



Facultad de Ciencias
Instituto de Estadística
Ingeniería en Estadística

Un método de Deep learning para el pronóstico espacio-temporal de eventos sísmicos

Angello Jorge Canales Olivares
29 de diciembre de 2021

Profesor Guía

Rodrigo Salas. Ph.D.

Escuela de Ingeniería C. Biomédica, Universidad de Valparaíso

Profesor Co-Guía

Daira Velandia Muñoz. Ph.D.

Instituto de Estadística, Universidad de Valparaíso

Proyecto de titulación para optar al:

grado académico de: *Licenciado en Estadística*

título profesional de: *Ingeniero en Estadística*

minor en: *Míneria de datos*

Resumen

Los eventos sísmicos son una de las catástrofes naturales más devastadoras para la civilización humana. Sus consecuencias son tan graves que pueden provocar daños estructurales notables, genera una gran cantidad de damnificados e incluso desencadenar otros tipos de tragedias como por ejemplo los tsunamis, esto sin mencionar la gran cantidad de muertes que pueden ocurrir producto de esto.

El momento exacto en el que puede ocurrir este tipo de catástrofe es completamente incierto, en propias palabras de Charles Richter: “Sólo los tontos, los charlatanes y los mentirosos predicen los terremotos”(Johnson 2020). Sin embargo, la ciencia dispone en la actualidad de instrumentos y técnicas matemáticas muy avanzadas para controlar las zonas sísmicas, a pesar de esto a ciencia cierta no es posible predecir la magnitud exacta y el momento de ocurrencia.

Chile es considerado, a nivel mundial, uno de los territorios que posee un alto riesgo de eventos sísmicos devastadores, siendo históricamente, el único país del mundo hasta la fecha que ha sufrido un cataclismo de magnitud 9,5 en el año 1960 en la Región de Magallanes, ocurrido las 15:11 horas del domingo 22 de mayo, según menciona un testimonio en la noticia ” Desde los relatos de sobrevivientes: 60 años del terremoto y maremoto de Valdivia”(22 de may. de 2020) Periódico Resumen, la tierra se “tragó” a algunas personas como si se hablase de una película de Hollywood. La magnitud de este evento fue tal que incluso lo describen como equiparable a 89 veces todo el arsenal nuclear existente.

La nación chilena, se ubica en la placa tectónica sudamericana, en su límite occidental donde convergen las placas de Nazca y Antártica y éstas generan una zona de subducción con la placa sudamericana.

El presente proyecto busca estimar la tasa de ocurrencia sísmica en la zona de subducción de las 2 placas, utilizando una arquitectura de red neuronal profunda (DNN), en base a un enfoque del modelo espacio-temporal ETAS (secuencias de réplicas de tipo epidémico). Este modelo corresponde a una función de intensidad condicional de procesos de puntos, la cual determina la tasa de ocurrencia sísmica espacio-temporal. Este modelo considera dos tipos de sismicidad, las cuales son, la sismicidad activa o de racimo y la sismicidad de fondo, donde los eventos de fondo desencadenan los eventos de racimo. El modelo ETAS propuesto se estima mediante una técnica semiparamétrica teniendo en cuenta los componentes paramétricos y no paramétricos correspondientes a la sismicidad activada y de fondo, respectivamente. Luego, el modelo se utiliza para predecir la incidencia sísmica temporal y espacial.

Dicha estimación de la tasa de ocurrencia sísmica se analizará de forma temporal, con una arquitectura de redes neuronales LSTM esta se utiliza para aprender de experiencias importantes y variaciones que han pasado en el tiempo; además de una arquitectura de red CNN, la cual se

utilizará para la predicción de la probabilidad de que la intensidad máxima se produzca en un lugar determinado.

Planteamiento del problema

La problemática nace sobre la necesidad de desarrollar herramientas de aprendizaje profundo que permitan descubrir el comportamiento la incidencia sísmica en distintas zonas. Esto debido al interés que representa este tipo de estudios para Chile debido a la gran cantidad de actividad sísmica que presenta esta región.

Hipótesis de trabajo

Es posible estimar la tasa de incidencia sísmica en la región de subducción entre la placa Sudamericana y la placa de Nazca mediante el uso de redes neuronales recurrentes y el preprocesamiento de un catálogo de eventos sísmicos con el modelo ETAS.

Objetivo general

Estimar con una alta probabilidad la intensidad y ubicación de eventos sísmicos en la zona de subducción de la placa de Nazca con la placa Sudamericana, a partir del uso de redes neuronales y de resultados obtenidos a través del modelo ETAS

Objetivos específicos

1. Obtener la tasa de ocurrencia sísmica diaria, en un periodo de tiempo de 2001-01-01 a 2021-01-01, a partir de el modelo ETAS espacio-temporal
2. Entrenar una red neuronal recurrente y una convolucional, para predecir la tasa de ocurrencia de eventos sísmicos y su ubicación en el espacio, en la zona de subducción de la placa de Nazca con la placa sudamericana.

Índice general

Resumen	3
Resumen	3
1. Estado del arte	5
1.1. Discusión Bibliográfica	5
2. Fundamentación teórica	7
2.1. Procesos puntuales	7
2.2. Propiedades de primer y segundo orden	8
2.3. Estacionariedad	9
3. Modelo de secuencia de réplicas de tipo epidémico	10
3.1. Variables	10
3.2. Modelo ETAS temporal	11
3.3. Modelo ETAS espacio-temporal	12
3.4. Estimación del modelo ETAS	14
3.5. Algoritmo 1: Davidon-Fletcher-Powell	16
3.6. Algoritmo 2: Desagrupamiento estocástico	17
4. Arquitecturas de redes neuronales	19
4.1. Red neuronal recurrente Long short term memory	19
4.2. Red neuronal Convolutacional	20
5. Resultados	22
5.1. Metodología	22
5.2. Base de datos	22
5.3. Análisis descriptivo	22
5.4. Resultados del modelo ETAS temporal	24
5.5. Resultados Red neuronal LSTM	24
5.6. Resultados modelo ETAS espacio-temporal	26
5.7. Análisis residual	27
5.8. Base de datos de imágenes	27
5.9. Resultados Red neuronal convolutacional	28
6. Conclusiones	31
Referencias	32

Capítulo 1

Estado del arte

1.1. Discusión Bibliográfica

[Harte \(2010\)](#) describe el paquete PtProcess que utiliza el lenguaje estadístico R. Además de explicar de manera detallada los antecedentes matemáticos correspondientes a un proceso de puntos espacio-temporal, proporciona un enfoque unificado para ajustar y simular una amplia variedad de modelos de procesos puntuales temporales o de procesos puntuales marcados.

En su trabajo [Nicolis \(2017\)](#), estima la función de intensidad espacio-temporal del proceso sísmico que describe la actividad chilena enfocado en 2 eventos principales, se consideró un periodo de tiempo en el cual se encuentran 3 eventos de fondo importantes. El terremoto 27F-2010, el terremoto de Iquique de 2014 y el terremoto de Illapel de 2015. Su principal objetivo fue obtener estimaciones diarias de la probabilidad de observar al menos un evento de magnitud mayor que la magnitud umbral m_0 , utilizando un enfoque llamado predicción de probabilidad directa para su estimación. Sin embargo, se puede concluir que un estudio de este tipo necesita mejores desarrollos para sus estimaciones referido al ajuste espacial. [? \(?\)](#) asegura que para ello, se necesitan versiones más flexibles del modelo ETAS, teniendo en cuenta, por ejemplo, las estructuras geofísicas espaciales observadas.

Por otra parte, [Jalilian \(2019\)](#) realiza una extensión en Rstudio, utilizando el enfoque de desagrupamiento estocástico para realizar un catálogo de eventos sísmicos y analizarlo por medio del modelo ETAS. También se hace uso de una técnica de segmentación poligonal para estudiar regiones geográficas específicas en el plano, con un catálogo Ht estructurado como $Ht = (ti, xi, yi, mi)$. Se exploran resultados de análisis realizados en Iran, Japón e Italia en donde se estudia la tasa de sismicidad en periodos de tiempo T específicos. Además, de estimar una ocurrencia de eventos con una probabilidad mayor a 0.95 en el territorio de estudio.

Una extensión al modelo ETAS es propuesta por [Ogata \(2006\)](#) por medio de la “Ley de Omori modificada” y “la ley de escala de Utsu-Seki del área de la réplica frente a la magnitud”, a fin de detectar la sismicidad anómala, como la inactividad y la activación en la región de Japón. Se propone una versión ampliada del mejor modelo de espacio-tiempo ajustado, obteniendo resultados significativamente mejores a la sismicidad en varias regiones en y alrededor de Japón.

Lo propuesto por [Plaza, 2019](#)) Corresponde a un estudio que combina los antecedentes estadísticos del modelo ETAS y el uso de técnicas de deep learnin. Se basa en 2 arquitecturas de redes

neuronales profundas para predecir la tasa sísmica en Chile. La primera de ellas corresponde a una red LSTM y la segunda corresponde a una arquitectura de red neuronal convolucional. En base a estimaciones hechas a partir del modelo de secuencias de réplicas de tipo epidémico, ya que estas proporcionan la información necesaria para entrenar la DNN (Deep neural network). En particular, el LSTM (Long Short Term Memory) se entrenó utilizando la intensidad máxima de 30 días como entrada para predecir la intensidad máxima del día siguiente, y la CNN se entrenó en las últimas 30 imágenes proporcionadas por la aplicación del modelo ETAS para predecir la probabilidad de que el próximo día el evento máximo será en cierta zona de Chile.

