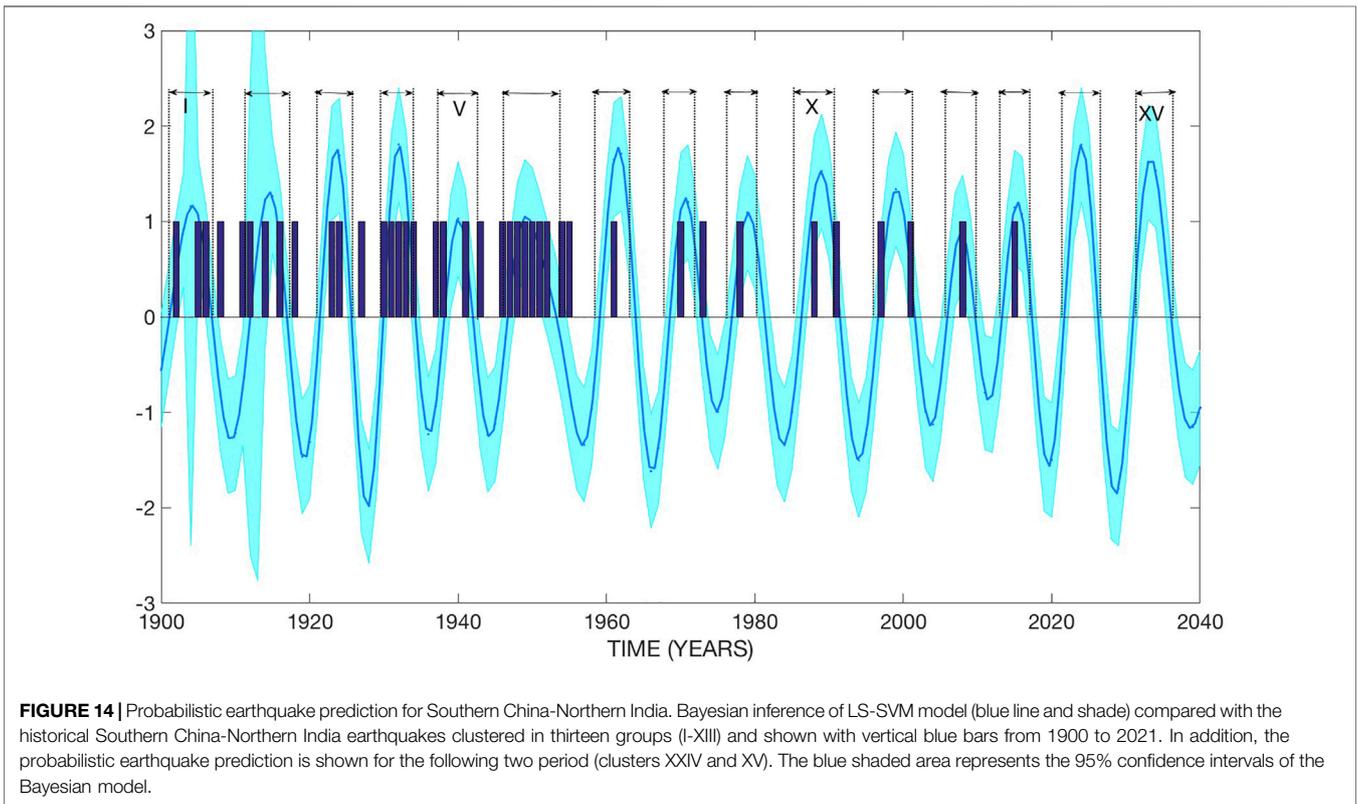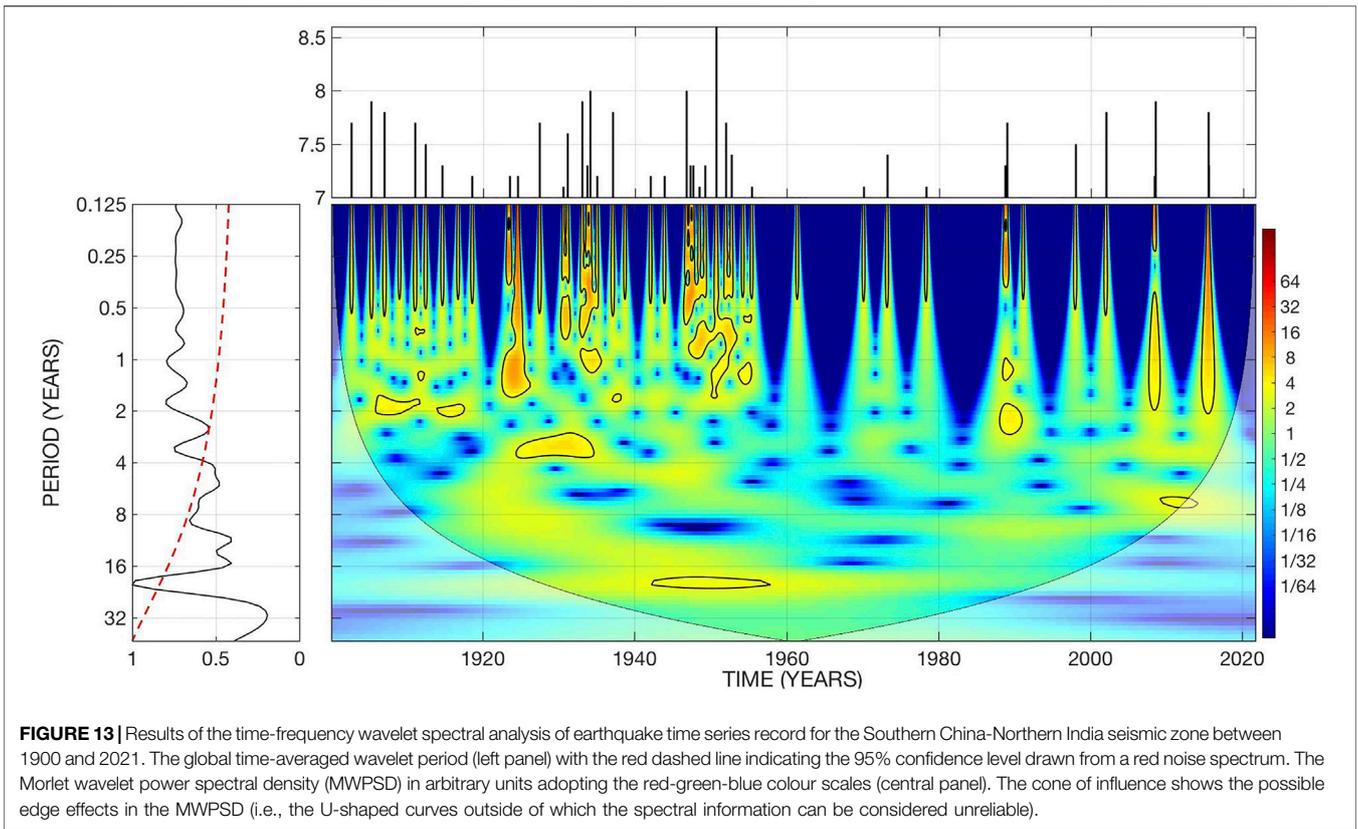
The relationships between stress fields and frictional responses (Lockner and Beeler, 2002) result from the rheological behavior of the materials that influence the viscosity of the affected plate contacts. The pressure by crust overloading and temperature conditions varies considerably from depth, affecting in consequence the origin, intensity and frequency of