



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

## ESTIMACIÓN DEL AZIMUT EN EVENTOS SÍSMICOS UTILIZANDO DEEP LEARNING

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

GONZALO JAVIER MAUREL ESCOBEDO

PROFESOR GUÍA:  
NÉSTOR BECERRA YOMA

MIEMBROS DE LA COMISIÓN:  
ANDRÉS CABA RUTTE  
FERNANDO HUENUPÁN QUINÁN

Este trabajo ha sido financiado por Fondef bajo el proyecto ID20i10212

SANTIAGO DE CHILE  
2023

## ESTIMACIÓN DEL AZIMUT EN EVENTOS SÍSMICOS UTILIZANDO DEEP LEARNING

La tarea de estimación de localización de eventos sísmicos es una labor de gran relevancia, sobre todo en el contexto de los Sistemas de Alerta Temprana. Usualmente, para realizar dicha estimación se utiliza la información registrada a partir de varias estaciones sismológicas, lo cual conlleva una pérdida de tiempo valioso. Considerando ello, el presente trabajo adopta un enfoque diferente, consistente en estimar la dirección de origen de un evento sísmico a partir de una sola estación, donde dicha dirección está definida por el ángulo back-azimut.

Para realizar la tarea de estimación del back-azimut, se emplean técnicas de Deep Learning. En particular, se emplea una arquitectura convolucional (CNN). Para reforzar el desempeño de dicha arquitectura, se emplea un mecanismo de auto-atención (Multi-Head Attention, abreviado MHAT), al igual que se añade información respecto a cuál estación es la que registra la señal (ID station). Al conjunto de esta red se le denomina CNN-MHAT. El modelo planteado hace uso de la polarización de la onda P, y solo requiere de la señal en los canales Z, N y E en un período que comprende un segundo antes de la llegada de esta onda, y tres segundos posterior a tal instante.

Para entrenar y evaluar el modelo, se dispone de un total de 3160 avistamientos sísmicos. Dicha base de datos se divide en conjuntos de entrenamiento, validación y prueba siguiendo una proporción aproximada de 60 %, 20 % y 20 % respectivamente. Adicionalmente, la base de datos se subdivide en 2: solo eventos con magnitud mayor o igual a 4, y avistamientos referidos a microsismicidad (distancia epicentral menor a 100 km y magnitud menor o igual a 2,5). Con ello, se busca revisar el desempeño del modelo frente al problema de localizar eventos en los contextos de alerta temprana y microsismicidad. En todos los casos, para evaluar el desempeño del modelo se emplea la métrica de error absoluto medio (MAE).

Los resultados obtenidos sobre el conjunto de prueba de la base de datos completa dan cuenta que tanto el uso del mecanismo atencional como el añadir información de la estación suponen una mejora en el rendimiento de la red. En efecto, sobre el conjunto de prueba de la base de datos completa la red CNN-MHAT reporta un valor MAE de  $10,06^\circ$ , mientras que las redes CNN-MHAT sin ID y CNN reportan valores MAE de  $11,44^\circ$  y  $10,91^\circ$  respectivamente. Adicionalmente, se compara el desempeño de la red con un método clásico de estimación de back-azimut, el cual arroja un error absoluto medio de  $23,85^\circ$ . Finalmente, al emplear la base de datos con eventos cuya magnitud es mayor o igual a 4, el error absoluto medio obtenido al evaluar la red CNN-MHAT sobre el conjunto de prueba es igual a  $8,51^\circ$ . Por otra parte, al utilizar la base de datos de microsismicidad, el modelo en cuestión reporta un error sobre el conjunto de prueba igual a  $15,42^\circ$ . A partir de lo anterior, se concluye que el problema enfrentado es dependiente del SNR de las señales.

*A mi familia, y en especial, a mi abuelita Eliana,  
por enseñarme el valor de la perseverancia.*

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes generales . . . . .	1
1.2. Identificación y formulación del problema . . . . .	2
1.3. Objetivos del trabajo de título . . . . .	2
1.4. Organización de la memoria . . . . .	2
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Deep Learning . . . . .	4
2.1.1. Perceptrón Multicapa . . . . .	4
2.1.2. Redes Convolucionales (1D) . . . . .	5
2.1.3. Redes LSTM . . . . .	6
2.1.4. Redes Transformer . . . . .	8
2.2. Eventos sísmicos . . . . .	11
2.3. Dirección de origen . . . . .	12
2.4. Estimación del back-azimut . . . . .	13
2.4.1. Métodos clásicos de estimación del back-azimut . . . . .	13
2.4.1.1. Estimación mediante la matriz de covarianza . . . . .	13
2.4.1.2. Estimación mediante Principal Component Analysis . . . . .	15
2.4.1.3. Estimación mediante correlaciones cruzadas . . . . .	15
2.4.1.4. Estimación mediante una sola muestra . . . . .	15
2.4.2. Métodos basados en Machine Learning . . . . .	17
2.4.3. Métodos basados en Deep Learning . . . . .	18
2.4.4. Consideraciones sobre los métodos de estimación de back-azimut revisados . . . . .	20
<b>3. Metodología</b>	<b>21</b>
3.1. Base de datos . . . . .	22
3.1.1. Estimación de SNR . . . . .	24
3.1.2. Base de datos de eventos con $M \geq 4.0$ . . . . .	25
3.1.3. Base de datos de micro-sismicidad local . . . . .	26
3.2. Pre-procesamiento de las señales . . . . .	28
3.3. Arquitecturas de red a evaluar . . . . .	29
3.3.1. Red Convolutiva . . . . .	30
3.3.2. Red Convolutiva con mecanismo atencional . . . . .	31
3.4. Entrenamiento de los modelos . . . . .	32
3.5. Evaluación de los modelos . . . . .	32
3.5.1. Reconstrucción del ángulo . . . . .	32

3.5.2.	Promedio circular . . . . .	33
3.5.3.	Métrica de evaluación a utilizar . . . . .	33
<b>4.</b>	<b>Resultados</b>	<b>35</b>
4.1.	Mejor configuración . . . . .	35
4.2.	Resultados en base de datos completa . . . . .	36
4.3.	Resultados en base de datos de microsismicidad . . . . .	40
4.4.	Resultados en base de datos de eventos con $M \geq 4$ . . . . .	42
4.5.	Análisis de resultados . . . . .	44
4.5.1.	Comparación de modelos de red en base de datos completa . . . . .	44
4.5.2.	Desempeño de red CNN-MHAT . . . . .	44
4.5.2.1.	Base de datos completa . . . . .	44
4.5.2.2.	Base de datos de microsismicidad . . . . .	45
4.5.2.3.	Base de datos $M \geq 4$ . . . . .	45
4.6.	Fortalezas y debilidades del modelo CNN-MHAT . . . . .	46
<b>5.</b>	<b>Conclusión</b>	<b>48</b>
5.1.	Trabajo futuro . . . . .	48
	<b>Bibliografía</b>	<b>50</b>

# Índice de Tablas

3.1.	Resumen de la cantidad de avistamientos de la base de datos descargada. . . .	22
3.2.	Resumen de la cantidad de avistamientos de la base de datos de eventos con $M \geq 4.0$ considerando las estaciones de referencia y secundarias. . . . .	25
3.3.	Resumen de la cantidad de avistamientos de la base de datos de microsismicidad considerando las estaciones de referencia y secundarias. . . . .	27
3.4.	Distribución de avistamientos de las bases de datos. . . . .	28
4.1.	Mejor conjunto de hiperparámetros para red CNN. . . . .	35
4.2.	Mejor conjunto de hiperparámetros para red CNN con MultiHead self-attention. . . . .	36
4.3.	Error absoluto medio obtenido al evaluar los modelos sobre los conjuntos de la base de datos completa. . . . .	36

# Índice de Ilustraciones

2.1.	Modelo no lineal de una neurona artificial. . . . .	4
2.2.	Ejemplo de perceptrón multicapa. . . . .	5
2.3.	Ejemplo de convolución en una dimensión. . . . .	6
2.4.	Esquema de celda de una red LSTM. Fuente: imagen adaptada de [5] (permiso bajo licencia CC-BY 4.0). . . . .	7
2.5.	Esquema de Multi-Head Attention planteado en [7]. . . . .	9
2.6.	Arquitectura Transformer considerando encoder (izquierda) y decoder (derecha). Fuente: imagen original de [9] (permiso bajo licencia CC-BY 4.0). . . . .	10
2.7.	Tipos de ondas sísmicas. Fuente: original de [14] (permiso bajo licencia CC-BY 3.0 CL). . . . .	11
2.8.	Ejemplo de ondas sísmicas registradas en un sismograma. Fuente: original de [15].	12
2.9.	Ejemplo de back-azimut. . . . .	12
2.10.	Distribución geográfica de los terremotos de la base de datos STEAD. Fuente: original de [33] (permiso bajo licencia CC-BY 4.0). . . . .	19
3.1.	Diagrama general de la metodología seguida. . . . .	21
3.2.	Mapa de epicentros y magnitudes de los eventos de la base de datos completa.	23
3.3.	Histograma de back-azimut y distancia epicentral de la base de datos completa.	24
3.4.	SNR promedio en los 3 canales. . . . .	24
3.5.	Histogramas de back-azimut y distancia para BD de eventos con $M \geq 4.0$ . . .	25
3.6.	Mapa de epicentros y magnitudes de los eventos de la base de datos $M \geq 4$ . .	26
3.7.	Histogramas de back-azimut y distancia al epicentro de la base de datos de microsismicidad. . . . .	27
3.8.	Mapa de epicentros y magnitudes de los eventos para la base de datos de microsismicidad local. . . . .	28
3.9.	Red convolucional con ID station . . . . .	30
3.10.	Bloque fully connected empleado al final de cada arquitectura. . . . .	30
3.11.	Bloques convolucionales empleados en la arquitectura. . . . .	31
3.12.	Red convolucional con mecanismo de auto-atención y ID station. . . . .	31
4.1.	Estimación vs valor real sobre el conjunto de validación de la base de datos completa, red CNN-MHAT. . . . .	37
4.2.	Histograma de errores sobre el conjunto de validación de la base de datos completa, red CNN-MHAT. . . . .	37
4.3.	Estimación vs valor real sobre el conjunto de prueba de la base de datos completa, red CNN-MHAT. . . . .	38
4.4.	Histograma de errores sobre el conjunto de prueba de la base de datos completa, red CNN-MHAT. . . . .	38
4.5.	Estimación vs valor real sobre el conjunto de validación de la base de datos completa, método de correlaciones cruzadas [23]. . . . .	39

4.6.	Estimación vs valor real sobre el conjunto de prueba de la base de datos completa, método de correlaciones cruzadas [23]. . . . .	39
4.7.	Estimación vs valor real sobre el conjunto de validación de la base de datos de microsismicidad, red CNN-MHAT. . . . .	40
4.8.	Histograma de errores sobre el conjunto de validación de la base de datos de microsismicidad, red CNN-MHAT. . . . .	40
4.9.	Estimación vs valor real sobre el conjunto de prueba de la base de datos de microsismicidad, red CNN-MHAT. . . . .	41
4.10.	Histograma de errores sobre el conjunto de prueba de la base de datos de microsismicidad, red CNN-MHAT. . . . .	41
4.11.	Estimación vs valor real sobre el conjunto de validación de la base de datos de eventos $M \geq 4$ , red CNN-MHAT. . . . .	42
4.12.	Histograma de errores sobre el conjunto de validación de la base de datos de eventos $M \geq 4$ , red CNN-MHAT. . . . .	42
4.13.	Estimación vs valor real sobre el conjunto de prueba de la base de datos de eventos $M \geq 4$ , red CNN-MHAT. . . . .	43
4.14.	Histograma de errores sobre el conjunto de prueba de la base de datos de eventos $M \geq 4$ , red CNN-MHAT. . . . .	43



# Capítulo 1

## Introducción

### 1.1. Antecedentes generales

El cinturón de fuego del Pacífico es una región que puede ser caracterizada como sismo-génica, lo cual se debe fundamentalmente a la interacción entre diversas placas tectónicas a través de una serie de procesos de convergencia y divergencia. Son diversos los países que están presentes en esta región, como por ejemplo sucede con Estados Unidos, Canadá, México, Japón, Taiwan, Malasia, Nueva Zelanda, Ecuador, Perú y Chile.

En el caso de Chile, el territorio continental del país se encuentra ubicado sobre la placa sudamericana, en un área de convergencia entre las placas de Nazca y Antártica, las cuales subducen por debajo. Por tal motivo, prácticamente la totalidad del territorio continental chileno es susceptible a la ocurrencia de algún evento sísmico. Un ejemplo de la alta sismicidad presente en el país, es el hecho de observar que solo durante el año 2021, el Centro Sismológico Nacional -organismo sismológico chileno encargado de registrar y entregar la información referida a eventos sísmicos a las oficinas de emergencia tales como la Oficina Nacional de Emergencias del Ministerio del Interior y el Servicio Hidrográfico y Oceanográfico de la Armada de Chile- localizó 7436 sismos con magnitudes entre 1,9 y 7,1.

También resulta relevante el considerar los eventos de gran magnitud que se han registrado en territorio chileno en los últimos 30 años. Algunos de dichos eventos no solo provocaron daños de manera directa a la infraestructura cercana, sino que también tuvieron como consecuencia la generación de un tsunami en las costas chilenas, lo cual da cuenta de la existencia de un riesgo adicional a los eventos sísmicos. Tal riesgo, debido al impacto que puede tener sobre la vida de una gran cantidad de personas, debe ser mitigado de manera eficaz. Para ello, cobra particular relevancia el que los organismos de emergencia puedan identificar de manera correcta y rápida si un evento sísmico tiene la capacidad de generar un tsunami, lo cual requiere de una estimación acertada respecto a la ubicación del sismo.

La tarea de estimar la localización de un evento sísmico puede ser abordada haciendo una distinción entre algoritmos clásicos y métodos que se han desarrollado en el último tiempo basados en técnicas pertenecientes al campo del Aprendizaje Profundo, las cuales involucran el uso de diversos tipos de redes neuronales artificiales.

## 1.2. Identificación y formulación del problema

Si bien todo el territorio continental chileno es susceptible a ser afectado por eventos sísmicos, existen zonas que presentan un nivel de riesgo mayor. Esto se debe a la existencia de áreas denominadas como lagunas sísmicas. Una laguna sísmica es un área dentro de una falla sísmica donde no han ocurrido eventos sísmicos de relevancia en el último tiempo pese a la sismicidad esperada, lo cual implica la existencia de un potencial sísmico importante. Algunas de las lagunas sísmicas de mayor relevancia están presentes en la zona norte de Chile, existiendo también el riesgo de generación de un tsunami.

Para saber si un terremoto tiene el potencial de producir un tsunami, es necesario no solo conocer la magnitud de este, sino también su localización, siendo una condición necesaria el que el sismo tenga área de ruptura mar adentro. Un problema a la hora de estimar la localización de un sismo, radica en que los algoritmos usados habitualmente emplean información proveniente de varias estaciones sismológicas, lo cual requiere de una cantidad valiosa de tiempo. Dicho tiempo, resulta vital en el caso de que sea necesario realizar una evacuación del borde costero para salvaguardar las vidas de las personas.

Una estrategia para ahorrar tiempo, consiste en el realizar una estimación en base a la información proveniente de alguna estación cercana al evento. Dicha estimación puede ser dividida entre estimar el ángulo que determina la dirección del epicentro del evento respecto a la estación sísmica y estimar la distancia entre el evento y la estación. En el caso particular de este trabajo, el problema abordado corresponde a estimar de manera rápida y eficaz el back-azimut -ángulo que define la dirección entre el epicentro del sismo y la estación- en eventos sísmicos empleando datos de solo una estación sismológica. Para abordar tal problema, se plantea el uso de técnicas de aprendizaje profundo. Debido al área de riesgo sísmico mencionada con anterioridad, es que los datos a utilizar corresponden a registros obtenidos por 49 estaciones sismológicas ubicadas en la zona centro-norte de Chile.

## 1.3. Objetivos del trabajo de título

En base a la problemática observada, se plantea como objetivo principal el estimar de manera adecuada el back-azimut en eventos sísmicos empleando técnicas de Deep Learning, considerando la información registrada desde una sola estación sismológica.

Frente a ello, se plantean los siguientes objetivos secundarios:

1. Generar una base de datos de registros sísmicos para el problema a enfrentar.
2. Desarrollar una arquitectura de red que permita resolver el problema planteado.
3. Evaluar el desempeño de la arquitectura de red.

## 1.4. Organización de la memoria

El presente informe se encuentra organizado en 5 capítulos, siendo el primero de ellos la introducción al trabajo realizado. En el capítulo 2 se presenta el marco teórico y el estado del arte, donde se da cuenta de diversos métodos que han sido planteados en el pasado para

estimar el back-azimut y/o la localización de eventos sísmicos a partir de una sola estación. Adicionalmente, se realiza una revisión de las arquitecturas de Deep Learning más conocidas y útiles para el problema a resolver. En el capítulo 3 se presenta la metodología utilizada para llevar a cabo el trabajo. En el capítulo 4 se presentan los resultados obtenidos y el análisis de estos, respectivamente. Finalmente, en el capítulo 5 se presentan las principales conclusiones derivadas del trabajo realizado, al igual que se da cuenta del trabajo a futuro por realizar.

# Capítulo 2

## Marco Teórico

### 2.1. Deep Learning

El *Deep Learning* es un subconjunto del *Machine Learning* en el cual, por medio del uso de diversas capas de redes neuronales artificiales, es posible aprender representaciones de los datos considerando múltiples niveles de abstracción [1]. En este tipo de modelos, el aprendizaje se realiza adaptando los parámetros de la red de manera de minimizar cierta función objetivo. La adaptación de dichos parámetros, los cuales se denominan pesos, se realiza a través de un algoritmo denominado *backpropagation*, a través del cual se computa el gradiente de la función objetivo respecto a los pesos de la red. A continuación se presentan algunos de los tipos de redes neuronales más conocidos.

#### 2.1.1. Perceptrón Multicapa

Para comprender lo que es un perceptrón multicapa, resulta necesario antes el caracterizar la unidad básica de procesamiento de este tipo de redes: la neurona. En la Figura 2.1 se presenta lo que es el modelo de una neurona artificial.