Capítulo 2

Fundamentación teórica

2.1. Procesos puntuales

Los puntos son la forma más básica de datos espaciales. El conjunto de localizaciones en las que observamos $Z(s)$ es aleatorio. Si medimos la magnitud de los sismos en Chile, estaremos trabajando con este tipo de procesos espaciales, puesto que la magnitud es una variable aleatoria y no tenemos control sobre las localizaciones de los eventos, es decir, de los epicentros.

Para entender la fundamentación del modelo de secuencia de réplicas de tipo epidémico (ETAS), necesitamos entender los teoremas de los procesos puntuales, los cuales se definen como un tipo de modelo estocástico el cual describe la distribución espacial de eventos u observaciones de interés. Según [Calahorra \(2017\)](#) esta tipología de datos y modelos resultan útiles para resolver las siguientes cuestiones en el ámbito científico:

- Averiguar si los datos del modelo espacial están agrupados, si se distribuyen de manera regular o de forma aleatoria;
- Hallar relaciones entre los diferentes tipos de eventos en los procesos marcados;
- Calcular funciones de densidad (número de eventos por unidad de área).

Los procesos puntuales espacio-temporales, se definen como modelos no determinados, sometidos al azar (estocásticos), los cuales describen la localización y el tiempo en el cual ocurre una observación o evento de interés. Estos modelos se centran en determinar la distribución espacial (y sus causas) de eventos específicos, cuyas ubicaciones se han obtenido dentro de un área de estudio, además de estudiar tendencias o patrones que pueden presentarse de manera temporal en él. Tales procesos espaciales y temporales forman un conjunto numerable, el cual se denota como:

$$\mathbb{P} = \{(s_i, t_i) : i = 1, 2, \dots\}, \quad (2.1)$$

donde

- $s_i \in \mathbb{R}^2$ corresponde a la localización espacial.

- $t_i \in \mathbb{R}^+$ corresponde al tiempo de ocurrencia.

El dominio por el cual se rigen los puntos es $W \times T$, donde W se puede definir como un cuadrilátero o un polígono cerrado y T un intervalo único cerrado.

Harte (2010) explica que los dos componentes más importantes de un modelo de este tipo son la función de intensidad de fondo y la distribución de las marcas. La función de intensidad del terreno describe el ritmo al que se producen los eventos en el tiempo, y no sólo está influenciada por el tiempo actual, también está relacionada con los eventos que ocurrieron antes del tiempo actual, es decir, la historia del proceso. El historial de eventos, posee la siguiente estructura:

$$\mathbf{H}_t = \{(t_i, x_i, y_i) | \forall i : t_i < t\}, \quad (2.2)$$

donde t_i es el tiempo del i -ésimo suceso y $s_i = (x_i, y_i)$ corresponde a las coordenadas espaciales. La función de intensidad condicional que describe la tasa instantánea de Poisson posee la siguiente forma:

$$\lambda(t, x, y | \mathbf{H}_t) = \lim_{\delta, \xi, \eta \rightarrow 0} \frac{1}{\delta \xi \eta} \mathbb{P} \{N_{\delta, \xi, \eta}(t, x, y) > 0 | \mathbf{H}_t\}, \quad (2.3)$$

donde $N_{\delta, \xi, \eta}(t, x, y)$ es el número de eventos que ocurren en $[t, t + \delta) \times [x, x + \delta) \times [y, y + \delta)$. La función de intensidad de fondo describe la tasa de Poisson instantánea en función del tiempo únicamente. Entonces ésta se define como

$$\lambda_g(t, \theta_1, \dots, \theta_m) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{P} \{N_\delta(t) > 0 | \mathbf{H}_t\}. \quad (2.4)$$

El vector $(\theta_1, \dots, \theta_m) \in \Theta_m$ corresponde a un vector de parámetros, siendo $\Theta_m \subseteq \mathbb{R}^m$.

2.2. Propiedades de primer y segundo orden

Calahorra (2017) explica que las propiedades de primer orden se describen por la intensidad del proceso, definida como

$$\lambda(s, t) = \lim_{|ds| \rightarrow 0, |dt| \rightarrow 0} \frac{\mathbb{E}[Y(ds, dt)]}{|ds| |dt|} \quad (2.5)$$

Se define a $\lambda(s, t)$ como el número medio de eventos por unidad de volumen (s, t) , ds se define como un subconjunto de s ($|ds|$ es su área) y dt como un pequeño subconjunto que contiene el tiempo t ($|dt|$ es su longitud); además, $Y(ds, dt)$ corresponde al número de eventos acontecidos en $ds \times dt$.

Las propiedades de segundo orden se describen como la relación entre el número de eventos en pares de subregiones dentro de $S \times T$. Se define como

$$\lambda_2((s_i, t_i), (s_j, t_j)) = \lim_{|D_i| \rightarrow 0, |D_j| \rightarrow 0} \frac{\mathbb{E}[Y(D_i, D_j)]}{|D_i||D_j|}, \quad (2.6)$$

Se define el producto los pares de (ds, dt) como $D_i = ds_i \times dt_i$ y $D_j = ds_j \times dt_j$ son cilindros que contienen los puntos $(s_i, t_i), (s_j, t_j)$, respectivamente.

Se define la densidad de la covarianza entre $(s_i, t_i), (s_j, t_j)$ como

$$\gamma((s_i, t_i), (s_j, t_j)) = \lambda_2((s_i, t_i), (s_j, t_j)) - \lambda(s_i, t_i)\lambda(s_j, t_j), \quad (2.7)$$

a su vez, se expresa la correlacion de pares como:

$$g((s_i, t_i), (s_j, t_j)) = \frac{\lambda_2((s_i, t_i), (s_j, t_j))}{\lambda(s_i, t_i)\lambda(s_j, t_j)} \quad (2.8)$$

La función de correlación de pares puede interpretarse como la densidad de probabilidad estándar aplicada a que un evento ocurra en cada uno de dos cilindros centrados en los puntos.

2.3. Estacionariedad

Calahorra (2017) explica que los procesos puntuales espacio-temporal $\{(s, t), s \in W, t \in T\}$ son estacionarios, tanto en primer y segundo orden con las siguientes propiedades

- Será estacionario en el espacio si: $\lambda(s, t) \equiv \lambda(t)$ y $\lambda_2((s, t), (s', t)) = \lambda_2(s - s', t)$
- Será estacionario en el tiempo si : $\lambda(s, t) \equiv \lambda(s)$ y $\lambda_2((s, t), (s, t')) = \lambda_2(s, t - t')$
- Será estacionario en el tiempo y el espacio si : $\lambda(s, t) \equiv \lambda$ y $\lambda_2((s, t), (s', t')) = \lambda_2(s - s', t - t')$

Además, un proceso espacio-temporal es isotrópico, es decir, sus propiedades en el espacio no dependen desde donde sea examinado, si $\lambda_2((s, t), (s', t')) = \lambda_2(u, v)$. El vector (u, v) corresponde a una diferencia espacio-temporal denotada por $u = \|s - s'\|$ y $v = |t - t'|$.

Capítulo 3

Modelo de secuencia de réplicas de tipo epidémico

En este capítulo se definirán los conceptos básicos del modelo de secuencia de réplicas de tipo epidémico, la cual es utilizada en este proyecto como preprocesamiento de los datos. Se destacan los conceptos de la función temporal, como espacio-temporal, además de esto se describe los distintos métodos y algoritmos de estimación paramétrica para este tipo de modelos.

El modelo ETAS se basa en la suposición de que todos los terremotos pueden provocar réplicas. Es por esto por lo que el modelo de secuencia de réplicas de tipo epidémico es ideal para identificar las anomalías de los enjambres. Estas anomalías de enjambre se representan por medio de los parámetros que poseen las funciones del modelo, estos representan características que posea el historial de eventos en la región S . Dicho esto para el análisis y aplicación de este modelo se catalogan los eventos sísmicos como, eventos desencadenados o eventos principales o de fondo.

[Harte \(2010\)](#) explica que el modelo asume que ciertas secuencias de réplicas de terremotos pueden modelarse como una “epidemia”, de grandes terremotos inducen más réplicas, como si nos refiriéramos a una tasa de infección en un intervalo de tiempo determinado y, además, la secuencia de réplicas se prolonga durante más tiempo después del después de la sacudida (sismo) principal.

El modelo de secuencia de réplicas de tipo epidémico (ETAS) se basa en los modelos estocásticos de procesos puntuales espacio-temporales. Este tipo de modelos necesita datos, en los cuales, la información se obtenga en base a una distribución de puntos en un espacio bidimensional. Se debe considerar al tiempo como una coordenada más y contemplar este proceso como un caso especial de un proceso espacial de dimensión superior.

