



# Implementation of the ETAS Model for the Simulation of Aftershocks on the Nazca Plate

## Investigación Estadística

*Docente: Ruiz Olorte Demetrio Antonio*

LIN CHIU CHEN YANG

**Abstract—** Este estudio tiene como objetivo desarrollar y aplicar el modelo de secuencia de réplicas de tipo epidémico (ETAS) para simular y analizar las secuencias de réplicas originadas entre la placa de Nazca y la Sudamericana, permitiendo así una evaluación más precisa del riesgo y una mejor planificación de la respuesta a emergencias en la región de Perú. La investigación comienza con una revisión teórica y metodológica de los modelos en sismología, centrándose en el modelo ETAS, su formulación matemática y los supuestos subyacentes. Se desarrolla una implementación computacional del modelo ETAS utilizando la técnica de máxima verosimilitud para la estimación de parámetros, la cual se valida mediante simulaciones y comparación con datos históricos de secuencias de réplicas. Luego, el modelo se aplica a un conjunto de datos reales de una región específica obtenidos del Instituto Geofísico del Perú (IGP), con un análisis de su precisión para predecir el momento y la ubicación de las réplicas, obteniendo así un mapa de calor de zonas con más riesgo sísmico. Finalmente, la capacidad predictiva del modelo sintético generado se evalúa comparando sus resultados con los datos reales, identificando las fortalezas y limitaciones del modelo ETAS en diferentes escenarios de actividad sísmica.

**Keywords:** *aftershock sequence, point process, hawkes process, etas model, earthquake forecasting*

## 1. Introducción

### 1.1. Contexto

Los terremotos representan uno de los fenómenos naturales más devastadores y difíciles de predecir en el mundo. A lo largo de la historia, estos eventos han causado pérdidas humanas y materiales significativas, generando la necesidad imperante de desarrollar métodos efectivos para su predicción y mitigación de riesgos. La predicción de terremotos es una rama de la ciencia de la sismología que estudia los terremotos, disciplina que llegó a ser científica a partir de la segunda mitad del siglo XIX. Es cuando el deseo de encontrar herramientas que permitan estudiar el fenómeno natural llevó en 1880 a John Milne, famoso ingeniero británico, a inventar el sismógrafo (Kossobokov and Shebalin, 2003). Se han planteado distintos métodos de predicción desde anomalías en el campo electromagnético de la tierra o perturbaciones de la ionosfera (Freund et al., 2009) hasta el comportamiento animal al sentir las ondas P momentos previos al fenómeno (Woith et al., 2018) pero no es hasta la década de los setenta que empezaron los estudios estadísticos representándolos como procesos estocásticos asimismo esta disciplina se ha extendido no solo en la geofísica sino también a través de múltiples campos de la ciencia como la sismología fotónica donde últimas investigaciones sugieren el uso de tecnologías emergentes a base de fibra óptica para su estudio (Lindsey and Martin, 2021).

¿Qué es la predicción de un terremoto? Según el Diccionario Oxford, las palabras predicción y pronóstico son sinónimos, mientras que para la sismología son términos que marcan diferencia. El pronóstico de terremotos se refiere a estimaciones a largo plazo durante un período de tiempo prolongado, que puede ir desde meses hasta décadas, en cambio, la predicción de terremotos generalmente se define como la especificación del tiempo, lugar y magnitud de un futuro terremoto dentro de límites establecidos a corto plazo pueden ocurrir en días, semanas o meses (Ismail-Zadeh, 2013); por su naturaleza estocástica la capacidad de predecir con precisión queda corta por no decir imposible, por eso ahora en adelante la investigación se centra netamente en el pronóstico de eventos sísmicos más no predicción.

Se han planteado varios métodos de pronóstico (Ogata, 1988) propone estudiar los eventos de forma epidemiológica planteando así el modelo ETAS de carácter temporal mas tarde en conjunto con (Zhuang et al., 2011) se desarrolló el modelo de tipo espacio-temporal que se detallará en las siguientes secciones del documento.

El Perú se encuentra ubicado en una región sísmicamente activa exactamente el cinturón de fuego del pacifico debido a la convergencia de dos placas tectónicas entre la Placa de Nazca y la Sudamericana. Este proceso llamado subducción, donde la Placa de Nazca se hunde debajo de la Sudamericana acumulando tensiones, es responsable de la generación de terremotos frecuentes y actividad volcánica en las costas del Pacifico incluyendo Colombia, Ecuador, Perú, Bolivia y Chile. En el artículo de (Villegas-Lanza et al., 2016) se detallan los grandes terremotos ocurridos a lo largo de la zona de subducción peruana en los segmentos centro y sur, como Lima-Callao de 1746 ( $8.8M_w$ ) y el terremoto de Pisco de 2007 ( $8.0M_w$ ) respectivamente.

### 1.2. Problema de Investigación

La comprensión de terremotos sigue siendo un desafío primario de la sismología; estas traen importantes implicaciones sociales y económicas. A pesar de décadas de investigación, aún no es posible determinar con exactitud cuándo, dónde y con qué intensidad ocurrirá un sismo. Esto se debe a la naturaleza compleja de los fenómenos sísmicos y a la interacción de múltiples factores geológicos, físicos u otros que lo preceden; lo cual se realiza el StatSei (The International Statistical Seismology conference) evento bienal que reúne a investigadores y expertos en el campo de la sismología estadística, esta investigación se apoya en varios de los artículos propuestos donde se abordará principalmente la posibilidad de su pronóstico mediante modelos estadísticos que simulan la sismicidad, acercándonos a la posición de (Geller, 1997) "¿Es posible la predicción?".