earthquakes. The velocity and the geometry of spatial orientations of the tectonic convergences also interact both in planar sense (magnitudes of the transcurrent components) and in depth (subduction dip angles).
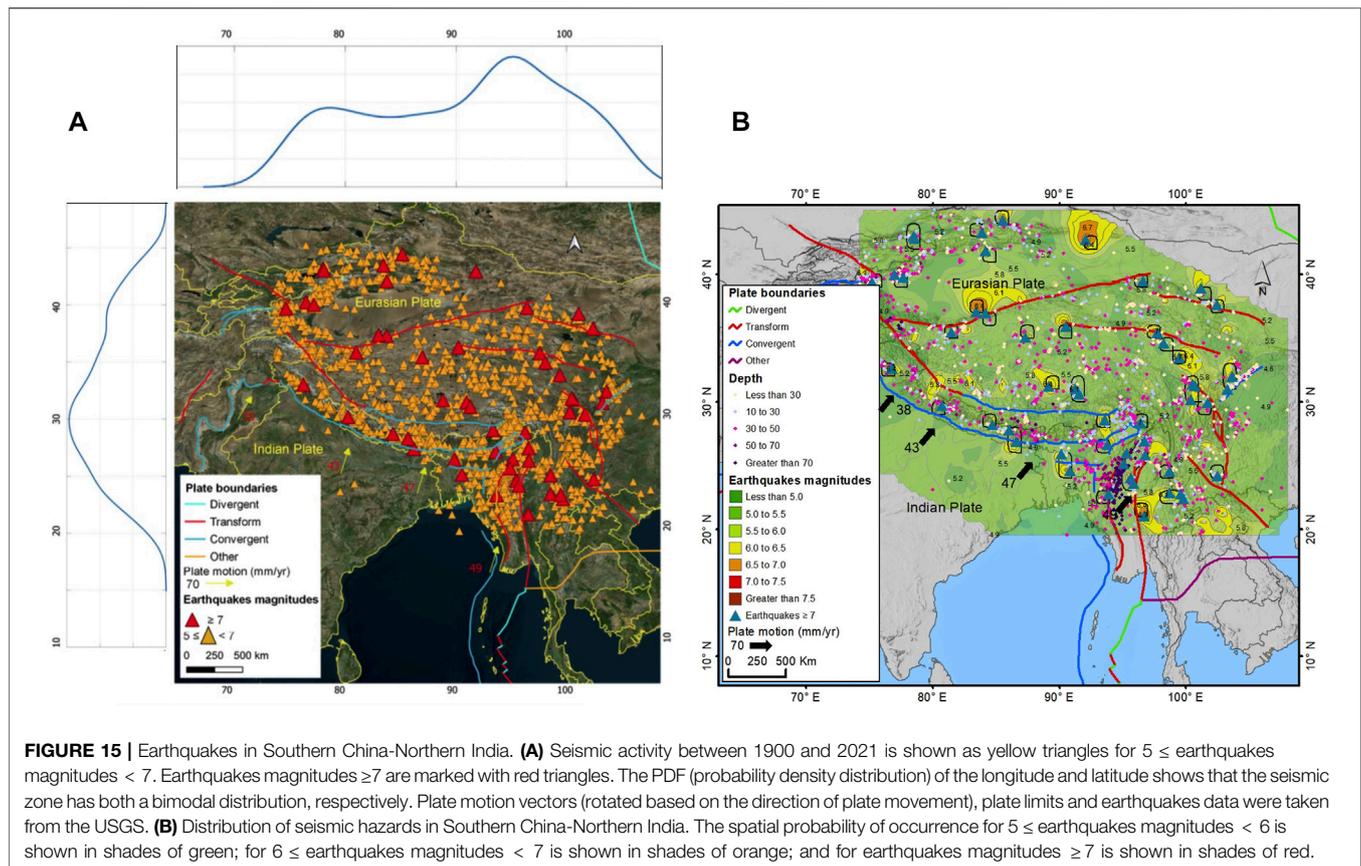
Other complementary surficial factors such as tidal friction of the Earth could also induce the earthquake's triggering (Zschau, 1986; Wilcock, 2001) due to the accumulation of energy for