3.1. Variables

En el modelo se considera a N como un catálogo de eventos sísmicos tomando $t_i, x_i, y_i, m_i, i = 1, \dots, N$

Estos datos se consideran como el historial de eventos sísmicos

$$H_t = \{(t_i, x_i, y_i) | \forall i : t_i < t\}. \quad (3.1)$$

Evento	Tiempo	Longitud	Latitud	Magnitud
1	t_1	x_1	y_1	m_1
2	t_2	x_2	y_2	m_2
\vdots	\vdots	\vdots	\vdots	\vdots
N	t_N	x_N	y_N	m_N

Cuadro 3.1: Estructura de las variables

Digamos que t_i corresponde a los días (hora, minuto y segundo) en los que ocurrió un evento sísmico, comienza con $t = 0$ correspondiente al origen en el tiempo específico, digamos 2000 – 01 – 01, 00 : 00 : 00, y un período de estudio de duración T días comenzó en el momento $t_i \geq 0$ es de interés. Solo eventos dentro de la región de estudio S y el período de estudio $[t_i, t_i + T]$ son considerados al momento de realizar estimaciones. El término magnitud fue introducido por Charles F. Richter inspirado por el concepto de magnitud estelar que se emplea en astronomía al momento de analizar el brillo intrínseco de una estrella. Los cambios que se generan al momento de visualizar una estrella son logarítmicos, y así es como Richter plantea su escala de magnitud [Peláez \(2011\)](#).

3.2. Modelo ETAS temporal

El modelo ETAS, en su versión simple (no espacial) es un ejemplo de una función de intensidad de fondo. [Harte \(2010\)](#) denota la función como

$$\lambda_\theta(t|H_t) = \mu + A \sum_{i:t_i < t} e^{\alpha(M_i - M_0)} \left(1 + \frac{t - t_i}{c}\right)^{-p}, \quad (3.2)$$

Los valores del vector de parámetros $\theta = (\mu, A, \alpha, c, p)$ son todos de carácter positivo, t_i es el momento del i -ésimo evento con magnitud M_i y M_0 es la magnitud de corte o magnitud umbral. En esta función se plantea que los sucesos se generan como un proceso de Poisson con tasa μ . El término $e^{\alpha(M_i - M_0)}$ dentro de la sumatoria, explica que los sucesos más grandes aumentan la intensidad más, y el último término determina la longitud (tiempo) de la secuencia de réplicas.

[Ogata \(1998\)](#) explica que la actividad de las réplicas está representada por un proceso de Poisson no estacionario según la fórmula de Omori modificada, de manera que la tasa de ocurrencia de las réplicas en el momento t que sigue al i -ésimo terremoto (t_i, m_i) viene dada por

$$v_i(t) = \frac{K_0}{(t - t_i + c)^p} \cdot e^{\alpha(M_i - M_0)}, \quad (3.3)$$

para $t > t_i$, donde los parámetros K_0, α, c, p son constantes comunes a todos los i .

3.3. Modelo ETAS espacio-temporal

Jalilian (2019) define el modelo de secuencia de réplicas de tipo epidémico, algebraicamente como

$$\lambda_{\beta,\theta}(t, x, y, m | \mathbf{H}_t) = \nu_{\beta}(m) \lambda_{\theta}(t, x, y | \mathbf{H}_t), \quad (3.4)$$

donde

$$\nu_{\beta}(m) = \beta \mathbf{exp}[-\beta(m - m_0)], \quad \beta > 0 \quad (3.5)$$

Corresponde a la función de densidad de probabilidad asociada a la magnitud de un evento, y además, la función de intensidad condicional se plantea de la siguiente forma:

$$\lambda_{\theta}(t, x, y | \mathbf{H}_t) = \tilde{u}(x, y) + \sum_{i:t_i < t} k_{a,\alpha}(m_i) g_{c,p}(t - t_i) f_{D,\gamma,q}(x - x_i, y - y_i; m_i). \quad (3.6)$$

donde

- t se expresa como el dominio asociado al tiempo.
- (x,y) corresponden a las coordenadas asociadas a la ubicación espacial.
- $\tilde{u}(x, y)$ se refiere a la tasa de sismicidad de fondo, que expresada en su forma semi-parametrica corresponde a $u\mu(x, y)$ con $\mu > 0$ y $u(x, y)$ que es una función suavizada en S .
- $k_{a,\alpha}$ corresponde a la esperanza e eventos desencadenados (réplicas) generados a partir de un evento de magnitud m_i : $A \mathbf{exp}[\alpha(m - m_0)]$; $\alpha, A > 0$.
- $g_{c,p}(t - t_i)$ denominada como "Ley de Omori". Según "The 150th Anniversary of Fusakichi Omori" (2018) "La ley de Omori establece que después de un fuerte terremoto, la frecuencia de las réplicas n , es decir, choques repetidos que siguen al choque principal, decaen con el tiempo t , en promedio, de acuerdo con el ley hiperbólica".
- $f_{D,\gamma,q}(x - x_i, y - y_i; m_i)$ Corresponde a la distribución espacial de los eventos desencadenados.

La "Ley de Omori." "Fórmula de Omori", se expresa de manera algebraica como:

$$g_{c,p}(t - t_i) = \begin{cases} \frac{p-1}{c} \left(1 + \frac{t-t_i}{c}\right) & t - t_i > 0 \\ 0 & t - t_i \leq 0, \end{cases} \quad (3.7)$$

Dicha fórmula corresponde a una relación empírica para la decadencia temporal de tasas de réplicas generado a partir de un evento de magnitud m_i ocurriendo en el tiempo t_i además, c y p

son parámetros correspondientes a la actividad sísmica, p es un parámetro considerado para la tasa de desintegración de la réplica. Se supone que la distribución de probabilidad del tiempo hasta la ocurrencia de un evento desencadenado es una función del lapso $t-t_i$ de su choque principal directo y es independiente de m_i . En base a la ley de Omori se asume que cada replica puede desencadenar un racimo de replicas o replicas provocadas interactivamente incluidos los que se encuentran en regiones cercanas a la falla. Es por esto por lo que la tasa de ocurrencia en el tiempo t se da por una superposición ponderada de la llamada ley de Omori, modificando funciones cambiadas en el tiempo. Ver (Ogata, 2006)

Luego se tiene la función de densidad de la distribución espacial de los eventos desencadenados, la cual se define como

$$f_{D,\gamma,q}(x - x_i, y - y_i; m_i) = \frac{q - 1}{\pi D \exp[\gamma(m_i - m_0)]} \left(1 + \frac{(x - x_i)^2 + (y - y_i)^2}{D \exp[\gamma(m_i - m_0)]} \right)^{-q}, \quad (3.8)$$

Hernandez (2012) explica esta distribución como una desintegración gaussiana de corto alcance, donde D es una constante, además se tiene que $D > 0$, $\gamma > 0$ y $q > 1$ parámetros desconocidos. En otras palabras, se refiere a la distribución o ubicación de la ocurrencia de un evento desencadenado a partir de las coordenadas espaciales (x, y) y una magnitud m_i

Jalilian (2019) explica que los eventos de fondo, o eventos principales, se generan a partir de un proceso Poisson el cual posee intensidad asociada a la tasa de sismicidad de fondo, es decir en $\tilde{\mu}(x, y)$, que teóricamente es estacionario en el tiempo. El parámetro μ en $\tilde{\mu}(x, y) = \mu u(x, y)$ se utiliza para fijar la convergencia del modelo. Teóricamente, se afirma que cualquier tipo de evento anterior, ya sea de fondo o desencadenado (réplicas), generan más eventos a partir de un proceso Poisson no estacionario

$$\sum_{i:t_i < t} k_{a,\alpha}(m_i) g_{c,p}(t - t_i) \times f_{D,\gamma,q}(x - x_i, y - y_i; m_i). \quad (3.9)$$

Si nos centramos en un evento principal, independiente de su magnitud, podemos obtener la esperanza de los eventos desencadenados como

$$\int_{m_0}^{\infty} k_{A,\alpha}(m) v_{\beta} dm = \frac{A\beta}{\beta - \alpha}. \quad (3.10)$$

Se dice que si $\frac{A\beta}{\beta - \alpha} < 1$ entonces es estable.

Si la condición de estabilidad se cumple el número esperado de eventos en un intervalo de longitud T es proporcional a T . Por lo tanto, el número esperado de eventos tiende a infinito. Teóricamente, la función de intensidad espacial total se define como

$$\Lambda(x, y) \approx \mu u(x, y) + \frac{1}{T} \sum_{i:t_i < T} k_{a,\alpha}(m_i) f_{D,\gamma,q}(x - x_i, y - y_i; m_i), \quad (3.11)$$

Finalmente, esto se utiliza para obtener el coeficiente de agrupación (activación), esto es para cuantificar el efecto de agrupación en relación con la intensidad espacial total en cualquier lugar $(x, y) \in S$.

$$\omega(x, y) = 1 - \frac{u(x, y)}{\Lambda(x, y)}. \quad (3.12)$$

[Jalilian \(2019\)](#) explica un enfoque probabilístico para desclasificar un catálogo de terremotos a fin de obtener la tasa de sismicidad de fondo. Sugiere utilizar lo siguiente

$$p_{ij} = \begin{cases} \frac{k_{a,\alpha}(m_i) g_{c,p}(t-t_i) f_{D,\gamma,q}(x-x_i, y-y_i; m_i)}{\lambda_\theta(t, x, y | H_t)} & , t_j > t_i \\ 0 & , t_j \leq t_i \end{cases}, \quad (3.13)$$

como la probabilidad de que el suceso j sea provocado por el suceso i . Dado esto, tenemos

$$p_{ij} = \sum_{i:t_i < t_j} p_{ij}, \quad j = 1, \dots, N, \quad (3.14)$$

es la probabilidad de que el suceso j sea provocado por un suceso anterior y

$$1 - p_{ij} = \frac{\mu u(x_i, y_i)}{\lambda_\theta(t_j, x_j)}. \quad (3.15)$$

3.4. Estimación del modelo ETAS

Para realizar la estimación del vector de parámetros en este modelo es posible hacer uso del estimador máximo verosímil. [Harte \(2010\)](#) sostiene la idea de derivar la función de verosimilitud para el modelo simple, ya que la función de intensidad condicional es sólo una función del tiempo, es decir, toma la forma de la función de intensidad del suelo en ecuación.