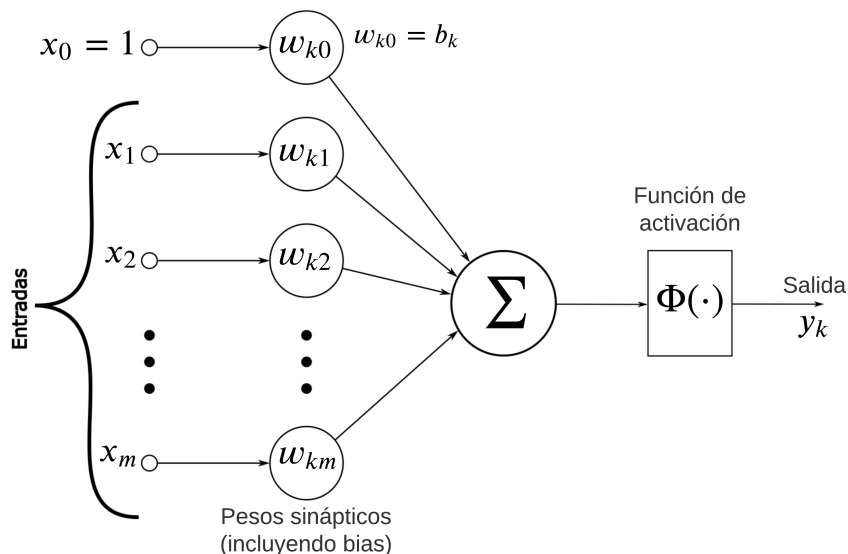


Figura 2.1: Modelo no lineal de una neurona artificial.

A partir de la Figura 2.1 es posible observar que la neurona artificial está compuesta, en general, por tres elementos básicos: sus pesos sinápticos, un combinador lineal y una función de activación. En efecto, si para la neurona  $k$ -ésima se consideran  $m$  entradas, denotadas como  $[x_1, \dots, x_m]$ , se tiene entonces igual número de pesos sinápticos, denotados como  $[w_{k1}, \dots, w_{km}]$ . Dichos pesos permiten ponderar las entradas con el fin de realizar una combinación lineal de ellas. Por otra parte, el *bias* es un valor que permite desplazar la combinación lineal. Este valor usualmente se escribe con notación de peso sináptico  $w_{k0}$ . Finalmente, si se considera cierta función de activación  $\Phi$ , entonces la salida de la neurona  $k$ -ésima viene dada por la Ecuación 2.1.

$$y_k = \Phi\left(\sum_{i=0}^m x_i \cdot w_{ki}\right) \quad (2.1)$$

Luego, un perceptrón multicapa es un tipo de red consistente en diversas capas de neuronas artificiales, las cuales se conectan en una sola dirección, desde la entrada hacia la salida, tal como se aprecia en la Figura 2.2, donde se presenta un perceptrón multicapa con 2 capas ocultas<sup>1</sup>.

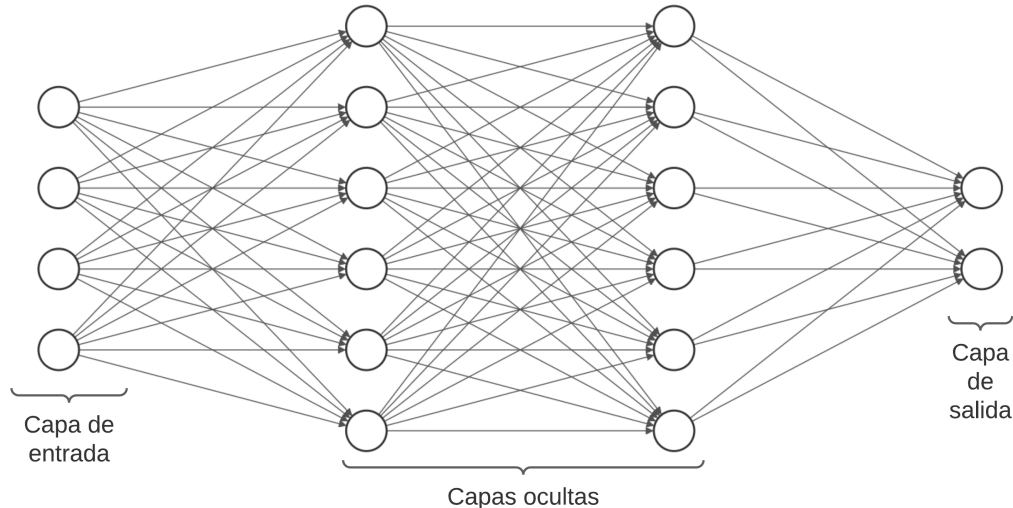


Figura 2.2: Ejemplo de perceptrón multicapa.

Una ventaja del perceptrón multicapa consiste en que, basta solamente una capa oculta y una función de activación adecuada para que este tipo de modelo se convierta en un aproximador universal [2]. Esto se refiere a que puede aproximar cualquier tipo de función real continua siempre y cuando se cuente con una cantidad suficiente de unidades neuronales.

### 2.1.2. Redes Convolucionales (1D)

Las redes convolucionales tienen su fuente de inspiración en el funcionamiento del sistema visual de los vertebrados. Este tipo de redes trabajan convolucionando las entradas con ciertos filtros, siendo los parámetros de estos últimos los que se aprenden con el entrenamiento. En efecto, el operador de convolución discreta en una dimensión entre una señal  $S$  y cierto kernel

<sup>1</sup> Se denomina capa oculta a aquellas que no están conectadas directamente a la entrada o a la salida de la red.

F de tamaño  $2N+1$  viene dado por la Ecuación 2.2.

$$(S * F)(k) = \sum_{i=-N/2}^{N/2} F(i) \cdot S(k - i) \quad (2.2)$$

Luego, la convolución entre la señal de entrada y los diversos filtros se realiza desplazando estos sobre la señal, tal como se puede observar en la Figura 2.3, donde se denomina “kernel\_size” al largo de la ventana de convolución. Una característica a destacar en el caso de las redes convolucionales 1D es el hecho de que el desplazamiento del filtro se realiza en solo una dirección. Por ejemplo, cuando se trabaja con imágenes se suele utilizar la convolución en dos dimensiones (kernel se desplaza a lo largo y a lo ancho), o incluso en tres dimensiones (kernel se desplaza a lo largo, ancho y en profundidad).

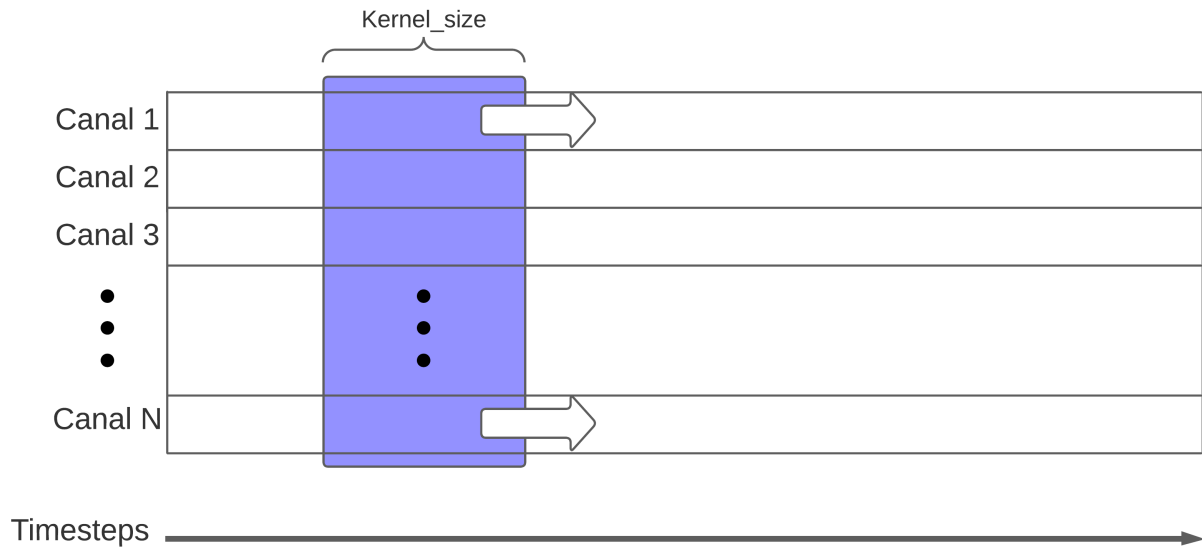


Figura 2.3: Ejemplo de convolución en una dimensión.

Una vez realizada la convolución, se aplica la función de activación. Cabe mencionar que resulta usual el uso de capas pooling de manera posterior a las capas convolucionales, pues permiten reducir la dimensionalidad de los datos (sub-sampling). En este aspecto, una de las técnicas más usadas es la de max-pooling (1D), donde la representación de entrada se va reduciendo de tamaño al tomar el máximo valor para cada canal considerando una ventana espacial de cierto tamaño que se va desplazando.

De esta manera, las redes convolucionales presentan la gran ventaja de que pueden ser empleadas para extracción de características, razón por la cual han cobrado gran relevancia en diversos campos de investigación. Además, se ha probado que las redes convolucionales profundas son también aproximadores universales [3].

### 2.1.3. Redes LSTM

Las Redes Neuronales Recurrentes son redes que se destacan por el uso de conexiones recurrentes dentro de su estructura, lo cual les permite trabajar de mejor manera con datos

secuenciales como puede suceder con señales o en tareas de procesamiento de lenguaje natural. Un problema del cual adolecían las primeras redes recurrentes era que resultaba difícil que estas aprendieran dependencias de largo plazo en secuencias de datos, ello debido a que el error de retropropagación tendía a desvanecerse o explotar durante el entrenamiento. Frente a tal problema es que en [4] se presentan las redes Long Short-Term Memory, abreviado LSTM.

Este tipo de red se destaca porque puede aprender dependencias de largo plazo, lo cual las vuelve aptas para trabajar con secuencias de tamaño considerable. Para lograr esto, es que las celdas de este tipo de red están compuestas por tres compuertas: la compuerta de olvido, la compuerta de entrada y la compuerta de salida. En la Figura 2.4 se presenta la estructura de la celda en cuestión.

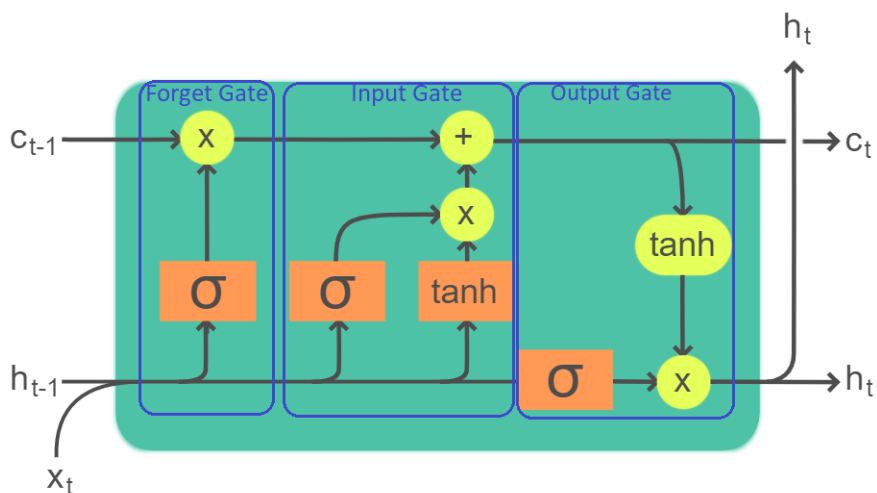


Figura 2.4: Esquema de celda de una red LSTM. Fuente: imagen adaptada de [5] (permiso bajo licencia CC-BY 4.0).

En la Figura 2.4 los elementos que presentan bordes rectangulares corresponden a capas neuronales feed-forward<sup>2</sup>, mientras que aquellos con bordes redondeados corresponden a operaciones elemento a elemento. Adicionalmente,  $C_{t-1}$  y  $h_{t-1}$  corresponden al estado de celda y el estado oculto respectivamente. Para comprender a que refiere cada uno, es posible interpretar que mientras el estado de la celda engloba la información global capturada hasta ahora (memoria de largo plazo), el estado oculto se refiere principalmente a una codificación de la información del último paso de tiempo (memoria de trabajo). A la salida de la celda se obtiene  $C_t$  y  $h_t$  los cuales son el estado de la celda y el estado oculto, actualizados.

En la celda de una red LSTM, cada compuerta juega un rol relevante. La compuerta de olvido, compuesta por una capa de red feed-forward con función de activación sigmoide ( $\sigma(z) = \frac{1}{1+e^{-z}}$ ), permite determinar qué información debe conservarse en el estado de la celda a partir de la información contenida en el estado oculto y la entrada actual.

La compuerta de entrada, compuesta por dos capas de red neuronal feed-forward, activa-

<sup>2</sup> El término feed-forward se refiere a que la red solo presenta conexiones en una sola dirección, desde la entrada a la salida.

das por funciones sigmoide y tangente hiperbólica ( $\tanh(z) = \frac{2}{1+e^{-2z}} - 1$ ) respectivamente, permite determinar qué información proveniente del estado oculto y la entrada actual debiesen ser añadidas al estado de la celda (memoria de largo plazo).

Finalmente, la compuerta de salida, compuesta por una capa de red neuronal feed-forward que recibe tanto el estado oculto como la entrada actual, permite actualizar el estado oculto, considerando solo la información necesaria y no toda la información que se ha guardado a partir de la memoria de largo plazo.

De esta manera, las redes LSTM son capaces de retener la información valiosa a partir de las diversas secuencias de datos ingresadas, lo cual les permite aprender dependencias de largo plazo. Es de destacar que las Redes Neuronales Recurrentes también son aproximadores universales [6].

### 2.1.4. Redes Transformer

La arquitectura transformer, presentada por primera vez en [7], destaca porque puede trabajar con datos secuenciales sin emplear mecanismos recurrentes o convolucionales. Para lograr esto, es que hace uso de un tipo de mecanismo atencional denominado auto-atención. Existen diversos tipos de atención, sin embargo, la idea central reside en que el modelo pueda enfocarse en los aspectos más relevantes de la información ingresada.

De manera general, en [7] se define a una función de atención como un “mapeo entre una query y un conjunto de pares key-value<sup>3</sup> respecto a una salida”, siendo todos ellos vectores. De esta forma, la salida se calcula como una suma ponderada de los valores, donde a su vez los ponderadores se calculan a través de cierta “función de compatibilidad entre la query y la key respectiva”.

Considerando lo anterior, se introduce en la Ecuación 2.3 la función de atención denominada *scaled dot product*. Dicha función, planteada por primera vez en [7], recibe queries y keys de tamaño  $d_k$  y valores de tamaño  $d_v$ , los cuales se agrupan en matrices Q, K y V respectivamente. La función de similaridad corresponde a la aplicación de la función Softmax<sup>4</sup> sobre el producto de las matrices Q y  $K^T$  (notar que entre más similares son 2 vectores, mayor es el valor de su producto punto). Dicho producto es escalado por un factor de  $1/\sqrt{d_k}$  para que los valores no sean demasiado grandes.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.3)$$

En el caso de la auto-atención empleada en [7], las queries, keys y values se calculan directamente a partir de la secuencia de entrada<sup>5</sup>. Para realizar esto, es que se emplean ciertas matrices de pesos entrenables, abreviadas  $W^Q$ ,  $W^K$  y  $W^V$  respectivamente. En efecto, si se considera cierta secuencia de entrada X, se calculan las matrices Q, K y V según las

<sup>3</sup> Se habla de query, key y value como una abstracción que hace analogía al funcionamiento de los sistemas de recuperación de información.

<sup>4</sup> Sea un vector  $z \in \mathbb{R}^N$ , entonces la función softmax está dada por  $\sigma(z)_j = \frac{e^{z_j}}{\sum_{t=1}^N e^{z_t}}$  con  $j \in \{1, \dots, N\}$ .

<sup>5</sup> Se denomina auto-atención debido a que los elementos de la entrada se comparan solamente entre sí.



Ecuaciones 2.4-2.6.

$$Q = XW^Q \quad (2.4)$$

$$K = XW^K \quad (2.5)$$

$$V = XW^V \quad (2.6)$$

Adicionalmente, en [7] se introduce un nuevo concepto, denominado como *Multi-Head Attention*, el cual apunta a calcular de manera paralela diversas instancias de la auto-atención descrita anteriormente. Para que no escale demasiado el costo computacional, se suele reducir la dimensionalidad de los vectores query, key y value (mediante la reducción del tamaño de las matrices entrenables  $W^Q$ ,  $W^K$  y  $W^V$ ) de manera que  $d_k = d_v = d_{model}/h$ , siendo  $d_{model}$  la dimensión del embedding asociado a cada elemento de la secuencia de entrada (esto es, la cantidad de features asociada a cada elemento de dicha secuencia) y  $h$  el número de cabezas considerado ( $h$  determina el número de instancias paralelas en que se calcula la atención).

Para cada cabeza de atención se calcula la instancia atencional según la Ecuación 2.3, con la diferencia de que las matrices Q, K y V se obtienen empleando matrices de pesos  $W^Q$ ,  $W^K$  y  $W^V$  diferentes según la cabeza en cuestión, tal como se presenta en la Ecuación 2.7.

$$head_i = Attention(XW_i^Q, XW_i^K, XW_i^V) \quad (2.7)$$

Una vez aplicada la función atencional para cada cabeza, los resultados se concatenan y se vuelven a proyectar linealmente. Así, la salida del mecanismo atencional *Multi-head Attention* viene dada por la Ecuación 2.8.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^o \quad (2.8)$$

Donde  $W^o \in \mathbb{R}^{hd_v \times d_{model}}$  es una matriz de pesos entrenable que multiplica al resultado de la concatenación de cada cabeza del mecanismo. En la Figura 2.5 se presenta el esquema de atención Multi-Head Attention indicado.

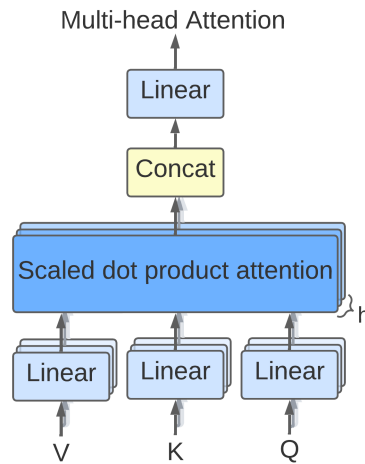


Figura 2.5: Esquema de Multi-Head Attention planteado en [7].

