NATIONAL UNIVERSITY OF ENGINEERING

COLLEGE OF ECONOMICS AND STATISTICAL ENGINEERING AND SOCIAL SCIENCES

STATISTICAL ENGINEERING PROGRAM

EF012 PROJECT WORKSHOP

Design of a Stochastic Processing System for the Analysis and Identification of Seismic Patterns in Peru Regions in the Period 2008-2018 Using a Semi-Markov Model

by:

Aldair Stefano CRUZ RAMIREZ
Jesus Francisco GARCÍA LOAYZA

Professor

Carlos RISCO FRANCO

LIMA – PERU

2021

Abstract

In recent years, the country has found itself in seismic silence, after the earthquake that occurred in Ica in August 2007, no earthquake with the same or similar destructive characteristics has been reported. This seismic silence brings with it harmful consequences in the future since the earth is not releasing energy and it is accumulating, which could cause a large earthquake to occur at any time and cause serious damage to the territory. Faced with the possible occurrence of a seism, be it of great magnitude or small magnitude but with high intensity, the national government as well as each regional government has been preparing the National Plan for the Prevention and Reduction of Disaster Risks from time to time, a document in which raises the main actions to be taken for the prevention of natural disasters and thus have better prepared its population.

This research has been developed in order to study the behavior patterns of earthquakes in the departments of Peru and thus support the relevant authorities in the development of prevention plans since they would have a finer view of the behavior of seismic events. In order to develop the objective, first a clustering by latitude and longitude will be carried out to be able to separate the earthquakes by zones, followed by the application of a Semi Markov model to the data of earthquakes with a magnitude greater than 5.5 MW from 2008 to 2018 based on the previously obtained clusters. For the purposes of applying the model, the earthquakes were grouped according to their magnitude in 3 states. The analysis presented is being carried out with the information provided by the Geophysical Institute of Peru (IGP).

The results are first shown at the level of the whole of Peru, thus showing the behavior of the 3 states of magnitude defined from 2008 to 2018, concluding that there are high levels of probability of seismic events which are aligned with what is stated in reports and previous studies. Now, applying the semi-Markov model to each of the areas resulting from the clustering, it is evident that the southern area of the country is the most prone to seismic occurrence due to the higher probabilities of change of state obtained. These results are highly aligned with reality since, as is known, the southern regions of the coast are the ones that usually present the highest seismic activity

General Index

Chapter 1

Introduction

1.1. Problem Statement

The Peruvian population is aware that a seismic movement can occur at any time, as it has been happening over the years. This is because in Peru there have always been seismic because it is located on the Pacific Ring of Fire, an area that concentrates around 85% of the seismic activity in the world and is also shared by the countries of Central America, Chile, Alaska, Japan, among others.

The geographical location is not the only reason for the notable seismic activity that occurs in Peru, but this is also due to the fact that the Peruvian coast is affected by the crossing of two tectonic plates: the Nazca and the South American, causing, in the moment of friction, the seismic phenomenon. This is one of the main reasons why in Peru there are constant drills to prevent any disaster that may occur.

In recent years, many of the seismic that have occurred have been of low magnitude, which does not rule out that a large earthquake may occur in the near future. For all the aforementioned, it can be observed that there is a need to know and identify in a more adequate way the behavior patterns of seismic activity in the different departments of Peru, since each territory has limited resources. Therefore, you need to properly manage those resources.

In the 1950s, the Semi-Markov processes were introduced to the literature by [Levy, 1954], with the purpose of generalizing the classic Markov processes. Practical applications of this type of process can be found in [Janssen and Manca, 2006].

[Masala, 2012] applies Semi-Markov processes to be able to estimate seismic occurrences according to the magnitude level they present, while [Sadeghian, 2012] applies Semi-Markov models to estimate seismic occurrence according to geographic location. The Semi-Markov process applied in earthquakes is a process that assumes the sequence of earthquakes as a markov process and also allows taking into consideration the more realistic assumption of the dependence of events in space and time. In this type of models, groups of behavior of the magnitude levels are usually generated.

1.2. Problem formulation

1.2.1. Practical motivations

How does the Semi-Markov model determine the pattern of seismic behavior of the departments of Peru?

1.2.2. Practical motivations

- According to the Semi-Markov model, which regions or zones have the highest seismic occurrence?
- According to the Semi-Markov model, which regions or zones have the highest level of seismic magnitude?
- According to the Semi-Markov model, what is the magnitude level with the highest probability of occurrence in the next periods in the Peruvian regions between 2008 and 2018?

1.3. Objective

1.3.1. General Objective

- Identify and study the behavior patterns of the seismic occurrence and magnitude in the territory of Peru between 2008 and 2018 using a Semi-Markov model.

1.3.2. Specific Objectives

- Identify the regions or zones with the highest seismic occurrence between 2008 and 2018 using the Semi-Markov model.
- Identify the regions or areas with the highest level of seismic magnitude between 2008 and 2018 using the semi-Markov model.

- Identify the magnitude level most likely to occur in the next periods in the regions of Peru between 2008 and 2018.

1.4. Research Objectives

1.4.1. General hypothesis

H_0: The semi-Markov model applied to studies on seismic events contributes to the identification of behavior patterns.

H_1: The semi-Markov model applied to studies on seismic events does not contribute to the identification of behavior patterns.

1.4.2. Specific hypothesis

For seismic occurrence in the regions:

H_0: According to the Semi-Markov model, the seismic occurrence in all regions or zones of Peru occurs at the same level.

H_1: According to the Semi-Markov model, there is at least one region that has a higher seismic occurrence compared to the other regions or zones of Peru.

For the magnitude level in the regions:

H_0: According to the Semi-Markov model, the magnitude in all regions or zones of Peru occurs at the same level.

H_1: According to the Semi-Markov model, there is at least one region that has a higher magnitude level compared to the other regions or zones of Peru.

For probability of occurrence in the next periods in the regions:

H_0: According to the Semi-Markov model, in the regions of Peru the magnitude levels have the same propensity of occurrence regardless of the period of time they are found.

H_1: According to the Semi-Markov model, in the regions of Peru there is a magnitude level that has a higher probability of occurrence compared to the other levels.

1.5. Research Justification

1.5.1. Practical motivations

Given the latent danger of the seismic occurrence in the country, thanks to the geographical location in which it is located, it is considered of utmost importance, both for the population and for the relevant authorities, to know the patterns that the seismic occurrence may follow. earthquakes in a differentiated way by department since by carrying out a correct analysis of these, disaster risk prevention and reduction plans for each of the regions or departments can be developed in a much more efficient way.

1.5.2. Methodological motivations

The purpose of this research work is to carry out a descriptive-explanatory study, since it seeks to describe the seismic occurrence throughout the Peruvian territory (including maritime territory) in a differentiated way based on the results of a previous clustering by Latitude and Length of seismic events.

1.5.3. Theoretical motivations

For the identification of the behavior seismic patterns, hidden semi-Markov models (HSMM) are usually applied, since these allow any arbitrary distribution for the residence times. These models assume that a sequence of seismic events is seen as a Markov process, in addition the time elapsed between two consecutive events is modeled as a general Weibull distribution.

Chapter 2

Theoretical Framework

2.1. Background

[Baisch et al., 2008] in their research “Seismic Cluster: What Can We Learn from Waveform Similarity?”, Uses two-dimensional synthetic wave field simulations in lateral heterogeneous media to investigate how the similarity of the hypocenter waveform changes. very close with the separation between events. That is, they seek to group seismic events based on the wavelength that is generated within a geographical area, study the so-called repetitive earthquakes, which in the geophysical literature are so called those earthquakes that occur in the same place, but at different times while sharing roughly the same origin process.

[Castillo Aedo, 1994] in their work "Seismic Danger in Peru" they explain the different factors that affect seismicity within the Peruvian territory, such as: The subduction of the Nazca plate under the South American continent and the changes that occur within the earth's crust as a result of morphology and iteration achieved by the Andean apparatus. In addition, they sought to predict probabilistically the maximum accelerations that could occur at any point in the country, using acceleration attenuation laws and correlating seismicity and tectonics to determine seismogenic sources and their respective seismological parameters. They also point out that the most risky areas are on the coast and this risk decreases as one moves east.

[Novianti et al., 2017] in their research “K-Means cluster analysis earthquake”, uses data from tectonic earthquakes in the Bengkulu province and surrounding areas to be able to more efficiently describe the characteristics of the earthquakes that occurred in said province. To achieve this, a grouping of these earthquakes is carried out through K-means clustering using the Euclidean distance method, having as variables the latitude, longitude and magnitude of the earthquakes.

[Masala, 2012] in his research work "Seismic occurrences estimation through a parametric semi-Markov approach", applies the semi-Markov discrete time model to estimate the seismic occurrence in Italy. To achieve this end, the earthquakes that occurred were grouped according to the value of the magnitude of the moment in 3 groups or states, which are: low (Mw < 4,7), medium (4,7 ≤ Mw < 5) and high (Mw ≥ 5) and the time elapsed between two consecutive events is modeled as a general Weibull distribution. Taking these considerations, he concludes that the probability that no event occurs within a given time interval can be quantified, conditional on knowing the state of the last earthquake.

[Votsi et al., 2014] in their research work "Hidden semi-Markov modeling for the estimation of seismic occurrence rates", apply the semi-Markov model to estimate the seismic occurrence in Greece. The data collected is from earthquakes with a magnitude greater than 6,5Mw and these were finally grouped according to the value of the magnitude at the moment in 3 states, which are: low (6,5 ≤ Mw ≤ 6,7), medium (6,7 < Mw ≤ 7,1) and high (Mw > 7,1). Taking these considerations, he concludes that the effects of a nearby earthquake are commonly associated with the changes in static and dynamic stress that it produces, but may also be related to processes that are set in motion by those stress changes, such as the flow of cortical fluid and plastic deformation.

2.2. General Definitions

- Seism: This is the name given to the process of generating and releasing energy to later propagate in waves through the interior of the earth. Upon reaching the surface, these waves are registered by seismic stations and perceived by the population and by structures.

- Magnitude: The magnitude of a seism is a number that seeks to characterize the size of a seism and the seismic energy released, this can be given in various measurement scales, the best known are the Richter scale and the scale of magnitude of moments, the latter being the most frequently used for earthquakes of great magnitude.

- Intensity: It refers to the measure of the effects produced by a seism in a particular place. Intensity values are denoted with Roman numerals on the modified Mercalli intensity scale [Wood and Neumann, 1931] which classifies seismic effects with twelve ascending levels in shaking severity.

- Depth: The depth of a seism is the distance from the point on the surface to the point inside the earth where the release of energy originated.
- Geographic location: it is the identification of a specific place on the planet, through the use of various tools such as maps, compasses, coordinates or geolocation systems. The most used coordinates to determine the geographical location are:
    - Latitude: Latitude refers to the angle located between the equatorial plane and a line through that point. Depending on your location, the latitude can be north or south.
    - Longitude: The longitude is the angle that make up the Greenwich meridian (also known as the reference meridian, or zero meridian) and the meridian that passes through the point on the earth's surface that you want to locate.
- IGP: These are the abbreviations of the Geophysical Institute of Peru, a decentralized public body, dependent on the Ministry of the Environment, which is in charge of detecting natural disasters of destructive magnitude, such as earthquakes, tsunamis, volcanic eruptions, among others.

2.3. Statistical Definitions

- Stochastic Process: a stochastic process is a family of random variables that we assume to be defined in the same probability space. For a fixed time t, Z_t constitutes a random variable whose values form the state space and the set of instants in time of the parametrial space. The mean and variance are defined as $E\{Z_t\} = m_t$ $V\{Z_t\} = s2t$
- Cluster analysis: multivariate statistical technique that seeks to group elements (or variables) trying to achieve maximum homogeneity in each group and the greatest difference between groups.
- Transition probability: In the theory of stochastic processes and, in particular, in the theory of Markov chains, the probability of transition, P_y, is called the probability that the system being in the state E_i at time n passes to state E_i at time n + 1.

2.4. Stochastic processes

A stochastic process with state space *E* is a collection $\{X(t), t \in T\}$ of random variables X_t defined in the same probability space and taking values in E. The set T is called the parameter set. If T is countable, especially if $T = \mathrm{N} = \{0, 1, . . .\}$, the process is said to be a discrete parameter process. Otherwise, if T is not countable, the process is said to have a continuous parameter. In the latter case, the usual examples are $T = \mathrm{R}_+ = [0, \infty)$ and $T = [a, b] \subset \mathrm{R} = (-\infty, \infty)$. is not countable, the process is said to have a continuous parameter. In the latter case, the usual examples are X_t as the "state" or "position" of the process at time t.

We call a stochastic process {N (t), t ≥ 0} a counting process, if N (t) represents the number of events that occurred up to time t.

A counting process N (t) must satisfy the following properties:

(i) $N(t) \geq 0$.

(ii) $N(t) \in \mathrm{Z}$.

(iii) If $s < t$, then $N(s) < N(t)$.

(iv) For $s < t$, $N(s) < N(t)$ is equal to the number of events that occur in the interval $(s, t]$.

2.5. Renewal Process

Let {N (t), t ≥ 0} be a counting process and let X_n be the time between event (n − 1 and event n of the process with n ≥ 1. If the sequence $\{X_1, X_2, . . .\}$ is independently and identically distributed, so we call the counting process N (t) a renewal process.

2.5.1. Markov renewal process

Let E be the state space. A Markov renewal process is a bivariate stochastic process (J_n, S_n), where J_n are the values of the state space E in the Markov chain and S_n are the jump times. We define $X_n := S_n - S_{n-1}$ as the time spent in the state.

The process must satisfy the following properties.

$$P(J_{n+1} = j, X_n \leq t | (J_0, S_0), (J_1, S_1), \ldots, (J_n = i, S_n)) = P(J_{n+1} = j, X_n \leq t | J_n = i), \tag{2.5.1}$$

$$P(J_0 = i, X_0 = 0) = P(J_0 = i), \tag{2.5.2}$$

for all $n \geq 0$ y $t \geq 0$ e $i, j \in E$.

2.5.2. Refresh matrix, refresh kernel

Let E be the state space and consider the Markov renewal process (J_n, S_n) as in the definition of the Markov Renewal Process. The matrix defined as

$$Q(t) = \{Q_{ij}(t) : i, j \in E\}, \tag{2.5.3}$$

$$Q_{ij}(t) := \mathrm{P}(J_{n+1} = j, X_n \leq t | J_n = i), \tag{2.5.4}$$

It is called the renewal matrix. We identify the refresh matrix Q (t) as the refresh kernel.

The Markov renewal matrix Q (t) satisfies the following conditions:

(i) For all $t \geq 0$ e $i, j \in E$, is true that $Q_{ij}(t) \geq 0$.

(ii) The functions $Q_{ij}(t)$ are continuous to the right.

(iii) For all $i, j \in E$, is true that $Q_{ij}(0) = 0$ y $Q_{ij}(t) \leq 1$ for all $t \geq 0$.

(iv) For all $i \in E$, it is held that $\operatorname{lím}_{t \to \infty} \sum_{j \in E} Q_{ij}(t) = 1.$

2.6. Markov Chains

A Markov process is a random process with the property that given the current value of the process X_t, the future values X_s for s > t are independent of the past values X_u for u < t. That is, if we have the information of the present state of the process, knowing how it got to the current state does not affect the probabilities of going to another state in the future.

2.6.1. Markov chain in discrete time

Let $\{X_n, n = 0, 1, 2, \ldots\}$ a stochastic process that takes a countable number of possible values in a set I. A Markov chain in discrete time is a stochastic process where the conditional distribution of a state future Xn+1 given the past states $X_0, X_1, \ldots, X_{n-1}$ and the current state X_n, is independent of the past states and only depends on the present state. The process satisfies the property

$$P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j|X_n = i), \tag{2.6.1}$$

for all $n \geq 0$ e $i_0, \ldots, i_{n1}, i, j \in I$.

If we have a Markov chain in discrete time and it is also true that the right-hand side of (2.6.1) is independent of time n, then the Markov chain in discrete time is said to have stationary or homogeneous transition probabilities. It follows that there is a fixed probability p_{ij} that, when the process begins in state i, it will be the next in state j. For the probabilities p_{ij}, it is true that

(i) $p_{ij} \geq 0$ for all $i, j \in E$
(ii) $\sum_{j=0}^{\infty} p_{ij} = 1$ for all $i = 0, 1, 2, \ldots$

2.6.2. Markov chain in continuous time

Let {X (t), t ≥ 0} be a continuous-time stochastic process that takes values in the set I of non-negative integers. A Markov chain in continuous time is a stochastic process with the property that the conditional distribution of the future X (t + s) given the present X (s) and the past X (u), 0 ≤ u <s, depends only of the present and is independent of the past. The process satisfies

$$P(X(t+s) = j|X(s) = i, X(u) = x(u), 0 \leq u < s) = P(X(t+s) = j|X(s) = i), \tag{2.6.2}$$

for all s, t ≥ 0 and non-negative integers i, j, x(u) ∈ I with 0 ≤ u < s.

If we have a continuous-time Markov chain and it is also true that the right-hand side of (2.6.2) is independent of time s, then the continuous-time Markov chain is said to have stationary or homogeneous transition probabilities.

2.7. Semi-Markov Model

This section presents the definitions of the semi-Markov model. First, the definitions for the Markov semi-process that we will apply in the rest of the work are summarized. After that, empirical estimators are presented for the quantities of a finite state space semi-Markov process. The asymptotic behavior of empirical estimators for some of the quantities is discussed.