$$\log L = \sum_{i:T_1 \leq t_i \leq T_2} \log \lambda_\theta(t_i | H_{t_i}) - \int_{T_1}^{T_2} \lambda_g(t | H_t) dt \quad (3.16)$$

Para el caso de la versión espacio-temporal, debido a la separabilidad entre β y el vector de parámetros θ , se tiene la siguiente función de verosimilitud

$$L(\beta, \theta | \mathbf{H}_t) = l_1(\beta | \mathbf{H}_t) + l_2(\theta | \mathbf{H}_t), \quad (3.17)$$

donde

$$l_1(\beta | \mathbf{H}_t) = \log(\beta) \sum_{i=1}^N \delta_i - \beta \sum_{i=1}^N \delta_i (m_i - m_0), \quad (3.18)$$

$$l_2(\theta | \mathbf{H}_t) = \sum_{i=1}^N \delta_i \log(\lambda_\theta(t_i, x_i, y_i | \mathbf{H}_{t_i})) - \int_{t_{start}}^{t_{start}+T} \int \int_S \lambda_\theta(t, x, y | \mathbf{H}_t) dx dy dt \quad (3.19)$$

[Jalilian \(2019\)](#) obtiene un estimador insesgado para β como

$$\hat{\beta} = \frac{N'}{\sum_{i=1}^N \delta_i (m_i - m_0)} \quad (3.20)$$

Para obtener estimaciones del parámetro θ , [Jalilian \(2019\)](#) hace uso de un algoritmo de llamado Davidon-Fletcher-Powell, para esto se requiere una aproximación numérica al parámetro μ sobre términos integrales en la región de estudio S y además un método de maximización numérica apropiado para $l_2(\theta | \mathbf{H}_t)$

[Ogata \(1998\)](#) sugirió una partición radial de la región geográfica S para aproximar eficazmente aproximar el término integral, la cual algebraicamente tiene la siguiente forma

$$\int \int_S f_{D,\gamma,q}(x - x_i, y - y_i; m_i) dx dy = \int \int_{S^{(i)}} f_{D,\gamma,q}(x - x_i, y - y_i; m_i) dx dy, \quad (3.21)$$

donde

$$S^{(i)} = (x - x_i, y - y_i) : (x, y) \in S, i = 1, \dots, N. \quad (3.22)$$

Esta integral asume que la región de estudio es un polígono con n_v subdivisiones de región $S_1^{(i)}, \dots, S_{n_v}^{(i)}$ por lo tanto la integral se realiza desde las coordenadas cartesianas (x, y) a las coordenadas polares (r, ϑ) .

[Jalilian \(2019\)](#) además explica un enfoque iterativo para la estimación simultánea de la función suavizada $u(x, y)$, $(x, y) \in S$, y el vector de parámetros en el marco de la desclasificación estocástica. Así consigue que la estimación de los parámetros converja a un valor luego de varias iteraciones.

El algoritmo comienza con una matriz hessiana de $\xi(\theta)$ en $\theta = \theta^*$, siendo $\xi(\theta) = -l_2(\theta|H_t)$ dado por

$$(H_k^{-1})_{ij} \rightarrow \frac{d^2}{d\theta_i d\theta_j} \xi(\theta)|_{\theta=\theta^*}. \quad (3.23)$$

La estructura del algoritmo propuesto por [Jalilian \(2019\)](#) es presentada a continuación:

3.5. Algoritmo 1: Davidon-Fletcher-Powell

1. Entrada:
 - $\hat{\theta}_0$: Estimación inicial de los parámetros del modelo
 - $\hat{u}_0(x, y)$: Tasa de sismicidad inicial
 - H_0 : Conjetura inicial para la inversa de la matriz hessiana
2. $k \leftarrow 0$
3. Repetir
4. ▪ $d_k \leftarrow -H_k \times \nabla \xi(\hat{\theta}_k)$: Dirección de ascenso
5. ▪ $\zeta_k \leftarrow \arg \min_{\zeta} \xi(\hat{\theta}_k + \zeta d_k)$: Búsqueda de líneas a lo largo de d_k
6. ▪ $\hat{\theta}_{k+1} \leftarrow \hat{\theta}_k + \zeta_k d_k$: Actualizar la estimación
7. ▪ $\eta_k \leftarrow \nabla \xi(\hat{\theta}_{k+1}) - \nabla \xi(\hat{\theta}_k)$
8. ▪ $c_1 \leftarrow \zeta_k / (d_k^\top \times \eta_k)$
9. ▪ $c_2 \leftarrow 1 / (\eta_k^\top \times H_k \times \eta_k)$
10. ▪ $H_{k+1} \leftarrow H_k + c_1 h_k \times d_k^\top - c_2 H_k \times \eta_k \times \eta_k^\top \times H_k$: Actualiza el hessiano inverso
11. ▪ $k \leftarrow k + 1$

12. Hasta que (Cr terio de convergencia)

- $\left\| \hat{\theta}_k - \hat{\theta}_{k-1} \right\| < \varepsilon$

 

- $\left\| \nabla \xi(\hat{\theta}_k) \right\| < \varepsilon$

13. Retornar $\hat{\theta}_k$ y H_k

La matriz H_k es estimada por $\Upsilon(\hat{\theta}_k)$

$$\Upsilon_{ij}(\theta) = \mathbf{E} \left[-\frac{d^2}{d\theta_i d\theta_j} l_2(\theta | \mathbf{H}_t) \right], i, j = 1, \dots, 8. \quad (3.24)$$

Adem s

$$\hat{u}(x, y) = \frac{1}{T} \sum_{j=1}^N (1 - \hat{p}_j) \varphi(x - x_j, y - y_j; h_j) \quad (3.25)$$

proporciona una mejor estimaci n para $u(x, y)$, donde

$$\varphi(x, y; h) = \frac{1}{2\pi h^2} \exp\left(-\frac{x^2 + y^2}{2h^2}\right) \quad (3.26)$$

3.6. Algoritmo 2: Desagrupamiento estoc stico

1. Entrada:

- $\hat{\theta}_0$: Estimaci n inicial de los par metros del modelo
- $\hat{u}_0(x, y)$: Tasa de sismicidad inicial

2. $H_0 \leftarrow \Upsilon$: Conjetura inicial para la inversa de la matriz hessiana

3. $k \leftarrow 0$

4. Repetir

5.

- utilice el Algoritmo 1 con entradas $\hat{\theta}_k, \hat{u}_k(x_i, y_i)$ y se obtiene $\hat{\theta}_{k+1}$ y H_{k+1}

6.

- $\hat{p}_j \leftarrow 1 - \hat{\mu}_{k+1} + \hat{u}_k(x_j, y_j) / \lambda_{\hat{\theta}_{k+1}}(t_j, x_j, y_j | \mathbf{H}_t)$: Probabilidad de decluster

7.

- $\hat{u}_{k+1}(x, y) \leftarrow \sum_{j=1}^N (1 - \hat{p}_j) \varphi(x - x_j, y - y_j; h_j) / T$: Actualiza la estimaci n de $u(x, y)$

8. ▪ $k \leftarrow k + 1$
9. ▪ $e_1 \leftarrow \max_l |\hat{\theta}_{l,k} / \hat{\theta}_{l,k-1} - 1|$: Cambio relativo máximo de los parámetros
10. ▪ $e_2 \leftarrow \max_j |\hat{\mu}_k(x_j, y_j) / \hat{\mu}_{k-1} - 1|$: Cambio relativo máximo de $\hat{\mu}(x_j, y_j)$
11. ▪ $e_3 \leftarrow |\xi(\hat{\theta}_k) / \xi(\hat{\theta}_{k-1}) - 1|$: Cambio relativo en la función de Verosimilitud
12. Hasta que (Criterio de convergencia)
 - $e_1 < \varepsilon$
 y
 - $e_2 < \varepsilon$
 y
 - $e_3 < \varepsilon$
13. Retornar $\hat{\theta}_k$ y H_k

Finalmente se tiene la estimación de la función de intensidad total, [Jalilian \(2019\)](#)

$$\hat{\Lambda}(x, y) = 1/T \sum_{j=1}^N \varphi(x - x_j, y - y_j; h_j) \quad (3.27)$$

Capítulo 4

Arquitecturas de redes neuronales

En el siguiente capítulo se describen las arquitecturas de redes neuronales las cuales utilizan los resultados obtenidos por el modelo ETAS.

4.1. Red neuronal recurrente Long short term memory

Las redes Long short term memory tienen un tipo primitivo de memoria en forma de capas recurrentes, estas se consideran como múltiples copias de la misma red, cada una de las cuales pasa un mensaje a un sucesor. Una red neuronal recurrente no tiene una estructura de capas definida, sino que permiten conexiones arbitrarias entre las neuronas, incluso pudiendo crear ciclos, con esto se consigue crear la temporalidad, permitiendo que la red tenga memoria.