En la Figura 2.6 se presenta la arquitectura transformer utilizada en [7], la cual está formada por dos bloques, los cuales se denominan *encoder* y *decoder*. En ambos, se destaca que solo se emplean mecanismos atencionales, conexiones residuales, capas feed-forward y capas de normalización. Las capas de normalización se denominan *Layer Normalization*. Este mecanismo fue introducido por primera vez en [8], y permite normalizar las activaciones de la anterior capa, donde las estadísticas usadas para la normalización se obtienen de las características asociadas a cada muestra, de manera independiente, en un batch <sup>6</sup>.

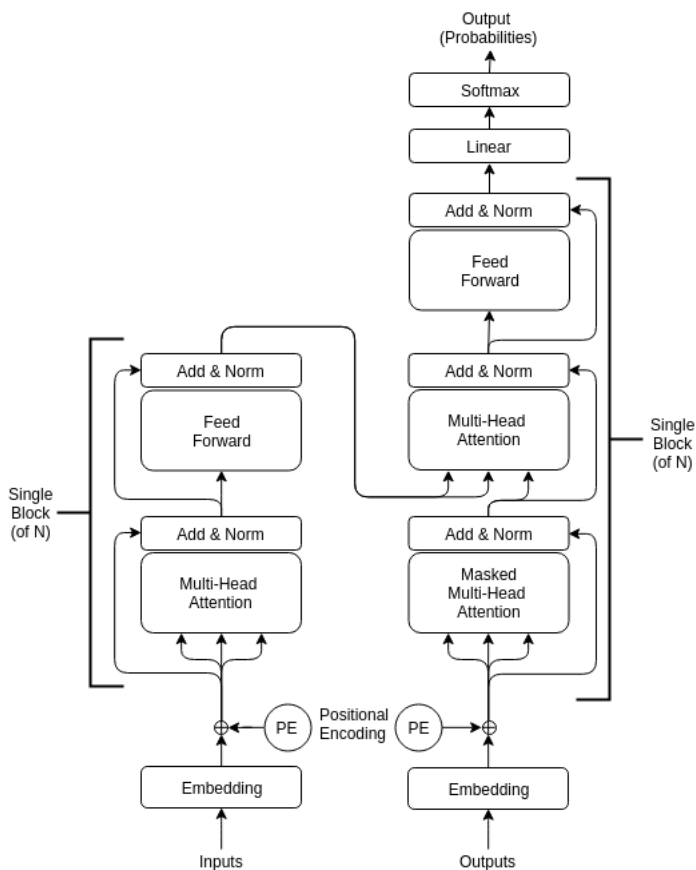


Figura 2.6: Arquitectura Transformer considerando encoder (izquierda) y decoder (derecha). Fuente: imagen original de [9] (permiso bajo licencia CC-BY 4.0).

Un aspecto sumamente relevante para el funcionamiento de la arquitectura transformer es el uso de positional encoding. Esto se debe a que al no emplearse elementos recurrentes, se pierde la información asociada al orden de cada secuencia. El positional encoding permite asignarle a cada elemento en la secuencia una representación única según su posición. En el caso de [7] se emplean las Ecuaciones 2.9 y 2.10 para realizar positional encoding, donde  $pos$  es la posición e  $i$  es la dimensión.

$$PE(pos, 2i) = \sin(pos/10000^{2i/d_{model}}) \quad (2.9)$$

$$PE(pos, 2i + 1) = \cos(pos/10000^{2i/d_{model}}) \quad (2.10)$$

<sup>6</sup> Se denomina batch o mini-batch a un grupo de muestras que son procesadas al mismo tiempo por la red durante su entrenamiento.

La arquitectura transformer presenta ventajas en comparación a las redes recurrentes, incluyendo a las de tipo LSTM. Una de ellas corresponde a la mayor rapidez en el entrenamiento que pueden alcanzar debido a su mayor grado de paralelización. Además, por sí solas pueden alcanzar altos niveles de desempeño en diversas tareas abordadas habitualmente mediante redes LSTM.

Se destaca que la arquitectura transformer, o variantes de ella, han sido utilizadas en diversas tareas que involucran el manejo de series de tiempo, como por ejemplo *forecasting*, detección de anomalías y labores de clasificación [10].

Adicionalmente, el corazón de la arquitectura transformer, esto es, el mecanismo Multi-Head Attention, ha sido utilizado de manera exitosa para reforzar y combinar la información proveniente de capas convolucionales [11, 12], esto siguiendo una estrategia de auto-atención.

## 2.2. Eventos sísmicos

Un sismo puede ser definido como un proceso de liberación de energía cuyo origen se remonta usualmente a procesos de interacción entre placas tectónicas, actividad de fallas geológicas o volcanismo [13]. Este proceso se manifiesta en forma de ondas sísmicas, las cuales pueden ser categorizadas en ondas de cuerpo y ondas superficiales.

Las ondas de cuerpo son aquellas que se desplazan por el interior de la Tierra. Dentro de esta categoría, se distinguen dos tipos de ondas, las ondas P y las ondas S. Las ondas P o Primarias se denominan así debido a que al poseer una velocidad de propagación mayor son las primeras en ser detectadas. Este tipo de ondas son de carácter longitudinal, por lo cual su dirección de propagación es similar a la dirección de oscilación. Por otra parte, las ondas S o Secundarias son de tipo transversal, lo cual se refiere a que la dirección de propagación es perpendicular a la dirección de oscilación.

Por otro lado, las ondas superficiales son aquellas que viajan a través de la superficie terrestre. Dentro de esta categoría se encuentran las ondas de Rayleigh y las ondas de Love. En la Figura 2.7 se presenta un ejemplo de los tipos de ondas sísmicas.

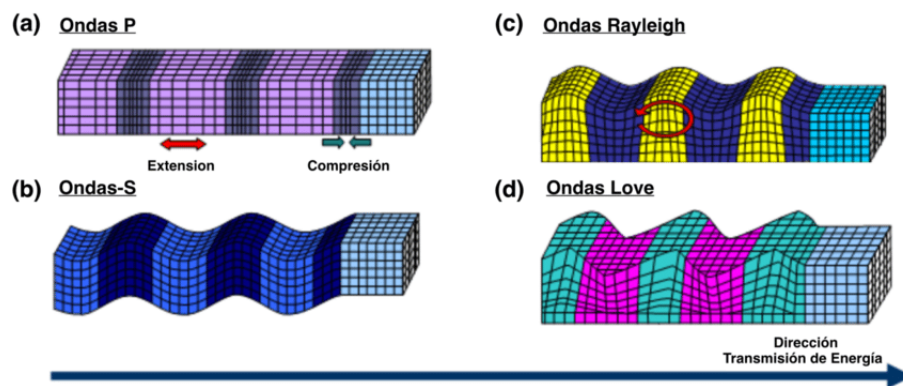


Figura 2.7: Tipos de ondas sísmicas. Fuente: original de [14] (permiso bajo licencia CC-BY 3.0 CL).

Para detectar y registrar las ondas sísmicas se utiliza un instrumento denominado sismó-

grafo. Dicho instrumento registra la señal en 3 canales o ejes: vertical (Z), norte-sur (N) y este-oeste (E). En general, la señal registrada proveniente de un sismo cuyo origen está relacionado a la interacción de placas tectónicas o fallas geológicas, tiende a tener cierta forma dada por la llegada de los diversos tipos de onda. En la Figura 2.8 se presenta un ejemplo de la forma adquirida por la señal registrada por un sismógrafo en presencia de un evento sísmico.

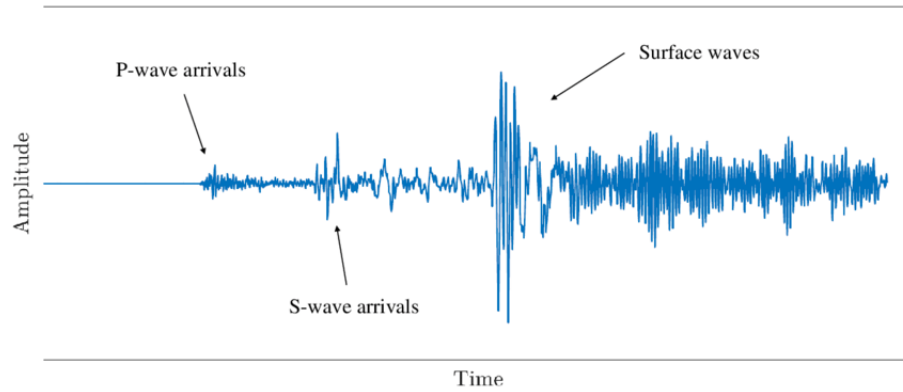


Figura 2.8: Ejemplo de ondas sísmicas registradas en un sismograma. Fuente: original de [15].

### 2.3. Dirección de origen

La dirección asociada a la llegada de ondas sísmicas a una estación determinada puede ser definida mediante el ángulo back-azimut. Este ángulo es medido en sentido horario y respecto al norte, y se encarga de definir la dirección desde la estación sismológica al epicentro<sup>7</sup> del evento sísmico. A partir de la Figura 2.9 se observa un ejemplo del ángulo mencionado.

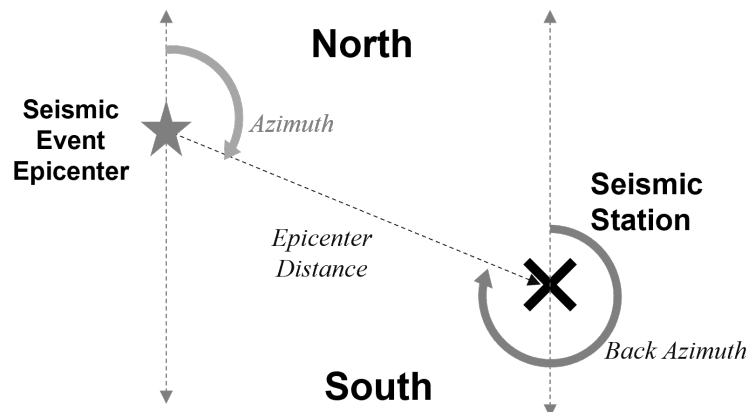


Figura 2.9: Ejemplo de back-azimut.

<sup>7</sup> Se denomina epicentro al punto en la superficie terrestre que resulta de la proyección vertical del punto donde se origina el sismo.

Si bien existen diversos métodos para calcular el back-azimut dadas las coordenadas de latitud y longitud del evento y la estación, uno de los más usados es el planteado en [16]. El algoritmo es iterativo y se denomina fórmula inversa de Vincenty. Dicho método permite calcular la distancia y los azimuts existentes entre dos puntos en la superficie terrestre (WGS84, sistema geodésico de coordenadas geográficas).

## 2.4. Estimación del back-azimut

A continuación se presentan diversos métodos de estimación del back-azimut que se han encontrado en la literatura revisada. Dichos métodos en general, pueden ser divididos en tres categorías: métodos clásicos, métodos que emplean machine learning y métodos basados en deep learning. Son estas dos últimas categorías las que abarcan el estado del arte. Es de relevancia mencionar que, debido al enfoque dado al presente trabajo, todos los métodos revisados consideran la información proveniente de una sola estación sísmológica.

### 2.4.1. Métodos clásicos de estimación del back-azimut

En general, los métodos asociados a la estimación del back-azimut en registros sísmicos están basados en características particulares de las ondas sísmicas. Por un lado, tanto las ondas P como las ondas S exhiben un alto grado de polarización lineal [17]. Adicionalmente, el azimut de propagación de la energía coincide con el movimiento de las partículas para fases compresionales (onda P), mientras que para fases transversales (onda S) el azimut es transversal al movimiento de las partículas [17]. De esta manera, se tiene entonces que mientras la onda P se encuentra polarizada en la dirección de propagación, la onda S presenta una polarización transversal [18].

En complemento a lo mencionado, es importante considerar que, seguido a la primera llegada correspondiente a la onda P, suelen aparecer pulsos, ondas superficiales y también ondas S [18]. Con todo ello, es de observar que la superposición de ondas y pulsos posterior a la primera llegada puede dificultar la estimación de direccionalidad (ya sea considerando información respecto a la polarización en la dirección de propagación o la polarización transversal).

#### 2.4.1.1. Estimación mediante la matriz de covarianza

Si se considera que la onda P se encuentra polarizada en la dirección de propagación, entonces se tiene que el realizar una estimación de la polarización de las primeras ondas de llegada permite obtener también una estimación de la dirección del origen del evento. Son diversos los trabajos en los cuales se plantea el uso de los vectores y valores propios de la matriz de covarianza para estimar la polarización, como se puede ver en [17–21].

En [21] se sugieren dos métodos para estimar el back-azimut que hacen uso de los vectores y valores propios mencionados. El primero de ellos, considera el uso de la señal registrada en los 3 canales durante el arribo de la fase P, mientras que el segundo solo emplea 2 canales (N y E), aunque de manera posterior se emplea el canal Z para remover cierta ambigüedad en la estimación. En ambos casos, se consideran asunciones tales como que la señal se encuentra polarizada de manera lineal, al igual que se asume la existencia de ruido aditivo gaussiano blanco afectando cada canal.

De esta manera, el método de 3 canales planteado en [21] se describe a continuación. En primer lugar, se estima la matriz de covarianza  $\hat{C}_3$  para cada muestra  $m$  dentro de la secuencia que contempla la primera llegada de ondas (P). En la Ecuación 2.11 se presenta la forma de la matriz de covarianza para cada muestra  $m$ , donde  $n$ ,  $e$  y  $z$  se refiere a los canales norte, este y vertical, respectivamente.

$$\hat{C}_3(m) = \begin{bmatrix} \hat{\sigma}_n^2(m) & \hat{\sigma}_{ne}(m) & \hat{\sigma}_{zn}(m) \\ \hat{\sigma}_{ne}(m) & \hat{\sigma}_e^2(m) & \hat{\sigma}_{ez}(m) \\ \hat{\sigma}_{zn}(m) & \hat{\sigma}_{ez}(m) & \hat{\sigma}_z^2(m) \end{bmatrix} \quad (2.11)$$

La estimación de cada elemento de la matriz de covarianza se realiza de manera recursiva, según las Ecuaciones 2.12-2.17, donde  $\gamma$  es un parámetro a través del cuál se busca estimar la matriz de covarianza tomando en cuenta la información pasada mediante un promedio ponderado exponencialmente. Se destaca también que  $n(m)$ ,  $e(m)$  y  $z(m)$  se refieren a la muestra  $m$  de la señal registrada en los canales norte, este y vertical respectivamente.

$$\hat{\sigma}_n^2(m) = \gamma \hat{\sigma}_n^2(m-1) + (1-\gamma)n^2(m) \quad (2.12)$$

$$\hat{\sigma}_e^2(m) = \gamma \hat{\sigma}_e^2(m-1) + (1-\gamma)e^2(m) \quad (2.13)$$

$$\hat{\sigma}_z^2(m) = \gamma \hat{\sigma}_z^2(m-1) + (1-\gamma)z^2(m) \quad (2.14)$$

$$\hat{\sigma}_{ne}(m) = \gamma \hat{\sigma}_{ne}^2(m-1) + (1-\gamma)n(m)e(m) \quad (2.15)$$

$$\hat{\sigma}_{ez}(m) = \gamma \hat{\sigma}_{ez}^2(m-1) + (1-\gamma)e(m)z(m) \quad (2.16)$$

$$\hat{\sigma}_{zn}(m) = \gamma \hat{\sigma}_{zn}^2(m-1) + (1-\gamma)z(m)n(m) \quad (2.17)$$

Luego, para cada matriz de covarianza  $\hat{C}_3$  se calculan sus vectores y valores propios, tomándose el vector propio asociado al mayor de los valores propios. Se calcula la proyección de tal vector propio en el plano N-E y se obtiene una estimación instantánea del ángulo back-azimut. Aunque este método es mencionado, cobra mayor relevancia su versión en 2 canales, también planteado en [21], puesto que requiere de un menor cómputo de variables, sin tener un impacto en cuanto a la calidad de la estimación. En efecto, en el método 2D, se estima para cada muestra la matriz de covarianza para las componentes N y E, según las Ecuaciones 2.12, 2.13 y 2.15

$$\hat{C}_2(m) = \begin{bmatrix} \hat{\sigma}_n^2(m) & \hat{\sigma}_{ne}(m) \\ \hat{\sigma}_{ne}(m) & \hat{\sigma}_e^2(m) \end{bmatrix} \quad (2.18)$$

Posteriormente, una estimación instantánea del back-azimut se obtiene según la Ecuación 2.19, donde  $\hat{\lambda}_{M2}(m)$  es el mayor valor propio de la matriz de covarianza registrada para la muestra  $m$ .

$$\phi(m) = \tan^{-1}\left(\frac{\hat{\lambda}_{M2}(m) - \hat{\sigma}_n^2(m)}{\hat{\sigma}_{ne}(m)}\right) \quad (2.19)$$

Para obtener un valor de estimación promedio, tanto el numerador como el denominador de la Ecuación 2.19 se promedian considerando los valores obtenidos para un período de 2

segundos, esto a partir de 1 segundo después de la primera llegada de ondas. El método tiene una ambigüedad de  $180^\circ$ , por lo que para deshacerse de esta es necesario considerar la polaridad de las diferentes correlaciones cruzadas  $\hat{\sigma}_{ne}(m)$ ,  $\hat{\sigma}_{zn}(m)$  y  $\hat{\sigma}_{ze}(m)$ , teniéndose el cuadrante según cierto criterio definido por los signos de dichas correlaciones.

#### 2.4.1.2. Estimación mediante Principal Component Analysis

Una manera de estimar el ángulo back-azimut es mediante el uso de Análisis de Componentes Principales (abreviado PCA), técnica a través de la cuál es posible encontrar la primera componente principal del movimiento de las partículas durante la llegada de la onda P. Cabe destacar, sin embargo, que para obtener las componentes principales se suele recurrir a los valores y vectores propios de la matriz de covarianza, por lo que este método podría caer en la anterior categoría.