2.7.1. Definition of Semi-Markov process

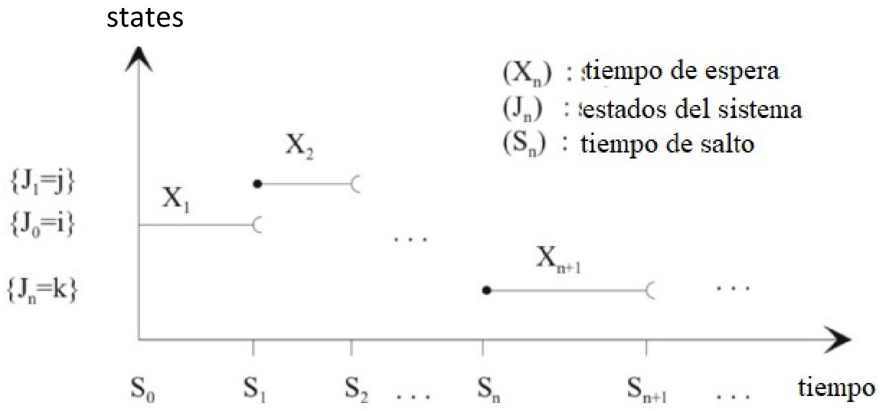
Define the stochastic process $Z_t := J_n$ for t ∈ [Sn, Sn+1). Then Z_t is called the semi-Markov process. The semi-Markov process is the process that evolves over time and all the realizations of the process have a defined state for a given moment. The semi-Markov process can be read as follows: Suppose a process can be in one of m states. Every time it enters a state i, it stays there for a random period of time, and then makes a transition from state i to state j with transition probability p_{ij}.

Furthermore, following the definitions of [Votsi et al., 2012]. Consider a Markov renewal process (Jn, Sn) defined in a complete probability space and with a state space E = {1, 2, . . . , m}. It is true that Jn are the values of the state space E in the Markov chain and Sn are the jump times for n ≥ 0. We assume that Sn takes values in [0, ∞). Define $X_n := S_n - S_{n1}$ as the time of permanence in the state and we assume that $X_0 = S_0 = 0$. Let $Z_t := J_{N(t)}$ for t ≥ 0 be the semi-Markov process where N (t) is the counting process of the semi-Markov process up to time t. It is true that N (t) := max{n ≥ 0 : S_n ≤ t}. In figure3.1.1 [Barbu and Limnios, 2009] we see an example of a sample route of the semi-Markov chain.

The semi-Markov process depends on its initial law, which we assume is equal to $\pi_i = P(J_0 = i)$, and its semi-Markov kernel.

$$\begin{aligned} Q_{ij} &:= \mathbb{P}(J_{n+1} = j, X_{n+1} \leq t | J_0, J_1, \ldots, J_n = i, X_1, X_2, \ldots, X_n) \\ &= (J_{n+1} = j, X_{n+1} \leq t | J_n = i), \end{aligned} \tag{2.7.1}$$

For all t ≥ 0 e i, j ∈ E. We consider that $Q_{ii}(t) \neq 0$ for all i ∈ E. Q_{ij} it will be called the renewal matrix.



The odds

$$\begin{aligned} p_{ij} :&= \lim_{x\to\infty} Q_{ij}(t) = Q_{ij}(\infty) \\ &= \mathbb{P}(J_{n+1} = j | J_n = i) \end{aligned} \tag{2.7.2}$$

are the transition probabilities from state i to state j of the embedded Markov chain {J_n, n = 0, 1, 2, . . .}. We will work under the assumption that the transition probabilities do not depend on time n.

It is observed that before entering state j, the process remains for a certain time x in state i. For this, it is appropriate to introduce the conditional cumulative distribution function of the waiting time in each state, since the state is occupied later. The following is then defined:

$$F_{ij}(t) := \mathbb{P}(X_{n+1} \leq t | J_n = i, J_{n+1} = j) \tag{2.7.3}$$

to be the distribution function associated with the time spent in state i, before moving to state j. From this definition we can derive the following result. The proof is given in the appendix.

Theorem 2.7.1. How is it true that

$$F_{ij}|(t) = \frac{Q_{ij}(t)}{p_{ij}}, \tag{2.7.4}$$

for all t ≥ 0 and i, j ∈ E.

The residence time distribution function, also called the waiting time, in state i is equal to

$$\begin{aligned} H_i(t) &= \mathbb{P}(X_{n+1} \leq t | J_n = i) \\ &= \sum_{j\in E} Q_{ij}(t) \end{aligned} \tag{2.7.5}$$

Equation 2.7.5 indicates the probability that the event that is in state i remains in the same state for at least one time of duration X.

It defines:

$$D_i(t) := 1 - H_i(t) \tag{2.7.6}$$

Equation 2.7.10 indicates the probability that the event that is in state i changes to a different state for at least one time of duration X.

The following probabilities are introduced which are specific to the discrete case:

$$B_{ij}(t) = P(J_{n+1} = j, S_{n+1} - S_n = t | J_n = i) \tag{2.7.7}$$

According to the definitions of Q it shows:

$$b_{ij}(t) = \begin{cases} Q_{ij}(0) = 0 \;\; si\ t = 0 \\ Q_{ij}(t) - Q_{ij}(t-1) \;\; si\ t > 0 \end{cases} \tag{2.7.8}$$

Finally, the semi-Markov discrete time process is presented Z = (Z(t), t ∈ N), where Z(t) = $J_{N(t)}$ y N (t) = max {n, S_n} denote the occupied state by the system at time t.

Then the transition probabilities are shown as:

$$\phi_{ij}(t) = P(Z_t = j | Z_0 = i)\ i, j = 1, \cdots, m \tag{2.7.9}$$

Similar to the continuous case, these probabilities satisfy the following equations for the evolution of the semi-Markov process:

$$\phi_{ij}(t) = \delta_{ij} \cdot (1 - H_i(t)) + \sum_{\beta=1}^{m} \sum_{\theta=1}^{t} b_{i\beta}(\theta) \cdot \phi_{\beta j}(t - \theta) \tag{2.7.10}$$

where δ_{ij} denotes the Kronecker delta. The evolution equations allow us to relate the transition probabilities of the matrix P with the distribution functions of the waiting time in each state.

2.7.2. Empirical estimators

Let T be the completion time of the process. For the semi-Markov kernel $Q_{ij}(t)$ we have the following empirical estimator

$$\hat{Q}_{ij}(t, T) := \frac{1}{N_i(T)} \sum_{n=1}^{N(t)} \mathbb{1}_{\{J_{n-1}=i, J_n=j, X_n \le t\}} \tag{2.7.11}$$

Since $F_{ij}(t) = Q_{ij}(t)/p_{ij}$, in a similar way we obtain that $\hat{F}_{ij}(t,T) = \hat{Q}_{ij}(t,T)/\hat{p}_{ij}(T)$ with:

$$\hat{F}_{ij}(t, T) := \frac{1}{N_{ij}(T)} \sum_{n=1}^{N(t)} \mathbb{1}_{\{J_{n-1}=i, J_n=j, X_n \le t\}},$$

$$\hat{p}_{ij}(T) := \frac{N_{ij}(T)}{N_i(T)}.$$

The quantities $\hat{F}_{ij}(t, T)$ y $\hat{p}_{ij}(T)$ are respectively the empirical estimators of the conditional transition functions and the transition probabilities. We see that for $\hat{p}_{ij}$ we divide the number of transitions from state i to state j by the (total) number of visits to state i.

We want to show that the empirical estimator $\hat{Q}_{ij}(t,T)$ (t, T) of the semi-Markov kernel is strongly consistent and asymptotically normal. Therefore, we need the following definition and thesis.

Definition 2.7.1. Let X_i, i = 1, . . . , n, an independent and identically distributed sequence of random variables with the distribution function F in R. Then the empirical distribution function is defined by

$$\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \tag{2.7.12}$$

Theorem 2.7.2. (Glivenko-Cantelli theorem) Let X_i, i = 1, . . . , n, be an independent and identically distributed sequence of random variables with the distribution function F in R. Then,

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| \to 0 \tag{2.7.13}$$

while n → ∞

Theorem 2.7.3. The empirical estimator $\hat{p}_{ij}(T)$ of p_{ij} for all i, j ∈ E is strongly consistent, that is

$$\hat{p}_{ij}(T) \to p_{ij} \tag{2.7.14}$$

while T → ∞

We observe that the empirical estimator $\hat{F}_{ij}(t, T)$ of the distribution function satisfies Theorem 2.7.2, because it meets the definition of the empirical distribution function. Now, we give the properties of the empirical estimator $\hat{Q}_{ij}(t, T)$.

Theorem 2.7.4. The empirical estimator $\hat{Q}_{ij}(t, T)$ of $Q_{ij}(t)$ for all i, j ∈ E is strongly consistent, that is

$$\underset{i,j\in E}{\text{máx}} \sup_{t\in[0,T)} \left|\hat{Q}_{ij}(t,T) - Q_{ij}(t)\right| \to 0 \tag{2.7.15}$$

while T → ∞

2.8. Weibull distribution function

The Weibull distribution is often used to represent component life. In the present work, it will be used to represent the inter-arrival time of the occurrence of one seismic event to another.

This application was used by [Masala, 2012] and [Votsi et al., 2014] in their different investigations for the application of times between earthquakes.

2.8.1. Density function

If X is a continuous random variable, we say that X has a Weibull distribution with parameters α, λ > 0 and we write:

$$f(x) = \lambda\alpha(\lambda x)^{\alpha-1}e^{-(\lambda x)^{\alpha}} \qquad x > 0 \tag{2.8.1}$$

where α is the shape parameter and λ is the scale parameter of the distribution.

2.8.2. Distribution Function

The cumulative distribution function of a random variable X ~ Weibull(α, λ) is:

$$F(x) = 1 - e^{-(\lambda x)^{\alpha}} \tag{2.8.2}$$

for x > 0.

2.8.3. Maximum Likelihood Function

The likelihood function for complete samples with Weibull distribution (λ, β) is

$$L(\alpha, \lambda) = (\alpha\lambda)^n \prod_{i=1}^{n} (\alpha T_i)^{\lambda} e^{-(\alpha T_i)} \tag{2.8.3}$$

and the log-likelihood function is

$$l(\alpha, \lambda) = n\log(\lambda) + \lambda n \log\alpha + (\lambda - 1)\sum_{i=1}^{n}\log T_i - \sum_{i=1}^{n}(\alpha T_i) \tag{2.8.4}$$

The maximum likelihood estimators are obtained by solving the equations resulting from setting the two partial derivatives of l(α, λ). equal to 0. The maximum likelihood estimator of λ is obtained by solving

$$\frac{\sum_{i=1}^{n} T_i^{\hat{\lambda}} \log T_i}{\sum_{i=1}^{n} T_i^{\hat{\lambda}}} - \frac{1}{\hat{\lambda}} - \frac{1}{n}\sum_{i=1}^{n}\log T_i = 0, \tag{2.8.5}$$

using numerical methods such as the Newton-Raphson method. Once the shape parameter has been obtained, an estimator for α will be obtained through the express ion

$$\hat{\alpha} = \left(\frac{n}{\sum_{i=1}^{n} T_i^{\hat{\lambda}}}\right)^{\frac{1}{\hat{\lambda}}} \tag{2.8.6}$$

2.9. Cluster K-Means Analysis

K-means cluster analysis seeks to divide the n individuals in a multivariate data set into K clusters, where each individual in the data set is fully assigned to a particular cluster. As a rigid partition algorithm, K-means cluster analysis is an iterative process. First, the data is initially divided. Each group calculates its means and then the data is divided again by assigning each data to its closest grouping position of means [Likas et al., 2003]. In its simplest form, this process consists of three stages:

- Partitioning objects in initial cluster K.
- Starting with nothing objects, determine an object in a group that has the closest centroid (mean). The distance is generally calculated using the Euclidean distance with standardized or non-standardized observation. Recalculate the cluster centroid to get a new object and cluster that missing object. The centroid of the group is calculated by calculating the mean value that is realized in 2.9.1

  $$C_{kj} = \frac{x_{1kj} + x_{2kj} + \cdots + x_{akj}}{a},)1, 2, \ldots, p \tag{2.9.1}$$

  where C_{kj} is the centroid of group-k, variable-j and a for the number of members of group k.

- The step is repeated until there is no more transfer.

The Euclidean distance is the distance that is selected with the most common type. Its simplicity is the geometric distance in multiple dimensions of space. Euclidean distance is generally calculated from raw data and not from standard data. This method has several advantages, including the distance of any two objects that are not affected by the addition of new objects to analyze, which can be an outlier.

However, the distance can become very large, caused only by the difference in scale. The Euclidean distance is calculated by 2.9.2:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2} \tag{2.9.2}$$

The goal of clustering is to divide a given data set into disjoint subsets to optimize specific clustering criteria.

Chapter 3

Materials, Methods and Procedures

3.1. Study materials

3.1.1. Source of information

The seismological data provided by the Geophysical Institute of Peru (IGP) was used. Which provides information related to the location (latitude and longitude), depth, magnitude, date and time of the seismic event.

3.1.2. Target population

All seismic events with a magnitude greater than 5.3 on the Magnitude of Moments (Mw) scale in the Peruvian territory (continental and maritime) from 2008 to 2018.

3.1.3. Kind of investigation

This research can be classified as Non-experimental, Longitudinal and cross-sectional, since the data is collected from the seismic occurrence of the Peruvian territory over the years. In addition to that, two other types of research will be described.

Descriptive:

The seismic occurrence in Peruvian territory is described through a cluster analysis with data from the earthquakes reported by the IGP throughout the continental and maritime territory of Peru, during the period from 2008 to 2018. To try to acquire a culture of prevention against These natural disasters, the government of Peru, as well as the governments of each department, draw up their Disaster Risk Prevention and Reduction Plan from time to time.

Explanatory:

It is known that currently the regional governments, as well as the national government of the country, are participants in the preparation of the Disaster Risk Prevention and Reduction Plan for each of their jurisdictions. There are various reports or articles that mention the great importance of having a culture of prevention before natural disasters and mainly before earthquakes, as well as the report by BBVA Research, which is entitled "Earthquakes in Peru: prevention is the best way to protect yourself".

In which a very important characteristic of the city of Lima is mentioned that makes its population vulnerable to the occurrence of strong earthquakes, said characteristic is that 70% of Lima's homes are vulnerable to a seism of great magnitude because they are built informally.

As is known, the data is about Lima, however, the reality of other cities or regions of the country may be very different from this, therefore, the need to identify behavior patterns of seismic events by applying the model semi-Markov for each of the departments of Peru, thus allowing a better view of which prevention measures should be implemented differently in each of the country's regions.

3.1.4. Analysis unit

Registered magnitudes of seismic events that occurred in Peru.

3.1.5. Study variables

For the purposes of this research, the variables Magnitude (which is measured in the magnitude scale of Mw moments), Depth (measured in kilometers) and the inter-arrival time between earthquakes (measured in days) are being used.

3.1.6. Sampling frame

The sampling frame is 425 records of seismic events in Peru from 2008 to 2018 with magnitudes greater than 5.3 in the moment magnitude scale.

3.1.7. Software used

Two programs have been used to prepare the work, which are Python and RStudio. The first of them was used to perform the graphical analysis of earthquakes, as well as to be able to perform the clustering by latitude and Longitude of seismic events at the national level, while the second (RStudio) was used to develop the semi-Markov model and clustering by magnitude and average time for each of the previously defined states.

3.2. Pattern identification technique

For the purposes of a better management of the seismic data, the earthquakes will be grouped into 3 states based on the magnitude levels. As a result of the classification of the magnitude levels, the 3 states are used for the application of the semi-Markov model, which will seek to identify behavior patterns of the seismic occurrence in the departments of Peru.

3.3. Formation of groups of magnitude states

Se The application of a semi-Markov model is required to be able to fulfill the proposed objective, which is to identify behavior patterns of seismic occurrence, and for this it is convenient to group the events into states, to be able to calculate the various variables involved. in the model. Due to the aforementioned, in this document earthquakes are classified according to their magnitude. Thus, 3 states of earthquakes are formed in the period of time from 2008 to 2018. Because earthquakes less than 5.3 M_w can become imperceptible or felt, but with very low intensity, not generating any damage of any kind, the first state covers earthquakes from 5.3 M_w to 5.5 M_w, the second covers earthquakes from 5.5 M_w to 6 M_w and finally, the third group includes earthquakes of magnitude greater than 6 M_w.

Table 1: Number of seismic events by State

State	Rank	Quantity
1	[5,3 − 5,5 >	229
2	[5,5 − 6,0 >	148
3	[6,0 − more	48

Where State 1 presents earthquakes that cause slight damage to buildings, State 2 contains earthquakes that cause greater damage to buildings and even roads, and finally in State 3 there are earthquakes that have a much more character. destructive.