centuries (Doglioni and Panza, 2015). Based on these multiple geophysical factors, we consider the main causes of the greater recurrency and intensity of earthquakes that are expected in long-lived tectonic contexts of near to orthogonal convergence and deeper subduction inclinations.

Kossobokov (2004) suggested that an apparent irregularity and a certain infrequency of earthquake occurrences may ultimately hinted that earthquakes are ultimately unpredictable phenomena rooted in stochastic processes beyond any deterministic rules or probabilities. But the appearance of stochastic nature or process may be due to the fact that abrupt or sporadic events such as earthquakes, and explosions, among others, must be analyzed in a different way than gradual processes such as temperature, atmospheric pressure, and other physical phenomena. Velasco Herrera et al. (2017) suggested that there is another type of natural phenomena that occurs only in a specific phase of a very particular oscillation such as strong earthquakes. The apparent irregularity of the strong or moderate earthquakes analyzed in this work could be observed in the heterogeneous distribution shown by the seismic events in the positive phase of each cluster obtained with Machine Learning.

The grouping of historical earthquakes makes it possible to find the periods of high and null seismic activity, but at the moment, it cannot answer the deeper questions posed by Kossobokov (2004): Why, where and when do earthquakes occur? In addition, Kossobokov (2004) raises a question that,

FIGURE 13 | Results of the time-frequency wavelet spectral analysis of earthquake time series record for the Southern China-Northern India seismic zone between 1900 and 2021. The global time-averaged wavelet period (left panel) with the red dashed line indicating the 95% confidence level drawn from a red noise spectrum. The Morlet wavelet power spectral density (MWPSD) in arbitrary units adopting the red-green-blue colour scales (central panel). The cone of influence shows the possible edge effects in the MWPSD (i.e., the U-shaped curves outside of which the spectral information can be considered unreliable).

FIGURE 14 | Probabilistic earthquake prediction for Southern China-Northern India. Bayesian inference of LS-SVM model (blue line and shade) compared with the historical Southern China-Northern India earthquakes clustered in thirteen groups (I-XIII) and shown with vertical blue bars from 1900 to 2021. In addition, the probabilistic earthquake prediction is shown for the following two period (clusters XXIV and XV). The blue shaded area represents the 95% confidence intervals of the Bayesian model.

FIGURE 15 | Earthquakes in Southern China-Northern India. **(A)** Seismic activity between 1900 and 2021 is shown as yellow triangles for 5 ≤ earthquakes magnitudes < 7. Earthquakes magnitudes ≥7 are marked with red triangles. The PDF (probability density distribution) of the longitude and latitude shows that the seismic zone has both a bimodal distribution, respectively. Plate motion vectors (rotated based on the direction of plate movement), plate limits and earthquakes data were taken from the USGS. **(B)** Distribution of seismic hazards in Southern China-Northern India. The spatial probability of occurrence for 5 ≤ earthquakes magnitudes < 6 is shown in shades of green; for 6 ≤ earthquakes magnitudes < 7 is shown in shades of orange; and for earthquakes magnitudes ≥7 is shown in shades of red.

to date, still does not have an answer. Are earthquakes predictable? Although a negative answer, according to Kossobokov (2004), is merely a guess. In addition, Kossobokov (2004) asks if there are intrinsic temporal and physical characteristics of earthquakes that can be used in other forecasting types to can be used to reduce human and economic losses? Different algorithms and models,[4,5] have been used to forecast earthquakes in different seismic zones (see, for example, Michael and Werner, 2018; Schorlemmer et al., 2018, for a full review). The medium-term prediction has been carried with the *M8* algorithm for strong earthquakes (Keilis-Borok and Kossobokov, 1990). In addition, the *M8* algorithm demonstrated that the largest earthquakes are not a random process (Kossobokov and Soloviev, 2021). Also, the strong earthquakes are forecasted with a limited precision in their ranges of time, space, and magnitude (Kossobokov and Soloviev, 2021). We propose a probabilistic algorithm for forecasting earthquakes (strong or moderate) that can be applied and implemented in different seismic zones. Different algorithms and methodologies for forecasting earthquakes require different tests. For example, we may note that the tests for the *M8* algorithm by V. Kossobokov and colleagues(see, Kossobokov and Soloviev, 2021) and for our Bayesian

Machine Learning models are very different (see the Methodology **Section 3.6.2**).

However, the proposed deterministic methodologies have not yield effective results nor robust conclusions. The study of seismic precursors has increased in recent years, but there is no reliable method to predict earthquakes over long time horizons/windows of multi-years or even decade to multidecades (Pulinets and Boyarchuk, 2005; Ouzounov et al., 2018; Pulinets and Ouzounov, 2018).

Predicting earthquakes must be one of the most significant challenges in modern science and technology. Earthquake prediction is necessary to minimize the enormous earthquake hazard risks in a seismic area. In particular the prediction of earthquakes is essential to minimize human tragedies and economic losses. The occurrences of an earthquake are complex and largely non-linear in character, which is why there is no deterministic model that can predict the exact location, magnitude, and time of an earthquake of any significant magnitudes.

There is currently a great debate in the scientific community about the origin of earthquakes. While some consider that it may not possible to predict earthquakes (Geller et al., 1997), others like us suggest that it could be a predictable phenomenon. Our point of view is informed by the seismic patterns found in the seismic zones analyzed adopting the methodology applied in this work.

Furthermore, we suggest that temporal forecasting of strong earthquakes should consist of forecasting magnitude widths/

---

[4]https://cseptesting.org/blog/
[5]https://www.scec.org/research

ranges rather than forecasting any "exact" magnitude. Moreover, from the geological point of view, the exact spatial forecast is not a correct concept since the geological faults are not point-like entities. In addition, the release of accumulated energy does not occur at a single point, nor is it instantaneous. Therefore, the energy released from an earthquake is an average of the inter-related cascade processes of rupture of the geological fault in a volumetric geographical area that occurs within a time interval. In other words, an earthquake does not occur at one point, nor is it a momentary event. We suggest that the problem of earthquake predictions should be viewed as a question of probability. One possible solution to earthquake forecasting is to calculate the predicted probabilities of future seismic cycles.

In our point of view, the key challenge in prediction of seismic activity today is to shift paradigms from deterministic forecasts to a probabilistic approach to predicting earthquakes with reliable estimates or quantifications of uncertainties. So earthquake prediction should be a multidisciplinary task that accounts for the recent advances in artificial intelligence and should be widely applied for earthquake forecasting (Beroza et al., 2021).

We have studied seismicity in North America, South America, Japan and Southern China and Northern India with Machine Learning by analyzing seismic patterns of variations for earthquakes with magnitude 7 or greater between 1900 and 2021. We then created a probabilistic earthquake prediction model for each seismic zone analyzed using the Bayesian Machine Learning method. Each model obtained groups the seismic of magnitude greater than 7 in clusters. Our result also partially explains the periods of earthquakes of magnitude 7 and the apparent seismic tranquillity for earthquakes of magnitude 7 or greater. We want to clarify that the periods where no earthquakes of magnitude 7 do not imply the total absence of seismic activity, indeed earthquakes of lesser magnitude must still occur during this period.