En [22], para estimar el back-azimut se hace uso de PCA sobre las componentes del desplazamiento registrado. Este trabajo plantea, a diferencia del uso convencional de una ventana de tiempo fija en torno a la llegada de la onda P, el que se considere una ventana de tiempo móvil que depende de la señal en sí. Tal ventana de tiempo abarca desde la primera llegada de la onda P hasta el primer cruce por cero. Considerando una base de datos de 1991 registros sísmicos, el algoritmo con ventana variable registra un error rms de  $49,0^\circ$ , mientras que al usarse una ventana fija optimizada de 0,6 segundos, el error rms corresponde a  $57,3^\circ$ .

#### 2.4.1.3. Estimación mediante correlaciones cruzadas

En [23] se estima el back-azimut a través de la correlación cruzada entre los canales vertical, norte y este. En efecto, sea  $y(t) = [y_n(t), y_e(t), y_z(t)]^T$  la señal registrada en los 3 canales en el instante t. Se asume que frente a la llegada de una onda P, esta señal se puede descomponer según las Ecuaciones 2.20-2.22, donde P describe la señal de la onda P;  $N_n$ ,  $N_e$  y  $N_z$  es ruido aleatorio no correlacionado en los canales norte, este y vertical respectivamente; Q describe a ruido correlacionado entre los canales norte y este; y a, b son coeficientes reales constantes para cierto arribo de onda P.

$$y_n(t) = aP(t) + Q(t) + N_n(t) \quad (2.20)$$

$$y_e(t) = bP(t) + cQ(t) + N_e(t) \quad (2.21)$$

$$y_z(t) = P(t) + N_z(t) \quad (2.22)$$

La estimación se realiza entonces según la Ecuación 2.23, donde  $\langle \rangle$  corresponde al promedio considerando cierta ventana de tiempo en la llegada de la onda P.

$$\tan(az) = \frac{-b}{-a} = \frac{-\langle y_e y_z \rangle}{-\langle y_n y_z \rangle} \quad (2.23)$$

#### 2.4.1.4. Estimación mediante una sola muestra

En el trabajo [24] se realiza una estimación de back-azimut basándose en el uso de una sola muestra. El período considerado abarca desde la primera detección de la onda P, dado por la muestra de índice  $n1$ , hasta el primer máximo significativo en la componente vertical, definido por la muestra cuyo índice se denota por  $n2$ . El criterio para hallar tal máximo corresponde al punto en que se excede en 10 veces la relación señal-ruido (SNR). En [24],

se define de manera particular la relación SNR como el mínimo entre el valor SNR para la componente vertical y el valor SNR para las componentes horizontales, según se describe en la Ecuación 2.24.

$$SNR_{eisermann} = \min\left(\sqrt{\frac{\frac{1}{nwin-n1} \sum_{n=n1}^{nwin} A_Z^2(n)}{\frac{1}{n1} \sum_{n=0}^{n1} A_Z^2(n)}}, \sqrt{\frac{\frac{1}{nwin-n1} \sum_{n=n1}^{nwin} [A_E^2(n) + A_N^2(n)]}{\frac{1}{n1} \sum_{n=0}^{n1} [A_E^2(n) + A_N^2(n)]}}\right) \quad (2.24)$$

En la ecuación anterior se tiene que  $nwin$  es el índice de la muestra escogida para realizar la estimación. El método para escoger tal muestra se presentará más adelante. Luego, el método de estimación de back-azimut (BAZ) planteado en [24] considera la existencia de una relación, mediante cierta función, entre las amplitudes registradas en el sismógrafo y el ángulo buscado. Esto queda reflejado en la ecuación 2.25.

$$BAZ = f[\phi(A_E, A_N), \text{sign}(A_N), \text{sign}(A_Z)] \quad (2.25)$$

Donde  $A_Z$ ,  $A_E$  y  $A_N$  corresponden a las amplitudes de la velocidad registrada en las componentes vertical, este y norte respectivamente. Además,  $\phi$  viene dada por la Ecuación 2.26.

$$\phi = \arctan\left(\frac{A_E}{A_N}\right) + \pi \quad (2.26)$$

Para resolver la ambigüedad en  $180^\circ$ , la función  $f$  depende del signo de la señal en la componente norte y vertical, según la Ecuación 2.27.

$$f(\phi) = \begin{cases} \phi + \pi & \text{si } A_Z \cdot A_N < 0 \\ \phi & \sim \end{cases} \quad (2.27)$$

Finalmente, para determinar qué punto considerar en la estimación, se emplea la condición dada por la Ecuación 2.28.

$$nwin = \begin{cases} \underset{n1 \leq n \leq n2}{\operatorname{argmax}} |A_N(n)| & \text{si } \left| \underset{n1 \leq n \leq n2}{\operatorname{argmax}} |A_N(n)| \right| > \left| \underset{n1 \leq n \leq n2}{\operatorname{argmax}} |A_E(n)| \right| \\ \underset{n1 \leq n \leq n2}{\operatorname{argmax}} |A_E(n)| & \sim \end{cases} \quad (2.28)$$

Se observa que la condición para escoger el punto de estimación corresponde a la muestra que posee la mayor amplitud en el canal norte o este.

El método anterior fue evaluado en [24] sobre un conjunto de 114 avistamientos sísmicos<sup>8</sup>, registrados en 3 canales. La estimación de back-azimut obtenida es tal que el 39% de las estimaciones presentan un error dentro dentro del rango de  $\pm 10^\circ$ , mientras que el 67% lo hace en el rango de  $\pm 20^\circ$ . Se reporta entonces que el error absoluto registrado presenta un promedio de  $24,4^\circ$ , con una mediana de  $12,8^\circ$  y una desviación estándar de  $42,1^\circ$ .

Otro algoritmo evaluado en [24] corresponde a la estimación mediante PCA. El resultado reporta un error absoluto promedio de  $25,4^\circ$ , con mediana de  $12,4^\circ$  y desviación estándar de

<sup>8</sup> El término avistamiento sísmico se emplea en la presente memoria para indicar el registro de un evento sísmico en cierta estación sismológica.



44,9°.

## 2.4.2. Métodos basados en Machine Learning

En el último tiempo se han presentado métodos basados en algoritmos de aprendizaje de máquinas para estimar el back-azimut. Es importante notar que, bajo el paradigma del Machine Learning clásico, no es el modelo el que decide qué características extraer de la información que se ingresa, sino que más bien es el usuario quién debe plantear tales descriptores, los cuales deben ser pensados de manera que aporten información valiosa en el problema a resolver. Por tal motivo, es usual que se requiera cierto conocimiento experto en el tema.

En concreto, en los trabajos exhibidos en [25] y [26] se hace uso de Máquinas de Vectores de Soporte para labores de regresión (SVR) y clasificación (SVM), respectivamente. En términos simples, este tipo de algoritmo intenta, en problemas de clasificación, encontrar el mejor hiperplano (o los mejores si es un problema multi-clase) que separe a las clases en el espacio de características o algún espacio distinto que deriva de este (se puede ampliar la dimensionalidad mediante el uso de funciones llamadas “kernel”). Por otra parte, en problemas de regresión, este tipo de modelos intenta encontrar el hiperplano que se adapte a la mayor cantidad de muestras, esto en el espacio de características.

En el caso de [25], se intenta estimar el back-azimut, para lo cual se usan descriptores asociados a la estimación de magnitud, distancia epicentral y back-azimut. Estos últimos descriptores guardan directa relación con estimar la polarización de la onda P, como lo son los máximos valores propios de la matriz de covarianza bidimensional descrita en los trabajos [20] y [21], considerando un período de tiempo de 5 segundos (se evaluaron 5, 10 y 15 segundos, escogiéndose el primer valor) a partir de la llegada de la onda P.

La base de datos utilizada en [25] corresponde a registros sísmicos en 3 canales asociados a 863 eventos. Esta base de datos presenta un sesgo importante en cuanto a back-azimut, observándose que cerca de 350 eventos presentan un ángulo cercano a 240°. De esta manera, los resultados obtenidos por el modelo, considerando solo eventos con magnitud  $M > 3$ , dan cuenta de un error promedio de 2,81° con una desviación estándar de 45,43°.

Por otra parte, en [26] se emplean 6 descriptores para estimar una serie de variables, entre ellas, el back-azimut. Estos descriptores se obtienen en base al canal Z y son los siguientes: el punto de inicio de la onda P, el intervalo S-P (tiempo entre la llegada de la onda S y la onda P), el punto de inicio de la onda S, la amplitud (absoluta) máxima de la señal, la amplitud promedio y el valor RMS del sismograma. Los resultados obtenidos en este caso no resultan satisfactorios para estimar el back-azimut (aunque si para estimar la magnitud, la distancia y el ángulo de elevación).

De esta manera, se observa la importancia de escoger adecuadamente las características que se ingresan al modelo (dentro del paradigma del Machine Learning más clásico), las cuales deben tener directa relación con el problema abordado. En el caso de [26], era esperable que los resultados para back-azimut no fueran satisfactorios, pues los descriptores eran calculados en base a solo el canal Z, y no guardaban relación con las características vistas en los métodos clásicos descritos anteriormente.

### 2.4.3. Métodos basados en Deep Learning

En general, la mayoría de los trabajos que, por medio de Deep Learning, realizan estimación de localización de sismos usando la información de solo una estación sismológica, se basan en el empleo de redes convolucionales. Por ejemplo, en [27] se hace uso de una red CNN denominada ConvNetQuake, la cual consta de 8 capas convolucionales seguida de una capa fully connected. Esta red se emplea para realizar detección de sismos y estimación de localización de estos. La última tarea mencionada se aborda como un problema de clasificación, donde previamente los sismos considerados en la base de datos son agrupados en 6 zonas o clústers mediante el algoritmo K-means. Cabe destacar que la red recibe como entrada la señal temporal en los tres canales (frecuencia de muestreo 100 Hz, 10 segundos de señal). Para evaluar el desempeño del modelo, se define como métrica el % de aciertos en la clase predicha, entonces se tiene un 74,5% de precisión en el conjunto de prueba utilizado.

En [28] se trabaja sobre el mismo set de datos utilizado en [27] al igual que el mismo problema. La red utilizada es, sin embargo, distinta. En concreto, se emplea una arquitectura que combina una capa LSTM en paralelo con una red fully convolutional (fcn). Se destaca que a la capa LSTM se le añade un mecanismo de atención aditiva [29]. Dicho mecanismo de atención permite prestar mayor atención a las entradas con información relevante. Por otra parte, a las primeras capas convolucionales de la red se le añade un bloque denominado como “squeeze and excitation”. Dicho bloque fue presentado por primera vez en [30] y permite escalar los mapas de features, dando énfasis a aquellos que resultan de mayor relevancia para la tarea de clasificación (este bloque también puede entenderse como un tipo de auto-atención de canal). Considerando la misma métrica usada en [27], la red empleada en [28] obtiene una precisión de 89,1%.

En [31] el problema abordado consta de detección de eventos sísmicos al igual que estimación de variables tales como magnitud, distancia, profundidad y azimut, usando para ello registros de una sola estación. Para esto, es que se emplea una variante de la red usada en [27], denominada ConvNetQuake\_INGV, la cual está compuesta por 9 capas convolucionales y 1 capa fully connected. En este trabajo la estimación de cada una de las variables se realiza abordando el problema como uno de clasificación, y teniendo como entrada la señal temporal (50 segundos, con 20 Hz de muestreo) en los 3 canales. En el caso particular del azimut, se consideran 36 clases, teniéndose que cada una de estas abarca 10°. En general, el error observado para azimut en la evaluación en el conjunto de prueba presenta un nivel importante de dispersión.

En el trabajo presentado en [32] se hace uso de dos redes neuronales bayesianas<sup>9</sup>, una para estimar distancia y otra para estimar back-azimut, esto considerando la información en 3 canales de una sola estación sismológica. La base de datos utilizada corresponde a un número aproximado de 150000 avistamientos sísmicos registrados, los cuales pertenecen a la base de datos STEAD [33]. Estos avistamientos están asociados a un aproximado de 113000 terremotos, los cuales están distribuidos en diversos países, entre ellos Chile. Se destaca que, en general, la base de datos STEAD posee una diversidad importante en cuanto a la ubicación

---

<sup>9</sup> Las redes neuronales bayesianas se diferencian de las redes convencionales en el sentido de que en vez de estimar un set de pesos óptimos, se procede a encontrar una distribución probabilística para estos. Este tipo de modelos permiten cuantificar de mejor forma la incertidumbre asociada a las predicciones.

de los eventos sísmicos, tal como se aprecia en la Figura 2.10.

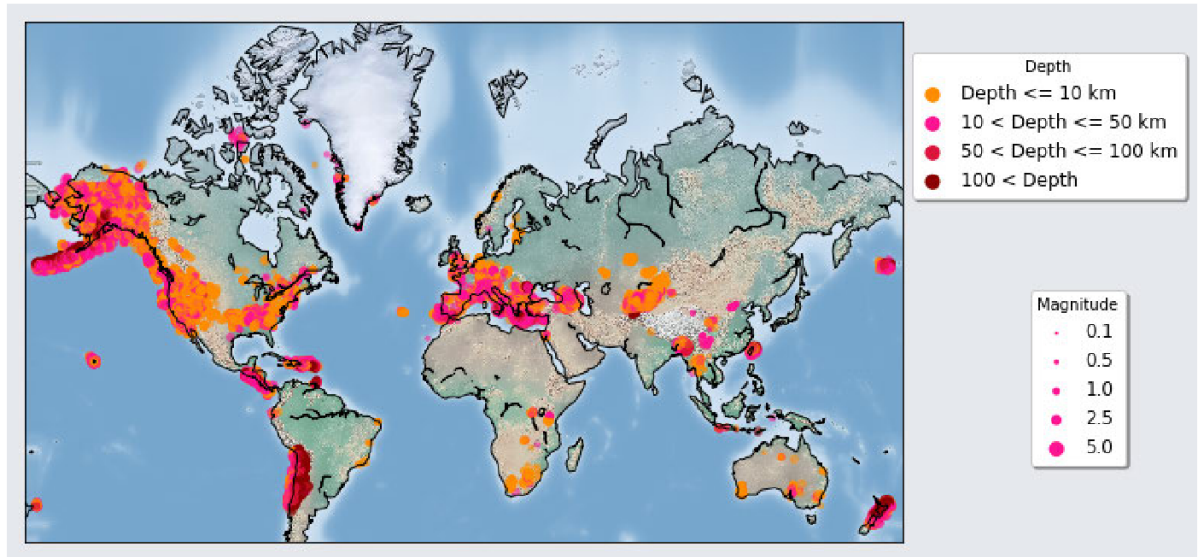


Figura 2.10: Distribución geográfica de los terremotos de la base de datos STEAD. Fuente: original de [33] (permiso bajo licencia CC-BY 4.0).

La red neuronal considerada para estimar back-azimut en el trabajo [32] corresponde a una red convolucional (CNN) profunda compuesta por cuatro capas convolucionales 1D y max-pooling 1D. La información que se ingresa a la red corresponde a la señal temporal en los tres canales, filtrada entre 1 y 45 Hz, en una ventana que contempla 0,5 segundos antes de la llegada de la onda P y 1 segundo posterior a esta. También se hace ingreso, de manera paralela, la matriz de covarianza calculada en la señal temporal, en la misma ventana de tiempo, junto a sus valores y vectores propios. Sobre esta entrada actúa una capa convolucional. Es de destacar entonces que las entradas tienen directa relación con la información considerada en los métodos clásicos de estimación, donde se ha visto que la información de la dirección está contenida en gran medida en los primeros segundos de la onda P. También se destaca que lo que se estima no es el ángulo directamente, sino el seno y el coseno de este.

Los resultados conseguidos en [32], los cuales se obtuvieron sobre un 20 % de la base de datos considerada, dan cuenta de un error promedio de  $1,07^\circ$  con una desviación estándar de  $35,50^\circ$ . Frente a estos resultados, es importante considerar que la base de datos utilizada considera solo avistamientos donde la distancia entre el epicentro y la estación es menor a 110 km, con una razón de señal-ruido (SNR) de 25 decibeles o más. Es relevante entonces mencionar que estas condiciones no siempre se cumplen.

Finalmente, aunque escapando del problema de estimación de back-azimut como tal, es de destacar que en [34] se aborda el problema de estimación de localización y magnitud de eventos sísmicos mediante información de múltiples estaciones, empleando para ello una arquitectura de red que incorpora una primera parte compuesta por una red convolucional, la cual permite extraer características de la señal temporal registrada en cada estación, para luego combinar tales características empleando una red transformer. Ello da cuenta que este tipo de redes son posibles de emplear en el problema de estimación de localización de eventos sísmicos.

#### 2.4.4. Consideraciones sobre los métodos de estimación de back-azimut revisados

En general, a partir de los diversos métodos revisados, tanto clásicos como del estado del arte, es posible observar que mayormente lo utilizado corresponde a la señal temporal registrada en los tres canales al momento del arribo de la onda P. Respecto al desempeño, se observa que los modelos basados en Deep Learning tienden a presentar un nivel de desempeño superior a los algoritmos clásicos. Así, este tipo de modelos cobra cada vez mayor relevancia, ello considerando su capacidad de generalización y la abundancia de datos disponibles.

También, es de destacar que la mayoría de los métodos expuestos fueron evaluados en bases de datos que no involucran sismos en el territorio nacional, por lo cual sus resultados podrían no ser extrapolables. El método que presentó mejor desempeño corresponde al planteado en [32], el cual fue entrenado y evaluado en una base de datos que sí contiene eventos presentes en el territorio nacional, y particularmente, en la zona norte del país, donde existe un potencial sísmico importante. Dicho método, sin embargo, fue entrenado y evaluado solo con señales que presentan un valor de SNR alto (mayor a 25 dB) y baja distancia epicentral ( $<110$  km), lo cual no siempre se cumple. De esta manera, resulta atingente el desarrollar un modelo que se base solo en eventos sísmicos registrados en territorio nacional, independientemente de la razón SNR.