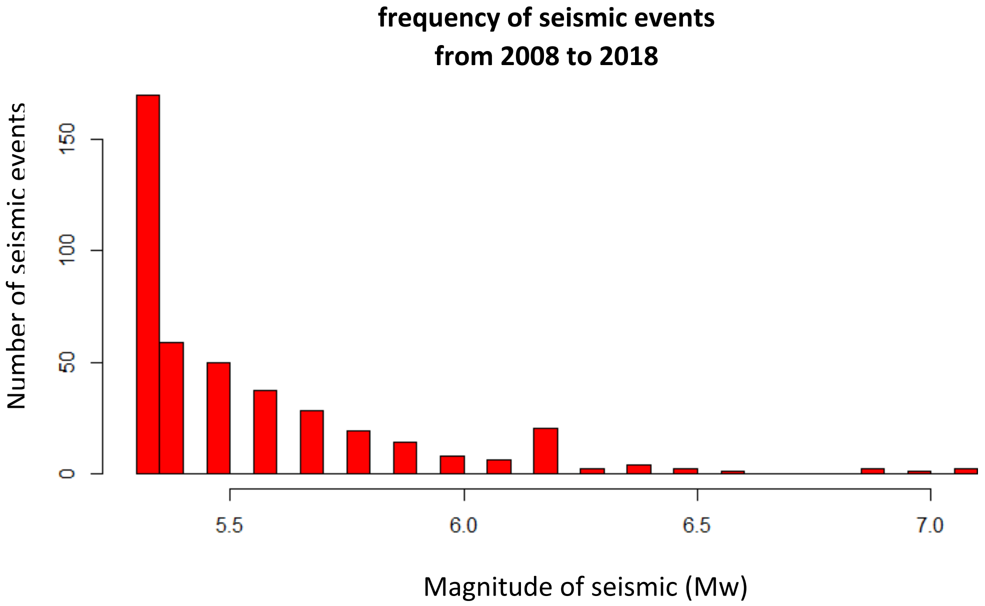


***Figure 3.1**: Distribution of the Magnitudes (M_w) of the seismic events of the sampling frame*

3.3.1. Description of seismicity in Peru according to magnitude

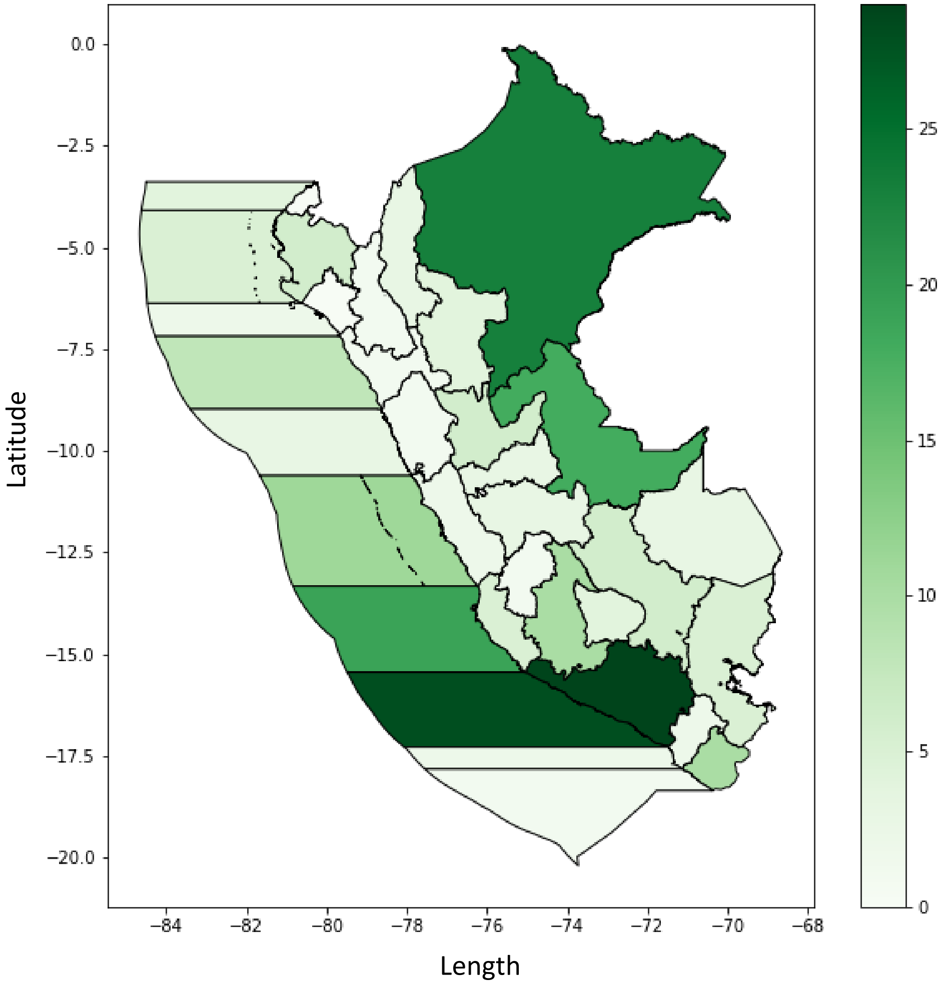
3.3.1.1. State 1

For the first state, there are earthquakes with magnitudes from 5.3 to 5.5 on the Magnitude of Moment (Mw) scale as of 2008. Earthquakes of these magnitudes are the most common, they can become perceptible by people and do not usually cause material damage.

It can be seen in figure 3.2 that the Arequipa area is the area with the highest concentration of seismic activity within the continental territory of the country with around 30 earthquakes, mainly due to the large number of geographic faults it has, among which we have volcanoes. that generate constant telluric movements of low magnitude and can cause ones of greater magnitude also causing the eruption of volcanoes.

In addition, we can also observe in the graph that regarding the continental territory of Peru we have a relatively high concentration of earthquakes in the Loreto area, within this large department the concentration of these earthquakes occurs mainly in the area near the Sierra de Our country where the Andes Mountain range and various geographical faults are located, and as mentioned previously, thanks to these faults, earthquakes of low to medium magnitudes can occur. Likewise, due to the internal deformation of the Nazca plate under the mountain range, a considerable number of earthquakes are caused in the area that can reach magnitudes from low to high magnitudes. Some effects of these low intensity earthquakes can be seen in some landslides on roads.

In the graph we can also see that the maritime areas have a considerable number of earthquakes that occurred during the analysis period, even reaching amounts of recorded earthquakes similar or greater than those already mentioned Arequipa and Loreto, within this maritime territory there are maritime areas off the coasts of Arequipa and Ica, which have had between 20 and 30 earthquakes with a magnitude of 5.3 to 5.5 on the Moment Magnitude (Mw). scale. This quantity of earthquakes presented is mainly due to the crossing of the great Nazca and Pacific plates right at sea, thus causing earthquakes that are also felt in the departments that are on the coast.



Source: Geophysical Institute of Peru

Figure 3.2: *State 1 - Earthquakes of magnitude between 5.3 M_w and 5.5 M_w*

3.3.1.2. State 2

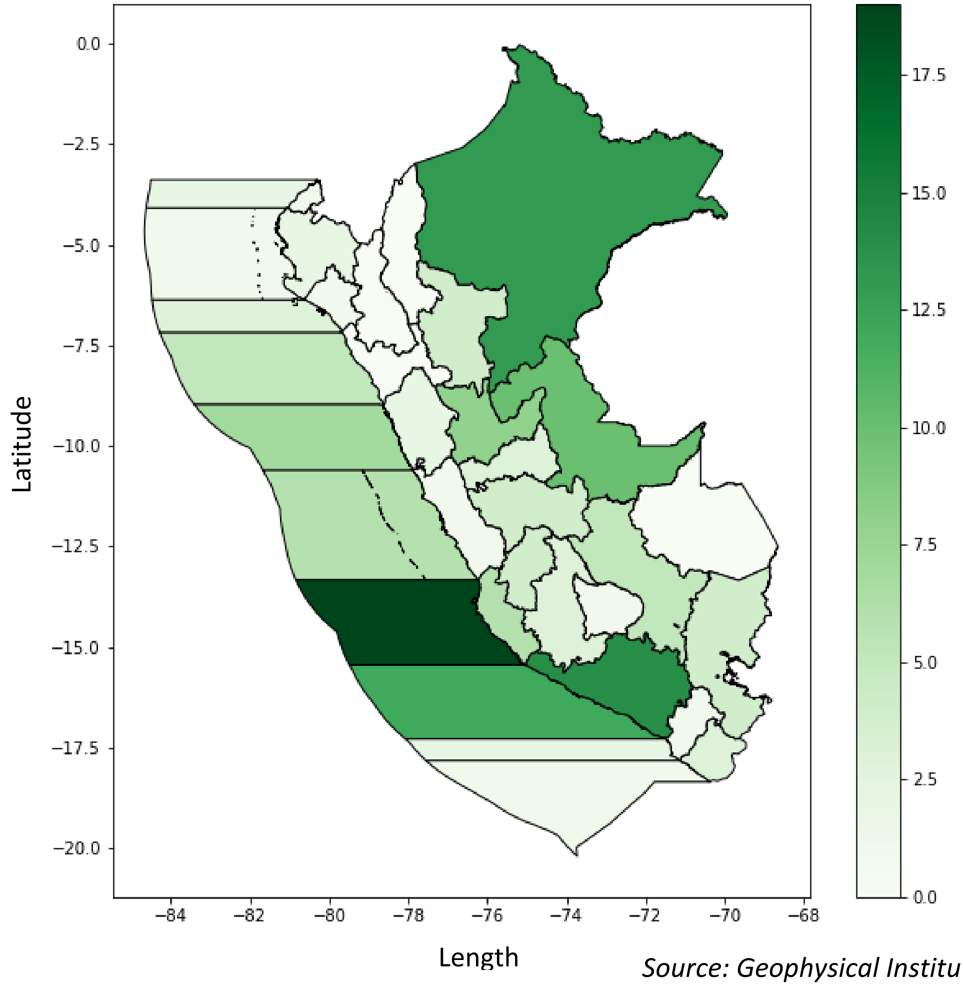
For the second group, earthquakes that have between 5.5 and 6 on the Magnitude of Moment (M_w) scale as of 2008 are represented. This type of magnitude becomes perceptible by people and also has the force necessary to cause minor material damage in buildings mainly.

In figure 3.3 shown it can be seen that there is a high concentration of earthquakes of state 2 in the department of Arequipa and Ucayali within the continental territory with around 12 earthquakes reported since 2008. These areas are of high seismic activity due to high number of faults that it presents in its geography, which is one of the reasons for the seismic occurrence.

Additionally, if the part of the maritime territory is observed, it can be seen that the areas with the highest concentration of seismic activity reported since 2008 are the areas of Arequipa and Ica with approximately 15 reported earthquakes. These earthquakes are recurrent in these areas due to the crossing of the Nazca and the South American plates, in addition to the fact that despite the fact that they occur in the sea, they also largely affect the departments that are in front of them, for example, many times they He has heard of earthquakes in Ica, however it seems curious that the map does not paint Ica as one of the main areas with earthquakes, this because the vast majority of earthquakes that are felt on the surface of the department have as epicenter the sea.

As can be seen, most of the earthquakes that are felt in the Country do not necessarily have the continental territory as their epicenter, but often occur off the coast and this in turn causes tsunami alerts, depending on the intensity of the earthquake, especially in the departments that are right in front of the epicenter.

On the other hand, we also have areas with very little or no number of earthquakes of these magnitudes, which are La Libertad, Lambayeque and Madre de Dios, which have only presented 1 or 0 earthquakes within this group of magnitudes throughout the analysis time.



Source: Geophysical Institute of Peru

***Figure 3.3**: State 2: Earthquakes between 5.5 to 6 M_w*

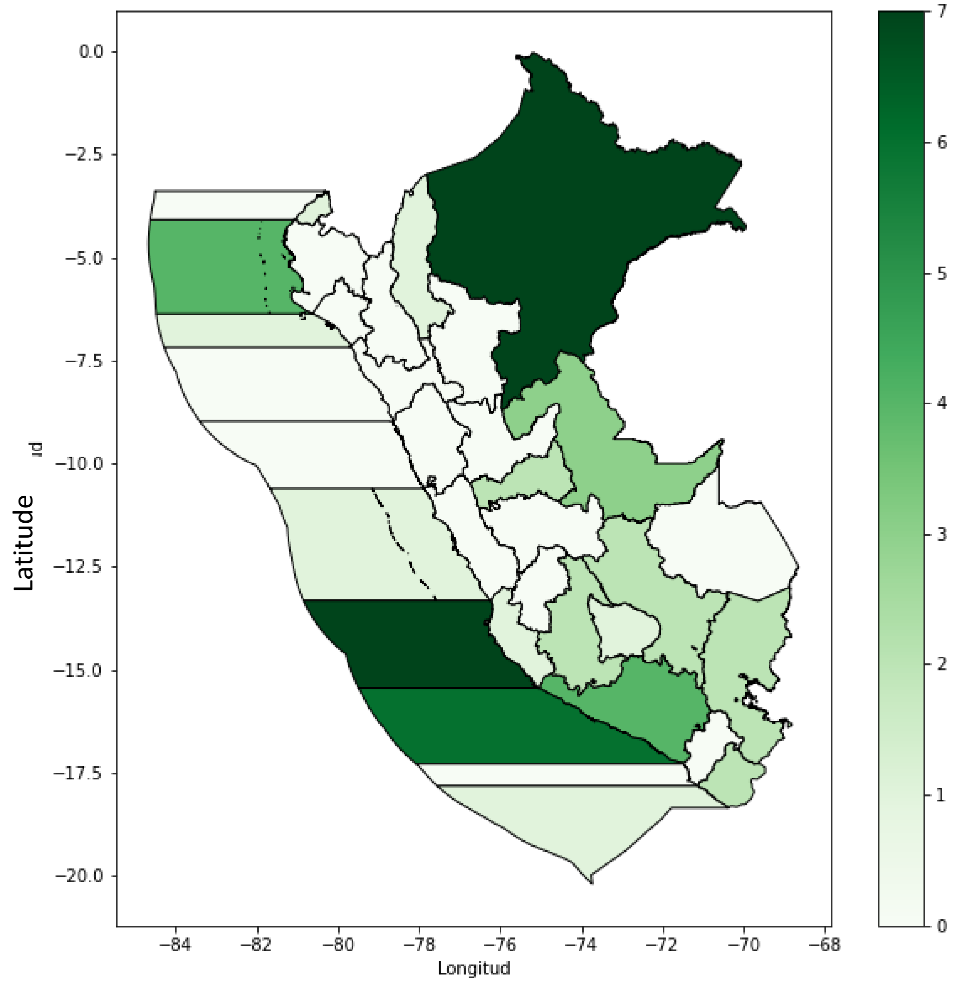
3.3.1.3. State 3

The earthquakes that are in this state usually cause a great number of losses in the population, since they usually generate a strong intensity in the earthquakes and can even bring as a consequence the occurrence of other natural disasters, as has been seen above. throughout the history of Peru.

It can be seen in figure 3.4 that the Loreto and Arequipa areas are the locations where the greatest amount of earthquake occurs within this state, with around 5 earthquakes reported between 2008 and 2018 each. These amounts are not so many compared to the previous groups, but they have a much greater destructive power. For example, taking a seism outside the period under study, we can remember the Pisco earthquake of 2007, which produced a large amount of material and human losses. At the epicenter, the intensity of this earthquake was VII MM, a degree of intensity so strong that it terribly affected areas such as Pisco, Ica, Chincha, Cañete, Yauyos, Huayrará and Castrovirreyna.

In addition, it can also be observed in figure 3.4 that the maritime zones of Ica and Arequipa once again have a great seismic concentration, even higher than the earthquakes reported in continental territory. Again, taking as an example a seism outside the study period, a great example is the earthquake that occurred in 2001 that had as its epicenter within the Arequipa Sea, producing a magnitude of 6.9 MW near the coastline. Both the Arequipa and Pisco earthquakes were so intense that they produced tsunamis up to 8 meters high that they reached the coast in an approximate time of 15 minutes.

This relatively high frequency of earthquakes of this magnitude explains why, unlike the North and Central regions of Peru, in the South region the earthquakes are distributed in depth from the pit, on an inclined plane of 30° to depths of 300 km. This change in the geometry of the Nazca plate is associated with the presence of the volcanic chain, from the Ayacucho region to the border with Chile.



Length

Source: Geophysical Institute of Peru

Figure 3.4: *State 3: Earthquakes greater than 6 Mw*

3.4. Depth

The Geophysical Institute of Peru manages to classify earthquakes according to their depths. In such a way that three groups or foci are formed: Superficial, Intermediate and Deep. Where the surface focus is between the earth's crust itself up to 70 km. The intermediate focus is between 70 km to 300 km. Finally, the deep focus is more than 300km away. It should be noted that the center of the Earth is located about 6,370 km deep. In Peru the concentration of earthquakes according to their depth is given as follows:

Table *2: Number of seismic events by State and Depth*

Depth	State 1	State 2	State 3
Superficial	118	84	27
Intermediate	111	64	21

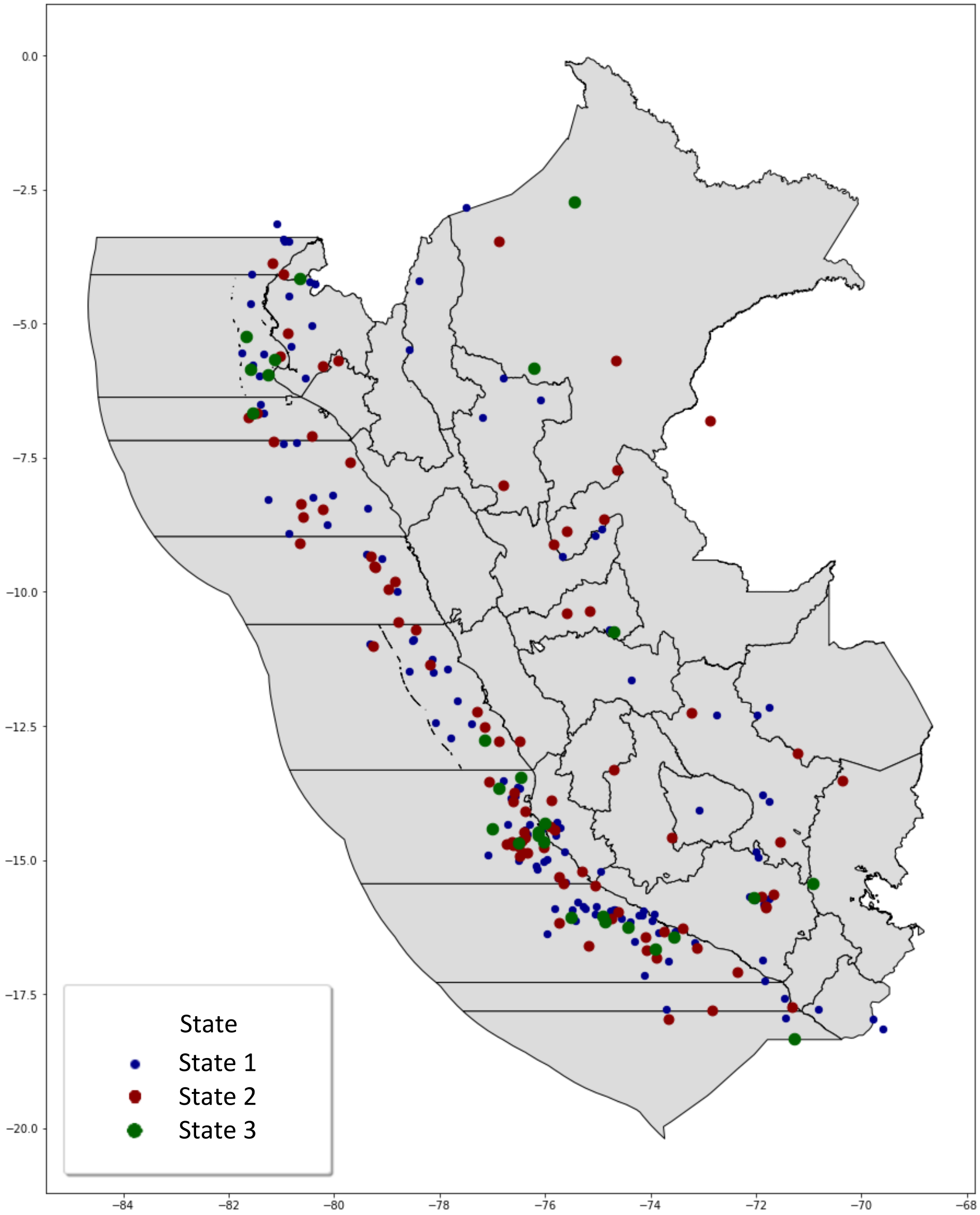
Where it can be seen that earthquake of a magnitude greater than 5.5 Mw with a hypocenter at a depth greater than 300 km below sea level have not been reported in the country for the period of time of analysis.