We suggest that if the dynamics of the tectonic plates have not changed in the last thousands of years, then no abrupt change in the tectonic movement would be expected, and therefore, the seismic pattern found should remain stationary, at least temporarily and spatially stable and unchanging, for the following next few decades.

We speculate that the seismic patterns found, that is, the periods of occurrence of earthquakes in each seismic zone analyzed could be interpreted as the period in which the energy accumulates and this energy releases through the rupture along faults and fractures near the plate tectonic boundaries.

As it can be seen in the results we obtained, each one of the analyzed areas has different characteristic frequencies for earthquakes of the analyzed magnitudes. This is probably due to the type of plate boundary, the lithological composition of the related plates and their consequent different friction coefficients, the geometry of the margin, the convergence angle of the plates, as well as the velocity of the approach. We leave this briefly expressed hypothesis awaiting further future studies.

As a highlight of our results, we note that the larger number of clusters are found and defined for the Himalayan collisional margin. We could speculate that this phenomenon is related

to the diverse lithological nature of the constituent materials of the two continental crustal plates that come into contact in a frontal collision with very different friction coefficients. The cases of the subduction convergence zones of Southwestern Mexico and South America seem to have similar frequencies, *sensu lato*, a fact that is not surprising because of the type of crust of the plates involved in the boundaries (oceanic crust versus continental crust).

To the contrary, convergence by subduction of two oceanic crust plates, as in the case of Japan, presents a notably higher frequency. This would be related to the physical properties of oceanic basalts and their response to imposed stresses.

In the Southwestern North American transform margin, we observed a very low frequency oscillation and modulation. In this case, the low frequency of earthquakes of magnitude greater than 7 could be related to the rectilinear geometry of the plate contact (plus its transcurrence) and the relation of this with the accumulation-release of elastic energy.

In addition, we would like to highlight the forecast for the earthquakes in the Southwestern United States of America since one of the biggest concerns is any major earthquake in the densely populated region of San Andres fault, such as the catastrophic event on 18 April 1906. This event is known as the San Francisco great earthquake. So after more than a hundred years, the "Big One" could occur according to our model between 2040 ± 5 and 2057 ± 5. Although this great earthquake may occur "tomorrow", we still have a little time to refine our forecasts of such strong earthquakes.

In summary, we have analyzed seismicity in North America, South America, Japan, and Southern China-Northern India for $M \geq 7$ earthquakes from 1900 to 2021. The primary seismic patterns found for $M \geq 7$ earthquakes in the analyzed seismic zones are 55, 3.7, 7.7, and 8.6 years, respectively, for the southwestern United States and northern Mexico, southwestern Mexico, South American, and Southern China-Northern India. In the Japanese zones, the primary seismic pattern for $7 \leq M < 8$ earthquake is 4.1 and 40 years for $M \geq 8$ earthquakes.

Our Machine Learning models show that there are periods where there are earthquakes magnitude ≥ 7 and periods without earthquakes with magnitude ≥ 7 in the analyzed seismic zones. In addition, our Machine Learning models predict a new seismically active phase for earthquakes magnitude ≥ 7 between 2040 ± 5 and 2057 ± 5, 2024 ± 1 and 2026 ± 1, 2026 ± 2 and 2031 ± 2, 2024 ± 2 and 2029 ± 2, and 2022 ± 1 and 2028 ± 2 for the five seismic zones in United States, Mexico, South America, Japan, and Southern China-Northern India, respectively. Finally, we note that our algorithms can be further applied to perform probabilistic forecasts in any seismic zone.

Our algorithm for analyzing strong earthquakes in extensive seismic areas can also be applied to smaller or specific seismic zones where moderate historical earthquakes with magnitudes between 5 and 7 occur, as is the case of the Parkfield section of the San Andreas fault (California, United States). Our analysis shows why a moderate earthquake could never occur in 1988 ± 5 as proposed by Bakun and Lindh (1985) and why the long-awaited characteristic Parkfield earthquake occurred in 2004. Furthermore, our Bayesian model of Machine Learning

adopting a periodicity of 35 years predicts that possible seismic events may occur between 2019 and 2031, with a high probability of event(s) around 2025 ± 2. The Parkfield section of the San Andreas fault is an excellent seismic laboratory for developing, testing, and demonstrating earthquake forecasts. In a few years, it will be possible to demonstrate whether our algorithm effectively forecasts strong and moderate earthquakes. We may note in anticipation that if some of our forecasts are not fulfilled in some of the analyzed seismic zones, it may not be the sole fault of Machine Learning algorithms. Instead the complex issues and problems may lie with our conjectures and the proposed models for each seismic zones, so we will have to carefully re-analyze again the seismic zone or zones where the forecast was not fulfilled. If all our forecasts for the next high season of earthquakes are fulfilled, then we must incorporate more elements such as the seismic precursors (see, for example, Pulinets and Boyarchuk, 2005; Ouzounov et al., 2018; Pulinets and Ouzounov, 2018, for a full review) of each zone analyzed in our models to give more accurate earthquake forecasts in order to provide earlier warnings and greater security to people living in these earthquake zones.

Spatial forecasting models Zechar and Jordan (2008) have been suggested, involving variables and metrics such as 1) Relative Intensity (RI) alarm function, 2) Pattern Informatics (PI), and 3) the United States Geological Survey National Seismic Hazard Map (NSHM). These models represent different assumptions about the spatial distribution of earthquakes. Concerning RI: the hypothesis is that future earthquakes are more likely to occur where seismicity is historically higher. On PI: the hypothesis is based on the fact that anomalous seismic activities indicate the locations of future earthquakes. Finally, on NSHM: the hypothesis suggests that future earthquakes will occur where previous earthquakes have occurred, and that some earthquakes may occur anywhere.

In our case, the possible zones where the following strong earthquakes in each seismic zone analyzed could occur have been essentially clustered and pre-determined according to the methodology described in **Section 3.1**. According to this spatial grouping, the fact that strong earthquakes occur probabilistically close to where strong earthquakes have historically occurred could indicate certain information about the tectonic plates. For example, precisely in those areas, there is greater fracturing compared to areas where strong earthquakes have not occurred. In addition, the probabilistic spatial clustering of **Figures 4**, **9**, **12**, **15** could show the areas with the highest probability of strong earthquakes. Therefore, it would be essential to analyze these highly probable areas of strong earthquakes with all the models and precursors that are currently known in order to minimize economic losses and human losses.