### 1.3. Motivación

Posterior al evento sísmico surgen las réplicas, también conocidas como aftershocks, ocurren después de un terremoto principal y pueden tener una magnitud menor, pero aún representan un riesgo significativo para la población y las infraestructuras afectadas (Darzi et al., 2022) las cuales pueden permanecer en pie frente al terremoto principal pero la llegada de oleadas de réplicas consecutivas sobrepasan el límite de estrés de la infraestructura haciendo que colapse sin previo aviso es por eso que el estudio de los modelos predictivos es indispensable para la sociedad, y su precisión conyeva al dilema de ¿alarmar? ¿O no alarmar? Advertir de un terremoto que no ocurre también conlleva un costo: no solo el costo civil y económico. Entidades publicas como el IGP realizan investigaciones científicas con finalidad de la prevención, reducción de riesgos de desastres orientando al Estado promover políticas para proteger la vida de la población. Por otro lado en (Mignan and Woessner, 2012) y (Gulia et al., 2012) ambos autores proponen la importancia de la completitud y consistencia de los datos para el análisis de sismicidad pero factores no controlables como la calibración instrumental, el ruido externo disminuyen la calidad de estos por eso la realización de simulaciones pueden ayudar a mejorarlos; tambien, en la revision de literatura no se encontro la implementacion del modelo en la región de la placa de Nazca por consiguiente esta investigación intenta replicar un modelo descriptivo propuesto por (Ogata, 2017) para comparar la sismicidad de la región y los estudios anteriores.

### 1.4. Objetivos de la Investigación

#### 1.4.1. Objetivos Generales

Desarrollar y replicar el modelo ETAS (Epidemic-Type Aftershock Sequence) para simular réplicas sísmicas seleccionando como región la costa del Pacífico, permitiendo una evaluación más precisa de focos del riesgo y una mejor planificación de respuestas ante emergencias.

#### 1.4.2. Objetivos Específicos

- **Revisión Teórica y Metodológica:** Realizar una revisión exhaustiva de la literatura sobre modelos temporales y de procesos espacio-puntuales aplicados a sismología, con énfasis en el modelo ETAS. Describir los fundamentos teóricos del modelo ETAS y sus variantes, incluyendo la formulación matemática y los supuestos subyacentes.
- **Implementación del Modelo ETAS:** Desarrollar una implementación computacional del modelo ETAS utilizando técnicas paramétricas para la estimación de parámetros. Validar la implementación mediante simulaciones y comparación con datos históricos de secuencias de réplicas.
- **Aplicación a Datos Reales:** Aplicar el modelo ETAS a un conjunto de datos reales de secuencias de réplicas sísmicas de una región específica. Analizar la precisión del modelo en la predicción de la ocurrencia de réplicas en términos de tiempo y ubicación.
- **Evaluación y Comparación:** Evaluar la capacidad predictiva del modelo ETAS comparando sus resultados con otros modelos existentes. Identificar las fortalezas y limitaciones del modelo ETAS en diferentes escenarios de actividad sísmica.

### 1.5. Hipótesis de Investigación

Se analizará si el modelo ETAS proporcionará simulaciones de réplicas sísmicas coherentes con los datos históricos en la región de la costa del Pacífico, permitiendo una mejor identificación de focos de riesgo. También si la inclusión de datos de sismicidad simulada mejorará la calidad y la completitud de los datos utilizados en el análisis de sismicidad, superando las limitaciones impuestas por factores no controlables.

## 2. Revisión de Literatura

### 2.1. Antecedentes

La pregunta es como se podrá desarrollar un modelo capaz de simular fenómenos tan caóticos como los sismos, gracias a su naturaleza estocástica se puede empezar planteando desde un modelo simple como poisson o de renovación pasando a modelos agrupados, a base de la fórmula empírica de Omori-Utsu nos conduce al modelo ETAS temporal propuesto por (Ogata, 1988) sin considerar el espacio todos estos son detallados por (Zhuang et al., 2012). A base de este se plantean variantes incluyendo el espacio transformándolos en modelos espacio-temporales todos detallados en (Zhuang et al., 2011) se detallará la forma de ellas en la siguiente sección.

Este modelo ha sido ampliamente aplicado en varios estudios como: análisis de sismicidad sur de California donde (Console et al., 2010) usa el modelo para desagrupar réplicas; (Omi et al., 2014) plantea la estimación de parámetros incluso con la incompletitud de datos; (Lombardi and Marzocchi, 2010) describe la sismicidad en la región de Italia; (Darzi et al., 2022) propone la estimación de parámetros de forma bayesiana analizando la sismicidad de Olfus Islandia; (lafmejani davood kazemi and Shabani, 2022) aplica en la región de Kermanshah Irán; (Nishikawa and Nishimura, 2023) incluyó al modelo el efecto de eventos de deslizamiento de la placa SSE; (Mizrahi et al., 2021) propone método para estimar parámetros frente a la incompletitud de datos.

### 2.2. Marco Conceptual

#### 2.2.1. Glosario de Términos

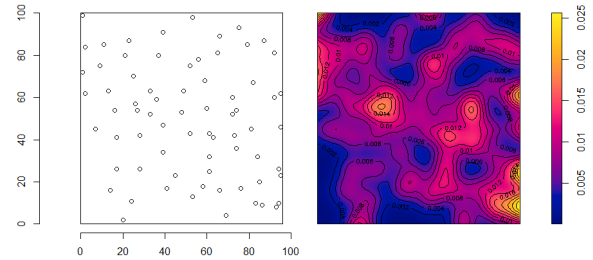
- **Aftershock o Réplica** es un terremoto más pequeño que sigue a un terremoto más grande.
- **Catálogo sísmico** Es una base de datos que contiene todos los parámetros que caracterizan a un sismo como el momento, la ubicación y la magnitud.
- **ETAS Modelo de secuencia de réplicas de tipo epidémico**: Es un modelo probabilístico utilizado para describir la ocurrencia de terremotos y réplicas (aftershocks). Este modelo se basa en la idea de que cada terremoto puede generar sus propias réplicas, y estas réplicas pueden, a su vez, generar más réplicas, creando una cadena de eventos similar a una epidemia.
- **Función de intensidad condicional** Concepto fundamental en la teoría de procesos puntuales es la probabilidad de que ocurra un evento bajo el historial del proceso hasta ese momento.
- **Hawkes process** Un tipo de proceso puntual autoexcitado, utilizado para modelar eventos que tienen la propiedad de incrementar la probabilidad de futuros eventos durante algún período de tiempo después de su ocurrencia.
- **MLE Estimación de máxima verosimilitud** Es un método estadístico utilizado para estimar los parámetros de un modelo probabilístico. Este enfoque se basa en maximizar la función de verosimilitud, que mide la probabilidad de observar los datos dados los parámetros del modelo.
- **Point process** Es un modelo matemático utilizado para describir y analizar la ocurrencia de eventos que se producen de forma aleatoria en el tiempo o en el espacio. Estos eventos se representan como puntos en una línea de tiempo, en un plano, o en un espacio de mayor dimensión.