3.4.1. Shallow depth

At this depth, which constitutes earthquakes with hypocenters ranging from 0 to 70 kilometers from the earth's surface, the vast majority of earthquakes occurred in the period of time being studied. As can be seen in the figure shown, the highest concentration of the points, which represent the epicenters of the earthquakes that occurred, occurs in the area of the southern coast of the country, the main affected departments being Ica and Arequipa, as well as the maritime territory in front of these. In these areas of high concentration of shallow depth earthquakes, we can also note that there is a high concentration of points corresponding to states 2 and 3 (magnitudes from 5.5M_w onwards).

Additionally, in the graph it can also be seen that in almost the entire surface of the country there are earthquakes of state 2, which range from 5.5 M_w to 6 M_w, with a slightly higher concentration of these on the coasts and the sea of Lima and Ancash in addition to the other areas already mentioned above. These earthquakes, having their hypocenter less than 70 kilometers away, are usually perceptible by the departments that are in front of them and some neighboring departments.

On the other hand, although high concentrations of state 1 earthquakes (from 5.5 M_w to 5.7 Mw), these are present in a large part of the entire coast and sea of Peru, having slight concentrations in the north, thus understanding that in those areas there are few earthquakes and in turn these earthquakes that occur are a magnitude that becomes Perceptible by people, but could cause slight damage to the infrastructure.



Source: Geophysical Institute of Peru

Figure 3.5: Shallow depth

3.4.2. Intermediate Depth

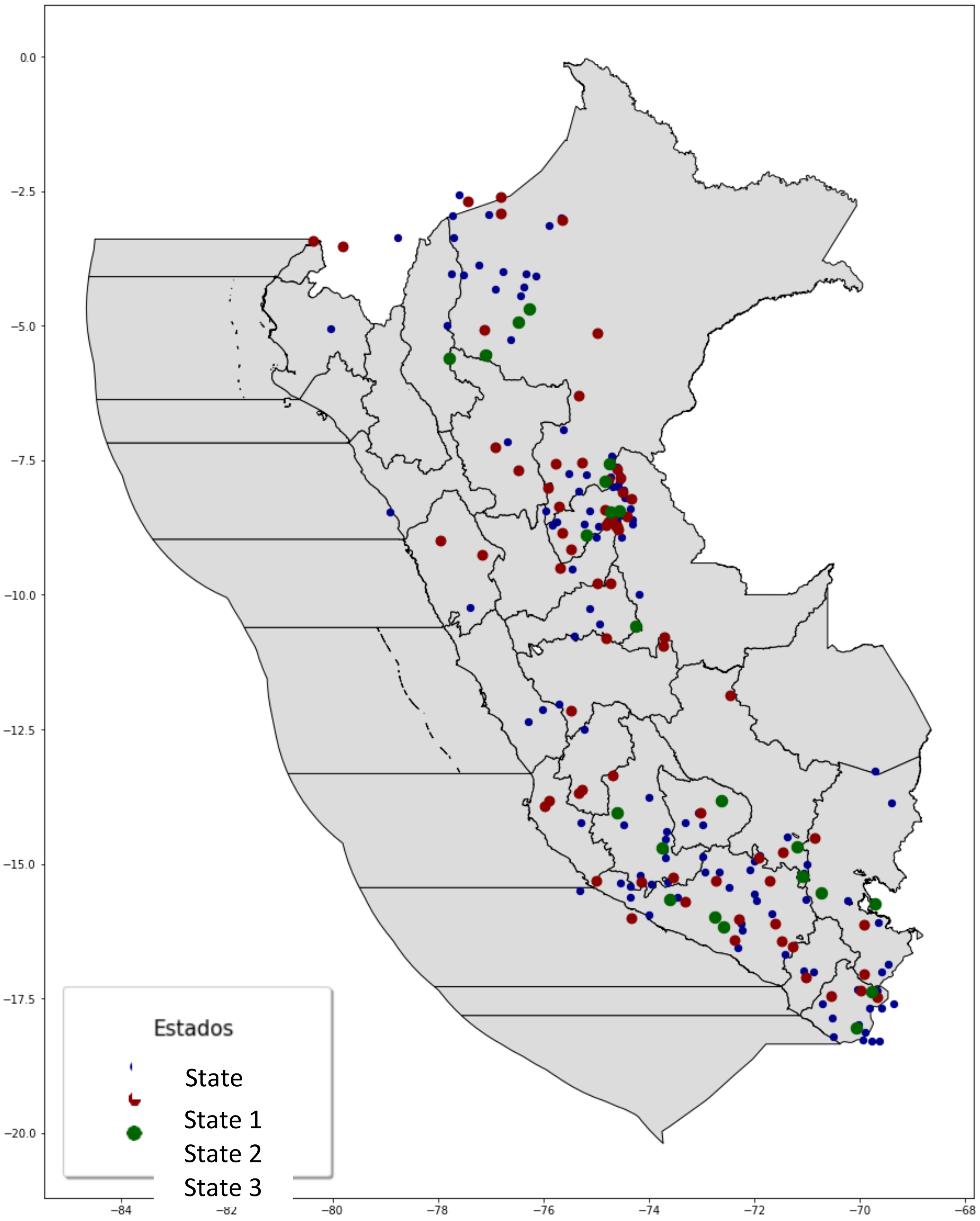
At intermediate depth, which constitutes earthquakes with hypocenters or foci between 70 to 300 kilometers from the earth's surface, it is observed that the amount of seismic occurrence is much less than for shallow depth. Mainly, what can be seen in figure 3.6 is that the highest concentration of earthquakes with intermediate depth is found in the department of Arequipa. There is a great concentration of both state 1, 2 and 3 earthquakes. As previously mentioned, the surface of this department complies with characteristics such as: Deformation in the continental crust and contact between the Nazca and South American plates.

Another area with a high concentration of intermediate depth earthquakes is the department of Ucayali, which registers earthquakes of state 2 and 3, as can be seen in the border with Loreto.

In addition, it should be noted that the most relevant earthquakes that have occurred below this depth is the recent earthquake of 2019 (outside the analysis period) in Loreto with a magnitude of 8 M_w. This occurred 135 kilometers from the earth's surface. It is the most powerful earthquake in Peru, 12 years after the 2007 earthquake with a magnitude of 7.9. And although the 2019 earthquake had the same, similar or even greater magnitude than the one that occurred in 2007, the intensity and damage generated by this earthquake were less, but had a great impact on the population. The first earthquake had an intensity of V-VI (IMM), while the second earthquake reached the intensity of IX on the Mercalli scale.

According to the Geophysical Institute of Peru (IGP), intermediate focus earthquakes are seldom perceptible on the surface, to be exact they are felt when they reach magnitudes ≥ 7 M_w and sometimes cause damage to homes and soil liquefaction processes. This is because there is a high relationship between the depth and the intensity of the seismic occurrence. Thus, the deeper the earthquakes, the intensity or perception thereof decreases as the energy produced dissipates before reaching the earth's surface.

Therefore, the difference in intensity of the Loreto earthquake in 2019 and the Pisco earthquake of 2007, may be related to the different depths between the two, since there are more factors that influence the intensity level of a seism.



Source: Geophysical Institute of Peru

Figure 3.6: *Intermediate Depth*

3.5. Treatment of observations

The database that is managed contains information on date, time, longitude, latitude, depth and magnitude. The first thing that was done was the classification of the magnitude levels of the seismic events in the three states. In addition, with the latitude and longitude information from the records, it was identified to which department each seismic event belongs to carry out the respective analysis by department. Finally, a historical ordering of seismic events was carried out (from the oldest to the most recent).

This is to identify what are the states of magnitude after seismic events.

For the application of the Semi-Markov model, states of magnitude and time will be used as main variables, but also variables such as latitude, longitude and depth that indicate the location of the seismic event will be used. It is important to point out that what is being sought is the estimation of the waiting time (or change) of the change in magnitude state, that is, what is the time that has to occur for the magnitude state of a seismic event to change to another state.

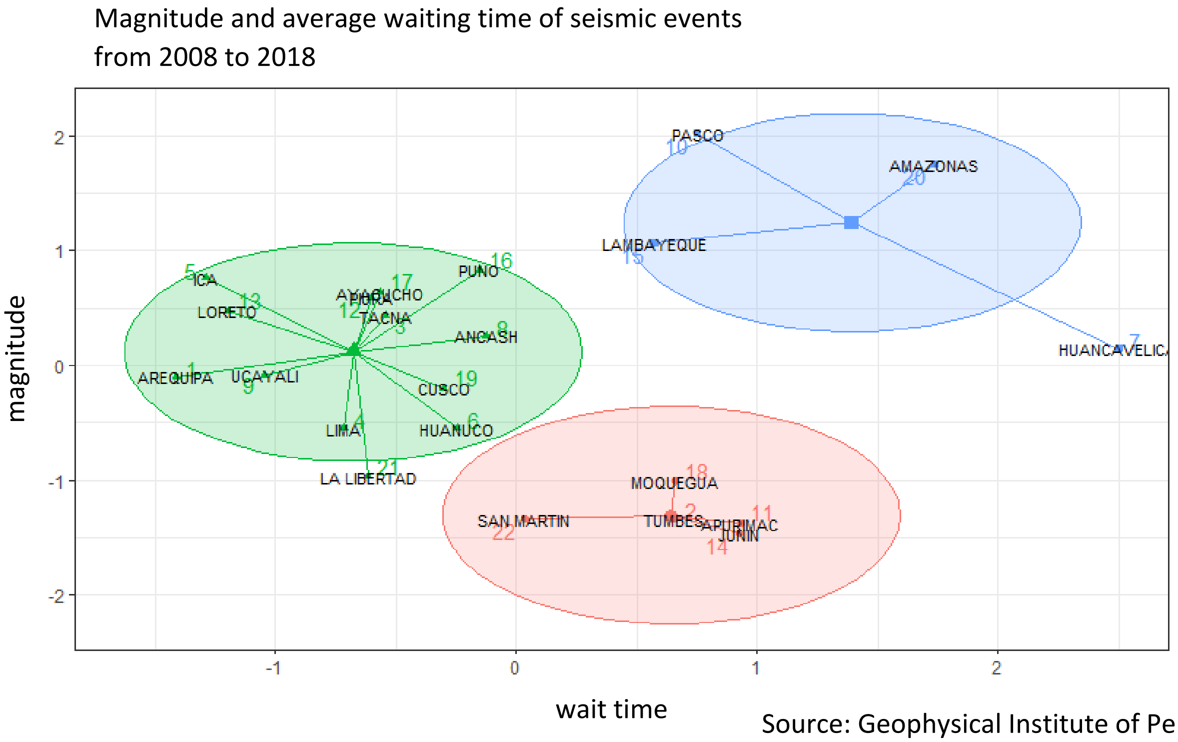
3.6. Cluster analysis of seismic events by magnitude and average waiting time

In this section, a clusterization of the departments of Peru is carried out, using as sources or cluster variables the average magnitude per department and the average waiting time per department, since it seeks to find the similarities of the departments with respect to time and magnitude. of seismic events. By applying the 'K-means' clustering technique, described in the previous section, the departments of Peru were able to identify 3 centroids that associate behavior groups for seismic events. The first thing that is observed from the cluster graph shown in Figure 3.7 is that the departments fall on a Cartesian plane, where the departments are grouped in 3 colors. Furthermore, the abscissa axis divides the departments into two views; to the right seismic events with a high waiting time level and to the left seismic events with low levels of average waiting time.

In addition, the ordinate axis indicates the level of average magnitude that seismic events have, where the departments shown in the upper part represent those with higher average magnitude levels and the departments that are in the lower part have the lowest levels.

Under these specifications, the departmental group of green color (which associates Ica, Loreto, Arequipa, Ucayali, Lima, La Libertad, Huánuco, Cusco, Ancash, Tacna, Piura, Ayacucho and Puna) tend to have low levels of average waiting times among the seismic events, also adding that the average magnitude of the departments is neither high nor low, but remains in a medium state. The departmental group of blue color (which associates Pasco, Lambayeque, Amazonas and Huancavelica) tend to have high average waiting times and in the same way they also have high levels of average magnitude, it means that after the occurrence of a seismic event the interval time for the next event becomes large and also the magnitude levels that present from event to event are high.

Finally, the departmental group of red color (San Martín, Moquegua, Tumbes, Apurímac and Junín) presents the average waiting time slightly high, however, its average magnitude levels of earthquakes are low, it means that the seismic occurrence in these departments becomes dim and the time between events does not become remarkable.


Magnitude and average waiting time of seismic events from 2008 to 2018

Source: Geophysical Institute of Peru

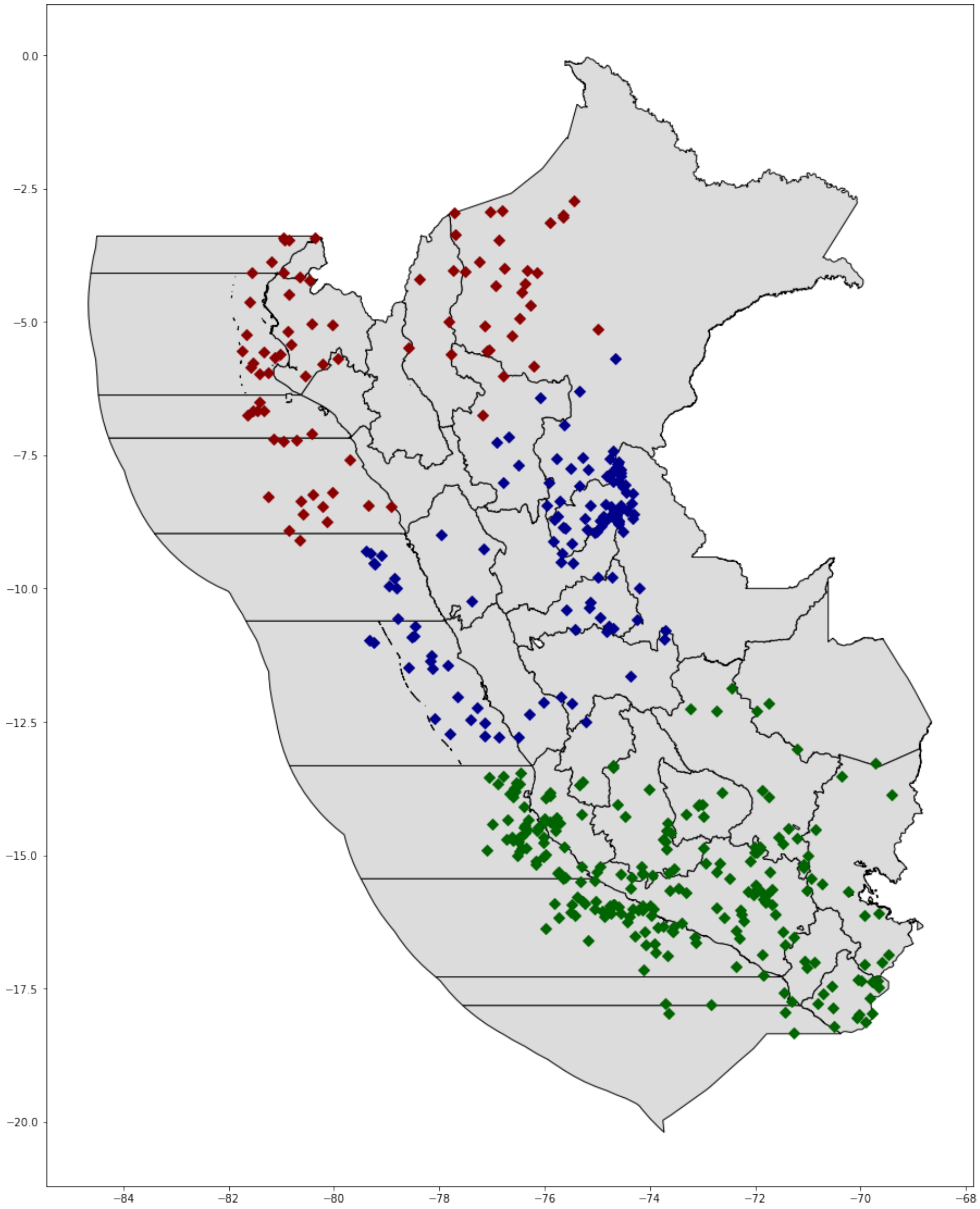
Figure 3.7: *Cluster analysis by department based on average magnitude and average wait time*

3.7. Clusterization by Latitude and Longitude

In this section, the clustering of seismic events in Peru by longitude and latitude is carried out, since we want to identify the groups of seismic events in the sampling frame that are related by geographic location, as shown in Figure 3.8. Having groups by geographic location will serve to analyze the behavior of earthquakes in a more detailed way. The groups identified can be described as: North (green), Center (red) and South (blue) of Peru. The departments that make up these groups would be:

- North: Tumbes, Piura, Lambayeque, Amazonas, Cajamarca, Loreto and La Libertad
- Center: San Martín, Ancash, Huánuco, Ucayali, Lima, Pasco and Junín.
- South: Ica, Huancavelica, Arequipa, Ayacucho, Apurímac, Cusco, Madre de Dios, Mo quegua, Tacna and Puno

It should be noted that Madre de Dios does not appear grouped since the records of its earthquakes are outside the sample frame (less than 5.5 Mw). In the next chapter the results of the departmental groups will be shown, to which the Semi-Markov model will be applied.