Dentro de una red LSTM cada línea lleva un vector completo, desde la salida de un nodo hasta las entradas de otros. Además, se producen operaciones puntuales entre, como la suma de vectores o algún tipo de producto entre ellos (Producto cruz o producto punto) En palabras simples, una red recurrente LSTM es capaz de “recortar” un dato relevante en la secuencia y “preservarlo” para ser enviado a la siguiente célula.

Tiene una estructura de parámetros similar a la de una red neuronal recurrente, además de una estructura de unidades de entrada ubicadas dentro de cada celda, definidas como $f(t)_i$ (para cada paso de tiempo t y celda i), que gestionan el flujo de información dentro y entre las celdas, luego se aplica una unidad sigmoidea para escalar los pesos en $[0-1]$ (Plaza, 2019), donde

$$f_i^{(t)} = \sigma \left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)} \right), \quad (4.1)$$

- $x^{(t)}$ es el vector de entrada;
- $h^{(t)}$ es el vector de la capa oculta, que contiene las salidas de todas las células;
- b^f, U^f, W^f corresponden a los pesos de entrada y los pesos recurrentes para las puertas de olvido, respectivamente,

- $\sigma(-)$ es una unidad sigmoide que trabaja como función de activación.

En (Plaza, 2019) se indica que el estado interno de cada célula se actualiza con un peso condicional de bucle propio $f_i^{(t)}$.

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma \left(b_i^f + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)} \right), \quad (4.2)$$

donde b , U y W denotan los sesgos, pesos de entrada y pesos recurrentes en cada celda, respectivamente. La unidad de la puerta de entrada externa $g_i^{(t)}$ se calcula de forma similar a la puerta de olvido. Está dada por

$$g_i^t = \sigma \left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)} \right), \quad (4.3)$$

donde $h_i^{(t)}$, denotado como la información que sale a través de la célula LSTM, puede procesarse a través de una tangente hiperbólica, que escala los valores de salida entre $(-1, 1)$, se define por

$$h_j^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)}, \quad (4.4)$$

Y también con una puerta de salida $q_i^{(t)}$ en base a la función sigmoidea, con ecuación dada por (4.5)

$$q_i^{(t)} = \sigma \left(b + \sum_j U_i^0 x_j^{(t)} + \sum_j W_i^0 h_j^{(t-1)} \right), \quad (4.5)$$

Con b^o , U^o y W^o sus sesgos, pesos y pesos recurrentes, respectivamente. Este tipo de arquitectura de red neuronal es ideal para analizar los resultados que se obtendrán a partir del modelo ETAS temporal, ya que las lecturas de datos a lo largo del tiempo permiten hallar patrones que sirvan para predecir averías o comportamientos.

4.2. Red neuronal Convolutiva

Las redes neuronales convolucionales o CNN están diseñadas para manejar datos de las imágenes. Sus principales usos se centran en tareas como la clasificación de imágenes, localización de objetos o la subtitulación de imágenes, entre otros.

Brownlee (2018) describe que las redes CNN operan directamente con datos brutos, como los valores brutos de los píxeles, de tal manera que el modelo aprende a extraer automáticamente las características de los datos brutos que son directamente útiles para el problema que se aborda.

Una CNN está compuesta por capas con filtros, que buscan patrones o características particulares en los datos. Cada filtro tiene un núcleo, que está representado por una matriz de números que se asemeja al patrón que el filtro está tratando de detectar (Plaza, 2019).

Una capa típica de CNN consta de tres fases:

1. Aplicaciones de los filtros basados en el kernel en paralelo para producir un conjunto de transformaciones matriciales lineales.
2. Cada activación lineal es decir, la operación matricial lineal que pasa por todos los píxeles de la imagen se ejecuta a través de una función de activación no lineal que podría ser una "relu", una "leaky-relu", "Sigmoid", etc.
3. Se utiliza una función de agrupación para modificar aún más la salida de la capa, para reducir los mapas de características de muestreo resumiendo la presencia de estas características.

La estructura de esta red se implementó a través de tensor Flow en Python.

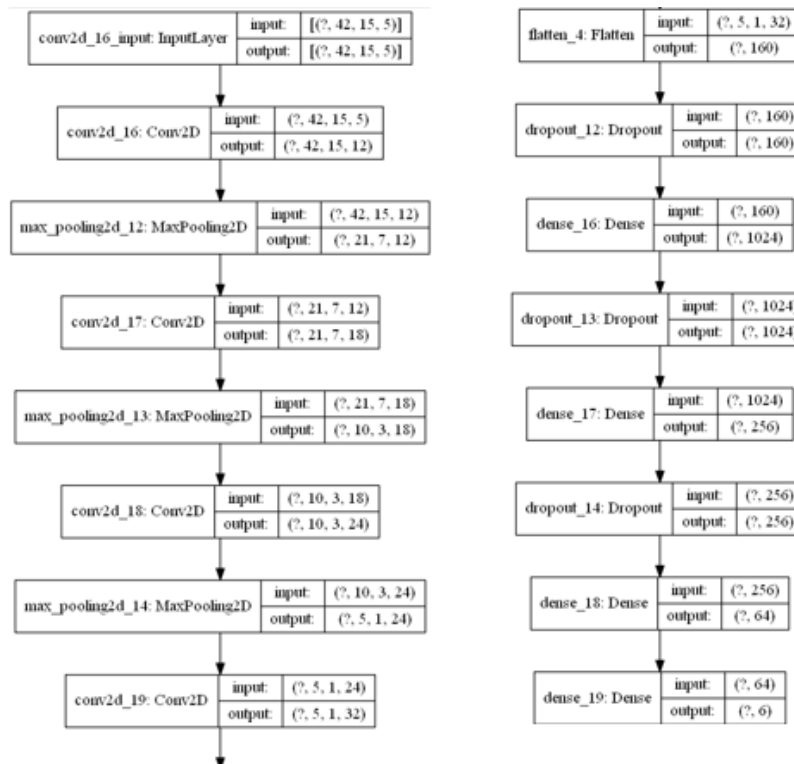


Figura 4.1: Arquitectura propuesta. Fuente: Elaboración propia. Salida de Python.

Capítulo 5

Resultados

5.1. Metodología

Se utiliza el modelo ETAS, a fin de estimar la función de intensidad condicionada, de la cual podemos obtener la tasa de ocurrencia de eventos sísmicos en la zona oeste de Sudamérica por medio de imágenes geoespaciales. Luego se utilizará una arquitectura de red neuronal profunda (CNN) para estimar la intensidad y la ubicación de los terremotos en la zona oeste de Sudamérica, en base a un entrenamiento formado a partir de las imágenes obtenidas del modelo ETAS.

5.2. Base de datos

Los datos corresponden a 13.187 observaciones correspondientes a registros del tiempo y ubicación en el que ocurrieron sismos, estos datos provienen del United States Geological Survey (USGS, www.earthquake.usgs.gov). Posee las mismas variables antes mencionadas en la sección 2.1. El tiempo es desde 2000-01-01 a 2021-01-01 y sus coordenadas van desde longitud -75, hasta -65; latitud -25, hasta -35.

5.3. Análisis descriptivo

Descriptivamente la variable magnitud posee el resumen que se presenta en el cuadro 5.1. Este análisis fue realizado en el software Rstudio.

Min.	1er. Qu.	Mediana	Media	3rd. Qu.	Max.
4.5	4.6	4.7	4.85	5	8.8

Cuadro 5.1: Resumen de magnitud. Fuente: Elaboración propia

Donde el terremoto de mas alta intensidad corresponde a la fecha 27 de febrero del 2010. En el cuadro 5.1 se ve que la media de la magnitud es cercano al mínimo valor, esto es indicador de que los sismos de magnitudes bajas, entre 4.5 y 5, son mucho mas recurrentes.

De forma espacial los sismos mayores a 6.5 en la escala Richter ocurrieron en los lugares observados en la figura 4.1 se muestra una gran cantidad de observaciones entre Perú y Chile, Y a su

vez también se observan eventos sísmicos entre Colombia y Ecuador, mas hacia el interior, alejados de la costa.