In conclusion, we believe that our results demonstrate that our methodology is a good alternative to traditional deterministic earthquake prediction. Thus, the problem of earthquake predictions should be considered as a question of probability. We believe the challenge in the study of seismic activity is to modify the forecast paradigm to a probabilistic earthquake prediction.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conceptualization, VV; Methodology, VV, ER, MO, LR-dlC, WS and EZ; Software, LR-dlC, EZ, and CV; Validation, VV, ER and MO; Formal Analysis, VV, WS and GV; Investigation, VV, WS, MO, ER and LA; Resources, GV; Data Curation, LA, LR-dlC and EZ; Writing—original draft preparation, VV, ER and MO; Writing—review and editing, VV, WS, ER and MO; Visualization, GV; Supervision, VV, ER, and MO; Project Administration, GV; Funding Acquisition, GV. All authors have read and agreed to the published version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ammirati, J.-B., Vargas, G., Rebolledo, S., Abrahami, R., Potin, B., Leyton, F., et al. (2019). The Crustal Seismicity of the Western Andean Thrust (Central Chile,

33°-34° S): Implications for Regional Tectonics and Seismic Hazard in the Santiago Area. *Bull. Seismol. Soc. Am.* 109 (5), 1985–1999. doi:10.1785/0120190082

Anagnostopoulos, G., Spyroglou, I., Rigas, A., Preka-Papadema, P., Mavromichalaki, H., and Kiosses, I. (2021). The Sun as a Significant Agent

Provoking Earthquakes. *Eur. Phys. J. Spec. Top.* 230, 287–333. doi:10.1140/epjst/e2020-000266-2

Assumpção, M., Dias, F. L., Zevallos, I., and Naliboff, J. B. (2016). Intraplate Stress Field in South america from Earthquake Focal Mechanisms. *J. S. Am. Earth Sci.* 71, 278–295. doi:10.1016/j.jsames.2016.07.005

Bakun, W. H., Aagaard, B., Dost, B., Ellsworth, W. L., Hardebeck, J. L., Harris, R. A., et al. (2005). Implications for Prediction and Hazard Assessment from the 2004 Parkfield Earthquake. *Nature* 437, 969–974. doi:10.1038/nature04067

Bakun, W. H., and Lindh, A. G. (1985). The Parkfield, California, Earthquake Prediction Experiment. *Science* 229, 619–624. doi:10.1126/science.229.4714.619

Bakun, W. H., and McEvilly, T. V. (1984). Recurrence Models and Parkfield, California, Earthquakes. *J. Geophys. Res.* 89, 3051–3058. doi:10.1029/jb089ib05p03051

Batakrushna, S., Bhaskar, K., and Shuanggen, J. (2022). Seismicity Modulation by External Stress Perturbations in Plate Boundary vs. Stable Plate Interior. *Geosci. Front.* 13, 101352. doi:10.1016/j.gsf.2022.101352

Bayes, T. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. *Philosophical Trans. R. Soc. Lond.* 53, 370–418.

Beroza, G. C., Segou, M., and Mostafa Mousavi, S. (2021). Machine Learning and Earthquake Forecasting-Next Steps. *Nat. Commun.* 12, 4761. doi:10.1038/s41467-021-24952-6

Bilham, R. (2019). Himalayan Earthquakes: a Review of Historical Seismicity and Early 21st Century Slip Potential. *Geol. Soc. Lond. Spec. Publ.* 483 (1), 423–482. doi:10.1144/sp483.16

Calais, E., Camelbeeck, T., Stein, S., Liu, M., and Craig, T. (2016). A New Paradigm for Large Earthquakes in Stable Continental Plate Interiors. *Geophys. Res. Lett.* 43, 10621–10637. doi:10.1002/2016gl070815

Carroll, J. D., Green, P. E., and Chaturvedi, A. (1997). *Mathematical Tools for Applied Multivariate Analysis*. Cambridge, MA, USA: Academic Press.

Castro, R. R., Stock, J. M., Hauksson, E., and Clayton, R. W. (2017). Active Tectonics in the Gulf of California and Seismicity (M > 3.0) for the Period 2002-2014. *Tectonophysics* 719-720, 4–16. doi:10.1016/j.tecto.2017.02.015

Dal Zilio, L., van Dinther, Y., Gerya, T., and Avouac, J. P. (2019). Bimodal Seismicity in the Himalaya Controlled by Fault Friction and Geometry. *Nat. Commun.* 10 (1), 48–11. doi:10.1038/s41467-018-07874-8

Dañobeitia, J., Bartolomé, R., Prada, M., Nuñez-Cornú, F., Córdoba, D., Bandy, W. L., et al. (2016). Crustal Architecture at the Collision Zone between Rivera and North American Plates at the Jalisco Block: Tsujal Project. *Pure Appl. Geophys.* 173, 3553–3573. doi:10.1007/s00024-016-1388-7

Davis, C., Keilis-Borok, V., Kossobokov, V., and Soloviev, A. (2012). Advance Prediction of the March 11, 2011 Great East japan Earthquake: A Missed Opportunity for Disaster Preparedness. *Int. J. Disaster Risk Reduct.* 1, 17–32. doi:10.1016/j.ijdrr.2012.03.001

Ding, Y. R., Zhao, H. B., and Li, Z. Y. (1998). A Method of Analyzing Incomplete Time Series with Application to Two Cataclysmic Variables. *Chin. Astronomy Astrophysics* 22, 235–242. doi:10.1016/s0275-1062(98)00032-0

Doglioni, C., and Panza, G. (2015). Polarized Plate Tectonics. *Adv. Geophys.* 56, 1. doi:10.1016/bs.agph.2014.12.001

Essam, Y., Kumar, P., Ahmed, A. N., Murti, M. A., and El-Shafie, A. (2021). Exploring the Reliability of Different Artificial Intelligence Techniques in Predicting Earthquake for malaysia. *Soil Dyn. Earthq. Eng.* 147, 106826. doi:10.1016/j.soildyn.2021.106826

Feynman, R. P., Leighton, R., and Sands, M. (2011b). *The Feynman Lectures on Physics, Volume 3: Quantum Mechanics*. Basic Books, A Member of the Perseus Books Group.