Source: Geophysical Institute of Peru

Figure 3.8: *Cluster by Latitude and Longitude*

3.8. Application of the Semi-Markov Model to seismic events in Peru

In this section we will proceed to develop the working methodology of the process to calculate the estimate of the waiting time of all seismic events at the Peruvian level from 2008 to 2018 applying the Semi Markov model. Given that the following section intends to detail the behavior of seismic events for the groups of departments, knowing the work methodology at the Peruvian level will allow us to extend the analysis of the estimation of the waiting time for seismic events at the departmental level. In the case of all of Peru, the earthquake in the sample frame took place on 02/19/2008 at 08:29:56 and the last earthquake at 12/06/2018

3.8.1. Estimate of the transition matrix

When ordering the seismic events at the level of Peru, the transition matrix or change of state is assembled, which can be seen in Table 3

State	1	2	3
1	125	82	21
2	82	50	16
3	21	16	11

Table 3: *Transition matrix of seismic events*

By observing Table 3, it is possible to understand in general terms how the changes of state of magnitude have been throughout the period under study. The largest number of seismic events that come to belong to state 3 have state 1 and 2 as the next transition state, this may be quite linked to the fact that after the occurrence of a state 3 seismic event a large amount of concentrated energy is released, which generates that the following seismic events occur with less energy. This premise would explain the reason that there are few seismic events that go from state 3 to state 3.

With the same idea, it is analyzed that the number of seismic events that go from state 1 to state 1 is high since the amount of energy emitted by state 1 events is released little by little. Regarding earthquakes that have state 2 as their initial state, they present a similar behavior to earthquakes with state 3 but with a greater amount of seismic occurrence.

By having the transition matrix, the transition probability matrix is calculated, which measures the percentage or probability of transition from a state i to a state j for all i, j ∈ E. The transition probability matrix is defined What

$$\hat{p}_{ij}(T) = \frac{N_{ij}(T)}{N_j(T)} \tag{3.8.1}$$

The evaluation of $N_i(T)$ for all $i \in E$ become:

$$N_1(T) = 125 + 82 + 21 = 228$$

$$N_2(T) = 82 + 50 + 16 = 148$$

$$N_3(T) = 21 + 16 + 11 = 48$$

Knowing the values of $N_i(T)$ we proceed to the calculation of the transition probability matrix:

State	1	2	3
1	54.82 %	35.97 %	9.21 %
2	55.40 %	33.78 %	10.81 %
3	43.75 %	33.33 %	22.92 %

Table 4: *Transition probability matrix*

$$\hat{P} = \begin{pmatrix} 0{,}5482 & 0{,}3597 & 0{,}0921 \\ 0{,}5540 & 0{,}3378 & 0{,}1081 \\ 0{,}4375 & 0{,}3333 & 0{,}2292 \end{pmatrix}$$

3.8.2. Parameter estimation of the waiting time distribution

After approximating the transition matrix, we proceed to estimate the parameters of the distribution of waiting times. For the present study, the Weibull distribution was used as a function of distribution of waiting times. The input data is the time (calculated in days) that has to elapse for the next seismic event to occur. Where α is the shape parameter and λ is the scale parameter.

Transition	Mean	SD	α	β
1-1	9.4278	9.8422	0.9345	9.1376
1-2	9.2841	12.318	0.7525	7.8224
1-3	11.231	7.9642	1.2014	11.839
2-1	8.2336	8.9290	0.8383	7.5365
2-2	7.8383	8.0911	0.8552	7.3096
2-3	10.823	8.1457	1.2584	11.579
3-1	12.157	13.691	0.8958	11.496
3-2	5.9382	7.9054	0.5148	3.5889
3-3	19.658	17.089	0.7329	17.461

Table 5: *Parameters of the Weibull distribution function*

Analyzing the results, it is observed that the transition with the longest average waiting time is the change from state 2 to state 3. This time state change has an average of 10 days. While the transition with the lowest average waiting time is 3 to 2, which has 5 days of waiting time on average.

3.8.3. Matrix estimation $F_{ij}(t)$

Since F_{ij} is defined as the distribution function associated with the time of stay of state i, before moving to state *j*. From 5 there are 9 pairs of parameters of the Weibull distribution. Using each distribution for each transition, each $F_{ij}(t)$ is calculated at time t. For example, the parameters of the transition pair 1 → 1 will help to calculate the $F_{11}(t)$ at time t, in the same way the parameters of the transition pair 1 → 2 will help to calculate the $F_{12}(t)$ in the time t and so on until we have all the $F_{ij}(t)$ at time *t*.

From Table 5 the estimates of the shape and scale parameters are obtained, which allow expressing the distribution function of each change of state as shown below:

$$
\begin{aligned}
&F_{11}(t) = 1 - e^{-(9{,}1376x)^{0{,}9345}}, && F_{12}(t) = 1 - e^{-(7{,}8224x)^{0{,}7525}}, \\
&F_{13}(t) = 1 - e^{-(11{,}839x)^{1{,}2014}}, && F_{21}(t) = 1 - e^{-(7{,}5365x)^{0{,}8383}}, \\
&F_{22}(t) = 1 - e^{-(7{,}3096x)^{0{,}8552}}, && F_{23}(t) = 1 - e^{-(11{,}579x)^{1{,}2584}} \\
&F_{31}(t) = 1 - e^{-(11{,}496x)^{0{,}8958}}, && F_{32}(t) = 1 - e^{-(3{,}5889x)^{0{,}5148}}, \\
&F_{33}(t) = 1 - e^{-(17{,}461x)^{0{,}7329}} &&
\end{aligned}
$$

Using the distribution functions of the waiting time and given that the waiting times are being worked on in days, the calculation of the matrix $F_{ij}(t)$ in the first 3 days will be carried out for t = 1, 2, 3. What will show the results of the state change distribution functions at time t = 1. Assuming that the distribution function at t = 0 becomes 0 (F (0) = 0).

State	1	2	3
1	0.1188279	0.1916113	0.05004320
2	0.1680256	0.1667826	0.04482269
3	0.1061255	0.4042878	0.11567797

Table 6: Matrix $F_{ij}(t = 1)$

State	1	2	3
1	0.2147676	0.3011714	0.1113598
2	0.2802763	0.2811341	0.1039066
3	0.1883974	0.5229285	0.1847946

Table7: Matrix $F_{ij}(t = 2)$

State	1	2	3
1	0.2975345	0.3850375	0.1748296
2	0.3699879	0.3730537	0.1670212
3	0.2593023	0.5982231	0.2404410

Table 8: Matrix $F_{ij}(t = 3)$

3.8.4. Matrix estimation $Q_{ij}(t)$

The matrix $Q_{ij}(t)$ is defined as $Q(t) = \hat{F}(t) \times \hat{P}$ which becomes the Kernel, which will allow the calculation of the probabilities of permanence and non-permanence of a state *i* in a time *t*.

State	1	2	3
1	0.1932038	0.1241510	0.04312765
2	0.2041358	0.1317166	0.04377846
3	0.3327892	0.2133110	0.07999104

Table 9: Matrix $Q_{ij}(t = 1)$

State	1	2	3
1	0.3333305	0.2161080	0.07786023
2	0.3548829	0.2304144	0.08001973
3	0.4738663	0.3060202	0.11623396

Table 10: Matrix $Q_{ij}(t = 2)$

State	1	2	3
1	0.4529415	0.2953648	0.1090953
2	0.4826080	0.3147712	0.1126837
3	0.5788022	0.3755072	0.1436569

Table 11: Matrix $Q_{ij}(t = 3)$

3.8.5. Matrix estimation $H_{ij}(t)$

Tables 12, 13 and 14 show the probability that the seismic event does not remain in state i during t. That is, the probabilities of leaving state i towards a new state at time t are shown t.

$$H(t) = \begin{cases} 0 \; si \; i \neq j \\ \sum_{j=1}^{m} Q_{ij}(t) \; si \; i = j \end{cases} \tag{3.8.2}$$

State	1	2	3
1	0.3604825	0	0
2	0	0.3796309	0
3	0	0	0.6260913

Table 12: Matrix $H_{ij}(t = 1)$

State	1	2	3
1	0.6272987	0	0
2	0	0.6653171	0
3	0	0	0.8961205

Table 13: Matrix $H_{ij}(t = 2)$

State	1	2	3
1	0.8574015	0	0
2	0	0.9100629	0
3	0	0	1

Tabla 14: Matriz $H_{ij}(t = 3)$

3.8.6. Matrix estimation $D_{ij}(t)$

$$D_{ij} = I - H_{ij}$$

Tables 15, 16 and 17 show the probability that the seismic event remains in state i during time t. That is, the probabilities of the non-occurrence of a new seismic event at time t are shown.

State	1	2	3
1	0.6395175	0	0
2	0	0.6203691	0
3	0	0	0.3739087

Table 15: Matrix $D_{ij}(t = 1)$

State	1	2	3
1	0.3727013	0	0
2	0	0.3346829	0
3	0	0	0.1038795

Table 16: Matrix $D_{ij}(t = 2)$

State	1	2	3
1	0.1425985	0	0
2	0	0.08993713	0
3	0	0	0

Table 17: Matrix $D_{ij}(t = 3)$

The results of the arrays H(t) and D(t), show how over time the probability of permanence of a seismic event decreases within Peruvian territory. As shown in the same tables, the probability that a seismic event remains in state 3 is low, which means that after some seismic event of very high magnitude, a new seismic event is likely to appear within the sample frame, I already know state 1, 2 or 3. Similarly, events with state 1, compared to the other two states, have the least probability of changing state to a new one.

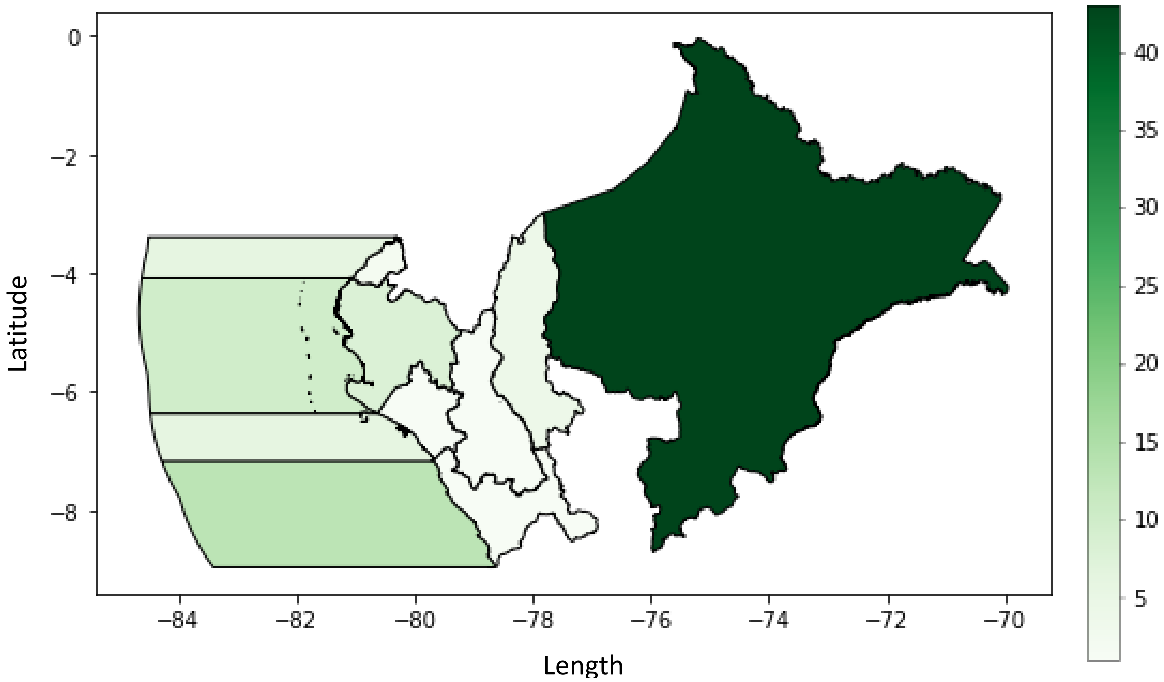
Chapter 4

Presentation and Analysis of Results

As a result of the clustering based on latitude and longitude, 3 clusters or sectors were obtained, which were called the North Zone, the Central Zone and the South Zone. After performing the analysis and application of the Semi Markov model in a differentiated way to each Zone, the following results were obtained for each of these

4.1. North Zone

Within this Zone are the departments of Loreto, Amazonas, Cajamarca, Tumbes, Piura, Lambayeque and La Libertad, as well as the maritime territory off the coasts of these last 4 departments mentioned.



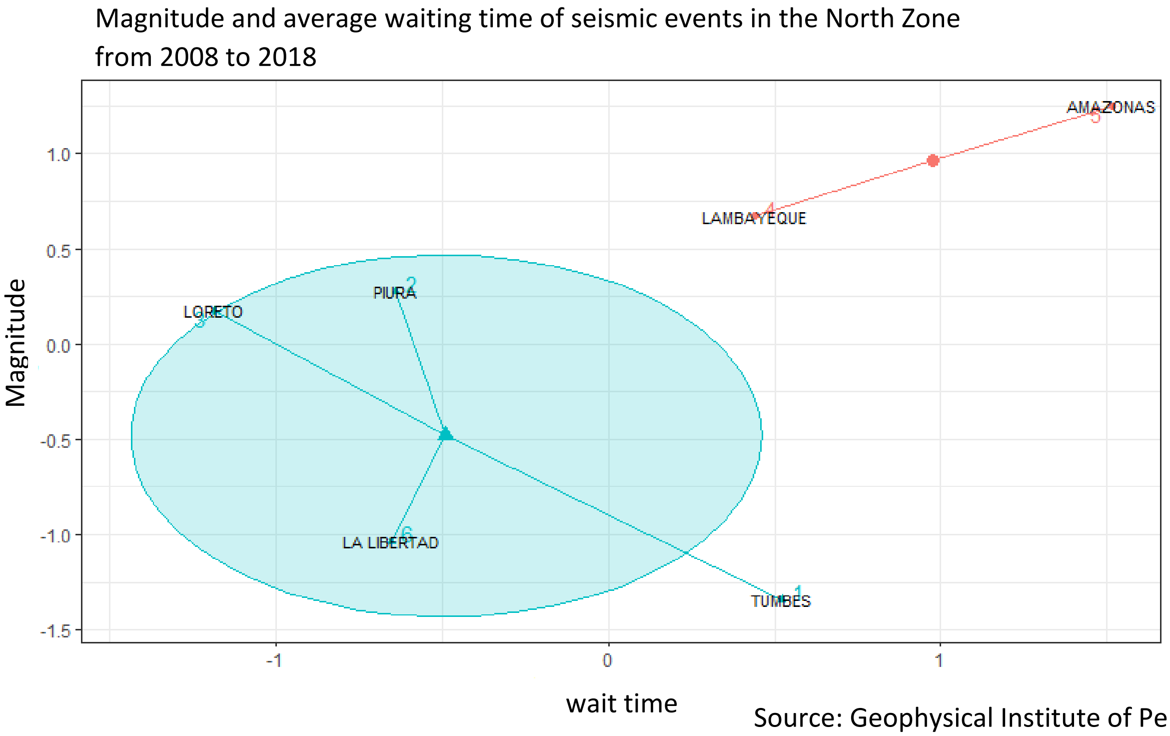
Source: Geophysical Institute of Peru

***Figure 4.1**: Distribution of earthquakes in the North Zone*

As can be seen in Figure 4.1, within this group, the department with the highest number of earthquakes reported in the period of time considered for the study is Loreto, it should be noted that most of the earthquakes that occur in this region are not of high intensity due to the high level of depth of its Hypocenter.

4.1.1. Cluster analysis of seismic events by magnitude and average waiting time in the Northern Zone of Peru

According to the methodology described in the previous chapter, the “K-means” technique will be applied to the northern part of Peru. In Figure 4.2 it is possible to identify two groups of earthquakes that meet similar characteristics within the groups, but different between the groups. The departmental group colored red (which associates Lambayeque and Amazonas) seems to tend to have on average high-magnitude seismic events, but these events do not occur on a recurring basis, which means that few seismic events occur in this group. On the other hand, the blue departmental group (which associates Loreto, Piura, La Libertad and Tumbes) seems to tend to have the most recurrent seismic events in this area, and also the average magnitudes that occur are much lower than those presented. in the red departmental group.


Magnitude and average waiting time of seismic events in the North Zone from 2008 to 2018

Source: Geophysical Institute of Peru

Figure 4.2: *Cluster analysis by department according to the average magnitude and average waiting time of the North Zone*

4.1.2. Estimate of the transition matrix

After ordering the seismic events of the Northern Zone of Peru, the transition matrix or change of state of said zone is assembled, which can be seen in Table 18

State	1	2	3
1	28	18	6
2	18	6	4
3	6	4	4

Table 18: *Transition matrix of North Zone seismic events*

The transitions or changes of state of the seismic events shown in Table 18 will be used to calculate the transition probability matrix.