### 2.3. Base Teórica

#### 2.3.1. Procesos Puntuales

Es un modelo matemático que se utiliza para describir eventos aleatorios que ocurren en el tiempo o el espacio. Por simplicidad, comenzaremos con la unidimensional, el proceso puntual temporal, poisson es uno de ellos, llamado también proceso de conteo, donde  $N(t)$  representa el número de eventos ocurridos hasta el tiempo  $t$  con tiempos de arribo independientes distribuidos exponencialmente de parámetro constante  $\lambda$  teniendo incremento estacionario. Si la tasa no es con-

stante y depende de  $t$  pasa a ser no-estacionario, de la misma forma trasladamos esta idea a dos dimensiones reemplazamos tiempo  $t$  por  $A$  una región en el espacio  $S$  entonces  $N(A)$  sería el número de eventos ocurridos en dicha región, el proceso puntual tendría como función de intensidad  $\lambda(S)$ , la media de eventos en  $A$  es  $\mu(A) = \int_A \lambda(S) dS$  si  $\lambda(S)$  es estocástica podemos expresarla como  $\Lambda(S)$ .



**Figure 1.** Observamos la distribución de la función de intensidad  $\Lambda(S)$  (derecha) del proceso puntual. Gráfico generado mediante spatstat de R

#### 2.3.2. Proceso de Hawkes

Para entender el proceso de Hawkes según (Kowshik.C, 2020) algunos procesos la ocurrencia de un evento aumenta las probabilidades de uno similar se llama procesos excitados un ejemplo simple sería la venta masiva de acciones aumenta las probabilidades de que se vendan más por pánico al contrario en algunos disminuyen la probabilidad de otro suceso estos son llamados procesos inhibidos un ejemplo sería luego de un incendio forestal se reducen las probabilidades de otro a corto plazo, ya que se ha combustionado la masa forestal por otro lado, tenemos procesos en los que la ocurrencia no afecta la ocurrencia de otro estos son los procesos poisson. Para modelarlos necesitamos de la función de intensidad condicional  $\lambda^*(t)$ , retornando a la dimensión temporal sería:

$$\lambda^*(t) = \lim_{\Delta t \rightarrow 0} \frac{E[N(t + \Delta t) - N(t) | H(t)]}{\Delta t} \quad (1)$$

esta está condicionada a la historia previa  $H(t)$ . El proceso de Hawkes es una variante de procesos puntuales autoexcitados planteado por Alan G. Hawkes. Utilizados para modelar eventos que están agrupados temporalmente como "terremotos" y "ventas de acciones por pánico" Definamos la función de intensidad condicional de Hawkes. Sea  $0 < t_1 < t_2 < \dots < t_k$  la secuencia de tiempos de llegada del proceso puntual hasta el tiempo  $t$ , la intensidad condicional de Hawkes es:

$$\lambda^*(t) = \mu(t) + \sum_{t_i < t} \phi(t - t_i) \quad (2)$$

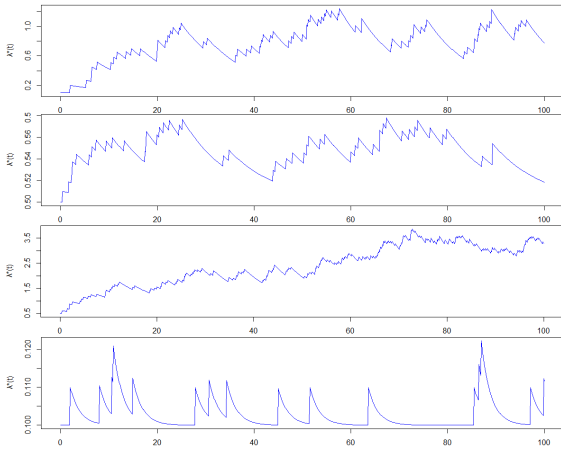
La constante  $\mu$  se denomina intensidad de fondo y  $\phi$  se denomina función de excitación. Si  $\phi$  es igual a cero, este proceso puntual autoexcitado se reduce a un simple proceso de Poisson estacionario. Existen múltiples funciones de excitación según el diseño del modelo, la más común es el decaimiento exponencial.

$$\phi(t) = \alpha e^{-\beta t}$$

Los parámetros  $\alpha$  y  $\beta$  son constantes. Donde  $\alpha$  es encargado de incrementar la intensidad luego de la llegada de un evento y con el tiempo esta decae con  $\beta$ .

Para comprender la función de intensidad condicional, simulemos un proceso de puntos de Hawkes simple donde  $\theta$  es el vector de parámetros  $\lambda, \alpha$  y  $\beta$ . En la figura 4 se observan cuatro gráficos de intensidad condicional frente al tiempo. Comenzamos con  $\theta = 0.1, 0.1, 0.1$  si incrementamos la intensidad  $\lambda$  a 0.5 en el segundo gráfico vemos que la intensidad de fondo se mantuvo por encima de 0.5. Para comprender el impacto de  $\alpha, \beta$ . Aumentemos  $\alpha$  a 0.5. En el tercer gráfico vemos que el número de eventos aumentó, porque en cada ocurrencia de un evento aumenta la intensidad de llegada del siguiente evento

en  $\alpha$ . Si aumentemos  $\beta$  a 0.5. En el cuarto gráfico se pueden ver decaimientos constantes.



**Figure 2.** Simulación del proceso de Hawkes con distintos valores de parámetros  $\lambda$ ,  $\alpha$  y  $\beta$  en R

### 2.3.3. Distribución de Réplicas

En todos los catálogos de hipocentros sísmicos incluyen la cinco parámetros principales las cuales son tiempo de ocurrencia, longitud y latitud del epicentro, profundidad focal y magnitud; en nuestro estudio se analiza la actividad sísmica en relación de cuatro dimensiones  $(t_n, x_n, y_n, M_n)$ ,  $n = 1, 2, \dots, N$ , ignorando la profundidad debido a la baja precisión en comparación con las otras coordenadas.