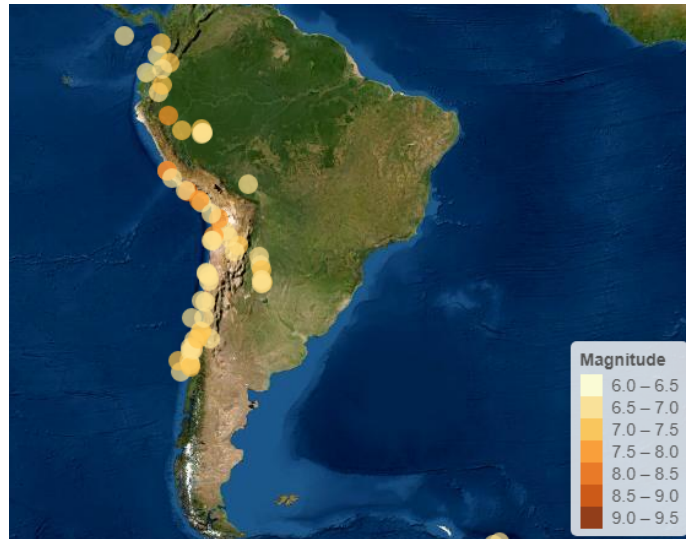


Figura 5.1: Sismos mayores a 6.5 en escala Richter. Fuente: Elaboración propia

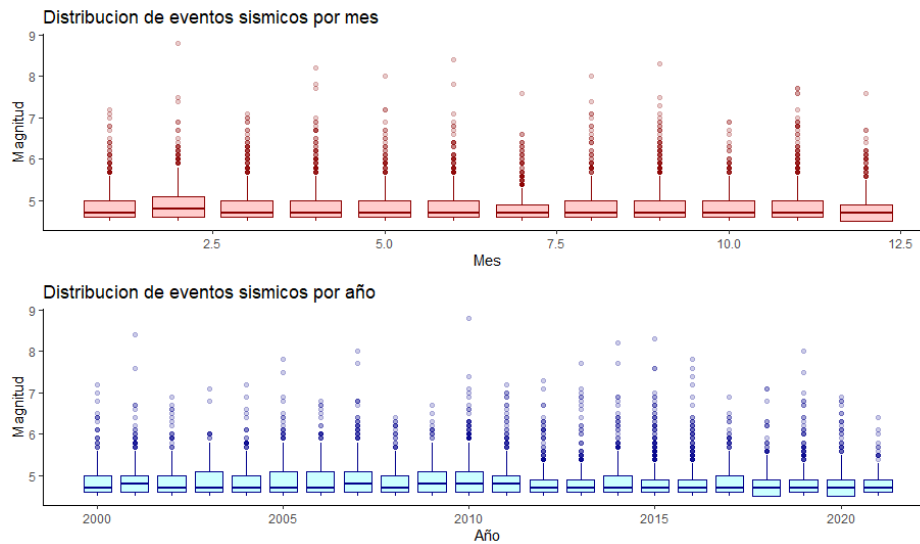


Figura 5.2: Distribución de eventos sísmicos por año y mes. Fuente: Elaboración propia

En la figura 4.2 si observamos la distribución por mes, podemos notar que ocurrieron eventos de gran magnitud teniendo datos atípicos mas altos. Para el caso de la distribución por año podemos que los años 2010 y 2015 poseen una gran densidad. En la figura 4.3 también notamos que existe mayo cantidad de eventos en cada rango entre los años 2010 y 2015.

La figura 5.3 muestra la cantidad de eventos sísmicos por cada año en rangos de 4 a 5, 5 a 6, 6 a 7, 7 a 8 y 8 a 9. Se destaca una alta cantidad de sismos en todos estos rangos en el año 2010, siendo este año el que posee as registros.

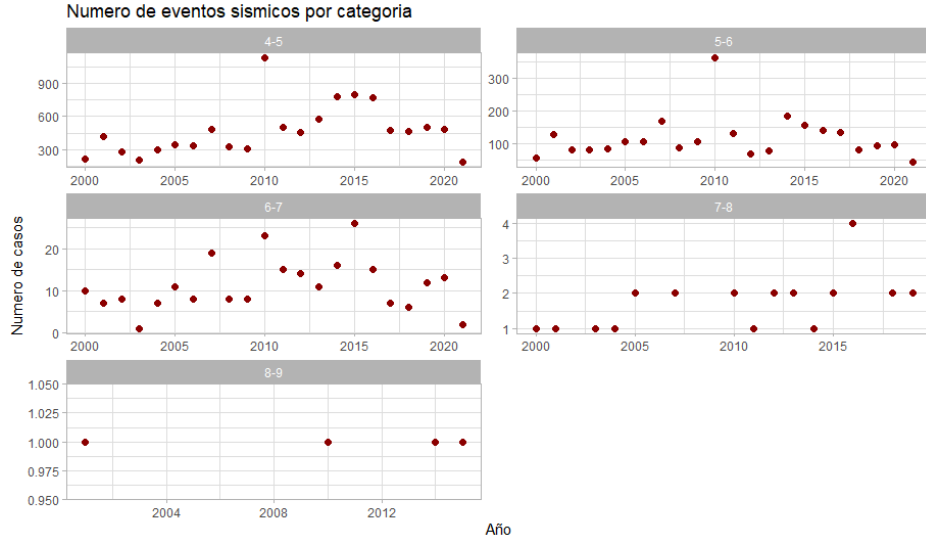


Figura 5.3: Número de eventos sísmicos por categoría. Fuente: Elaboración propia

5.4. Resultados del modelo ETAS temporal

Para una mayor precisión al momento de realizar las estimaciones se eligieron valores iniciales sugeridos por Ogata (1998), $\theta = (0.0015, 0.01, 0.01, 1, 1.3, 0.01, 2, 1)$

Se muestra en la figura 5.4 y 5.5 la estimación de la función de intensidad condicional $\lambda_\theta(t|H_t)$ que representa la tasa de incidencia sísmica, estas poseen distinta magnitud de corte. Como se observa, en el centro de la figura aparece un pico bastante alto, esta observación de gran magnitud corresponde al terremoto del 27 de Febrero del 2010, el cual podemos notar que desencadenó réplicas con magnitudes bastante altas, las cuales además perduraron bastante en el tiempo. Esta estimación se realizó sobre una región $S' \subseteq S$ la cual corresponde a las longitudes $-75, -65$ y latitudes $-40, -25$.

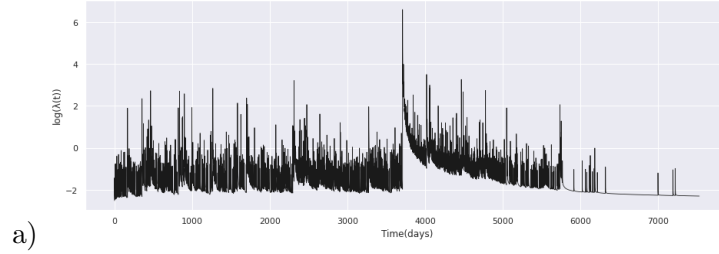
α	\mathbf{c}	\mathbf{p}	\mathbf{K}	μ
1.8451	0.1574	1.1865	0.2002	0.0807

Cuadro 5.2: Estimación de los parámetros

Estos parámetros de la tabla 5.1 representan en conjunto las características de la actividad sísmica de la region S . Los parámetros pueden correlacionarse con las condiciones tectofísicas. Por lo tanto, varían espacialmente, y también temporalmente en algunos casos. El valor de \mathbf{p} indica la tasa de decaimiento de las réplicas, y el valor alfa mide una eficiencia de magnitud. Ogata (2006) explica el parámetro \mathbf{K} (sacudidas diarias) en la función \mathbf{v} representa la cantidad estandarizada por $e^{a(M_i M_c)}$, que mide la productividad de la actividad de las réplicas durante un breve período justo después de la réplica principal.

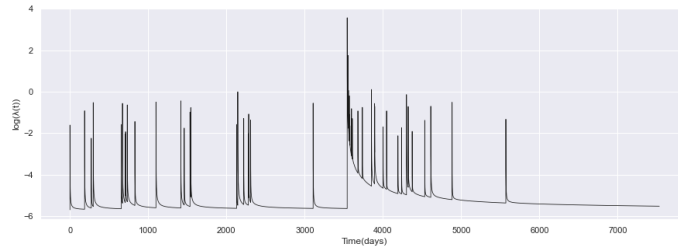
5.5. Resultados Red neuronal LSTM

Para realizar estimaciones con la arquitectura de red neuronal LSTM, subdividieron los datos en un conjunto de entrenamiento de 4670 días y otro de prueba 3127 días, el primero para entrenar a la red y el segundo para probar su desempeño. Estos corresponden al 60% y el 40% de los datos



a)

Figura 5.4: Función de intensidad $M_c = 4.5$



b)

Figura 5.5: Función de intensidad $M_c = 6.0$

(los cuales poseen 13187 observaciones) respectivamente para el conjunto de entrenamiento y el conjunto de prueba. Las intensidades máximas de los últimos 30 días se utilizaron como entradas para que la LSTM predijera la intensidad máxima del día siguiente (salida). La red compuesta por 50 células LSTM en las que se utilizó la tangente hiperbólica como función de activación. Se utilizó un número de 100 para entrenar el modelo.

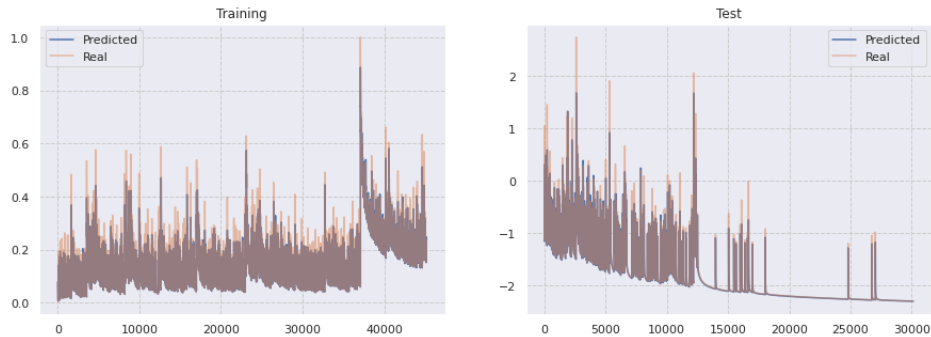


Figura 5.6: Función de intensidad de fondo. Fuente: Elaboración propia

MSE	MAE	RMSE	R2
0.013235	0.019079	0.115043	0.938020

Cuadro 5.3: Métricas de desempeño (Elaboración propia)

el modelo tiene un **R2** bastante alto, lo cual es indicador de que el modelo explica la variabilidad de los datos de forma eficiente. Además, el error cuadrático medio **RMSE** no indica que los valores estimados y los valores reales poseen una distancia promedio pequeña, lo que es indicador de una estimación asertiva por parte del modelo. En la figura 5.6 se corrobora lo indicado en el cuadro 5.2,

ya que visualmente los valores estimados se asemejan a la variabilidad de la información real, salvo por los picos que representan valores de incidencia sísmica alta. Por lo tanto la arquitectura LSTM es efectiva para el análisis de la tasa de ocurrencia sísmica temporal.