From the transition matrix, the transition probability matrix is obtained, which is presented in Table 19.

State	1	2	3
1	53.85 %	34.61 %	11.54 %
2	64.28 %	21.43 %	14.29 %
3	42.86 %	28.57 %	28.57 %

Table 19: Matriz de probabilidad de transición Zona Norte

4.1.3. Parameter estimation of the waiting time distribution of the Northern Zone

The statistics and the estimation of the parameters of the waiting time distribution (Weibull Distribution). are shown in Table 20.

Transition	Mean	SD	α	β
1-1	14.422	15.708	0.6843	11.707
1-2	16.713	14.406	0.9676	16.506
1-3	10.099	12.663	1.0277	10.237
2-1	15.791	9.9370	1.7529	17.854
2-2	5.2060	3.1330	1.9144	5.8858
2-3	2.0730	1.5200	1.7060	2.3454
3-1	9.8340	8.4230	1.2428	10.553
3-2	7.4260	6.9770	1.1931	7.8900
3-3	13.705	8.9960	1.9178	15.596

Table 20: *Parameters of the Weibull distribution function for the waiting times of the North Zone*

In Table 20 you can find the average transition times of seismic events, where the shortest average transition time is 2 days that occurs in the change of state from 2 to 3, in addition, it contains the smallest standard deviation compared to the rest of states. On the other hand, the longest average time occurs in the transition from 1 to 2.

4.1.4. Matrix estimation F$_{ij}$(t)

From Table 20 the estimates of the shape and scale parameters are obtained, which allow expressing the distribution function of each change of state as shown below:

$$
\begin{array}{ll}
F_{11}(t) = 1 - e^{-(11{,}707x)^{0{,}6843}}, & F_{12}(t) = 1 - e^{-(16{,}506x)^{0{,}9676}}, \\
F_{13}(t) = 1 - e^{-(10{,}237x)^{1{,}0277}}, & F_{21}(t) = 1 - e^{-(17{,}854x)^{1{,}7529}}, \\
F_{22}(t) = 1 - e^{-(5{,}8858x)^{1{,}9144}}, & F_{23}(t) = 1 - e^{-(2{,}3454x)^{1{,}7060}} \\
F_{31}(t) = 1 - e^{-(10{,}553x)^{1{,}2428}}, & F_{32}(t) = 1 - e^{-(7{,}8900x)^{1{,}1931}}, \\
F_{33}(t) = 1 - e^{-(15{,}596x)^{1{,}9178}} &
\end{array}
$$

Times are analyzed t = 1, 2, 3.

State	1	2	3
1	0.1073666	0.2351698	0.07557583
2	0.1414523	0.2601101	0.16170334
3	0.3519335	0.2007477	0.08238706

Table 21: Matrix $F_{ij}(t = 1)$

State	1	2	3
1	0.25803685	0.1216800	0.17033349
2	0.02132071	0.1189579	0.53326122
3	0.11887947	0.1767227	0.01928112

Table 22: Matrix $F_{ij}(t = 2)$

State	1	2	3
1	0.32557048	0.1747558	0.24667748
2	0.04292018	0.2406073	0.78168872
3	0.18899290	0.2705418	0.04148533

Table 23: Matrix $F_{ij}(t = 3)$

4.1.5. Matrix estimation $Q_{ij}(t)$

The matrix $Q_{ij}(t)$ is defined as $Q(t) = F(t) \times \hat{P}$

State	1	2	3
1	0.16678805	0.11182335	0.04260904
2	0.11292320	0.08288136	0.05188982
3	0.07597172	0.04798658	0.01478848

Table 24: Matrix $Q_{ij}(t = 1)$

State	1	2	3
1	0.2834058	0.1906887	0.07595591
2	0.3108999	0.2256102	0.13702976
3	0.3639141	0.1088856	0.03447318

Table 25: Matrix $Q_{ij}(t = 2)$

State	1	2	3
1	0.3832381	0.2583561	0.10540952
2	0.4988291	0.3572854	0.20910177
3	0.2716592	0.1731988	0.05616206

Table 26: Matrix $Q_{ij}(t = 3)$

4.1.6. Matrix estimation $H_{ij}(t)$

$$H(t) = \begin{cases} 0 \; si \; i \neq j \\ \sum_{j=1}^{m} Q_{ij}(t) \; si \; i = j \end{cases} \tag{4.1.1}$$

State	1	2	3
1	0.3212204	0	0
2	0	0.2476944	0
3	0	0	0.1387468

Table 27: Matrix $H_{ij}(t = 1)$

State	1	2	3
1	0.5500504	0	0
2	0	0.6735398	0
3	0	0	0.3148833

Table 28: Matrix $H_{ij}(t = 2)$

State	1	2	3
1	0.7470037	0	0
2	0	1	0
3	0	0	0.5010201

Table 29: Matrix $H_{ij}(t = 3)$

4.1.7. Matrix estimation $D_{ij}(t)$

$$D_{ij} = I - H_{ij}$$

State	1	2	3
1	0.6395175	0	0
2	0	0.6203691	0
3	0	0	0.8612532

Table 30: Matrix $D_{ij}(t = 1)$

State	1	2	3
1	0.4499496	0	0
2	0	0.3264602	0
3	0	0	0.6851167

Table 31: Matrix $D_{ij}(t = 2)$

State	1	2	3
1	0.2529963	0	0
2	0	0	0
3	0	0	0.4989799

Table 32: Matriz $D_{ij}(t = 3)$

4.1.8. Situation of seismic activity in the North Zone

According to Table 19, the highest percentage of transitions have state 1 as their final state. In addition, the mean inter-arrival times vary markedly when comparing the 9 state transitions, this suggests that the times of the following seismic events are closely linked to the first state. For example, according to the results of Table 20, the average transition time 2-1 is approximately 16 days, transition 2-2 has approximately 5 days, and finally transition 2-3 has an average transition time of 2 days. This allows us to think that after the occurrence of the event of state 2 it is expected that as the days increase, the next event to occur will be more likely to be in state 1. The same table also shows the estimated parameters of the Weibull distribution for each one of the 9 transitions that allowed the application of the semi-Markov model.

The results obtained when applying the semi-Markov model in the Northern Zone of Peru and for times t = 1, t = 2 and t = 3 are represented in the arrays $H_{ij}(t)$ shown in Tables 21, 22 and 23, which indicate the probability the event changes state. It is known that by increasing the number of days it is more likely that a new seismic event occurs within the territory, in the case of the North Zone the probability increases very quickly, it is observed that already on day 3, having as the initial state the state 2, the probability of state change becomes 1. This converges with Table 20, which shows the average level of days for a state change to occur. In the same way, it is understood as the probability of staying in the same state ($D_{ij}(t)$ decreases over the days.

It should be remembered that the study works with seismic events of magnitudes greater than or equal to 5.3, which present slight damage in the best of cases and are terribly destructive in the worst of cases, but are always perceived, that is, that this area, according to the results of the Semi-Markov model, it presents a high recurrence of seismic events of considerable magnitude. This is demonstrated by the prevention policies presented by each department according to its level of seismic risk, which are managed by the authorities themselves. For example:

- **Tumbes**: Among the actions that are proposed within the plan for this region, we can mainly find: (1) Preparation of risk assessment studies in critical areas, (2) Implement communication campaigns for the prevention and reduction of disaster risk, directed to the high and very high-risk population and (3) Promote the updating of the Economic - Ecological Zoning studies of the department of Tumbes, incorporating territorial studies of disaster risk.

- **Piura:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Generate technical information for the Prevention and Reduction of Disaster Risk, (2) Strengthen capacities for impact evaluation socioeconomic in the disaster scenario and (3) Develop mechanisms to implement prevention measures and reduction of Regional Impact.

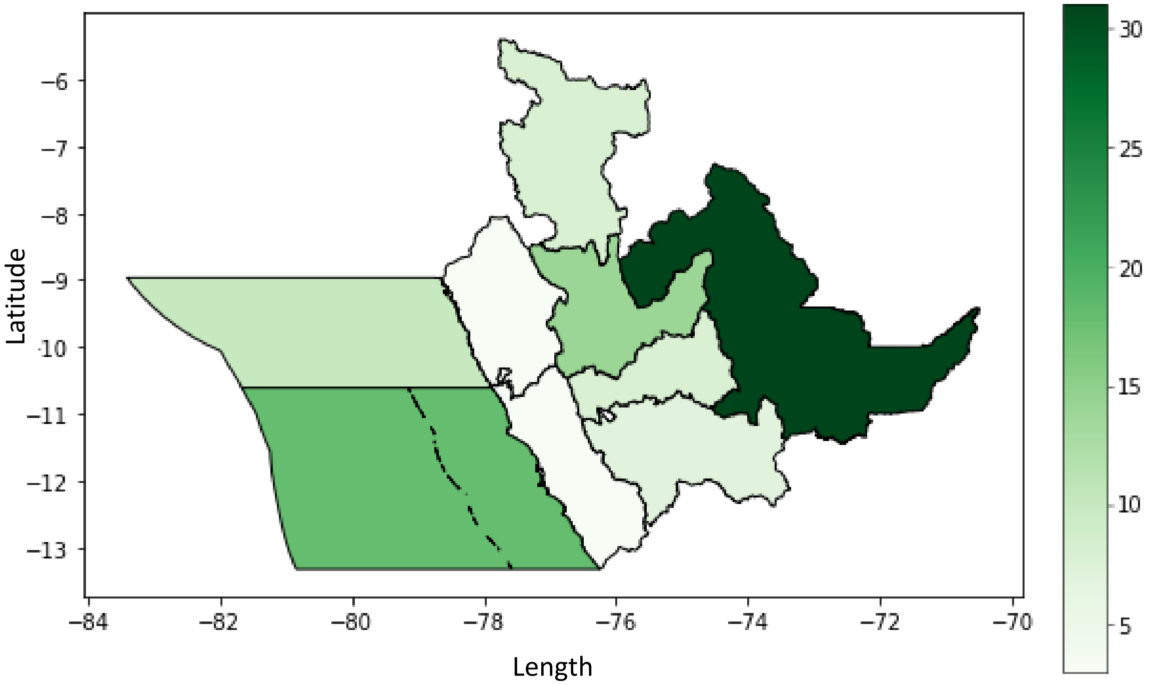
- **Lambayeque:** Among the actions that are proposed within the plan of this region can be found mainly: (1) Carry out seismic microzoning studies in the districts with the highest urban population concentration, (2) Encourage the formulation of urban development plans and zoning schemes urban at the district level, that regulate the use of land restricting its use based on water and seismic risk and (3) Implement training for the population to comply with technical construction standards as a safety measure.

- **Amazonas:** Among the actions that are proposed within the risk prevention and reduction plan for this region we can mainly find: (1) Promote a culture of prevention and citizen participation to promote prospective, corrective and reactive management of Disaster Risk Management (DRM) in the Amazonas department, (2) Priority risk prevention and reduction works in the Amazonas department and (3) Promote a culture of prevention and citizen participation to promote prospective, corrective and reactive management of Disaster Risk Management (DRM) in the department of Amazonas.

- **La Libertad:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Promote territorial planning and management processes that do not generate new disaster risks, (2) Strengthen the culture of prevention in civil society for sustainable and healthy urban development and (3) Promote at the level of regular basic education and higher level the incorporation in the curricular programming of contents on prospective and corrective DRM.

4.2. Central Zone

Within this Zone are the departments of Ucayali, San Martín, Huánuco, Pasco, Junín, Ancash and Lima, as well as the maritime territory off the coasts of these last 2 departments mentioned.



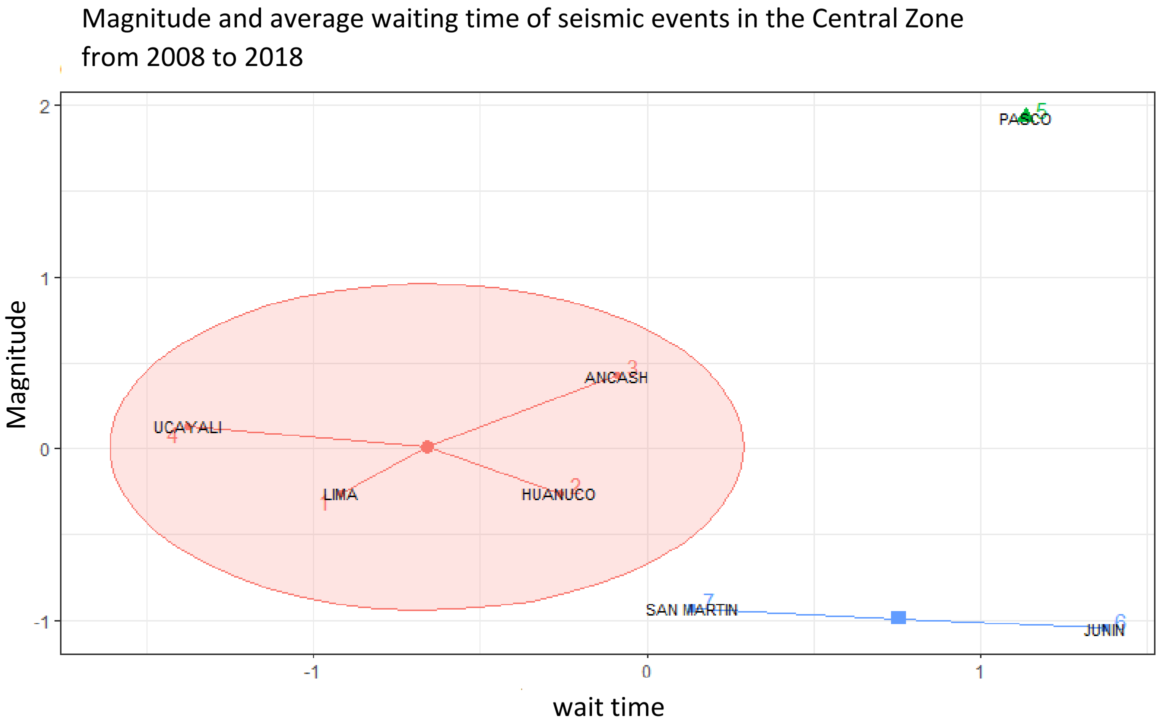
Source: Geophysical Institute of Peru

Figure 4.3: *Distribution of earthquakes in the Central Zone*

As can be seen in Figure 4.3, within this group, the department with the highest number of earthquakes reported in the period of time considered for the study is Ucayali, it should be noted that the concentration of earthquakes that occur in this region is in the zone attached to Huánuco and that these tend to have very high depth levels.

4.2.1. Cluster analysis of seismic events by magnitude and average waiting time

According to the methodology described in the previous chapter, the “K-means” technique will be applied to the central zone of Peru. In Figure 4.4 it is possible to identify 3 groups of earthquakes that meet similar characteristics within the groups, but different between the groups. The departmental group of red color (which associates Ucayali, Lima, Huánuco and Ancash) seems that it tends to have on average seismic events with levels of medium magnitude, that is to say that on average the magnitudes of these departments are at a level above the departments with lower magnitudes and below the departments with very high magnitudes, in addition, these departments tend to have the highest seismic event recurrence in the area. The department of Pasco is only green, since there is no other department in the central zone that presents the same average behavior of magnitude and waiting time. The departmental group of blue color (which associates San Martin and Junín) come to present high levels of average inter-arrival time, that is to say that it is not very recurrent that earthquakes occur in these departments, in addition to that the magnitude levels of earthquakes reach not be tall.



Source: Geophysical Institute of Peru

***Figure 4.4:** Cluster analysis by department according to the average magnitude and the average waiting time of the Central Zone*

4.2.2. Estimate of the transition matrix

After ordering the seismic events of the Central Zone of Peru, the transition matrix or change of state of said zone is assembled, which can be seen in Table 33.

State	1	2	3
1	26	22	3
2	22	19	3
3	3	3	0

***Table 33:** Transition matrix of the Central Zone seismic events*

The transitions or changes of state of the seismic events shown in Table 33 will be used to calculate the transition probability matrix.

From the transition matrix, the transition probability matrix is obtained, which is presented in Table 34.

State	1	2	3
1	50.98 %	43.14 %	5.88 %
2	50.00 %	43.18 %	6.82 %
3	50.00 %	50.00 %	0 %

Table 34: *Transition probability matrix of the Central Zone*

4.2.3. Parameter estimation of the waiting time distribution in the Central Zone

The statistics and the estimation of the parameters of the waiting time distribution (Weibull Distribution). are shown in Table 35.

Transition	Mean	SD	α	β
1-1	15.241	15.334	1.0095	15.306
1-2	10.648	15.014	0.6118	7.4339
1-3	11.294	6.4260	2.3578	12.743
2-1	12.093	10.461	1.0595	12.349
2-2	15.462	12.004	0.9346	15.115
2-3	14.925	5.0140	3.8983	16.513
3-1	19.427	10.490	2.4832	22.026
3-2	5.8980	5.0680	1.4844	6.5671

Table 35: Parameters of the Weibull distribution function for the Central Zone wait times

In Table 35 you can find the average transition times of seismic events, where the shortest average transition time is 6 days that occurs in the change of state from 3 to 2, in addition, it contains the smallest standard deviation compared to the rest of states. On the other hand, the longest average time occurs in the transition from 3 to 1, with an average time of 19 days.