La probabilidad de que ocurra un terremoto en un tiempo  $t$ , una ubicación  $(x, y)$  de magnitud  $M$ , condicionado a los eventos ocurridos en el pasado  $H_t = \{(t_i, x_i, y_i, M_i); t_i < t\}$  se representa en la siguiente ecuación:

$$\frac{P(\text{evento en } [t, t + \Delta t] \times [x, x + \Delta x] \times [y, y + \Delta y] \times [M, M + \Delta M] | H_t)}{\Delta t \Delta x \Delta y \Delta M} \approx \lambda(t, x, y, M | H_t) \quad (3)$$

donde  $\lambda$  es función de intensidad condicional del proceso puntual que intentamos modelar que representa la probabilidad de que ocurra eventos de magnitudes  $\Delta M$  en el intervalo de tiempo  $t$  y una región definida por  $\Delta xy$ . Las tasas y magnitudes de las réplicas siguen varias leyes empíricas bien establecidas; como la de Gutenberg-Richter donde  $\lambda(M)$  es la intensidad de terremotos de magnitud menor o igual a  $M$  en una región en un período de tiempo con constantes  $a$  y  $b = \beta$  que varían según la región de estudio.

$$\lambda(M) = 10^{a-bM} = Ae^{-\beta M} \quad (4)$$

También la ley de Omori, 1894 y (Utsu, 1961) proponen el decaimiento de las réplicas según el tiempo después del terremoto principal expresado como:

$$n(t) = \frac{k}{(c + t)^p} \quad (5)$$

teniendo como constantes  $k$ ,  $c$  y  $p$ .

### 2.3.4. Modelo Espacio Temporal ETAS

El modelo ETAS es un proceso de puntos de ramificación (también conocidos como Hawkes o procesos de puntos autoexcitantes) la ocurrencia de un terremoto excita la actividad sísmica en curso, aumentando la probabilidad de ocurrencia de nuevos terremotos en el futuro cercano donde la sismicidad es el resultado de la suma de dos componentes: la sismicidad de fondo (eventos independientes) y la sismicidad desencadenada (eventos agrupados). Matemáticamente, esto conduce a la siguiente expresión para la intensidad sísmica condi-

cional:

$$\lambda(t, x, y, M | H_t) = S(M) \left[ \mu(x, y) + \sum_{i: t_i < t} k(M_i) g(t - t_i) f(x - x_i, y - y_i) \right] \quad (6)$$

donde

- $H_t$  es la historia sísmica hasta el tiempo  $t$ ;
- $\mu$  representa la tasa de sismicidad de fondo
- $S(M)$  función de magnitud G-R representado por la ley exponencial de Gutenberg-Richter
- $k(M) = Ae^{\alpha M}$  productividad directa de Utsu
- $g(t, M)$  se define como la función de excitación en el contexto de los procesos de Hawkes

### 2.3.5. Estimación de Parámetros

#### • Estimación mediante MLE

Los parámetros a estimar son  $\mu, K, \alpha, c, p$  donde  $K, \alpha$  son constantes de la ley de productividad de las réplicas de Utsu y  $c$  y  $p$  son constantes de la ley de Omori-Utsu modificada. Los cinco parámetros se pueden estimar mediante el método de estimación de máxima verosimilitud (MLE) la probabilidad logarítmica se puede expresar de la siguiente manera:

$$\log L(\theta) = \sum_{i=1}^N \log \lambda(t_i, \theta) - \int_0^T \lambda(t, \theta) dt \quad (7)$$

donde  $\theta$  es el conjunto de parámetros a estimar y  $[0, T]$  es el intervalo de tiempo para el cual se deben estimar los parámetros. Encontrar el máximo de una función es un problema de optimización típico, que en la mayoría de los casos debe resolverse numéricamente por métodos Quasi-Newton.

#### • Estimación por inferencia Bayesiana

Según (Molkenthin et al., 2022) el enfoque bayesiano de la estimación de parámetros ETAS el teorema de Bayes dice, la distribución de probabilidad posterior  $p(\theta | X)$  del conjunto de parámetros  $\theta$  dado los datos  $X$  se puede expresar utilizando la probabilidad  $L(\theta | X)$  y la distribución de probabilidad previa  $p(\theta)$  del conjunto de parámetros  $\theta | X$  de la siguiente manera:

$$p(\theta | X) \propto L(\theta | X) p(\theta) \quad (8)$$

### 2.3.6. Selección del Modelo

Criterios de selección:

- **Criterio de Información Akaike (AIC):** Evalúa la calidad del ajuste del modelo penalizando por el número de parámetros. Se calcula como:

$$AIC = 2k 2\log(L) \quad (9)$$

donde  $k$  es el número de parámetros y  $L$  es la verosimilitud máxima.

- **Criterio de Información Bayesiano (BIC):** Similar al AIC, pero penaliza más fuertemente modelos con más parámetros. Se calcula como:

$$BIC = k \log(n) 2\log(L) \quad (10)$$

donde  $n$  es el número de datos.

## 3. Metodología

En el presente proyecto de investigación está orientado al desarrollo de modelos estadísticos lo que convida el uso de la metodología metodología CRISP-ML(Q) (Studer et al., 2021) un proceso estructurado que guía el desarrollo de modelos de aprendizaje automático de manera eficiente y confiable basado en el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), pero extendido para incluir tareas y metodologías de aseguramiento de la calidad específicas. Esta se divide en 6 fases:



3.1. Comprensión de Datos

Un catálogo sísmico es un registro de eventos sísmicos dentro de una región específica o a nivel mundial. Funciona como una base de datos histórica que recopila información sobre terremotos. Esta información normalmente incluye: fecha y hora del evento, ubicación en coordenadas geográficas (latitud y longitud), magnitud en la escala de Richter y profundidad; algunos catálogos también pueden incluir detalles adicionales; se puede consultar en Hauksson et al., 2010.Podemos un ejemplar en la Tabla 1

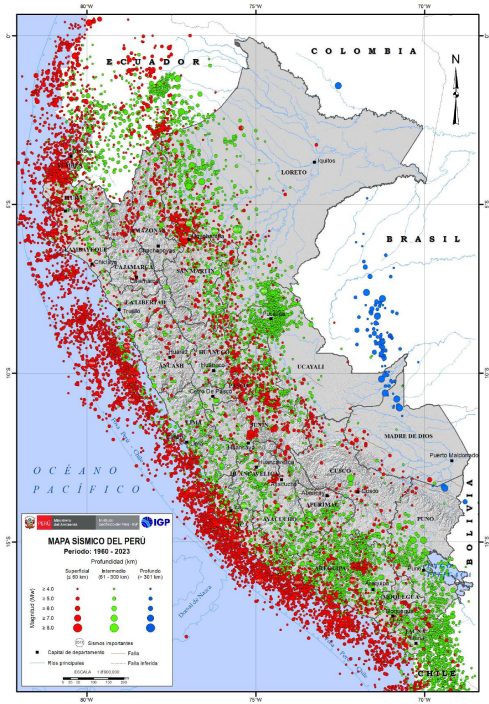


Figure 3. Mapa Sísmico del Perú 1960 - 2023 CENSIS, 2023.