Feynman, R. P., Leighton, R., and Sands, M. (2011a). *The Feynman Lectures on Physics, Volume I: Mainly Mechanics, Radiation, and Heat*. Basic Books, A Member of the Perseus Books Group.

Frick, P., Baliunas, S. L., Galyagin, D., Sokoloff, D., and Soon, W. (1997). Wavelet Analysis of Stellar Chromospheric Activity Variations. *Astrophysical J.* 483, 426–434. doi:10.1086/304206

Frick, P., Grossmann, A., and Tchamitchian, P. (1998). Wavelet Analysis of Signals with Gaps. *J. Math. Phys.* 39, 4091–4107. doi:10.1063/1.532485

García, D., Singh, S. K., Herráiz, M., Ordaz, M., and Pacheco, J. F. (2005). Inslab Earthquakes of Central mexico: Peak Ground-Motion Parameters and Response Spectra. *Bull Seismol. Soc. Am.* 95, 2272–2282. doi:10.1785/0120050072

Geller, R. J., Jackson, D. D., Kagan, Y. Y., and Mulargia, F. (1997). Earthquakes Cannot Be Predicted. *Science* 275 (5306), 1616. doi:10.1126/science.275.5306.1616

Gelman, A., and Meng, X. L. (2005). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. Hoboken, NJ, USA: John Wiley & Sons.

Gilman, D. L., Fuglister, F. J., and Mitchell, J. M. (1963). On the Power Spectrum of "Red Noise". *J. Atmos. Sci.* 20, 182–184. doi:10.1175/1520-0469(1963)020<0182:otpson>2.0.co;2

Gitis, V., and Derendyaev, A. (2020). The Method of the Minimum Area of Alarm for Earthquake Magnitude Prediction. *Front. Earth Sci.* 11, 585317. doi:10.3389/feart.2020.585317

Hainzl, S., Kraft, T., Wassermann, J., Igel, H., and Schmedes, E. (2006). Evidence for Rainfall-Triggered Earthquake Activity. *Geophys. Res. Lett.* 33, L193003. doi:10.1029/2006gl027642

Hampel, A., Hetzel, R., and Densmore, A. L. (2007). Postglacial Slip-Rate Increase on the Teton Normal Fault, Northern Basin and Range Province, Caused by Melting of the Yellowstone Ice Cap and Deglaciation of the Teton Range? *Geol* 35 (12), 1107. doi:10.1130/g24093a.1

Heidbach, O., Rajabi, M., Cui, X., Fuchs, K., Müller, B., and Reinecker, J. (2016). The World Stress Map Database Release 2016: Crustal Stress Pattern across Scales. *Tectonophysics* 744, 484–498. doi:10.1016/j.tecto.2018.07.007

Heki, K. (2019). Snow Load and Seasonal Variation of Earthquake Occurrence in japan, Earth Planet. *Sci. Lett.* 46, 13730–13736.

Jain, R., Nayyar, A., Arora, S., and Gupta, A. (2021). A Comprehensive Analysis and Prediction of Earthquake Magnitude Based on Position and Depth Parameters Using Machine and Deep Learning Models. *Multimed. Tools Appl.* 80, 28419–28438. doi:10.1007/s11042-021-11001-z

Jara, J. M., López, M. G., Olmos, B. A., and Jara, M. (2015). Engineering Demand Functions for Rc Medium Length Span Bridges. *Bull. Earthq. Eng.* 13, 679–702. doi:10.1007/s10518-014-9604-2

Jopek, T. J., and Kaňuchová, Z. (2017). IAU Meteor Data Center-The Shower Database: A Status Report. *Planet. Space Sci.* 143, 3–6. doi:10.1016/j.pss.2016.11.003

Kanamori, H., and Brodsky, E. E. (2004). The Physics of Earthquakes. *Rep. Prog. Phys.* 67, 1429–1496. doi:10.1088/0034-4885/67/8/r03

Keilis-Borok, V. I., and Kossobokov, V. G. (1990). Premonitory Activation of Earthquake Flow: Algorithm M8. *Phys. Earth Planet. Interiors* 61, 73–83. doi:10.1016/0031-9201(90)90096-g

Kossobokov, V. G. (2004). Earthquake Prediction: Basics, Achievements, Perspectives. *Acta Geod. Geophys. Hung.* 39, 205–221. doi:10.1556/ageod.39.2004.2-3.6

Kossobokov, V. G., and Soloviev, A. A. (2021). Testing Earthquake Prediction Algorithms. *J. Geol. Soc. India* 97, 1514–1519. doi:10.1007/s12594-021-1907-8

Kossobokov, V., Peresan, A., and Panza, G. (2015). On Operational Earthquake Forecast and Prediction Problems. *Seismol. Res. Lett.* 96, 287–290. doi:10.1785/0220140202

Kossobokov, V., and Soloviev, A. (2018). Pattern Recognition in Problems of Seismic Hazard Assessment. *Chebyshevskii Sb.* 19, 53–88.

Kossobokov, V., and Soloviev, A. (2008). Prediction of Extreme Events: Fundamentals and Prerequisites of Verification. *Russ. J. Earth Sci.* 10, ES2005. doi:10.2205/2007es000251

Kostoglodov, V., and Bandy, W. (1995). Seismotectonic Constraints on the Convergence Rate between the Rivera and North American Plates. *J. Geophys. Res.* 100 (B9), 17977–17989. doi:10.1029/95jb01484

Lambert, S., and Sottili, G. (2019). Is There an Influence of the Pole Tide on Volcanism? Insights from Mount Etna Recent Activity. *Geophys. Res. Lett.* 46, 13730–13736. doi:10.1029/2019gl085525

Landau, L., and Lifshitz, E. (1988a). *Course of Theoreticcal Physics: Mechanics, Volume 1*. Moskow: Nauka.

Landau, L., and Lifshitz, E. (1988b). *Course of Theoreticcal Physics: Quantum Mechanics: Non-relativistic Theory, Volume 3*. Moskow: Nauka.