5.6. Resultados modelo ETAS espacio-temporal

Para una mayor precisión al momento de realizar las estimaciones se eligieron valores iniciales sugeridos por [Ogata \(1998\)](#), $\theta = (0.0015, 0.01, 0.01, 1, 1.3, 0.01, 2, 1)$, donde μ es calculado por $\frac{N}{(4T|S|)}$. Lo que se puede observar en la figura 5.7 corresponde a la función de intensidad de fondo $\lambda_{\beta,\theta}(t, x, y, m|H_t)$ y en cada una de las 4 imágenes corresponden a las ecuaciones antes mencionadas en el capítulo 2, como lo es la tasa de sismicidad de fondo $\tilde{u}(x, y)$, la tasa de sismicidad total espacial A , el coeficiente de agrupamiento $f_{D,\gamma,q}(x - x_i, y - y_i; m_i)$ y la función de intensidad pero en el final del periodo de estudio. Se nota que existe un alto coeficiente de agrupamiento en la zona central de Chile y en parte de Perú.

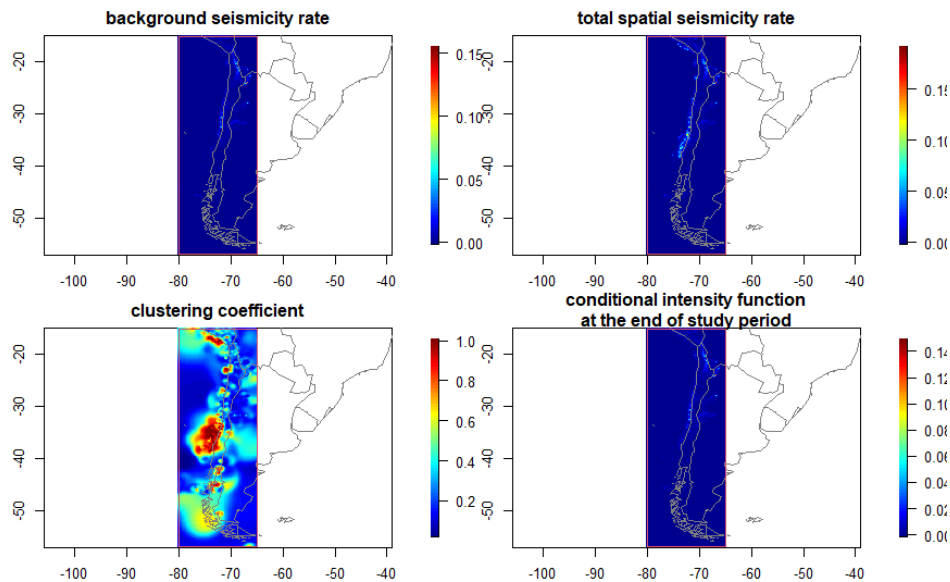


Figura 5.7: Función de intensidad de fondo. Rstudio “ETAS” [Jalilian \(2019\)](#). Fuente: Elaboración propia.

Lo que se observa en la figura 5.7 es que los lugares con mayor tasa de sismicidad de fondo se centran en el norte de Chile hasta aproximadamente la región del maule, lo mismo podemos decir de la tasa de sismicidad total, solo que en este caso se centra mas en la costa de Chile hasta el sur de Perú. Respecto al coeficiente de agrupamiento, presenta un gran agrupamiento de eventos de racimo en la zona central de Chile el cual se encuentra entre la región del maule y el BíoBío, además se presenta otro agrupamiento más pequeño entre el norte del Chile y el sur de Perú.

Como podemos notar estas estimaciones difieren ligeramente de las mostradas en la tabla 4.1 correspondientes a el modelo que es solo temporal: Estas diferencias se deben a que en el modelo espacial la región de estudio S es mas grande qe la región \hat{S}

	Est.	Error
β	2.832	0.0016
μ	0.948	0.0063
K	0.263	0.0158
c	0.035	0.0322
α	1.5632	0.0045
p	1.1241	0.0026
D	0.011	0.033
q	2.1187	0.0116
γ	0.952	0.0113

Cuadro 5.4: Estimación de los parámetros

5.7. Análisis residual

En la figura 5.8 tenemos los residuos temporales (arriba izquierda), los residuos espaciales suavizados (arriba a la derecha), los tiempos transformados τ_i contra i (abajo a la izquierda) y el gráfico Q-Q de U_i (abajo a la derecha).

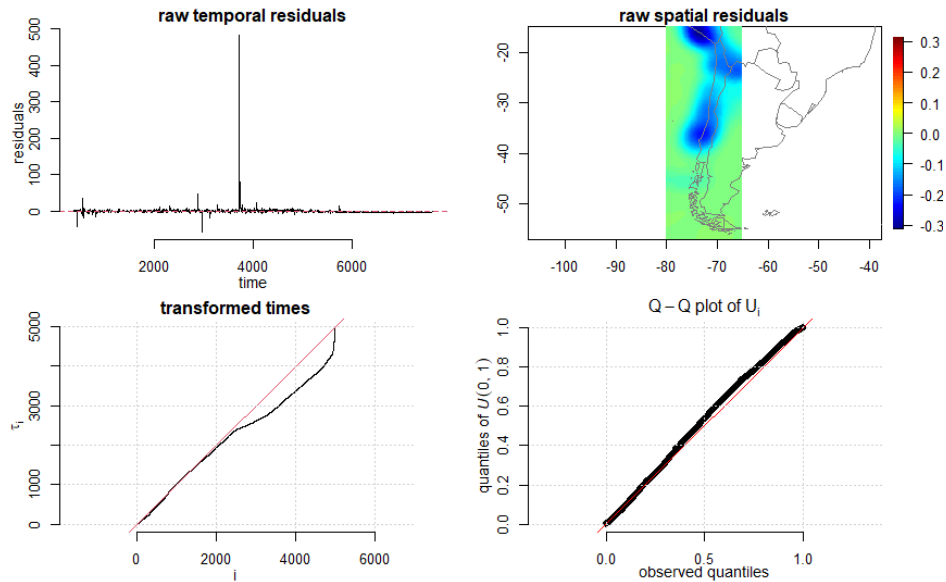


Figura 5.8: Residuos. Jalilian (2019)

5.8. Base de datos de imágenes

Se realiza un grillado sobre $\lambda_\theta(t|H_t)$ en dentro de la longitud -75, hasta -65; latitud -25, hasta -35; para realizar una estimación píxel a píxel. Por medio de esto se obtuvieron 7797 imágenes correspondiente a la función de intensidad espacial diaria desde 2000-01-01 hasta 2021-01-01. En la figura 5.9 tenemos la función de intensidad de 4 días distintos, en donde la ultima imagen corresponde al último día del periodo de estudio.

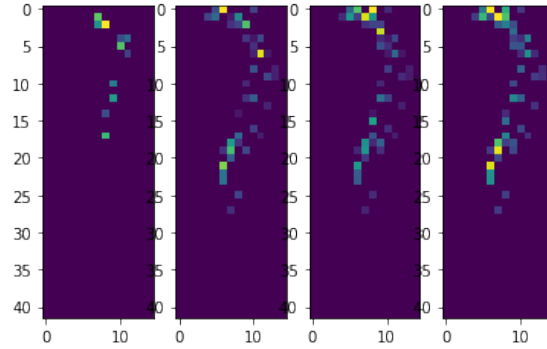


Figura 5.9: Función de intensidad de fondo diaria. Plaza (2019)

Esta base de datos nos indica la distribución espacial de la función lambda en cada t_i . En la figura 5.9 se observa que el enjambre sísmico dentro del periodo de estudio T comienza desde la zona norte de Chile y el sur de Perú, exactamente a partir del terremoto del 27 de febrero del 2010, es cuando en las observaciones comienzan a aparecer enjambres sísmicos en la zona central de las imágenes. Esta sucesión de imágenes se utiliza como base de datos para el modelo de red neuronal CNN, esto para predecir la probabilidad de que la intensidad máxima se produzca en un lugar determinado.

5.9. Resultados Red neuronal convolucional

Plaza (2019) propone subdividir la región de estudio S en 6 macrozonas para predecir la ubicación de la intensidad máxima por medio de la red CNN. La zona 1 corresponde al norte de Chile y algunas partes del sur de Perú y Bolivia, la zona 2 corresponde a la región de Atacama, la zona 3 a las regiones centrales de Chile, la zona 4 desde Rancagua a la región del Bio Bio, la zona 5 al sur, desde la Araucanía hasta la región de Los Lagos y la zona 6 a Aysén. (Ver cuadro 5.4 para más detalles de la latitud y longitud de cada zona).