Los datos sísmicos (CENSIS, 2024) se adquirieron del Centro Sismológico Nacional (CENSIS) servicio desarrollado por el Instituto Geofísico del Perú (IGP) institución oficial del Estado que se encarga de monitorear la ocurrencia de sismos en el territorio nacional para ello, se dispone de datos provenientes de la Red Sísmica Nacional (RSN) cuyos sensores se encuentran distribuidos en todo el territorio nacional. La base de datos sísmicos para el Perú ha sido dividida en Sismos Históricos para los eventos sísmicos ocurridos entre los años 1471 y 1959; y Sismos Instrumentales para los ocurridos después del año 1960, La elección del ultimo se basa en las siguientes razones: la precisión de los instrumentos de medición, completitud y consistencia, el incremento notable de estaciones sísmicas a partir del año 2000 nos dan la disponibilidad de datos mas robustos. El catálogo nos provee datos de los sismos ocurridos en una region rectangular que abarca el territorio y el mar peruano, con latitud sur desde 1.396° hasta 25.701° y longitud oeste desde 87.382° hasta 65.624°.

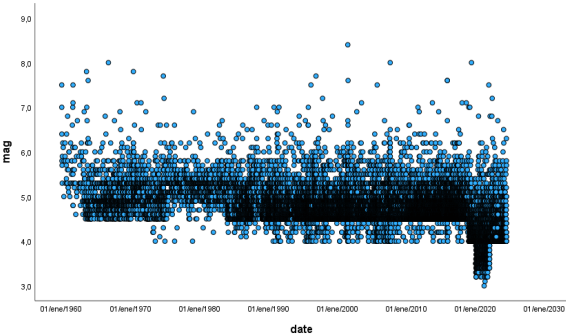


Figure 4. Diagrama de dispersión de magnitudes de eventos con el tiempo.

3.2. Preparación de Datos

Procedemos a realizar un análisis exploratorio de los datos contenidos en el catálogo tomando como referencia a (Mark Naylor, 2010), lo primero que se analizará es la magnitud. Si bien la magnitud más conocida es la escala Richter esta también tiene varios tipos, no confundir con la escala de intensidad representado por la escala de Mercalli, como: Richter Magnitude Scale ( $M_L$ ); Body Wave Magnitude Scale ( $m_b$ ); Duration Magnitude Scale ( $M_d$ ); Moment Magnitude ( $M_W$ ); Mueller, 2018 Describe la metodología desarrollada en los EE.UU. por el U.S. Geological Survey (USGS) para la elaboración de catálogos sísmicos; con este mismo criterio el IGP opto por homogenizar los valores a magnitud momento ( $M_w$ ) de forma propuesta por (Scordilis, 2006)

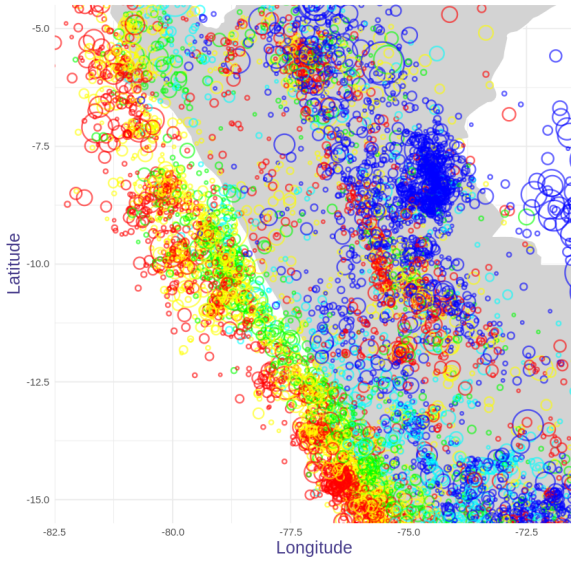
El catálogo consta de 23 944 eventos una magnitud media de 4.71. En la figura 4 podemos ver que casi no hay eventos de magnitud menor a 4.0 esto se debe a la sencibilidad de los instrumentos o la depuración humana y hay pocos eventos de gran magnitud, la región intermedia es la más representativa donde se visualiza mejor en una escala log-lineal como en la figura 5 Esto muestra claramente una relación lineal, entre las magnitudes 4 y 7, cumpliendose la ley de Gutenberg-Richter.

Se empleará la libreria *ETAS* de (Jalilian, 2019) en *R* puede adaptar el modelo para un catálogo de terremotos utilizando el enfoque de desagrupación estocástica propuesto por Zhuang y Ogata. Definimos nuestra ventana de estudio tomando una porción de nuestro catálogo; esta será nuestro evento principal, independiente o de fondo; los datos fuera de esta ventana o externo al poligono de estudio, que se explicará más adelante, son eventos secundarios o de ruido, `time.begin` como el comienzo del lapso de tiempo del catálogo, `time.start` es el inicio del período de estudio, `time.end` es el final del período de estudio, tener en cuenta el formato de las fechas como `yyyy-mm-dd`.

Lo siguiente es definir el región de estudio puede ser rectangular o poligonal en nuestro caso se opto por el ultimo, se puede especificar mediante el argumento `region.poly` que contiene coordenadas de los vértices del polígono, ver en la Tabla?? deben enumerarse en orden antihorario

Table 1. Ejemplo de un catálogo de terremotos

Event	Date	Time	Longitude	Latitude	Depth	Magnitude
1	1998-01-06	15:39:31.00	-75.82	-14.15	49	4.20
2	1998-01-06	20:01:50.90	-70.34	-18.10	188	4.80
3	1998-01-10	17:02:56.00	-80.28	-4.19	36	5.10
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	1998-07-25	08:57:26.90	-69.14	-16.83	67	4.90
101	1998-07-29	13:34:14.00	-74.65	-14.98	96	4.70



**Figure 5.** Gráfico de epicentros de terremotos en un mapa base coloreado por profundidad y con tamaño proporcional a la magnitud. Los colores cálidos indican profundidades bajas y los colores fríos eventos más profundos. Se puede encontrar el código para producir este gráfico en los anexos.