# Capítulo 3

## Metodología

La metodología empleada en el presente trabajo puede ser dividida en 4 etapas, las cuales se presentan en la Figura 3.1. Adicionalmente, se resumen a continuación:

1. **Descarga y división de la base de datos:** se realiza un proceso de búsqueda y descarga de avistamientos sísmicos registrados en estaciones de la zona centro-norte de Chile. Para realizar los experimentos, esta base de datos se divide de manera aleatoria en 3 conjuntos: entrenamiento, validación y prueba, siguiendo una proporción aproximada de 60 %, 20 % y 20 %. Adicionalmente, manteniendo fijos estos conjuntos, la base de datos se subdivide en 2: un set de datos enfocado a sistemas de alerta temprana para tsunami (avistamientos con magnitud  $M \geq 4$ ) y otro set de de datos enfocado a micro-sismicidad (distancia epicentral  $\leq 100$  km y  $M \leq 2,5$ ).
2. **Pre-procesamiento:** todas las trazas son pre-procesadas según diversos aspectos, tales como remoción de respuesta instrumental, filtrado pasa-bandas, equiparar las frecuencias de muestreo, normalización de las señales, entre otros.
3. **Entrenamiento de los modelos:** los modelos desarrollados son entrenados y evaluados según los respectivos conjuntos asociados a las bases de datos completa,  $M \geq 4$  y micro-sismicidad.
4. **Análisis y comparación de los resultados obtenidos:** los resultados obtenidos se comparan con el fin de escoger el modelo con el mejor desempeño sobre el conjunto de prueba, empleándose para ello la métrica MAE (error absoluto medio).

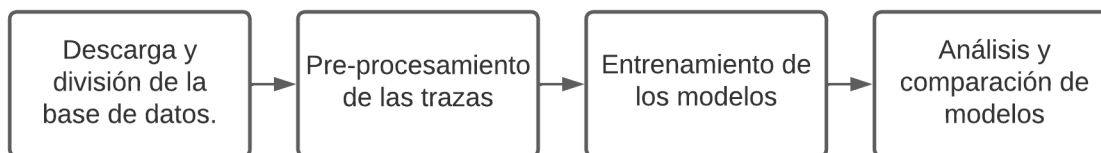


Figura 3.1: Diagrama general de la metodología seguida.

### 3.1. Base de datos

La base de datos empleada para la construcción del sistema consta de un total de 3160 avistamientos sísmicos, obtenidos desde la plataforma “Incorporated Research Institutions for Seismology” (IRIS <sup>10</sup>), empleando las redes sismológicas habilitadas en el país y usadas por el CSN (Centro Sismológico Nacional). Para cada una de las señales se marcó de manera manual el instante de llegada de las ondas P y S.

La base de datos se compone de un total de 49 estaciones sismológicas, las cuales pueden ser divididas en 2 grupos: 10 estaciones de referencia, que son estaciones usadas por recomendación del CSN debido a su buena calidad en la relación señal ruido, amplio rango de muestreo de datos y una distribución uniforme en el territorio entre la frontera norte y la capital; y 39 estaciones secundarias, las cuales generalmente no poseen la misma calidad de registro que las primeras.

Los eventos disponibles ocurrieron entre los años 2014 y 2022, entre las latitudes  $-17,51^\circ$  hasta  $-34,18^\circ$  y las longitudes  $-72,95^\circ$  hasta  $-66,68^\circ$ . Estos cuentan con un rango de magnitudes que va desde los M1.8 hasta los M8.1. En la Tabla 3.1 se presenta un resumen por rango de magnitud de la base de datos descargada, la cual fue dividida en conjuntos de entrenamiento, validación y prueba.

Tabla 3.1: Resumen de la cantidad de avistamientos de la base de datos descargada.

Rango $M$	Entrenamiento	Validación	Test	Cantidad Total
$M < 2.5$	262	77	93	432
$2.5 \leq M < 3.5$	420	139	124	683
$3.5 \leq M < 4.5$	675	251	254	1180
$4.5 \leq M < 5.5$	454	137	155	746
$5.5 \leq M < 6.5$	71	16	18	105
$M \geq 6.5$	11	0	3	14
TOTAL	1893	620	647	3160

Para cada uno de los avistamientos sísmicos, se cuenta con información referida a la localización del epicentro respecto a la estación de registro (esto es, en términos de distancia y back-azimut). En la Figura 3.2 se presenta la distribución de localizaciones de los epicentros, donde la tonalidad da cuenta de la magnitud del evento y las cruces representan estaciones sismológicas. Adicionalmente, en las Figuras 3.3.a y 3.3.b se presentan los histogramas de distancia epicentral y back-azimut de toda la base de datos, respectivamente. Para obtener tanto la distancia epicentral como el back-azimut se empleó el algoritmo de la fórmula inversa de Vincenty [16], el cual se encuentra implementado en la librería ObsPy.

<sup>10</sup> <https://www.iris.edu/hq/>

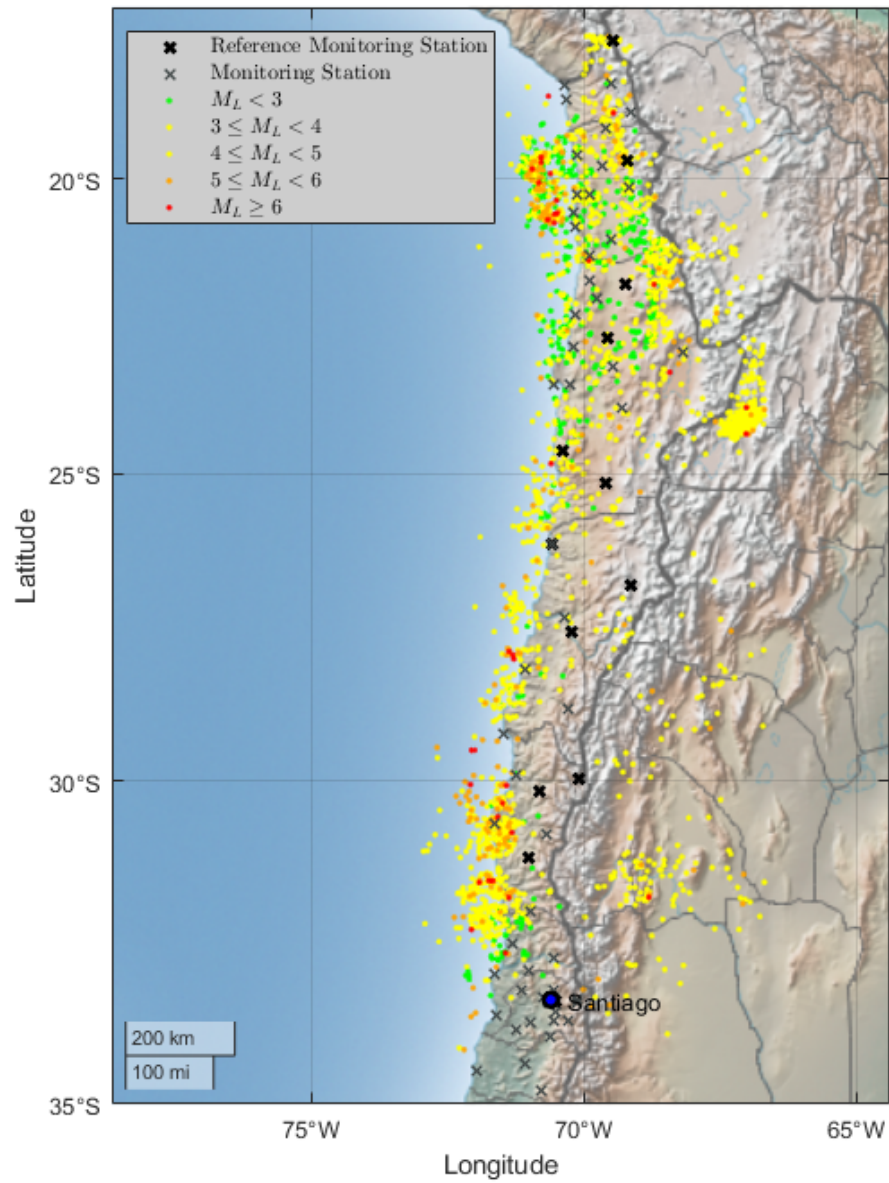


Figura 3.2: Mapa de epicentros y magnitudes de los eventos de la base de datos completa.

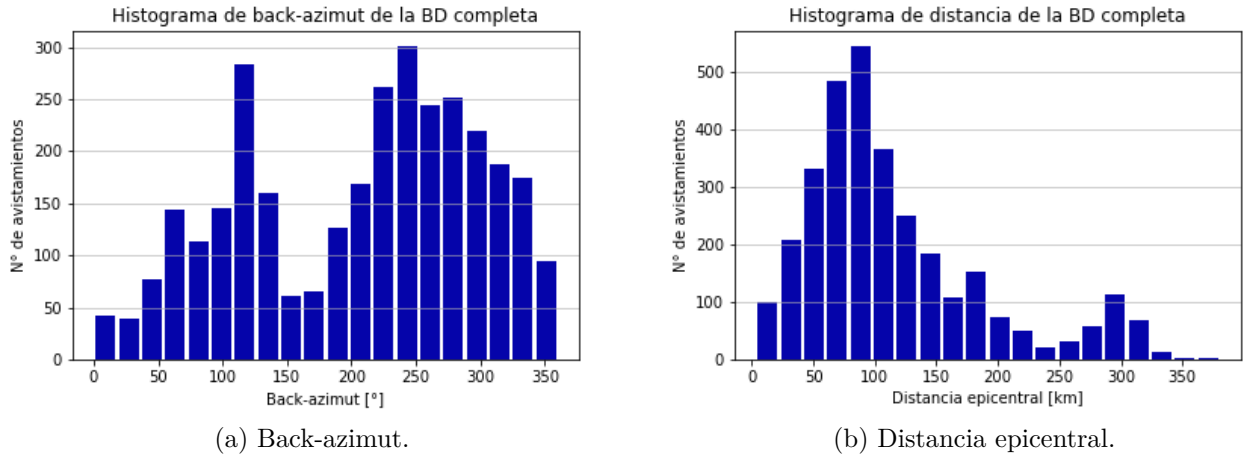


Figura 3.3: Histograma de back-azimut y distancia epicentral de la base de datos completa.

### 3.1.1. Estimación de SNR

El cálculo del SNR se llevó a cabo siguiendo un procedimiento similar al descrito por [33]. Para ello, todas las señales fueron homologadas a una frecuencia de muestreo de 40 Hz<sup>11</sup> y se les removió su media. Posteriormente, el SNR se calculó según la Ecuación 3.1, para cada canal de manera independiente.

$$SNR = 10 \cdot \log_{10} \left( \frac{\|S_{window}\|_2^2}{\|N_{window}\|_2^2} \right) \quad (3.1)$$

Donde  $S_{window}$  corresponde a aquellas amplitudes que superan el percentil 95 en una ventana de 5 segundos posterior a la llegada de la onda S; y  $N_{window}$  corresponde a aquellas amplitudes que superan el percentil 95 en una ventana de 5 segundos anterior a la llegada de la onda P. En la Figura 3.4 se presenta un histograma con los promedios del SNR en los tres canales para los avistamientos de la base de datos completa.

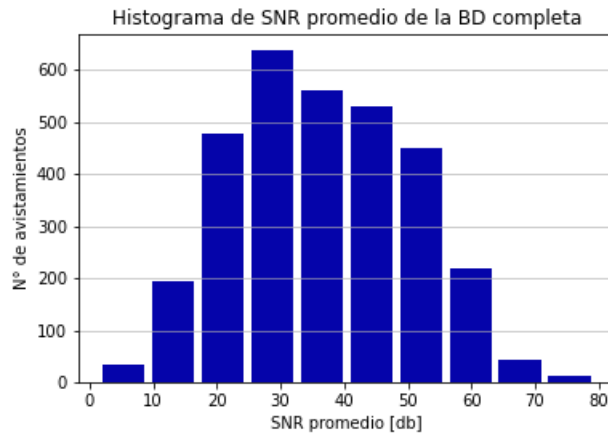


Figura 3.4: SNR promedio en los 3 canales.

<sup>11</sup> Algunas estaciones registran con frecuencia de muestreo de 100 Hz y otras en 40 Hz.



### 3.1.2. Base de datos de eventos con $M \geq 4.0$

Para este sub-grupo de la base de datos, se consideraron solo registros sísmicos de eventos con magnitud  $M \geq 4$ . En la Tabla 3.2 se presenta un resumen de la cantidad de avistamientos perteneciente a esta base de datos. En las Figuras 3.5.a y 3.5.b se presenta la distribución de back-azimut y distancia epicentral de la base de datos  $M \geq 4$ .

Tabla 3.2: Resumen de la cantidad de avistamientos de la base de datos de eventos con  $M \geq 4.0$  considerando las estaciones de referencia y secundarias.

Rango $M$	Entrenamiento	Validación	Test	Cantidad Total
$M < 2.5$	0	0	0	0
$2.5 \leq M < 3.5$	0	0	0	0
$3.5 \leq M < 4.5$	543	196	211	950
$4.5 \leq M < 5.5$	454	137	155	746
$5.5 \leq M < 6.5$	71	16	18	105
$M \geq 6.5$	11	0	3	14
TOTAL	1079	349	387	1815

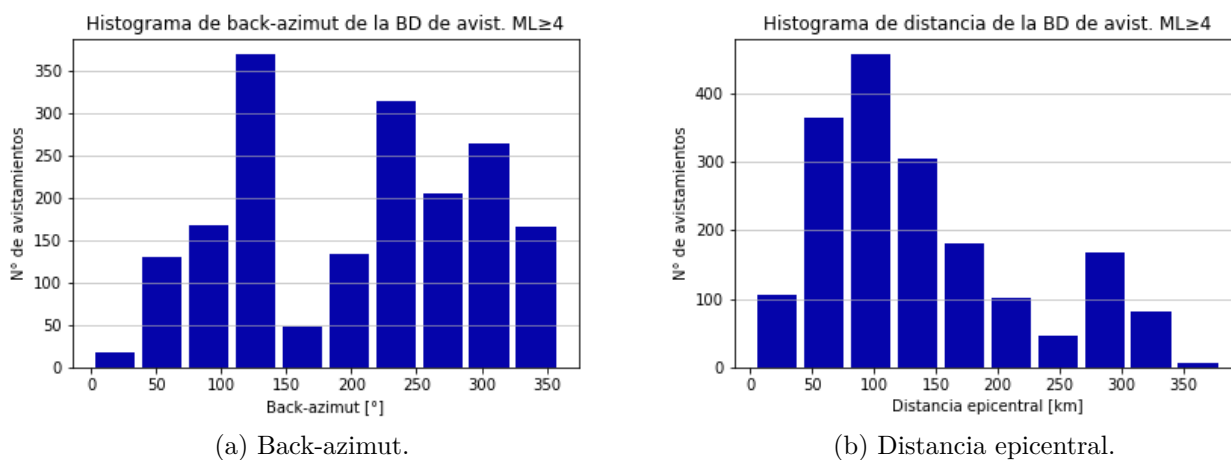


Figura 3.5: Histogramas de back-azimut y distancia para BD de eventos con  $M \geq 4.0$ .

Adicionalmente, en la Figura 3.6 se presenta la distribución geográfica de epicentros de la base de datos de eventos con magnitud  $M \geq 4$ .

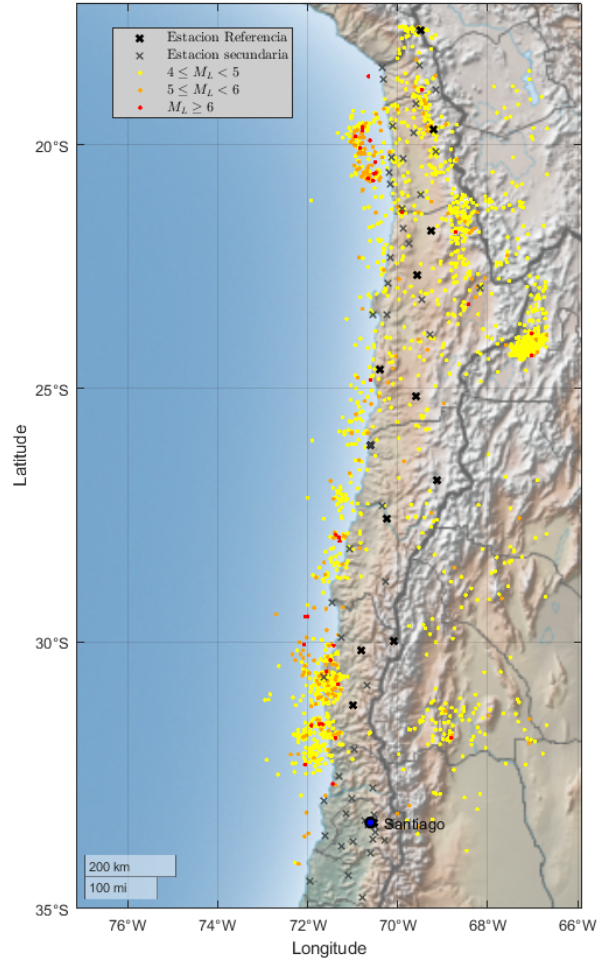


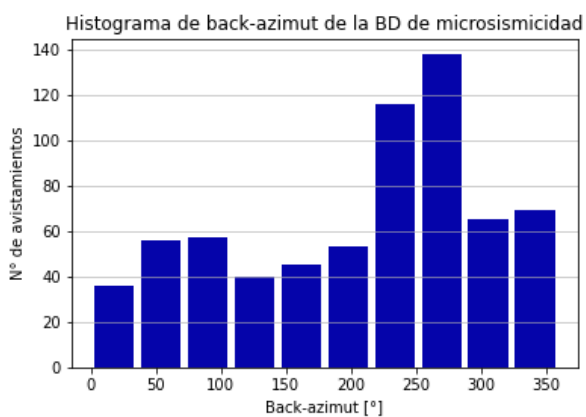
Figura 3.6: Mapa de epicentros y magnitudes de los eventos de la base de datos  $M \geq 4$ .