Lin, A., Chen, P., Satsukawa, T., Sado, K., Takahashi, N., and Hirata, S. (2017). Millennium Recurrence Interval of Morphogenic Earthquakes on the Seismogenic Fault Zone that Triggered the 2016 Mw 7.1 Kumamoto Earthquake, Southwest Japan. *Bull. Seismol. Soc. Am.* 107 (6), 2687–2702. doi:10.1785/0120170149

Lin, A. (2018). Late Pleistocene-Holocene Activity and Paleoseismicity of the Nojima Fault in the Northern Awaji Island, Southwest japan. *Tectonophysics* 747-748, 402–415. doi:10.1016/j.tecto.2018.10.009

Liu, C., Linde, A. T., and Sacks, I. S. (2009). Slow Earthquakes Triggered by Typhoons. *Nature* 459, 833–836. doi:10.1038/nature08042

Lockner, D., and Beeler, N. (2002). *Rock Failure and Earthquakes*. Cambridge, MA, USA: International Handbook of Earthquakes & Engineering Seismology. Academic Press.

Maoz, D., Sternberg, A., and Leibowitz, E. M. E. (1997). *Astronomical Time Series*. Berlin, Germany: Springer.

Mendoza, B., Velasco, V. M., and Valdés-Galicia, J. F. (2006). Mid-term Periodicities in the Solar Magnetic Flux. *Sol. Phys.* 233, 319–330. doi:10.1007/s11207-006-4122-2

Michael, A. J., and Werner, M. J. (2018). Preface to the Focus Section on the Collaboratory for the Study of Earthquake Predictability (Csep): New Results and Future Directions. *Seismol. Res. Lett.* 89, 1226–1228. doi:10.1785/0220180161

Michel, S., Jolivet, R., Rollins, C., Jara, J., and Zilio, L. D. (2021). Seismogenic Potential of the Main Himalayan Thrust Constrained by Coupling Segmentation and Earthquake Scaling. *Geophys. Res. Lett.* 2021, e2021GL093106. doi:10.1029/2021gl093106

Moradia, S., Fadaeib, H., Adl, A., and Ataellahid, S. (2020). "Interpolation Methods in Identification Seismic Space Risk of Earthquake Case Study: 50km Radius of Sarpol-E Zahab City, Kermanshah Province," in International Congress on Engineering, Technology & Innovation eti'20, 1–21.

Murray, V. (2021). Hazard Information Profiles: Supplement to Undrr-Isc Hazard Definition & Classification Review: Technical Report. *U. N. Office Disaster Risk Reduct.* 144, 1–827.

Novelo-Casanova, D., Suárez, G., Cabral-Cano, E., Fernández-Torres, E. A., Fuentes-Mariles, O. A., and Havazli, E. (2013). The Risk Atlas of mexico City, mexico: a Tool for Decision-Making and Disaster Prevention. *Nat. Hazards* 111, 411–437. doi:10.1007/s11069-021-05059-z

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *J. Am. Stat. Assoc.* 83 (401), 9–27. doi:10.1080/01621459.1988.10478560

Ouzounov, D., Pulinets, S., Hattori, K., and Taylor, P. E. (2018). *Pre-Earthquake Processes: A Multi-Disciplinary Approach to Earthquake Prediction Studies*. Hoboken, NJ, USA: American Geophysical Union and John Wiley & Sons, Inc.

Panda, D., Kundu, B., Gahalaut, V. K., Bürgmann, R., Jha, B., Asaithambi, R., et al. (2020). Reply to "A Warning against Over-interpretation of Seasonal Signals Measured by the Global Navigation Satellite System". *Nat. Commun.* 11 (1), 1–2. doi:10.1038/s41467-020-15103-4

Pardo, M., and Suárez, G. (1995). Shape of the Subducted Rivera and Cocos Plates in Southern mexico: Seismic and Tectonic Implications. *J. Geophys. Res.* 100 (B7), 12357–12373. doi:10.1029/95jb00919

Pulinets, S., and Boyarchuk, K. (2005). *Ionospheric Precursors of Earthquakes*. Berlin: Springer-Verlag Berlin Heidelberg.

Pulinets, S., and Ouzounov, D. (2018). *The Possibility of Earthquake Forecasting: Learning from Nature*. Bristol, England: Institute of Physics Books, IOP Publishing.

Ramírez-Rojas, A., Di G. Sigalotti, L., Flores Márquez, E., and Rendón, O. (2019). *Time Series Analysis in Seismology*. Amsterdam, Netherlands: Elsevier.

Rivas, C., Ortiz, G., Alvarado, P., Podesta, M., and Martin, A. (2019). Modern Crustal Seismicity in the Northern Andean Precordillera, argentina. *Tectonophysics* 762 (4), 144–158. doi:10.1016/j.tecto.2019.04.019

Rossello, E. A., Heit, B., and Bianchi, M. (2020). Shallow Intraplate Seismicity in the Buenos Aires Province (argentina) and Surrounding Areas: Is it Related to the Quilmes Trough? *Bol. Geol.* 42 (2), 31–48. doi:10.18273/revbol.v42n2-2020002

Salcedo, G. E., Porto, R. F., and Morettin, P. A. (2012). Comparing Non-stationary and Irregularly Spaced Time Series. *Comput. Statistics Data Analysis* 56, 3921–3934. doi:10.1016/j.csda.2012.05.022

Sawires, R., Santoyo, M. A., Peláez, J. A., and Henares, J. (2021). Western Mexico Seismic Source Model for the Seismic Hazard Assessment of the Jalisco-Colima-Michoacán Region. *Nat. Hazards* 105, 2819–2867. doi:10.1007/s11069-020-04426-6

Scargle, J. D., Norris, J. P., Jackson, B., and Chiang, J. (2013). Studies in Astronomical Time Series Analysis. Vi. Bayesian Block Representations. *ApJ* 764, 167. doi:10.1088/0004-637x/764/2/167

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., et al. (2018). The Collaboratory for the Study of Earthquake Predictability: Achievements and Priorities. *Seismol. Res. Lett.* 89, 1305–1313. doi:10.1785/0220180053