### 3.3. Ajuste del Modelo

Sea  $\theta$  el conjunto de parámetros a estimar en el intervalo de tiempo  $[0, T]$ . En la librería de (Jalilian, 2019) se utiliza el método Davidon-Fletcher-Powell DFS el algoritmo genera iterativamente una aproximación local de la función de log-verosimilitud por una función cuadrática hasta que converge a su máximo.

#### Algoritmo 1 Davidon-Fletcher-Powell algorithm.

A1.entradas

$\hat{\theta}_0$  parámetros iniciales del modelo

$\hat{\mu}_0(x_j, y_j)$   $j = 1, 2, \dots, N$  tasa de fondo inicial

$H_0$  la inversa de la matriz de Hessiana

A2.  $k \leftarrow 0$

A3. repetir

$d_k \leftarrow -H_k \times \nabla \xi(\hat{\theta}_k)$  dirección ascendente

$\zeta_k \leftarrow \argmin_{\zeta} \xi(\hat{\theta}_k + \zeta d_k)$

$\hat{\theta}_{k+1} \leftarrow \hat{\theta}_k + \zeta_k d_k$  actualizar la estimación

$\eta_k \leftarrow \nabla \xi(\hat{\theta}_{k+1}) - \nabla \xi(\hat{\theta}_k)$

$c1 \leftarrow \zeta_k / (d_k^T \times \eta_k)$

$c2 \leftarrow 1 / (\eta_k^T \times H_k \times \eta_k)$   
 $H_{k+1} \leftarrow H_k + c_1 \eta_k \times d_k^T - c_2 H_k \times \eta_k \times \eta_k^T \times H_k$  actualizar el hessiano  
 inverso  
 $k \leftarrow k + 1$   
 A4. hasta  $\|\hat{\theta}_k - \hat{\theta}_{k-1}\| < \epsilon$  o  $\|\nabla \xi(\hat{\theta}_k)\| < \epsilon$   
 A5. retornar  $\hat{\theta}_k$  y  $H_k$

## 4. Resultados

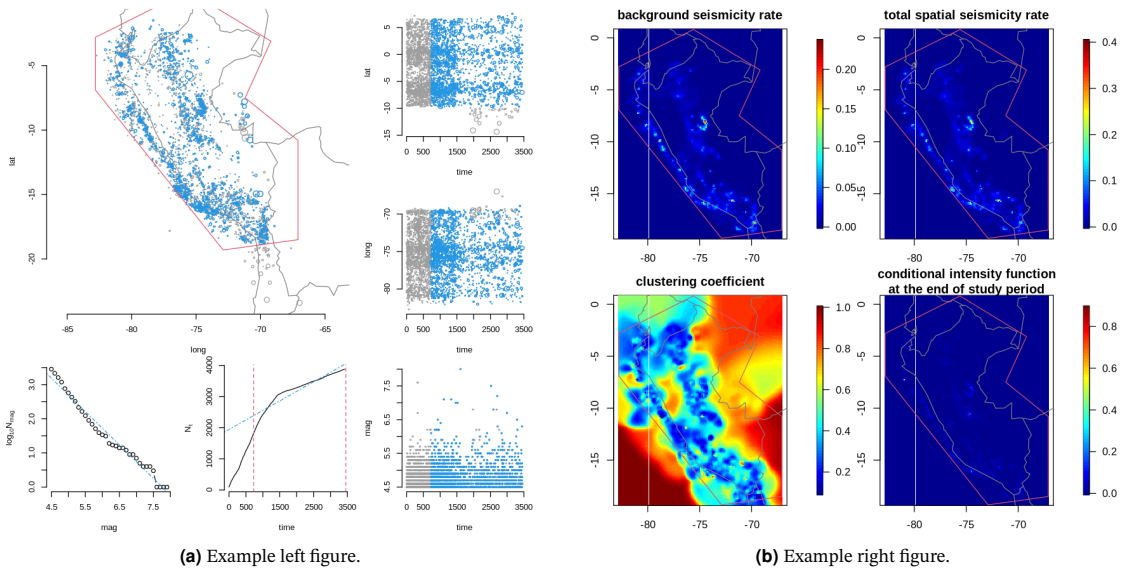
Luego de ajustar nuestros parámetros el modelo ETAS: ajustado mediante el método iterativo de desagrupación estocástica convergió después de 5 iteraciones tiempo de ejecución transcurrido 18,4 minutos procesado paralelamente en un procesador octacore Intel Core i5 decima generación:

**Table 2.** Resultados del ajuste de parámetros

ML estimates of model parameters:							
	$\beta$	$\mu$	A	c	$\alpha$	p	D
$\hat{\theta}$	3.3243	0.5497	1.8560	0.0453	0.8895	1.0132	0.0023
SE	0.0053	0.0084	0.0276	0.0662	0.0213	0.0003	0.0732
Decustering Probabilities							
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
	0.0000	0.5188	0.7382	0.6490	0.8512	1.0000	
log-likelihood:	-10361.76			AIC:	20739.53		

## 5. Conclusión

De los puntos planteados anteriormente se pudo desarrollar con éxito el modelo ETAS pero según los resultados la prueba de bondad de ajuste de Kolmogorov-Smirnov indica que  $U_j$  no sigue una distribución  $U(0, 1)$  con un p-value de  $8.438e-12$  el modelo ajustado aun no es capaz de describir con detalle la compleja agrupación espacial de los terremotos podría existir anomalía en la sismicidad de la región se sospecha que son originados por los datos depurados. En el punto de vista de la literatura la revisión teórica y metodologías empleados por otros investigadores se aconseja modificar método de estimación de parámetros como la bayesiana o encontrar otra variante de función de activación. Por último la sismicidad de cada región es distinta por lo cual existe múltiples métodos para modelarla.



**Figure 6.** Example figure that covers the width of the page obtained from PGFPlots “PGFPlots - A LaTeX package to create plots.” n.d.