Zona	Rango de latitud
1	(−44,00; −39,67)
2	(−39,67; −35,34)
3	(−35,34; −31,00)
4	(−31,00; −26,67)
5	(−26,67; −22,34)
6	(−22,33; −18,00)

Cuadro 5.5: Rangos de cada una de las Subdivisiones de S

Las métricas de desempeño de la red CNN mostradas en la figura 5.10 corresponden a un caso conocido como “underfit” lo que se refiere a una adaptación insuficiente del modelo hacia los datos, debido a que la red no puede tener un valor de error suficientemente bajo en el conjunto de entrenamiento Brownlee (2019). Esto puede notarse en las curvas de aprendizaje de la figura 5.10, ya que las curvas de entrenamiento y validación tanto en la precisión como la pérdida se encuentran alejadas entre sí. Este tipo de resultado nos da a concluir que el modelo no es capaz de aprender de los datos.

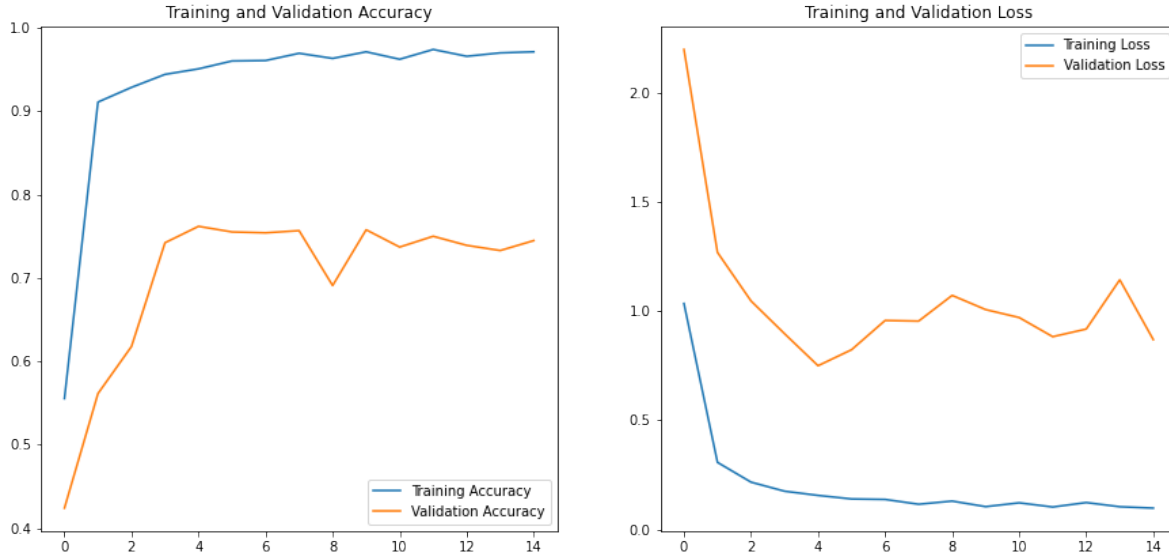


Figura 5.10: Métricas. Fuente: Elaboración propia

Uno de los motivos probables que genera un mal desempeño se deba al tamaño de la región de estudio espacial, ya que en ella se pueden ver mas de 1 enjambre sísmico en la región S. Una posible solución podría ser el análisis de la arquitectura CNN en una región de datos mas acotada a un solo enjambre.

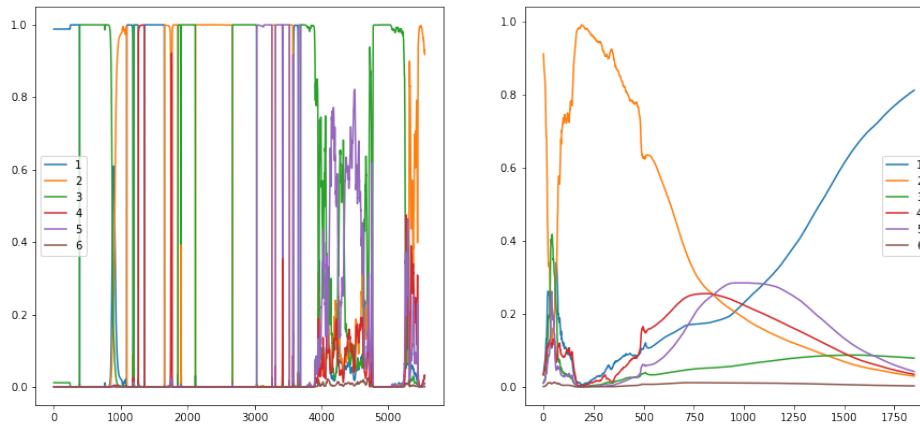


Figura 5.11: Resultados red CNN en las 6 macrozonas. Fuente : Elaboración propia

En la Figura 5.11, a la izquierda tenemos los valores predichos en el conjunto de entrenamiento y en la imagen a la derecha tenemos los valores predichos en el conjunto de prueba. Esta imagen muestra la probabilidad sísmica en las localizaciones de las distintas macrozonas descritas en el cuadro 5.5, esto para actividades sísmicas mayores a 4.5. Al igual que Plaza (2019) se consideraron de 30 a 210 días para predecir la ocurrencia de terremotos con una ventana de tiempo de 30 días. Como podemos notar en la imagen de la derecha (la predicción del conjunto de prueba) la probabilidad más alta para la siguiente semana posterior a la ventana de tiempo, la posee la zona 2 correspondiente a la Región de Atacama, y al final del periodo de tiempo la zona 1 posee la mayor probabilidad hacia el final del tiempo en el conjunto de prueba. El conjunto de prueba comienza

desde la fecha 14-10-2012 y finaliza en 07-05-2021. A pesar del mal desempeño de la red neuronal CNN (mostrado en la figura 5.10) los resultados mostrados no difieren de la realidad ya que, en 2014 hubo un terremoto en el norte de Chile el día 1 abril. A su vez, en junio de este año 2021 ocurrió un sismo de magnitud 8 en Perú, lo cual puede verse en la grafica izquierda de la figura 5.11 la cual muestra una alta probabilidad de incidencia sísmica en la zona 1 la que justamente corresponde al sur de Perú y el norte de Chile.

Capítulo 6

Conclusiones

Es sabido que las predicciones de eventos sísmicos constituyen un área el cual necesita un mayor desarrollo, el presente proyecto constituye un análisis sobre un conjunto de métodos de Deep learning, como lo es una arquitectura LSTM y CNN. Asimismo, este trabajo establece una línea base a partir de la cual se podría mejorar el modelo de Deep Learning, propuesto incorporando algunas variables exógenas como la profundidad sísmica, el desplazamiento de la corteza, la distancia a la falla principal y otras variables geológicas, especialmente para la predicción de choques principales donde normalmente los modelos ETAS fallan. Además este trabajo proporciona un preciso análisis diario espacio-temporal el cual puede ser utilizado para otros estudios en el área.

Otras mejoras podrían realizarse mejorando la resolución del mapa y la técnica de grillado que se utilizó, además de escoger una mejor área de estudio espacial, en donde se tenga un clúster de eventos sísmicos más acotado, a fin para mejorar las estimaciones de la red CNN, ya que por falta de recursos y tiempo no se pudieron mejorar. A pesar de esto los resultados vistos en el conjunto de prueba no difieren del todo con la incidencia sísmica de algunas zonas en específico de la región S.

Respecto a la arquitectura LSTM, esta es capaz predecir la intensidad máxima de los eventos con un R2 pequeño, podemos concluir que este es el mejor modelo para analizar enjambres sísmicos en el tiempo, ajustándose de buena forma a las estimaciones entregadas por el modelo ETAS.

Referencias

- The 150th anniversary of fusakichi omori. (2018). *Russian Academy of Sciences*.
- Brownlee, J. (2018). *Deep learning for time series forecasting*. Machine learning mastery.
- Brownlee, J. (2019, Aug). *How to use learning curves to diagnose machine learning model performance*. Descargado de <https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/>
- Calahorra, J. (2017). Procesos puntuales espacio-temporales con aplicación a la modelización de accidentes de tráfico. *Universitat Jaume*.
- Harte, D. (2010). Ptprocess: An r package for modelling marked point processes indexed by time. *Journal of Statistical Software*. doi: JournalofStatisticalSoftware
- Hernandez, N. (2012). Bayesian point process modelling of earthquake occurrences. *Ruprecht-Karls-Universitat Heidelberg*.
- 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). doi: 10.18637/jss.v088.c01
- Nicolis, O. (2017). Space-time forecasting of seismic events in chile. *Intechopen*.
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences.
- Ogata, Y. (2006). Space-time etas models and an improved extension. *Elsevier B.V.* doi: <https://doi.org/10.1016/j.tecto.2005.10.016>
- Peláez, J. (2011). Sobre las escalas de magnitud. *Universidad de Jaén*.
- Plaza, F. (2019). Prediction of intensity and location of seismic events using deep learning. *Elsevier B.V.* Descargado de <https://www.sciencedirect.com/science/article/abs/pii/S2211675320300361?via%3Dihub> doi: <https://doi.org/10.1016/j.spasta.2020.100442>