4.2.4. Matrix estimation $F_{ij}(t)$

From Table 35 the estimates of the shape and scale parameters are obtained, which allow expressing the distribution function of each change of state as shown below:

$$\begin{array}{ll} F_{11}(t) = 1 - e^{-(15{,}306x)^{1{,}0095}}, & F_{12}(t) = 1 - e^{-(7{,}4339x)^{0{,}6118}}, \\ F_{13}(t) = 1 - e^{-(12{,}743x)^{2{,}3578}}, & F_{21}(t) = 1 - e^{-(12{,}349x)^{1{,}0595}}, \\ F_{22}(t) = 1 - e^{-(15{,}115x)^{0{,}9346}}, & F_{23}(t) = 1 - e^{-(16{,}513x)^{3{,}8983}} \\ F_{31}(t) = 1 - e^{-(22{,}026x)^{2{,}4832}}, & F_{32}(t) = 1 - e^{-(6{,}5671x)^{1{,}4844}} \end{array}$$

Times are analyzed t = 1, 2, 3

State	1	2	3
1	0.06167515	0.25401783	0.002473881
2	0.06735911	0.07598170	0.000017888
3	0.00046253	0.05936012	0

Table 36: Matrix $F_{ij}(t = 1)$

State	1	2	3
1	0.12028855	0.3609987	0.0126166406
2	0.13527061	0.1401853	0.0002666951
3	0.00258343	0.1573643	0

Table 37: Matrix $F_{ij}(t = 2)$

State	1	2	3
1	0.175506227	0.4367028	0.032489353
2	0.200147587	0.1979829	0.001294919
3	0.007054855	0.2684322	0

Table 38: Matrix $F_{ij}(t = 3)$

4.2.5. Matrix estimation $Q_{ij}(t)$

The matrix $Q_{ij}(t)$ is defined as $Q(t) = F(t) \times \hat{P}$

State	1	2	3
1	0.15968809	0.1375314	0.020947348
2	0.07233974	0.0618761	0.009142872
3	0.02991586	0.0258323	0.004074489

Table 39: Matrix $Q_{ij}(t = 1)$

State	1	2	3
1	0.24813126	0.21408332	0.03168935
2	0.13918747	0.11901992	0.01751518
3	0.07999918	0.06906718	0.01088135

Table 40: Matrix $Q_{ij}(t = 2)$

State	1	2	3
1	0.3240698	0.2805295	0.04009909
2	0.2016750	0.1724783	0.02527222
3	0.1378127	0.1189572	0.01871718

Table 41: Matrix $Q_{ij}(t = 3)$

4.2.6. Matrix estimation $H_{ij}(t)$

$$H(t) = \begin{cases} 0 \; si \; i \neq j \\ \sum_{j=1}^{m} Q_{ij}(t) \; si \; i = j \end{cases} \tag{4.2.1}$$

State	1	2	3
1	0.3181669	0	0
2	0	0.1433587	0
3	0	0	0.05982266

Table 42: Matrix $H_{ij}(t = 1)$

State	1	2	3
1	0.4939039	0	0
2	0	0.2757226	0
3	0	0	0.1599477

Table 43: Matrix $H_{ij}(t = 2)$

State	1	2	3
1	0.6446984	0	0
2	0	0.3994255	0
3	0	0	0.275487

Table 44: Matrix $H_{ij}(t = 3)$

4.2.7. Matrix estimation $D_{ij}(t)$

$$D_{ij} = I - H_{ij}$$

State	1	2	3
1	0.6818331	0	0
2	0	0.8566413	0
3	0	0	0.9401773

Table 45: Matrix $D_{ij}(t = 1)$

State	1	2	3
1	0.5060961	0	0
2	0	0.7242774	0
3	0	0	0.8400523

Table 46: Matrix $D_{ij}(t = 2)$

State	1	2	3
1	0.3553016	0	0
2	0	0.6005745	0
3	0	0	0.724513

Table 47: Matrix $D_{ij}(t = 3)$

4.2.8. Situation of seismic activity in the Central Zone

According to Table 34, the highest percentage of transitions have state 1 and 2 as their final state, which, as shown even in Table 34, is common that after some simian event, the next state is 1 or 2, during the time period of the study frame there has not been any seismic event that has a 3-3 transition.

The mean of the inter-arrival times between the 8 transition states has an almost similar behavior in each transition with the exception of the 3-2 transition as shown in Table 35. Based on this detail, it is observed that the transitions with initial state 3, they are very different from each other. For example, there are no 3-3 state transitions, the 3-2 state transition has an average time of 6 days, and lastly, the 3-1 state transitions have about 19 days average time. This allows us to think that after several days of the seismic event have passed, the next event is likely to be found in state 1.

The same table also shows the estimated parameters of the Weibull distribution for each of the 8 transitions that will allow the application of the semi-Markov model.

The results obtained when applying the semi-Markov model in the Central Zone of Peru and for the times t = 1, t = 2 and t = 3 are represented in the arrays Hij(t) are listed in Tables 36, 37 and 38, which indicate the probability that the event changes state. It is known that by increasing the number of days it is more likely that a new seismic event occurs within the territory, in the case of the Central Zone the probability increases more slowly than in the North Zone. It is observed that state 1 is the most prone to change state with probability 0.6446984, while state 3 is the least prone to change state, with probability 0.275487. This indicates that if there is any seismic occurrence in the Central Zone, the probabilities will be low in the first three days, in other words, the frequency of earthquakes (or the time between earthquakes) is much lower (greater) compared to the North Zone. In the same way, the probabilities that a seism will remain in the same state ($D_{ij}(t)$) are relatively high the first 3 days.

It should be remembered that the study works with seismic events of magnitudes greater than or equal to 5.3, which present slight damage in the best of cases and are terribly destructive in the worst of cases, although they are always perceived by people, that is, that this area, according to the results of the Semi-Markov model, presents low recurrence of seismic events of considerable magnitude, although that does not imply that there are prevention plans, especially in the departments that are on the coastal part, since there is still a high risk since the Central Zone can be affected by the intensity of earthquakes emitted by the South and North Zone. This is demonstrated by the prevention policies presented by each department according to its level of seismic risk, which are managed by the authorities themselves. For example:

- **San Martín:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Preparation of emergency technical sheets for rehabilitation and reconstruction of public infrastructures and (2) Installation of the early warning service in the region.

- **Ancash:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Promote studies and research related to DRM, (2) Prepare and implement the territorial ordering and management process and related and

(3) Strengthen and implement Early Warning Systems by type of hazard to prevent disaster risks.

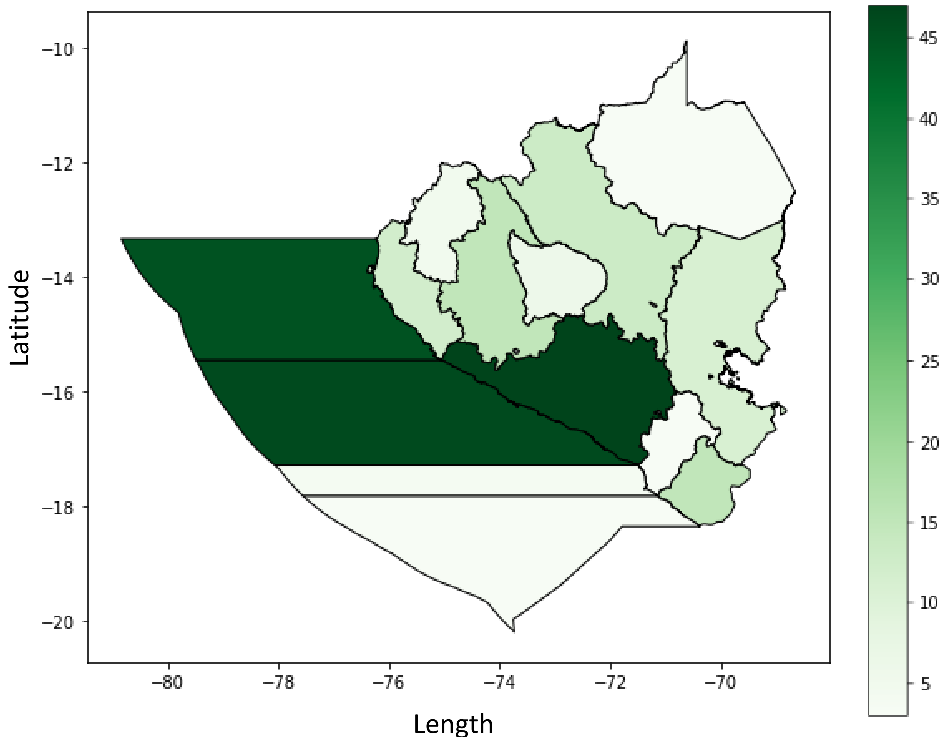
- **Ucayali:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Prepare risk assessment reports for the critical areas identified in the Ucayali region according to type of danger, prioritizing intervention on the areas that present greater susceptibility and greater exposure and (2) Incorporate specific technical criteria for the evaluation and prioritization of investment projects related to disaster risk management.

- **Lima:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Manage the proper use and occupation of the territory incorporating disaster risk management, (2) Promote participation of the population in disaster risk management and (3) Develop studies to establish the level of disaster risk at the territorial level in the face of the possible impact of the main recurring dangers.

4.3. South Zone

Within this Zone are the departments of Madre de Dios, Cusco, Puno, Apurímac, Ayacucho, Huancavelica, Ica, Arequipa, Moquegua and Tacna, as well as the maritime territory off the coasts of these last 4 departments mentioned.

As can be seen in Figure 4.5, within this group, the department with the highest number of earthquakes reported in the period of time considered for the study is Arequipa, followed by the maritime territory off the coasts of Ica and Arequipa, this group is the one that concentrates the highest amount of seismic occurrence with a magnitude greater than 5.3 Mw reported from 2008 to 2018. It should be noted that although it is true that the territory of Ica is not portrayed as one of the locations with the highest seismic activity, it has always been heard that such a department frequently presents earthquakes, this is due to the fact that a large percentage of the earthquakes that are felt in Ica they occur in the sea off its coast.

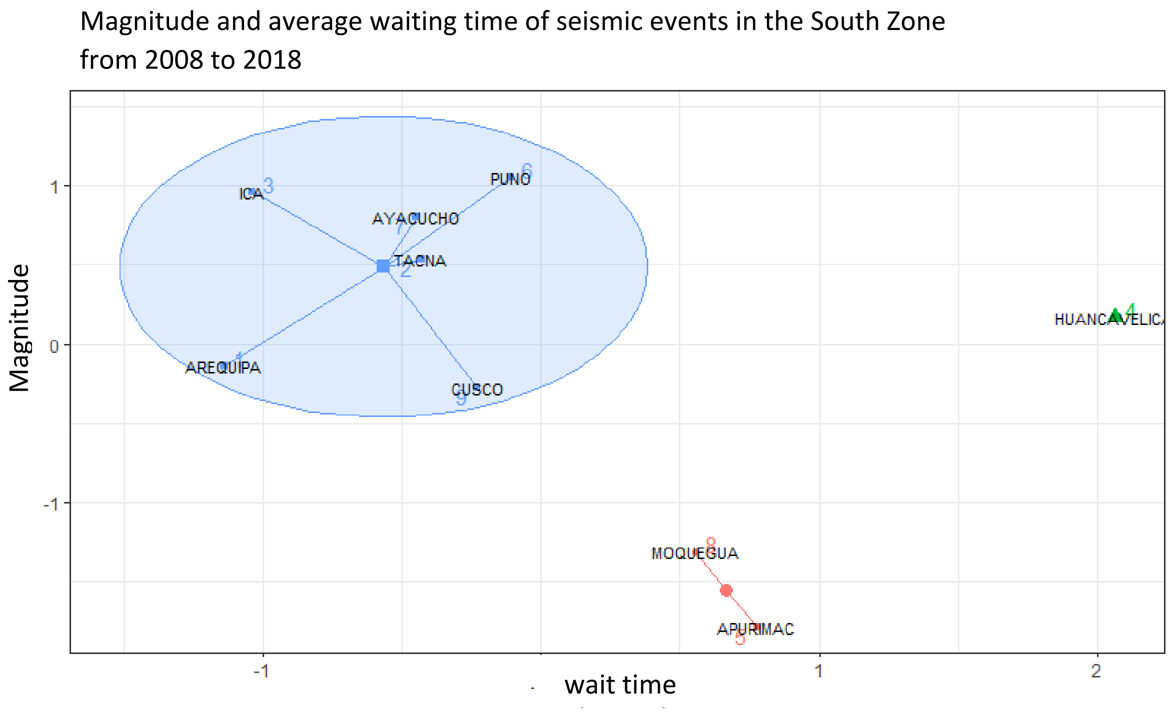


Source: Geophysical Institute of Peru

Figure 4.5: *Distribution of earthquakes in the South Zone*

4.3.1. Cluster analysis of seismic events by magnitude and average waiting time

According to the methodology described in the previous chapter, the “K-means” technique will be applied to the southern zone of Peru. In Figure 4.6 it is possible to identify 3 groups of earthquakes that meet similar characteristics within the groups, but different between the groups. The departmental group of blue color (which associates Ica, Arequipa, Tacna, Ayacucho, Puno and Cusco) seems to have on average seismic events with high magnitude levels compared to the other groups, in addition to having a low average level of time of change of state, this means that this group presents earthquakes with high levels and at the same time recurring, which is very dangerous. The department of Huancavelica is only green, since there is no other department in the southern zone that presents the same average behavior of magnitude and waiting time, this department does not have high levels of magnitude and also the recurrence of earthquakes is low. The departmental group of red color (which associates Moquegua and Apurímac) comes to present low levels of average magnitude, to be more exact they are the lowest levels of the entire southern zone, in addition to also having low recurrence of seismic events.



Source: Geophysical Institute of Peru

Figure 4.6: *Cluster analysis by department based on average magnitude and average wait time*

4.3.2. Estimate of the transition matrix

After ordering the seismic events of the Central Zone of Peru, the transition matrix or change of state of said zone is assembled, which can be seen in Table 48.

State	1	2	3
1	71	36	14
2	37	30	8
3	13	9	6

Table 48: *Transition matrix of South Zone seismic events*

The transitions or changes of state of the seismic events shown in Table 33 will be used to calculate the transition probability matrix. From the transition matrix, the transition probability matrix is obtained, which is presented in Table 34.

State	1	2	3
1	58.678 %	29.75 %	11.57 %
2	49.33 %	40.00 %	10.67 %
3	46.43 %	32.14 %	21.43 %

Table 49: *Transition probability matrix of the South Zone*

4.3.3. Parameter estimation of the waiting time distribution

The statistics and the estimation of the parameters of the waiting time distribution (Weibull Distribution). are shown in Table 50.

Transition	Mean	SD	α	β
1-1	10.044	11.764	0.7234	8.4066
1-2	10.280	8.5490	0.8822	9.8012
1-3	6.2690	7.4490	0.5618	4.3550
2-1	12.486	12.741	0.9178	12.009
2-2	14.520	14.123	0.8785	13.700
2-3	6.5770	5.5150	1.0893	6.7648
3-1	12.322	12.387	0.9549	12.075
3-2	13.641	11.078	1.2827	14.759
3-3	11.266	14.105	0.5702	7.4987

Table 50: *Parameters of the Weibull distribution function for the South Zone wait times*

In Table 50 you can find the average transition times of the seismic events of the South zone, where the shortest average transition time is 6 days that occurs in the change of state of 1-3 compared to the rest of the states. On the other hand, the longest average time occurs in the transition from 2 to 2, with an average time of 19 days. These transition averages are the lowest of the 3 Zones.

4.3.4. Matrix estimation $F_{ij}(t)$

From Table 50 the estimates of the shape and scale parameters are obtained, which allow expressing the distribution function of each change of state as shown below:

$$
\begin{array}{ll}
F_{11}(t) = 1 - e^{-(8{,}4066x)^{0{,}7234}}, & F_{12}(t) = 1 - e^{-(9{,}8012x)^{0{,}8822}}, \\
F_{13}(t) = 1 - e^{-(4{,}3550x)^{0{,}5618}}, & F_{21}(t) = 1 - e^{-(12{,}009x)^{0{,}9178}}, \\
F_{22}(t) = 1 - e^{-(13{,}700x)^{0{,}8785}}, & F_{23}(t) = 1 - e^{-(6{,}7648x)^{1{,}0893}} \\
F_{31}(t) = 1 - e^{-(12{,}075x)^{0{,}9549}}, & F_{32}(t) = 1 - e^{-(14{,}759x)^{1{,}2827}} \\
F_{33}(t) = 1 - e^{-(7{,}4987x)^{0{,}5702}} &
\end{array}
$$

The times t = 1, 2, 3. are analyzed

State	1	2	3
1	0.19291708	0.12497155	0.3543866
2	0.09709723	0.09544241	0.1171613
3	0.08848368	0.03115713	0.2716977

Table 51: Matrix $F_{ij}(t = 1)$

State	1	2	3
1	0.2980440	0.21812896	0.4757924
2	0.1754947	0.16841479	0.2329084
3	0.1643924	0.07412056	0.3754391

Table 52: Matrix $F_{ij}(t = 2)$

State	1	2	3
1	0.3778170	0.2966391	0.5556251
2	0.2441959	0.2315194	0.3379385
3	0.2324250	0.1215080	0.4474063

Table 53: Matrix $F_{ij}(t = 3)$

4.3.6. Matrix estimation $H_{ij}(t)$

$$H(t) = \begin{cases} 0 \; si \; i \neq j \\ \sum_{j=1}^{m} Q_{ij}(t) \; si \; i = j \end{cases} \tag{4.3.1}$$

State	1	2	3
1	0.6722753	0	0
2	0	0.309701	0
3	0	0	0.3913385

Table 57: Matrix $H_{ij}(t = 1)$

State	1	2	3
1	0.9919654	0	0
2	0	0.5768178	0
3	0	0	0.6139521

Table 58: Matrix $H_{ij}(t = 2)$

State	1	2	3
1	1	0	0
2	0	0.8136539	0
3	0	0	0.8013394

Table 59: Matrix $H_{ij}(t = 3)$

4.3.7. Matrix estimation $D_{ij}(t)$

$$D_{ij} = I - H_{ij}$$

State	1	2	3
1	0.3277247	0	0
2	0	0.690299	0
3	0	0	0.6086615

Table 60: Matrix $D_{ij}(t = 1)$

State	1	2	3
1	0.008034572	0	0
2	0	0.4231822	0
3	0	0	0.3860479

Table 61: Matrix $D_{ij}(t = 2)$

State	1	2	3
1	0	0	0
2	0	0.1863461	0
3	0	0	0.1986606

Table 62: Matrix $D_{ij}(t = 3)$

4.3.8. Situation of seismic activity in the South Zone

According to Table 49, the highest percentage of transitions have state 1 and 2 as their final state, which, as shown even in Table 34, is common that after some seismic event, the next state is 1, with the exception of transitions with initial state 2, where the recurrence 2-1 and 2-2 is similar.