## References

- Omori, F. (1894). On the after-shocks of earthquakes [application/pdf]. *The journal of the College of Science, Imperial University, Japan* = ., 7, 111–200. <https://doi.org/10.15083/00037562>
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401), 9–27. <https://doi.org/10.1080/01621459.1988.10478560>
- Geller, R. J. (1997). Earthquake prediction: A critical review. *Geophysical Journal International*, 131(3), 425–450. <https://doi.org/10.1111/j.1365-246x.1997.tb06588.x>
- Kossobokov, V., & Shebalin, P. (2003). Earthquake prediction. In *Nonlinear dynamics of the lithosphere and earthquake prediction* (pp. 141–207). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-05298-3\\_4](https://doi.org/10.1007/978-3-662-05298-3_4)
- Scordilis, E. M. (2006). Empirical global relations converting m s and m b to moment magnitude. *Journal of Seismology*, 10(2), 225–236. <https://doi.org/10.1007/s10950-006-9012-4>
- Freund, F. T., Kulahci, I. G., Cyr, G., Ling, J., Winnick, M., Tregloan-Reed, J., & Freund, M. M. (2009). Air ionization at rock surfaces and pre-earthquake signals. *Journal of Atmospheric and Solar-Terrestrial Physics*, 71(17–18), 1824–1834. <https://doi.org/10.1016/j.jastp.2009.07.013>
- Console, R., Jackson, D. D., & Kagan, Y. Y. (2010). Using the etas model for catalog declustering and seismic background assessment. *Pure and Applied Geophysics*, 167(6–7), 819–830. <https://doi.org/10.1007/s00024-010-0065-5>
- Hauksson, E., Hardebeck, J. L., & Woessner, J. (2010). What is an instrumental seismicity catalog? <https://doi.org/10.5078/CORSSA-38784307>
- Lombardi, A. M., & Marzocchi, W. (2010). The etas model for daily forecasting of italian seismicity in the csep experiment. *Annals of Geophysics*, 53(3). <https://doi.org/10.4401/ag-4848>
- Mark Naylor, D. H., Katerina Orfanogiannaki. (2010). Exploratory data analysis: Magnitude, space and time. <https://doi.org/10.5078/CORSSA-92330203>
- Zhuang, J., Werner, M. J., Zhou, S., Harte, D., & Hainzl, S. (2011). Basic models of seismicity: Spatiotemporal models. <https://doi.org/10.5078/CORSSA-07487583>
- Gulia, L., Wiemer, S., & Wyss, M. (2012). Catalog artifacts and quality control. <https://doi.org/10.5078/CORSSA-93722864>
- Mignan, A., & Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. <https://doi.org/10.5078/CORSSA-00180805>
- Zhuang, J., Werner, M. J., Zhou, S., Hainzl, S., & Harte, D. (2012). Basic models of seismicity: Temporal models. <https://doi.org/10.5078/CORSSA-79905851>
- Ismail-Zadeh, A. T. (2013). Earthquake prediction and forecasting. In *Encyclopedia of earth sciences series* (pp. 225–231). Springer Netherlands. [https://doi.org/10.1007/978-1-4020-4399-4\\_106](https://doi.org/10.1007/978-1-4020-4399-4_106)
- Omi, T., Ogata, Y., Hirata, Y., & Aihara, K. (2014). Estimating the etas model from an early aftershock sequence. *Geophysical Research Letters*, 41(3), 850–857. <https://doi.org/10.1002/2013gl058958>
- Villegas-Lanza, J. C., Chlieh, M., Cavalié, O., Tavera, H., Baby, P., Chire-Chira, J., & Nocquet, J.-M. (2016). Active tectonics of peru: Heterogeneous interseismic coupling along the nazca megathrust, rigid motion of the peruvian sliver, and subandean shortening accommodation. *Journal of Geophysical Research: Solid Earth*, 121(10), 7371–7394. <https://doi.org/10.1002/2016jb013080>
- Ogata, Y. (2017). Statistics of earthquake activity: Models and methods for earthquake predictability studies. *Annual Review of Earth and Planetary Sciences*, 45(1), 497–527. <https://doi.org/10.1146/annurev-earth-063016-015918>
- Mueller, C. S. (2018). Earthquake catalogs for the usgs national seismic hazard maps. *Seismological Research Letters*, 90(1), 251–261. <https://doi.org/10.1785/0220170108>
- Woith, H., Petersen, G. M., Hainzl, S., & Dahm, T. (2018). Review: Can animals predict earthquakes? *Bulletin of the Seismological Society of America*, 108(3A), 1031–1045. <https://doi.org/10.1785/0120170313>
- Jalilian, A. (2019). Etas: An r package for fitting the space-time etas model to earthquake data. *Journal of Statistical Software*, 88(Code Snippet 1). <https://doi.org/10.18637/jss.v088.c01>
- Kowshik, C. (2020). Understanding point processes. <https://kowshikchilamkurthy.medium.com/understanding-point-processes-6e3d2f6c5480>
- Lindsey, N. J., & Martin, E. R. (2021). Fiber-optic seismology. *Annual Review of Earth and Planetary Sciences*, 49(Volume 49, 2021), 309–336. <https://doi.org/10.1146/annurev-earth-072420-065213>
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021). Embracing data incompleteness for better earthquake forecasting. *Journal of Geophysical Research: Solid Earth*, 126(12). <https://doi.org/10.1029/2021jb022379>
- Studer, S., Bui, T. B., Drescher, C., Hanuschkina, A., Winkler, L., Peters, S., & Müller, K.-R. (2021). Towards crisp-ml(q): A machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, 3(2), 392–413. <https://doi.org/10.3390/make3020020>
- Darzi, A., Halldorsson, B., Hrafnkelsson, B., & Vogfjörð, K. S. (2022). Short-term bayesian etas spatiotemporal forecasting of the ölfus 2008 earthquake sequence in iceland. *Tectonophysics*, 839, 229522. <https://doi.org/10.1016/j.tecto.2022.229522>
- lafmejani davood kazemi, d. k., & Shabani, E. (2022). Modeling earthquake data using etas model to forecast aftershock subsequences applying different parameterizations in kerman-shah region, iran. <https://doi.org/10.21203/rs.3.rs-1996634/v1>
- Molkenthin, C., Donner, C., Reich, S., Zöller, G., Hainzl, S., Holschneider, M., & Opper, M. (2022). Gp-etis: Semiparametric bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing*, 32(2). <https://doi.org/10.1007/s11222-022-10085-3>
- CENSIS. (2023). Mapasismico2023. [https://ultimosismo.igp.gob.pe/img/mapas\\_sismicos/mp\\_2023.jpg?v=3221881](https://ultimosismo.igp.gob.pe/img/mapas_sismicos/mp_2023.jpg?v=3221881)
- Nishikawa, T., & Nishimura, T. (2023). Development of an epidemic-type aftershock-sequence model explicitly incorporating the seismicity-triggering effects of slow slip events. *Journal of Geophysical Research: Solid Earth*, 128(5). <https://doi.org/10.1029/2023jb026457>
- CENSIS. (2024). Igp\_datos\_sismicos. <https://ultimosismo.igp.gob.pe/descargar-datos-sismicos>
- PGFPlots - A LaTeX package to create plots. (n.d.). <https://pgfplots.sourceforge.net/>