### 3.1.3. Base de datos de micro-sismicidad local

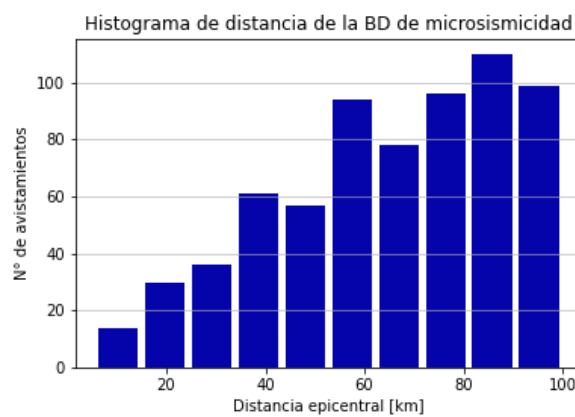
Para este sub-grupo de la base de datos se consideraron avistamientos cuya magnitud  $M \leq 2.5$  al mismo tiempo que su distancia al epicentro es menor o igual a 100 km. En la Tabla 3.3 se presenta un resumen de la cantidad de avistamientos perteneciente a esta base de datos. En las Figuras 3.7.a y 3.7.b se presenta la distribución de back-azimut y distancia epicentral de la base de datos de microsismicidad, respectivamente.

Tabla 3.3: Resumen de la cantidad de avistamientos de la base de datos de microsismicidad considerando las estaciones de referencia y secundarias.

Rango $M$	Entrenamiento	Validación	Test	Cantidad Total
$M < 2.5$	221	64	80	365
$2.5 \leq M < 3.5$	203	50	57	310
$3.5 \leq M < 4.5$	0	0	0	0
$4.5 \leq M < 5.5$	0	0	0	0
$5.5 \leq M < 6.5$	0	0	0	0
$M \geq 6.5$	0	0	0	0
TOTAL	424	114	137	675



(a) Back-azimut.



(b) Distancia epicentral.

Figura 3.7: Histogramas de back-azimut y distancia al epicentro de la base de datos de microsismicidad.

Por otra parte, en la Figura 3.7 se presenta la distribución geográfica de los epicentros de la base de datos de microsismicidad. Adicionalmente, en la Tabla 3.4 se presenta un resumen de las diversas bases de datos consideradas, incluyéndose el SNR promedio registrado para cada una de ellas.

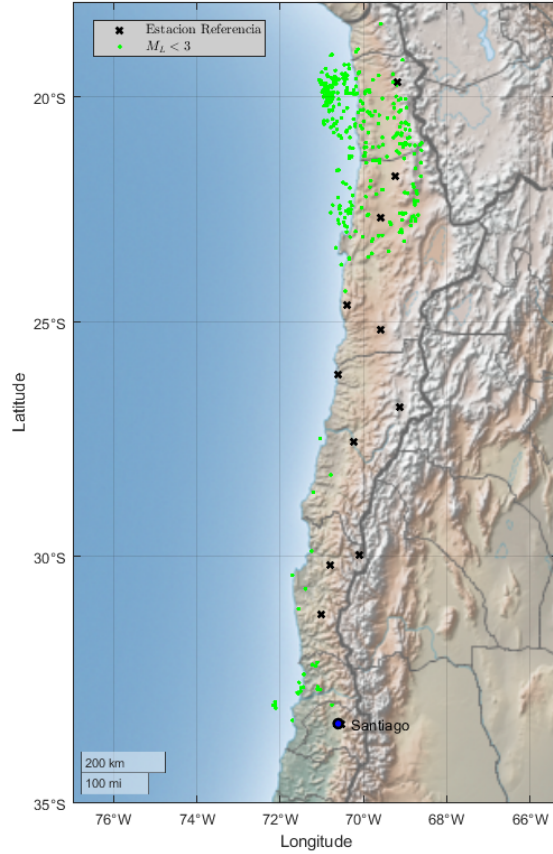


Figura 3.8: Mapa de epicentros y magnitudes de los eventos para la base de datos de microsismicidad local.

Tabla 3.4: Distribución de avistamientos de las bases de datos.

	Train	Val	Test	$SNR_{mean}$
Base de datos completa	1893	620	647	38,34
Base $M \geq 4$	1079	349	387	42,90
Base microsismicidad	424	114	137	25,27

## 3.2. Pre-procesamiento de las señales

En el caso de este trabajo, son empleadas las señales registradas en los tres canales (vertical, norte y este). Debido a las diferencias en frecuencia de muestreo entre diversas estaciones, se optó por homologar todas las señales a una frecuencia de muestreo de 40 Hz. Adicionalmente, son diversas las acciones llevadas a cabo, las cuales se pueden listar a continuación:

1. **Remoción de la respuesta instrumental:** a todas las trazas se les remueve la respuesta instrumental asociada a la estación de registro. Asimismo, cada señal es transformada a velocidad.

2. **Filtrado pasa-banda:** En base a experimentación, se ha observado que la información del back-azimut tiende a estar concentrada en bandas de baja frecuencia. Por tal motivo, en este caso se aplica un filtrado pasa-bandas con frecuencias de corte de 0,5 Hz y 10 Hz.
3. **Segmentación de las señales:** Debido a que se ha observado que es posible estimar el back-azimut solamente empleando los datos asociados a la primera llegada de la onda P, es que se realiza un segmentado de las señales, considerando 1 segundo antes de la llegada de la onda P y 3 segundos después de esta. Dicha llegada fue marcada y revisada de manera manual para cada uno de los avistamientos de la base de datos.
4. **Normalización de las señales segmentadas:** Para cada avistamiento se normaliza, de manera independiente<sup>12</sup>, las señales asociadas a sus tres canales en base a la máxima amplitud registrada. Lo anterior queda reflejado en la Ecuación 3.2, donde las tres señales segmentadas son agrupadas en una matriz tal que  $signal \in \mathbb{R}^{timesteps \times channels} = \mathbb{R}^{160 \times 3}$ .

$$signal_{norm}(i, j) = signal(i, j) / \max(abs(signal)) \quad (3.2)$$

### 3.3. Arquitecturas de red a evaluar

Debido al buen desempeño alcanzado por las redes convolucionales en la tarea de estimación de localización, en este trabajo también se hace uso de este tipo de arquitectura. Una fuente de inspiración importante es la red de estimación de back-azimut exhibida en [32], donde no se estima directamente el ángulo sino el seno y el coseno de este. Ello permite representar el back-azimut como un punto perteneciente a la circunferencia unitaria, con lo cual es posible evadir la discontinuidad existente entre 0 y 360°.

Adicionalmente, se propone el uso de dos elementos que podrían mejorar el desempeño de la red:

- **ID station:** identificador de la estación que registra la señal, a través de lo cual se busca poder discernir e identificar posibles singularidades de las estaciones empleadas. Para ingresar esta información se utiliza la técnica de codificación denominada *One Hot Encoding*. De esta manera, si se tienen 49 estaciones, para identificar la estación  $i$ -ésima se emplea un vector  $v \in \mathbb{R}^{49}$  tal que  $v_k = 0, \forall k \neq i$  y  $v_i = 1$ . Se ha observado anteriormente que añadir información respecto a la estación utilizada puede ser beneficioso [35] (esto al menos en el contexto de estimación de magnitud).
- **Uso de mecanismo de auto-atención:** reforzar las características extraídas por la red convolucional mediante el uso del mecanismo de auto-atención empleado en [7], a través de un bloque Multi-Head Attention según lo indicado en la Figura 2.5, donde las proyecciones lineales de Q (queries), K (keys) y V (values) se calculan en base a la salida de la estructura convolucional.

De esta manera, a continuación se detallan las variantes de red planteadas.

---

<sup>12</sup>La normalización se realizó de manera independiente entre avistamientos debido a que la diferencia de amplitud entre avistamientos de baja y alta magnitud puede ser considerable, lo cual puede originar valores normalizados notablemente pequeños.

### 3.3.1. Red Convolucional

En la Figura 3.9 se presenta el diagrama de la red convolucional propuesta que incluye el uso de ID station. En este caso, las características extraídas a partir de la estructura convolucional son aplanadas y concatenadas con el vector asociado a la información del ID station. Dicho vector es posteriormente ingresado a un bloque fully connected, representado en la Figura 3.10. Tal bloque consiste en una capa Densa con función de activación ReLU ( $ReLU(x) = \max(0, x)$ ) y Dropout<sup>13</sup> [36], conectado luego con la capa de salida. Es a través de esta última capa que se obtiene una predicción del seno y el coseno del back-azimut buscado.

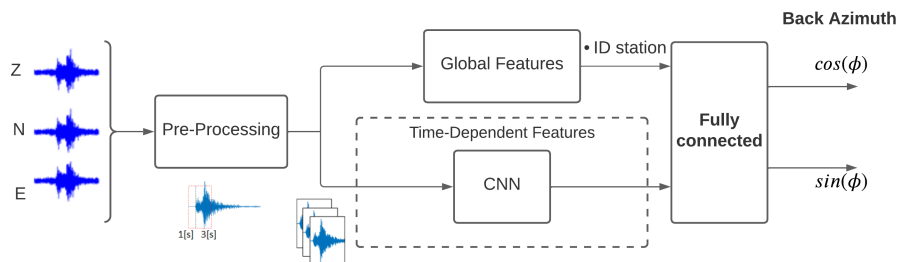


Figura 3.9: Red convolucional con ID station

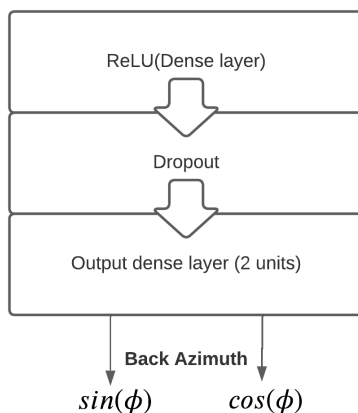


Figura 3.10: Bloque fully connected empleado al final de cada arquitectura.

Es de destacar que la estructura convolucional (elemento CNN en Figura 3.9) está compuesta por un cierto número de bloques convolucionales, los cuales están formados por una capa convolucional 1D con función de activación ReLU, Dropout y una capa max-pooling 1D, tal como se presenta en la Figura 3.11. Debido a la cantidad limitada de datos de entrenamiento con que se cuenta, es de interés que la red diseñada posea pocos parámetros. Por ello, particular relevancia cobran las capas max-pooling 1D empleadas, pues permiten reducir el tamaño de la secuencia proveniente de la capa anterior.

<sup>13</sup> Técnica de regularización consistente en ignorar de manera aleatoria durante el entrenamiento cierta fracción de las unidades neuronales.

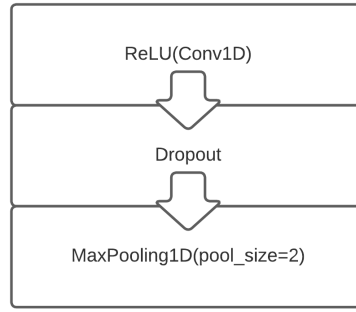


Figura 3.11: Bloques convolucionales empleados en la arquitectura.

### 3.3.2. Red Convolutiva con mecanismo atencional

En la Figura 3.12 se presenta la arquitectura convolutiva reforzada con un mecanismo de auto-atención que es propuesta, la cual se denomina como CNN-MHAT. En esta arquitectura se usa el mismo tipo de bloque convolutivo descrito en la Figura 3.11, al igual que al final se usa la misma estructura fully connected presentada en la Figura 3.10.

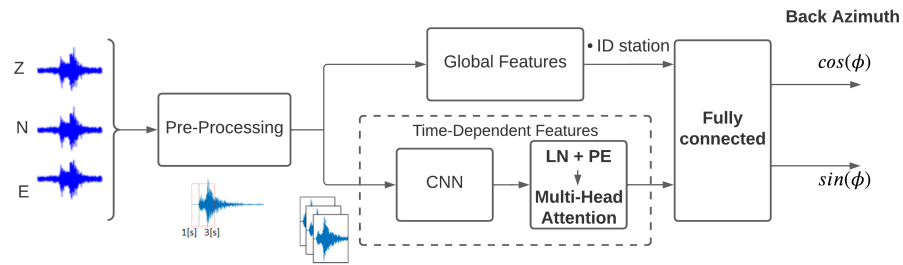


Figura 3.12: Red convolutiva con mecanismo de auto-atención y ID station.

Se destaca que LN denota el uso de *layer normalization* para normalizar la salida del último bloque convolutivo empleado. Una vez normalizados los vectores de la secuencia obtenida, a estos se les agrega *positional encoding*, denotado como PE. Dicha codificación corresponde al uso de funciones seno y coseno según lo dictado en las Ecuaciones 2.9 y 2.10, donde  $d_{model}$  es igual a la cantidad de filtros que emplea la última capa convolutiva.

Posteriormente, la secuencia es ingresada al mecanismo atencional descrito en la Figura 2.5, donde se realizan proyecciones lineales de las entradas para luego aplicar la función de atención *scaled dot product*. Esto se realiza en diversas instancias de manera paralela, según la cantidad de cabezas de atención usadas. Los resultados obtenidos son concatenados y finalmente proyectados por una sub-capta lineal, teniéndose que la salida posee la misma dimensionalidad que la secuencia ingresada al mecanismo atencional.

A continuación, la secuencia obtenida es aplanada y concatenada con el vector asociado al ID station. Luego, el vector resultante se hace ingresar a la red fully connected. Para

cuantificar el efecto del identificador ID station, también se entrena y evalúa el modelo sin utilizar esta característica.

### 3.4. Entrenamiento de los modelos

Considerando cierta base de datos, para entrenar los modelos se emplea el respectivo conjunto de entrenamiento. Para todos los modelos, se usa como optimizador el algoritmo ADAM [37]. Por otra parte, la función a optimizar corresponde al error cuadrático medio, presentado en la Ecuación 3.3, donde  $N$  es la cantidad de salidas (en este caso  $N = 2$ ),  $M$  corresponde a la cantidad de avistamientos en un mini-batch, e  $y_{true}, y_{pred} \in \mathbb{R}^{M \times N}$  son los targets y las predicciones respectivamente.

$$MSE(y_{true}, y_{pred}) = \frac{1}{M} \left( \sum_{i=1}^M \frac{1}{N} \left( \sum_{j=1}^N (y_{ij}^{true} - y_{ij}^{pred})^2 \right) \right) \quad (3.3)$$

Para evitar el sobreajuste en las redes, se emplea la técnica denominada *earlystopping*, donde durante el entrenamiento de la red, al final de cada *epoch*<sup>14</sup> se evalúa el modelo sobre el conjunto de validación, deteniéndose el entrenamiento cuando, a partir de cierta cantidad de *epochs*, la función de pérdida sobre este conjunto deja de disminuir.

### 3.5. Evaluación de los modelos

Una vez entrenados los modelos, la evaluación de estos no es directa, puesto que las predicciones no corresponden directamente al ángulo back-azimut, sino más bien al seno y coseno de este. Adicionalmente, es necesario establecer cierta métrica a partir de la cual se pueda comparar el desempeño de los modelos. Todo esto se aborda a continuación.

#### 3.5.1. Reconstrucción del ángulo

Si se consideran predicciones de la red  $y_1, y_2$  tales que apuntan al seno y al coseno del ángulo, respectivamente, entonces la reconstrucción del back-azimut se puede realizar a través de la función arcotangente de 2 argumentos, presentada en la Ecuación 3.4. Esta función es similar a la función arcotangente, con la diferencia de que toma en cuenta el signo del numerador y denominador para determinar el cuadrante al que pertenece el ángulo.

---

<sup>14</sup>El término *epoch* se refiere a una pasada completa de todo el set de datos de entrenamiento por la red neuronal.



$$BAZ = \arctan2(y_1, y_2) = \begin{cases} \arctan\left(\frac{y_1}{y_2}\right) & \text{si } y_2 > 0 \\ \arctan\left(\frac{y_1}{y_2}\right) + \pi & \text{si } y_2 < 0 \text{ y } y_1 \geq 0 \\ \arctan\left(\frac{y_1}{y_2}\right) - \pi & \text{si } y_2 < 0 \text{ y } y_1 < 0 \\ +\frac{\pi}{2} & \text{si } y_2 = 0 \text{ y } y_1 > 0 \\ -\frac{\pi}{2} & \text{si } y_2 = 0 \text{ y } y_1 < 0 \\ \text{indefinido} & \text{si } y_2 = 0 \text{ y } y_1 = 0 \end{cases} \quad (3.4)$$

### 3.5.2. Promedio circular

Para promediar los resultados obtenidos a partir de diversas instancias del mismo modelo frente a un mismo avistamiento (esto es, que es entrenado en diversas ocasiones, lo cual arroja resultados diferentes frente al mismo avistamiento), es que se utiliza el promedio circular, el cual se presenta en la Ecuación 3.5, donde K es la cantidad de ángulos a promediar.

$$B\hat{A}Z_{mean} = \arctan2\left(\frac{1}{K} \sum_{j=1}^K (\sin(B\hat{A}Z_j)), \frac{1}{K} \sum_{j=1}^K (\cos(B\hat{A}Z_j))\right) \quad (3.5)$$

### 3.5.3. Métrica de evaluación a utilizar

Para evaluar y comparar el desempeño de los diversos modelos entrenados, se propone el uso de la métrica *MAE*, acrónimo referido al error absoluto medio (Mean Average Error). Dicha métrica se presenta en la Ecuación 3.6, donde N es la cantidad de avistamientos,  $BAZ_i$  el ángulo real del avistamiento i-ésimo y  $B\hat{A}Z_i^{corregido}$  el ángulo estimado para el mismo avistamiento. Dicho ángulo se encuentra corregido, de manera de obtener la menor diferencia angular con el valor real. La corrección en cuestión no afecta en absoluto a la dirección de origen estimada, pues solamente consiste en sumar o restar  $2\pi$  al ángulo, o mantenerlo igual, considerando que este se encuentra en radianes<sup>15</sup>. Para ejemplificar el motivo de esta corrección, considerar los ángulos  $358^\circ$  y  $2^\circ$ , la diferencia directa son  $356^\circ$ , sin embargo, la menor diferencia angular corresponde a  $4^\circ$ .