The average of the times between the 9 transition states has an almost similar behavior in each transition with the exception of transitions 1-3 and 2-3 as shown in Table 50, although the difference is not much. The South Zone presents the most stable average levels of each transition compared to the North Zone and the Central Zone. For example, the average transition time from an event with initial state 1 is approximately 10 days, when the initial state is 2 the average time is approximately 12 days. Finally, the same happens when the initial state is 3, the average time is approximately 12 days. The same table also shows the estimated parameters of the Weibull distribution for each of the 9 transitions of the South Zone that will allow the application of the semi-Markov model.

The results obtained by applying the semi-Markov model in the Southern Zone of Peru for the times t = 1, t = 2 and t = 3 are represented in the arrays $H_{ij(}$t) shown in Tables 51, 52 and 53, which indicate the probability the event changes state. It is known that as the number of days increases, a new seismic event is more likely to occur within the territory, in the case of the South Zone, the probabilities of change of state increase very quickly over the days, even faster than in the North Zone. It is observed that on day 2, the probabilities of non-permanence are quite high, which means that for day 3 the probability that state 1 changes state is 1. This indicates that by day 3 some change is already expected state, or in other words, some new seismic occurrence, regardless of whether the initial state is 1, 2 or 3 In the same way, it is understood that the probabilities that a seism remains in the same state (D_{ij}(t)) they are quite low, even from day 2 (t = 2).

It should be remembered that the study works with seismic events of magnitudes greater than or equal to 5.3, which present slight damage in the best of cases and are terribly destructive in the worst of cases. In particular, according to the results of the Semi-Markov model, the South zone presents the most recurrent levels of earthquakes in the entire region, surpassing the Central zone and the North zone. This area presents high levels of frequency of seismic events of considerable magnitude, and it is because this area is affected by geological faults and the collision of the tectonic plates: the Mediterranean plate and the Nazca plate. That is why, even more so, there are different public policies that are implemented in the departments of this area, especially those that are located in the coastal area since a tsunami can occur at any time. Public policies are managed by the authorities themselves. For example:

- **Ica:** Among the actions that are proposed within the risk prevention and reduction plan for this region we can find mainly: (1) Creation of the Early Warning System in the Ica Region, (2) Implement earthquake-resistant norms in design and construction of housing and (3) Promote studies that allow the identification of hazards and vulnerabilities to calculate the risk.

- **Arequipa:** Among the actions that are proposed within the risk prevention and reduction plan for this region, we can mainly find: (1) Comprehensive management of the territory and relocate populations located in vulnerable areas to rural urban centers, (2) Implement policies of resilient cities at the municipal level and promote the adoption of technologies that reduce vulnerability for the construction of infrastructure and (3) develop disaster risk management content in the Arequipa region in the school and higher curricula.

- **Ayacucho:** Among the actions that are proposed within the risk prevention and reduction plan for this region we can mainly find: (1) Promote the specialization of the technical staff of the municipality in Disaster Risk Assessment and (2) Raise awareness and awareness among the population exposed and vulnerable with respect to the risks to which it is exposed and the mechanisms of citizen participation.

- **Apurímac:** Among the actions that are proposed within the plan for this region we can find mainly: (1) Formulate the Community Education Plan in DRM and (2) Implement the contents of DRM in the basic and higher education curriculum.

- **Madre de Dios:** Among the actions that are proposed within the plan for this region we can mainly find: (1) Formulate technical normative instruments for the management and occupation of the territory, (2) Inform and sensitize the exposed and vulnerable population to the impact of the identified dangers and studied and (3) Build working groups for DRM.

- **Moquegua:** Among the actions that are proposed within the plan for this region we can mainly find: (1) Develop knowledge of risk and (2) Avoid and reduce risk conditions for the livelihoods of the population with a territorial approach.

- **Tacna:** Among the actions that are proposed within the plan for this region, we can mainly find: (1) Execute earthquake risk assessment studies and (2) Avoid and reduce risk conditions for the population's livelihoods with a focus on territorial.

- **Puno:** Among the actions that are proposed within the plan for this region we can mainly find: (1) Formulate the Community Education Plan in DRM and (2) Identify and evaluate the existing risks in the Puno region in the face of the possible impact of the main dangers recurring.

Conclusions

The results obtained from the seismic events throughout Peru show the behavior of the magnitude states between 2008 and 2018. The transition matrix (Table 3) allowed us to observe that there have been a large number of seismic events. strong magnitudes, since it is working with earthquakes of magnitude greater than and equal to 5.3. The transition matrix (Table 3) together with the frequency matrix of the time of stay (Table 6, 7, 8) generated the probabilities of change of state for (t = 1, t = 2 and t = 3), which show as the probability of state change increases as the days progress. State 1, has probabilities of 0.36048, 0.62729 and 0.85740 for the next 3 days respectively. This indicates that there is a high probability that a new seismic event will occur on the third day. In a simulated way, it occurs with state 2 and 3, where its probabilities on day 3 are 0.91 and 1 respectively, which indicates that it is almost certain that on the third day (after a state earthquake has occurred) a new seismic event will appear, and with much more reason this applies to state 3 which indicates that it is totally probable that in Peru. These high levels of probability of seismic events are closely aligned with those indicated by [Castillo Aedo, 1994], which indicates that the seismic risk of Peru is given by the characteristics of the territory such as the land morphology and the location in which it is found. Therefore, the Semi-Markov model has given good results for the seismic behavior of Peru.

For the North, Center and South Zones, the results after the application of the Semi-Markov model showed that the South Zone is the one with the highest seismic occurrence, since it contains the highest probabilities of significant change of state with the passing of the days (Table 57, 58, 59), state 1 is the most common in this area, although the other two states have relatively similar results. The South Zone contains precisely the departments of Ica and Arequipa, which are those with the highest degree of seismic occurrence, especially in the sea, as shown in Figures 3.8. This area has approximately 10 earthquakes per day (Table 50). That is why the prevention plans in the South Zone are the strictest in Peru.

The descriptive analysis used in Figure 3.4 shows that the area that receives the highest levels of seismic magnitude is the southern area of Peru, which is in line with the results obtained by the Semi-Markov model. The southern zone transition matrix (Table 48) contains the highest number of state 3 seismic events (magnitude greater than 6) compared to the other two zones. If we add to it the high seismic recurrence that it has, the degree of danger that the seismic events of the southern zone of Peru contain is observed. This is mainly due to the fact that the most significant surface faults in the territory are found in the southern zone. It should be remembered that in Ica the seismic events are so strong in intensity that they are even felt in Lima, as is the case of the Pisco earthquake of 2007 that caused great disasters. In addition, without considering the Pisco earthquake, the department of Ica continues to be part of the area with higher magnitude seismic events.

The results of the H_{ij} tables of the Semi-Markov model found in Peru and in each corresponding zone of Peru, allow us to indicate which would be the states most prone to change. In the case of the southern area, which is the area with the highest levels of occurrence, state changes are very prone, especially state 1, which by the second day is expected to change state, that is, to occur a new seismic event. In the central zone, the levels of propensity for change of state are low, that is, the next seismic event is expected to take time. Finally, the northern zone of Peru has high probabilities of change of state, although less than the southern zone.

The adjustment of the Semi-Markov model has made it possible to understand the behavior of earthquakes both for Peru in general, as well as for the 3 areas that comprise it. It was emphasized how the elapsed time affects the probabilities of change of state, for some areas quickly and for others more slowly, but in the end, they are always affected by time.

Bibliography

[Baisch et al., 2008] Baisch, S., Ceranna, L., and Harjes, H.-P. (2008). Seismic Cluster: what can we learn from waveform similarity? *Bulletin of the Seismological Society of America*, 98(6):2806–2814.

[Barbu and Limnios, 2009] Barbu, V. S. and Limnios, N. (2009). *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*, volume 191. Springer Science & Business Media.

[Castillo Aedo, 1994] Castillo Aedo, J. L. (1994). Seismic hazard in Peru.

[ISO, 2015] ISO, E. (2015). Vocabulary and symbols. part 1. general statistical terms and terms used in the calculation of probabilities. (iso 3534-1: 2015). *International Organization for Standardization*.

[Janssen and Manca, 2006] Janssen, J. and Manca, R. (2006). *Applied semi-Markov processes*. Springer Science & Business Media.

[Levy, 1954] Levy, P. (1954). Processus semi-Markoviens. In *Proc. Int. Congress. Math. III, Amsterdam, 1954*.

[Likas et al., 2003] Likas, A., Vlassis, N., and Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461.

[Masala, 2012] Masala, G. (2012). Earthquakes occurrences estimation through a parametric semi-Markov approach. *Journal of Applied Statistics*, 39(1):81–96.

[Novianti et al., 2017] Novianti, P., Setyorini, D., and Rafflesia, U. (2017). K-means clusteranalysis in seismic epicenter clustering. *International Journal of Advances in Intelligent Informatics*, 3(2):81–89.

[Sadeghian, 2012] Sadeghian, R. (2012). Forecasting time and place of earthquakes using asemi-Markov model (with case study in tehran province). *Journal of Industrial Engineering International*, 8(1):1–7.

[Votsi et al., 2012] Votsi, I., Limnios, N., Tsaklidis, G., and Papadimitriou, E. (2012). Estimation of the expected number of seismic occurrences based on semi-Markov models.*Methodology and Computing in Applied Probability*, 14(3):685–703.

[Votsi et al., 2014] Votsi, I., Limnios, N., Tsaklidis, G., and Papadimitriou, E. (2014). Hidden semi-Markov modeling for the estimation of seismic occurrence rates. *Communications in Statistics-Theory and Methods*, 43(7):1484–1502.

[Wood and Neumann, 1931] Wood, H. O. and Neumann, F. (1931). Modified mercalli intensity scale of 1931. *Bulletin of the Seismological Society of America*, 21(4):277–283.

Appendix

APPENDIX A: Project development schedule

	December		January				February				March	
Introduction												▨
Problem Statement	▨	▨										
Contextual Theoretical Framework			▨	▨								
Methodology					▨	▨	▨					
Analysis and Results								▨	▨			
Conclusions										▨	▨	

APPENDIX B: Demonstrations

In this section the demonstrations of the equations shown in the previous sections will be carried out.

7.2.1. Equation 2.7.4

From the definition of conditional probability, it is shown that:

$$F_{ij}(t) = P(X_{n+1} \leq t | J_n = i, J_{n+1} = j)$$

$$F_{ij}(t) = \frac{P(X_{n+1} \leq t, J_n = i, J_{n+1} = j)}{P(J_n = i, J_{n+1} = j)}$$

$$F_{ij}(t) = \frac{P(X_{n+1} \leq t, J_n = i, J_{n+1} = j)}{P(J_n = i)} . \frac{P(J_n = i)}{P(J_n = i, J_{n+1} = j)}$$

$$F_{ij}(t) = \frac{P(J_{n+1} = j, X_{n+1} \leq t | J_n = i)}{P(J_{n+1} = j | J_n = i)}$$

$$F_{ij}(t) = \frac{Q_{ij}}{p_{ij}}$$

APPENDIX C: Tables and figures according to APA Standards

The APA Norms look for a set of standards to fully unify the writing of research papers and be at an international level. In this work, you will follow the standards suggested by the APA Standard when constructing the statistical tables and graphs.

- **Tables:**

The tables under the APA Standards model do not have vertical margins, the horizontal margins are the upper parts (where the titles are separated) and the lower parts (which separate the source). The table must be numbered, have a short and concise title, have a heading for each column, a note (if it is necessary to clarify the content) and the source.

Table [number] : [Table title]

Factor name	**Variable Name**	
	Label 1	**Label 2**
Category 1	X	X
Category 2	X	X
Category 3	X	X
Category 4	X	X
Category 5	X	X
Category 6	X	X
Total	X	X

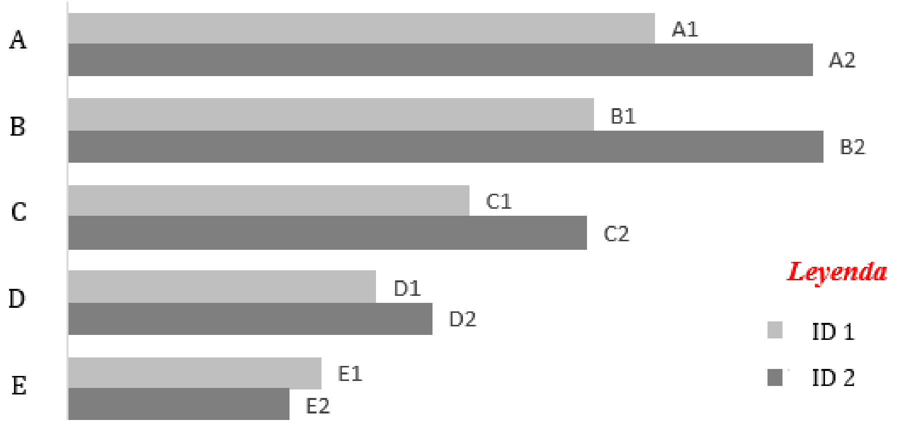
Source: [Source Name]

It is important to bear in mind that the tables do not duplicate the information that has been developed in the work but, on the contrary, complement the information that is being developed.

- **Figures:**

According to the APA Standards, the figures or graphics must be numbered, have a short and concise title, have a legend (which makes clear the signs that will be used in the figure), have a note (add a comment that reinforces the idea in case it cannot be understood by the title) and the source.

Gráfico [número]: [título del gráfico]



Fuente: [Nombre de la fuente]

Both the tables and the figures must be accompanied by a paragraph where the message to be transmitted is summarized, these paragraphs can go before or after the tables or figures.

APPENDIX D: Standards for Statistical Definitions, Vocabulary and Symbols Iso-3534-1 [ISO, 2015]

- **Population:** Totality of the elements taken into consideration.
- **Descriptive statistics:** Graphical, numerical or other summary representation of the observed values.
- **Frequency distribution:** empirical relationship between classes and their number of occurrences or observed values.
- **Sample mean:** Arithmetic mean sum of the random variables of a random sample divided by the number of elements in the sum.
- **Sample variance:** Sum of the squares of the deviations of the random variables of a random sample with respect to its sample mean, divided by the number of terms in the sum minus 1.
- **Estimator:** statistic used to estimate a parameter. An estimator could be the sample mean to estimate the population mean, which could be denoted by µ. For a distribution such as the normal distribution, the “natural” estimator of the population mean µ is the sample mean.

- **Maximum likelihood estimator:** estimator assigned to the parameter value when the likelihood function reaches or approaches its highest value.
- **Sample space:** set of all possible outcomes.
- **Event:** subset of the sample space.
- **Probability of an event:** real number of the closed interval [0,1] assigned to an event.
- **Probability distribution:** probability measure induced by a random variable.
- **Distribution function of a random variable X:** F (x) function of x that gives the probability of the event (−∞, x]. In general, the distribution functions are classified into discrete distribution functions and continuous distribution functions
- **Continuous distribution:** probability distribution (2.11) for which the distribution function evaluated at x can be expressed as an integral of a non-negative function from −∞ a x.
- **Weibull distribution:** continuous distribution that has the distribution function.

$$F(x) = 1 - e^{-\left(\frac{x-a}{b}\right)^k}$$

where x > a with parameters−∞ < a < ∞, b > 0, k > 0. Being a location parameter in the sense that it is the minimum value that the Weibull distribution can reach. The parameter b is a scale parameter [related to the standard deviation of the Weibull distribution]. The parameter k is a shape parameter.

APPENDIX E: NIST GCR 11-917-12 Standards of Seismic Safety for Existing Federally Owned and Leased Buildings - ICSSC Recommended Practice 8 (RP 8)

The intent of this standard is to provide a common minimum and higher standard for the evaluation and mitigation of seismic risks in their existing owned or leased buildings and in privately-owned buildings to ensure that all agencies have balanced, agency-conceived and -controlled seismic safety programs.

The standards establish procedures and criteria intended to provide a low risk of earthquake-related death or life-threatening injury. The standards also provide criteria suitable for certain essential facilities for use by the agencies when they address such buildings in their inventories.

APPENDIX F: Limitations

F.1. Limitations of the Methodology

- **Information consistency**

The Geophysical Institute of Peru reports all earthquakes that occur within the continental territory as maritime Peru, these have all kinds of magnitudes, from very low that can become imperceptible, to the highest that generate great levels of destruction in the territory and even bring with them the occurrence of other natural phenomena, among the most common are the tsunamis that occur after the occurrence of a strong earthquake with an epicenter in the maritime territory. By doing an analysis of the frequency of occurrence of seismic events reported by the IGP by magnitudes, it can be observed that most of these seismic events present not so high magnitudes, this translates to magnitudes of 4 to 5 on the magnitude scale of moments (M_w).

If seismic events are taken for the application of the semi-Markov model, the results would not be as close to reality, for the reason that the group of earthquakes from 4 M_w to 5 M_w is very large. For this reason, it is that for the purposes of a correct application of the model and obtaining results that are more adjusted to reality in this study, seismic events that have registered a magnitude greater than 5.5 M_w will be taken.

- **Updated and current information**

The database obtained from the website of the Geophysical Institute of Peru (IGP) contains information on earthquakes from 1990 to the present, if the totality of these is reviewed it can be seen that in the past, earthquakes were not reported as adequately as performed now in the last century. In view of this, in the present study, earthquakes will be taken within the period from 2008 to 2018, this also in order that there are not very large differences between earthquakes since for the application of the semi-Markov model they are considered the inter-arrival times, which in this case would be the times between the seismic occurrence calculated in days. As these numbers are very large, they could alter the results and they may not be representative.