## Anexos

### Codigos

Se ejecutó las siguientes funciones en lenguaje R versión 4.4.1 con un procesador octacore Intel i5 decima generación

Los codigos empleados para el desarrollo del modelo.

```

1 # importando datos
2 dataf <- read.csv("IGP_peru.csv", sep = ";")
3 head(dataf)
4
5 # adecuando nuestros datos
6 df <- dataf[, c("date", "time", "long", "lat", "
7 mag")]
8 df$long <- as.numeric(gsub(",", ".", df$long))
9 df$lat <- as.numeric(gsub(",", ".", df$lat))
10 df$date <- as.Date(df$date, format = "%d/%m/%Y")
11 df$date <- as.factor(df$date)
12 head(df)
13
14 # instalar los paquetes necesarios
15 install.packages('ETAS')
16 library(ETAS)
17
18 # specifying a polygonal geographical region
19 ppoly <- list(
20   long=c(-82.8, -82.8, -72.9, -67.1, -67.1,
21     -71.2, -69.2, -75.6),
22   lat=c(-2.8, -6.9, -19.3, -18.5, -10.8, -7.5,
23     -3.1, 0.8)
24 )
25
26 # creating a catalog with polygonal study region
27 peru.cat <- catalog(df, time.begin="2015-01-01",
28   study.start="2017-01-01", study.end="2024-06-16"
29   ,
30   region.poly=ppoly, mag.threshold=4.5)
31 plot(peru.cat)
32
33 param01 <- c(0.5, 0.2, 0.05, 2.7, 1.2, 0.02,
34   2.3, 0.03)
35
36 # ajustando el modelo
37 peru.fit <- etas(peru.cat, param0 = param01)

```

Code 1. Codigos empleados.

los graficos generados son:

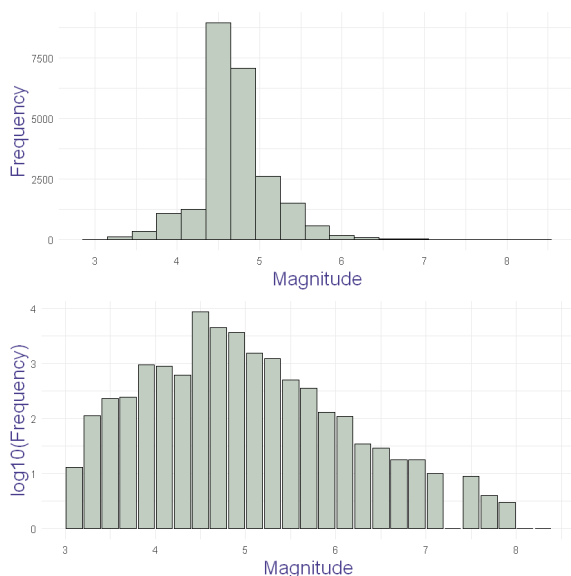


Figure 7. se demuestra que se cumple la ley de G-R Gutenberg-Richter.

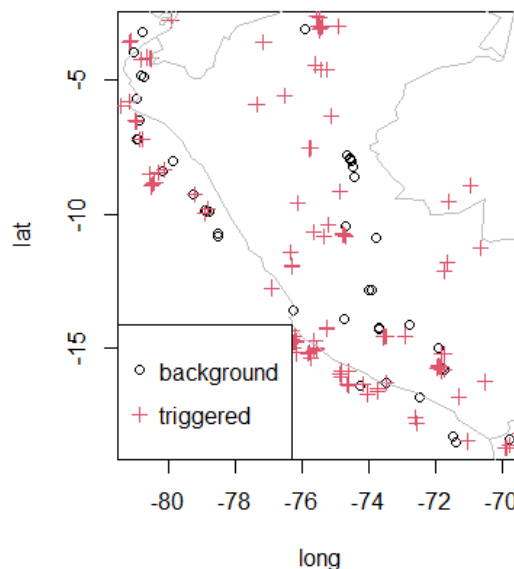


Figure 8. Eventos provocados en la simulación.

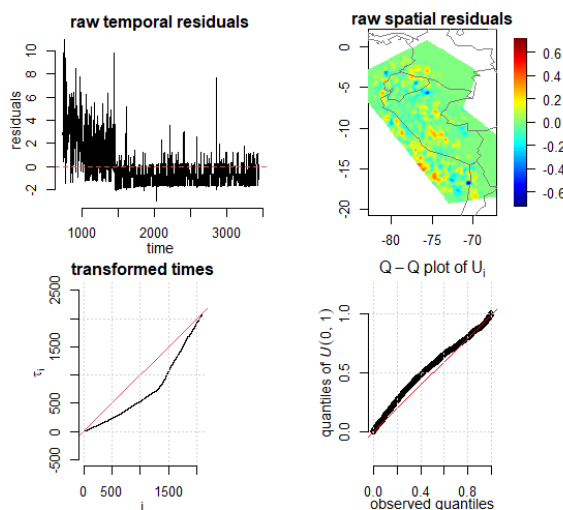


Figure 9. Análisis de Residuos.

## Agradecimientos

Al docente del curso por impartir conocimientos necesarios para desarrollar el presente trabajo. A CORSSA: the Community Online Resource for Statistical Seismicity Analysis por proporcionar recursos valiosos en el análisis estadístico de la sismicidad. A Kowshik chilamkurthy por proporcionar una vista más simple de los procesos Hawkes. Finalmente al IGP por proporcionar datos para el presente estudio.