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^N \left| BAZ_i - B\hat{A}Z_i^{corregido} \right| \quad (3.6)$$

Para expresar claramente la corrección mencionada, considerar las variables auxiliares presentadas en las ecuaciones 3.7, 3.8 y 3.9. Se tiene que la corrección del ángulo estimado para el avistamiento i-ésimo viene dado por la ecuación 3.10.

$$Ae_i^{cor1} = |BAZ_i - B\hat{A}Z_i| \quad (3.7)$$

<sup>15</sup> De manera posterior, para transformar un ángulo que está en radianes a grados, solo basta multiplicarlo por  $\frac{180}{\pi}$ .

$$Ae_i^{cor2} = |BAZ_i - (\hat{B}\hat{A}Z_i - 2\pi)| \quad (3.8)$$

$$Ae_i^{cor3} = |BAZ_i - (\hat{B}\hat{A}Z_i + 2\pi)| \quad (3.9)$$

$$\hat{B}\hat{A}Z_i^{corregido} = \begin{cases} \hat{B}\hat{A}Z_i, & \text{si } Ae_i^{cor1} < \min(Ae_i^{cor2}, Ae_i^{cor3}) \\ \hat{B}\hat{A}Z_i - 2\pi, & \text{si } Ae_i^{cor2} < \min(Ae_i^{cor1}, Ae_i^{cor3}) \\ \hat{B}\hat{A}Z_i + 2\pi, & \text{si } Ae_i^{cor3} < \min(Ae_i^{cor1}, Ae_i^{cor3}) \end{cases} \quad (3.10)$$

# Capítulo 4

## Resultados

En este capítulo se presentan los resultados obtenidos para los modelos planteados. Para ello, los modelos emplean la mejor configuración de hiperparámetros hallada. La búsqueda de tales hiperparámetros se realizó entrenando y evaluando los modelos sobre los conjuntos respectivos de la base de datos completa.

De esta manera, a continuación se presenta el conjunto de hiperparámetros utilizado para cada modelo de manera que se obtienen los mejores resultados.

### 4.1. Mejor configuración

En la Tabla 4.1 se presenta el conjunto de hiperparámetros de la red convolucional de la Figura 3.9. Adicionalmente, en la Tabla 4.2 se presenta el conjunto de hiperparámetros de la red convolucional que emplea un mecanismo de auto-atención (CNN-MHAT). Se destaca que tanto la tasa de aprendizaje como el tamaño de batch resultan ser de un valor bajo, lo cual es consistente con la afirmación de que reducir el tamaño de batch y la tasa de aprendizaje puede permitir a la red entrenar de mejor manera [38].

Tabla 4.1: Mejor conjunto de hiperparámetros para red CNN.

Hiperparámetros	Valor
Cantidad de bloques Convolucionales	4
Filtros por capa Conv1D	24, 16, 16, 16
Kernel size capas Conv1D	3
% Dropout bloques Conv.	20 %
Nº de unidades en capa densa	20
% Dropout capa densa	20 %
Batch Size	1
Learning Rate	0,00015

Tabla 4.2: Mejor conjunto de hiperparámetros para red CNN con MultiHead self-attention.

Hiperparámetros	Valor
Cantidad de bloques Convolucionales	4
Filtros por capa Conv1D	24,16,16,16
Kernel size capas Conv1D	3
% Dropout bloques Conv.	20 %
Nº de unidades en capa densa	20
% Dropout capa densa	20 %
Nº de cabezas en MultiHeadAtt	2
Dimensión de proyección $d_k = d_v$	8
Batch Size	1
Learning Rate	0,00015

## 4.2. Resultados en base de datos completa

Debido al proceso probabilístico detrás de la inicialización de los pesos y el entrenamiento de la red, los modelos fueron entrenados 30 veces. Luego, las respectivas predicciones asociadas a cada avistamiento fueron promediadas circularmente según la Ecuación 3.5.

En la Tabla 4.3 se presenta el error absoluto medio obtenido por el modelo convolucional que emplea el mecanismo atencional Multi-Head Attention (CNN-MHAT) al ser evaluado sobre los diversos conjuntos pertenecientes a la base de datos completa. Adicionalmente, se comparan dichos resultados con los obtenidos por la variante que no emplea ID station, al igual que aquella que no utiliza el mecanismo de atención (CNN).

Tabla 4.3: Error absoluto medio obtenido al evaluar los modelos sobre los conjuntos de la base de datos completa.

Conjuntos BD completa	Error absoluto medio (MAE) [°]		
	CNN-MHAT	CNN-MHAT sin id	CNN
Train	6,4419	8,1141	6,1799
Val	9,4507	11,1247	9,9941
Test	10,0573	11,4372	10,9102

En las Figuras 4.1 y 4.2 se presentan las estimaciones promedio de back-azimut y el histograma de errores de predicción al evaluar el modelo CNN-MHAT sobre el conjunto de validación, respectivamente.

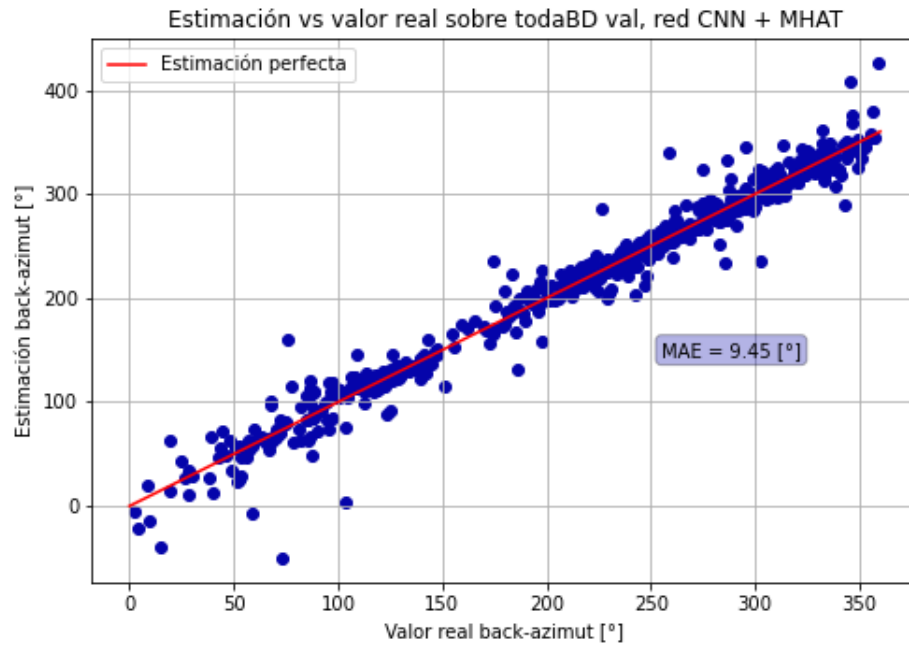


Figura 4.1: Estimación vs valor real sobre el conjunto de validación de la base de datos completa, red CNN-MHAT.

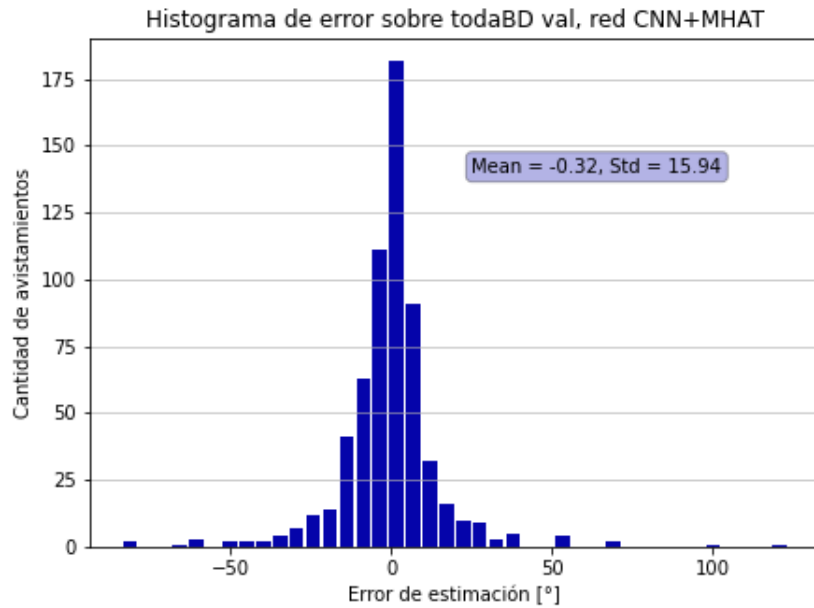


Figura 4.2: Histograma de errores sobre el conjunto de validación de la base de datos completa, red CNN-MHAT.

En las Figuras 4.3 y 4.4 se presentan las estimaciones promedio y el histograma de errores de predicción al evaluar el modelo CNN-MHAT sobre el conjunto de prueba, respectivamente.

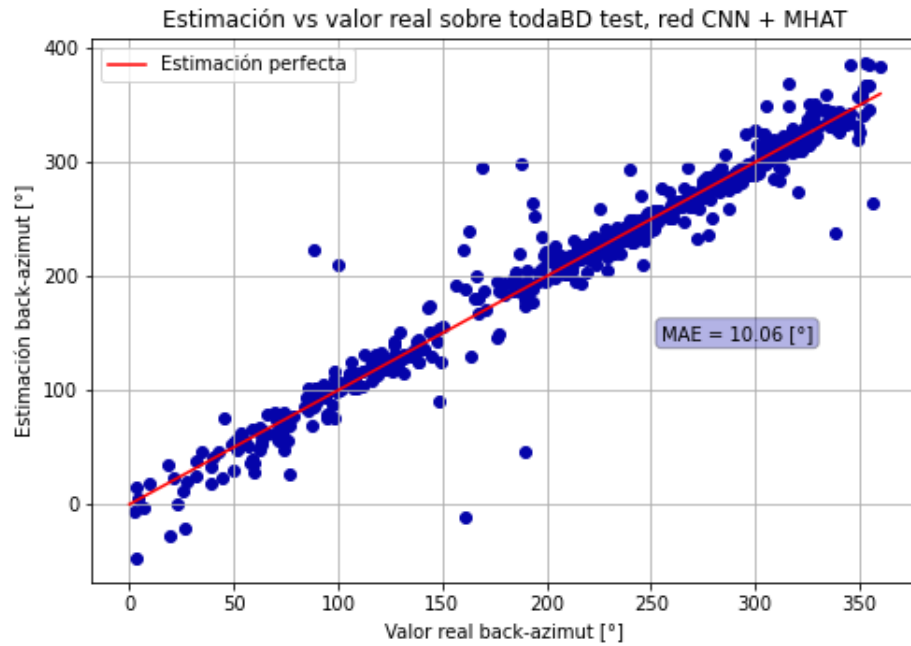


Figura 4.3: Estimación vs valor real sobre el conjunto de prueba de la base de datos completa, red CNN-MHAT.

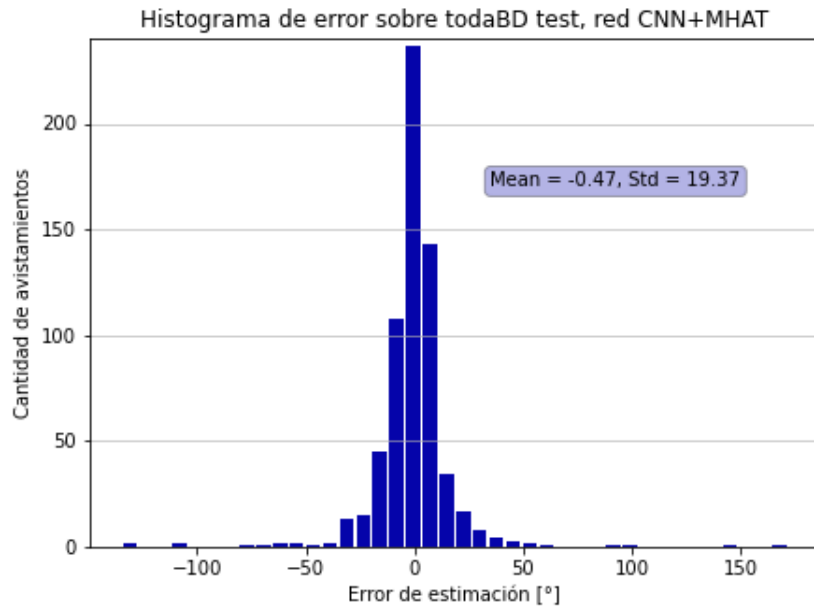


Figura 4.4: Histograma de errores sobre el conjunto de prueba de la base de datos completa, red CNN-MHAT.

Adicionalmente, a modo de comparación, se obtienen estimaciones del back-azimut al emplear el método de correlaciones cruzadas planteado en [23]. Para ello, se considera una ventana de tiempo que contempla desde el inicio de la onda P hasta 3 segundos a partir de tal instante. Los mejores resultados se obtuvieron utilizando un filtrado pasabandas entre 0.5

y 3 [Hz]. De esta manera, en la Figura 4.5 se presentan las estimaciones sobre el conjunto de validación, y en la Figura 4.6, las estimaciones sobre el conjunto de prueba.

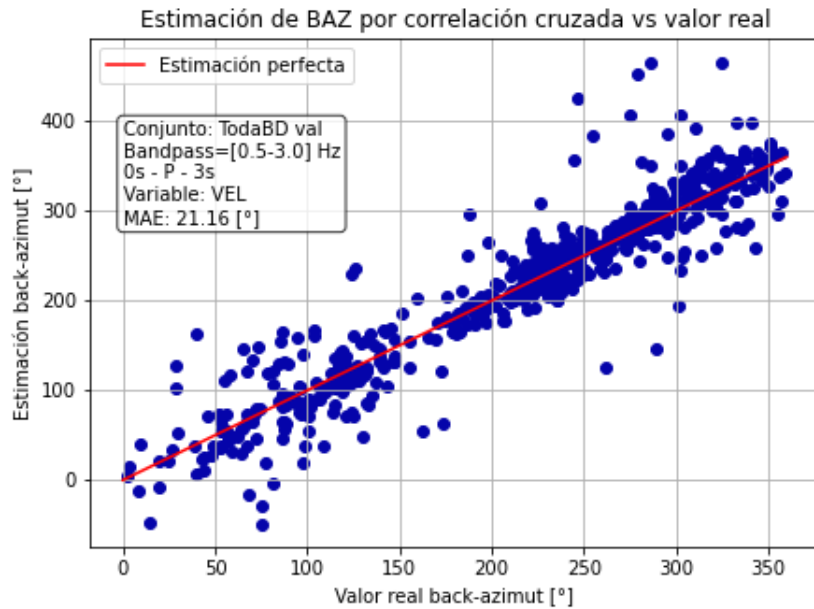


Figura 4.5: Estimación vs valor real sobre el conjunto de validación de la base de datos completa, método de correlaciones cruzadas [23].

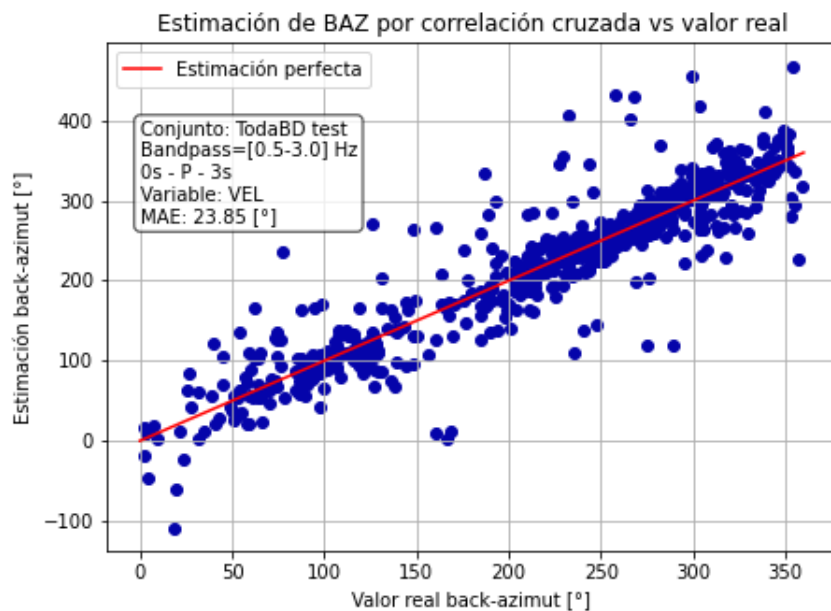


Figura 4.6: Estimación vs valor real sobre el conjunto de prueba de la base de datos completa, método de correlaciones cruzadas [23].

### 4.3. Resultados en base de datos de microsismicidad

Luego de entrenar 30 veces el modelo CNN-MHAT sobre el conjunto de entrenamiento de la base de datos de microsismicidad, en la Figura 4.7 se presentan las estimaciones promedio obtenidas al evaluar el modelo sobre el respectivo conjunto de validación. En la Figura 4.8 se presenta el correspondiente histograma de los errores de predicción.

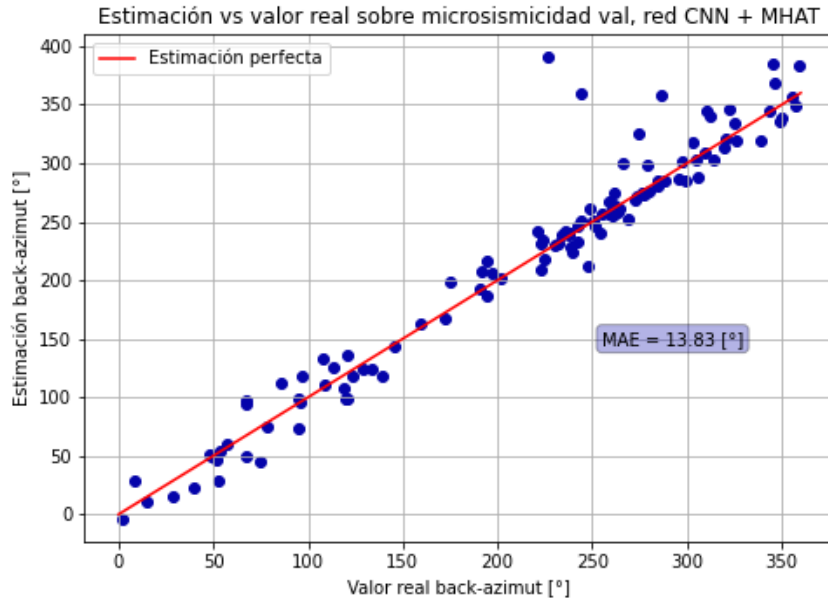


Figura 4.7: Estimación vs valor real sobre el conjunto de validación de la base de datos de microsismicidad, red CNN-MHAT.