Senapati, B., Kundu, B., and Jin, S. (2022). Seismicity Modulation by External Stress Perturbations in Plate Boundary vs. Stable Plate Interior. *Geosci. Front.* 13, 101352. doi:10.1016/j.gsf.2022.101352

Shcherbakov, R., Zhuang, J., Zöller, G., and Ogata, Y. (2019). Forecasting the Magnitude of the Largest Expected Earthquake. *Nat. Commun.* 10, 4051. doi:10.1038/s41467-019-11958-4

Shen, Z.-K., Wang, Q., Burgmann, R., Wan, Y., and Ning, J. (2005). Pole-tide Modulation of Slow Slip Events at Circum-Pacific Subduction Zones. *Bull. Seismol. Soc. Am.* 95 (5), 2009–2015. doi:10.1785/0120050020

Singh, S. K., and Pardo, M. (1993). Geometry of the Benioff Zone and State of Stress in the Overriding Plate in Central mexico. *Geophys. Res. Lett.* 20, 1483–1486. doi:10.1029/93gl01310

Soon, W., Dutta, K., Legates, D. R., Velasco, V., and Zhang, W. (2011). Variation in Surface Air Temperature of china during the 20th Century. *J. Atmos. Solar-Terrestrial Phys.* 73, 2331–2344. doi:10.1016/j.jastp.2011.07.007

Soon, W., Velasco Herrera, V. M., Cionco, R. G., Qiu, S., Baliunas, S., Egeland, R., et al. (2019). Covariations of Chromospheric and Photometric Variability of the Young Sun Analogue HD 30495: Evidence for and Interpretation of Mid-term Periodicities. *MNRAS* 483, 2748–2757. doi:10.1093/mnras/sty3290

Sturges, W. (1983). On Interpolating Gappy Records for Time-Series Analysis. *J. Geophys. Res.* 88, 9736–9740. doi:10.1029/jc088ic14p09736

Suárez, G., Monfret, T., Wittlinger, G., and David, C. (1990). Geometry of Subduction and Depth of the Seismogenic Zone in the Guerrero Gap. *Nature* 345, 336–338.

Suykens, J., Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. (2005). *Least Squares Support Vector Machines*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Tapponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R., and Cobbold, P. (1982). Propagating Extrusion Tectonics in Asia: New Insights from Simple Experiments with Plasticine. *Geol* 10 (12), 611–616. doi:10.1130/0091-7613(1982)10<611:petian>2.0.co;2

Teves-Costa, P., Batlló, J., Matias, L., Catita, C., Jiménez, M. J., and García-Fernández, M. (2019). Maximum Intensity Maps (Mim) for portugal Mainland. *J. Seismol.* 23 (3), 417–440. doi:10.1007/s10950-019-09814-5

Tiwari, D. K., Jha, B., Kundu, B., Gahalaut, V. K., and Vissa, N. K. (2021). Groundwater Extraction-Induced Seismicity Around Delhi Region, India. *Sci. Rep.* 11, 10097. doi:10.1038/s41598-021-89527-3

Torrence, C., and Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.* 79, 61–78. doi:10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2

Türker, T., and Bayrak, Y. (2018). "Creating of Probability Maps of Earthquake Occurrences Using Kriging Method with the Geographic Information Systems (Gis): Estimates for 3 Section of the Nafz (Western, Central, Eastern)-Part 2," in International Conference on Advanced Technologies, Computer Engineering and Science ICATCES'18), 547–549.

Uyeda, S. (2013). On Earthquake Prediction in japan. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 89 (9), 391–400. doi:10.2183/pjab.89.391

Velasco Herrera, V. M., Soon, W., Knoska, S., Perez-Peraza, J. A., Cionco, R. G., Kudryavtsev, S. M., et al. (2022c). The New Composite Solar Flare Index from Solar Cycle 17 to Cycle 24 (1937-2020). *Sol. Phys.* in Press.

Velasco Herrera, V. M., Mendoza, B., and Velasco Herrera, G. (2015). Reconstruction and Prediction of the Total Solar Irradiance: From the Medieval Warm Period to the 21st Century. *New Astron.* 34, 221–233. doi:10.1016/j.newast.2014.07.009

Velasco Herrera, V. M., Soon, W., Velasco Herrera, G., Traversi, R., and Horiuchi, K. (2017). Generalization of the Cross-Wavelet Function. *New Astron.* 56, 86–93. doi:10.1016/j.newast.2017.04.012

Velasco Herrera, V., Soon, W., and Legates, D. (2021). Does Machine Learning Reconstruct Missing Sunspots and Forecast a New Solar Minimum? *Adv. Space Res.* 68, 1485–1501.

Velasco Herrera, V., Soon, W., Legates, D., Hoyt, D., and Muraközy, J. (2022a). Group Sunspot Numbers: A New Reconstruction of Sunspot Activity

Variations from Historical Sunspot Records Using Algorithms from Machine Learning. *Sol. Phys.* 297, 1485–1501. doi:10.1007/s11207-021-01926-x

Velasco Herrera, V., Soon, W., Pérez-Moreno, C., Velasco Herrera, G., Martell-Dubois, R., Rosique-de la Cruz, L., et al. (2022b). Past and Future of Wildfires in Northern Hemisphere's Boreal Forests. *For. Ecol. Manag* 504, 119859.

Wigner, E. (1967). *Symmetries and Reflections*. Bloomington, IN, USA: Indiana University Press.

Wilcock, W. S. D. (2001). Tidal triggering of microearthquakes on the juan de fuca ridge. *Geophys. Res. Lett.* 28 (20), 3999–4002. doi:10.1029/2001gl013370

Yousefzadeh, M., Hosseini, S. A., and Farnaghi, M. (2021). Spatiotemporally Explicit Earthquake Prediction Using Deep Neural Network. *Soil Dyn. Earthq. Eng.* 144, 106663. doi:10.1016/j.soildyn.2021.106663

Zechar, J. D., and Jordan, T. H. (2008). Testing Alarm-Based Earthquake Predictions. *Geophys. J. Int.* 172, 715–724. doi:10.1111/j.1365-246x.2007.03676.x

Zschau, J. (1986). *Tidal Friction in the Solid Earth: Constrains from the Chandler Wobble Period*. Greenbelt, MD: Space geodesy and geodynamics.