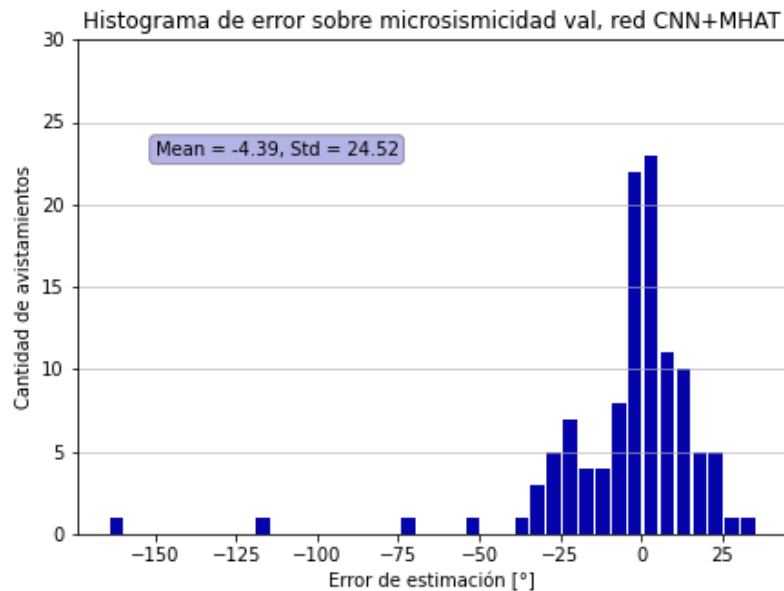


Figura 4.8: Histograma de errores sobre el conjunto de validación de la base de datos de microsismicidad, red CNN-MHAT.



En la Figura 4.9 se presentan las estimaciones promedio obtenidas al evaluar el modelo CNN-MHAT sobre el conjunto de prueba de la base de datos de microsismicidad. En la Figura 4.10 se presenta el respectivo histograma de los errores de predicción.

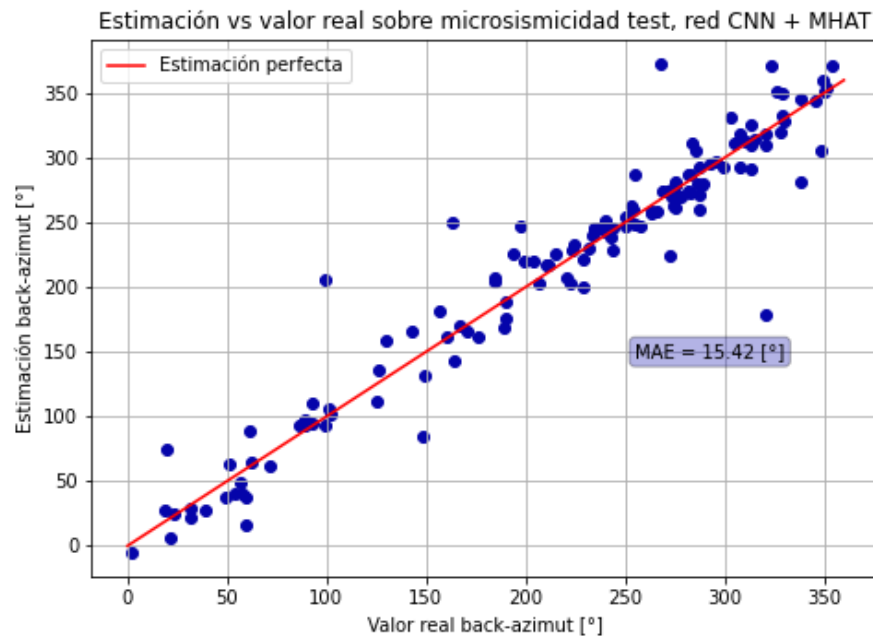


Figura 4.9: Estimación vs valor real sobre el conjunto de prueba de la base de datos de microsismicidad, red CNN-MHAT.

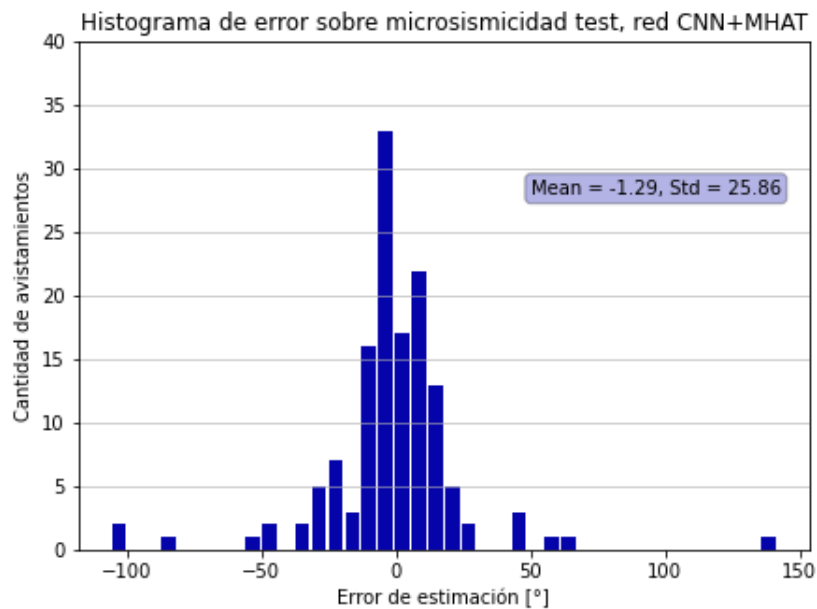


Figura 4.10: Histograma de errores sobre el conjunto de prueba de la base de datos de microsismicidad, red CNN-MHAT.

## 4.4. Resultados en base de datos de eventos con $M \geq 4$

Después de entrenar 30 veces el modelo CNN-MHAT sobre el conjunto de entrenamiento de la base de datos  $M \geq 4$ , en la Figura 4.11 se presentan las estimaciones promedio obtenidas al evaluar el modelo sobre el conjunto de validación de dicha base de datos. En la Figura 4.12 se presenta el respectivo histograma de los errores de predicción.

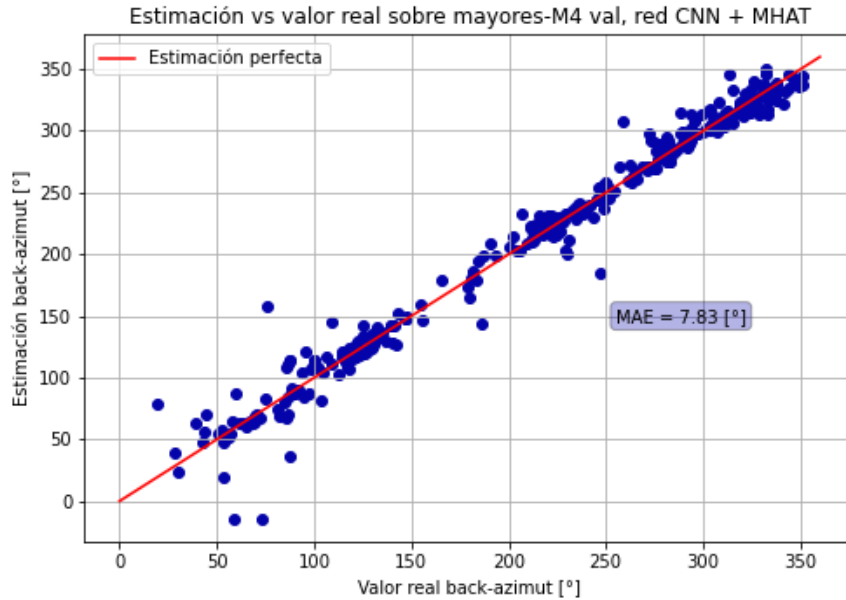


Figura 4.11: Estimación vs valor real sobre el conjunto de validación de la base de datos de eventos  $M \geq 4$ , red CNN-MHAT.

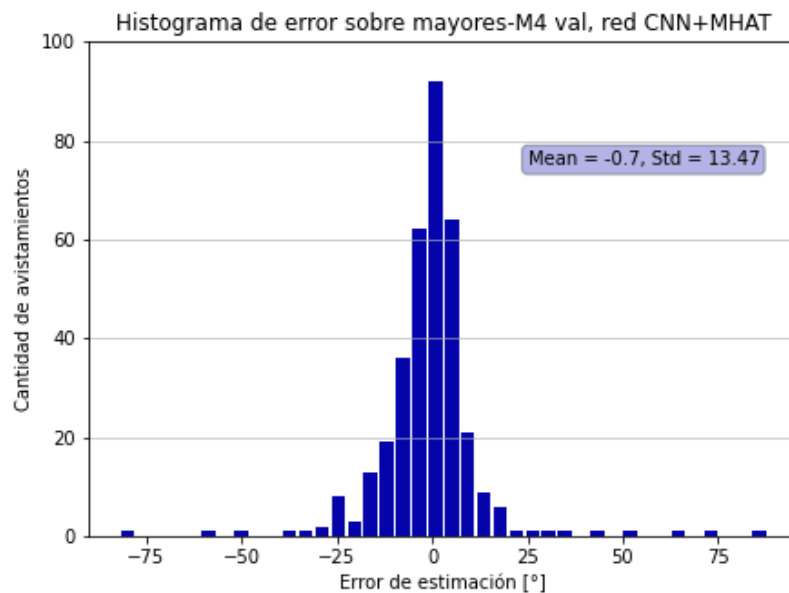


Figura 4.12: Histograma de errores sobre el conjunto de validación de la base de datos de eventos  $M \geq 4$ , red CNN-MHAT.

En la Figura 4.13 se presentan las estimaciones promedio obtenidas al evaluar el modelo CNN-MHAT sobre el conjunto de prueba de la base de datos  $M \geq 4$ . En la Figura 4.14 se presenta el respectivo histograma de los errores de predicción.

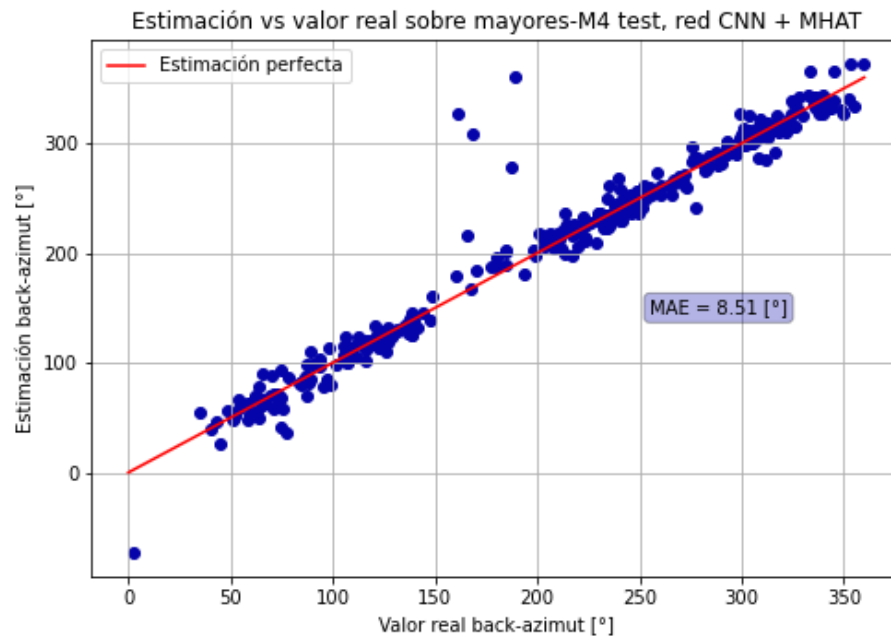


Figura 4.13: Estimación vs valor real sobre el conjunto de prueba de la base de datos de eventos  $M \geq 4$ , red CNN-MHAT.

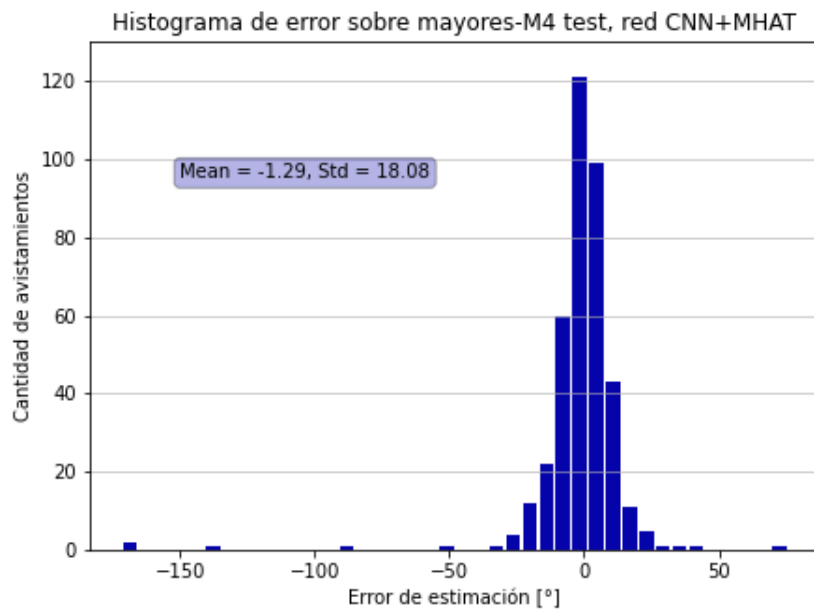


Figura 4.14: Histograma de errores sobre el conjunto de prueba de la base de datos de eventos  $M \geq 4$ , red CNN-MHAT.

## 4.5. Análisis de resultados

### 4.5.1. Comparación de modelos de red en base de datos completa

A partir de los resultados presentados en la Tabla 4.3 se observa que, frente al conjunto de entrenamiento de la base de datos completa, todos los modelos se entrenaron de manera adecuada. Se destaca que el modelo CNN-MHAT es aquel que presenta el mejor desempeño en términos de error absoluto medio, esto tanto para el conjunto de validación como para el conjunto de prueba. Esto significa que el modelo CNN-MHAT es aquel que mejor se comporta frente a avistamientos nuevos, distintos a los utilizados para su entrenamiento.

Es de notar que, si se retira la característica global ID station, el error absoluto medio se incrementa en todos los conjuntos. En particular, en el conjunto de prueba dicho error aumenta en aproximadamente  $1,38^\circ$ . En base a lo anterior, es posible afirmar que la característica ID station si aporta información valiosa para la red.

Por otro lado, al retirarse el mecanismo atencional, se observa que el error absoluto medio se incrementa tanto en el conjunto de validación como en el conjunto de prueba. En particular, en este último el error absoluto medio se incrementa en aproximadamente  $0,85^\circ$ . Debido a lo anterior, es posible afirmar que el bloque atencional es beneficioso para el funcionamiento de la red en la labor requerida.

### 4.5.2. Desempeño de red CNN-MHAT

En base a lo observado, la red CNN-MHAT es aquella que presenta el mejor rendimiento. Por tal motivo, a continuación se analiza de manera detallada los resultados obtenidos por esta red considerando los conjuntos de validación y prueba para la base de datos completa, microsismicidad y eventos con magnitud mayor o igual a 4 ( $M \geq 4$ ).

#### 4.5.2.1. Base de datos completa

En la Figura 4.1 se presenta el gráfico de estimación versus el valor real de back-azimut para los avistamientos del conjunto de validación, donde la estimación perfecta viene dada por la recta (diagonal) de color rojo. A partir de dicha figura se observa que, en general, las estimaciones son bastante cercanas a su valor real, lo cual se refleja en el error absoluto medio registrado. A partir del histograma de errores asociado al conjunto de validación, presentado en la Figura 4.2, se observa que el promedio del error es cercano a cero mientras que la desviación estándar es de solo  $15,94^\circ$ .

En la Figura 4.3 se presentan las estimaciones versus el valor real para los elementos del conjunto de prueba. Al igual que para el conjunto de validación, la mayoría de las estimaciones son cercanas a su valor real, aunque se aprecia que existe una zona entre aproximadamente los ángulos  $150^\circ$  y  $200^\circ$  donde una cantidad considerable de puntos se alejan de la diagonal (estimación perfecta). Esta zona donde el error de estimación se ve incrementado, se explica debido a que no existen demasiados avistamientos cuyo valor de back-azimut se encuentre en dicho rango, tal como se aprecia en la Figura 3.3.a, lo cual limita la cantidad de datos de entrenamiento para tal segmento. A partir de la Figura 4.4 se aprecia que la desviación estándar del error presenta un valor bajo ( $19,37^\circ$ ), aunque es superior a la dispersión observada para el conjunto de validación.

Adicionalmente, a modo de comparación, se presenta en las Figuras 4.5 y 4.6 las estimaciones obtenidas al emplear el método de estimación por correlaciones cruzadas indicado en la Ecuación 2.23. El que las frecuencias de corte óptimas del filtrado pasa-bandas fuesen 0,5 Hz y 3 Hz, entrega indicios que la información respecto a la polarización de la onda P esta mejor reflejada en bandas de baja frecuencia. Al comparar los resultados obtenidos con los del modelo de red neuronal planteado, se aprecia que el error absoluto medio aumenta en más de dos veces, esto tanto en el conjunto de validación como en el de prueba. Por ende, el modelo CNN-MHAT resulta ser muy superior en términos de desempeño respecto al método de estimación planteado en [23].

#### 4.5.2.2. Base de datos de microsismicidad

A partir de la Figura 4.7 se observa que el error absoluto medio registrado al evaluar el modelo CNN-MHAT sobre el conjunto de validación de la base de datos de microsismicidad es de  $13,83^\circ$ . Dicho error resulta superior al registrado frente al conjunto de validación de la base de datos completa. Esta diferencia puede explicarse debido a la disparidad en términos de SNR entre las bases de datos. En particular, la base de datos de microsismicidad posee un SNR promedio de 25,27 dB, mientras que en el caso de la base de datos completa este valor es de 38,34 dB.

Adicionalmente, el conjunto de entrenamiento de la base de datos de microsismicidad posee una cantidad considerablemente inferior de avistamientos respecto a la base de datos completa. Ello se traduce en que la red posee una menor cantidad de información de la cual aprender. Al observar el histograma de errores de la Figura 4.8, se aprecia que la desviación estándar, con un valor de  $24,52^\circ$ , es mayor respecto a la reportada para el conjunto de validación de la base de datos completa. Ello se condice con el mayor error absoluto medio obtenido, dando cuenta que existe un mayor grado de dispersión en las estimaciones (respecto a la diagonal, la estimación perfecta).

Por otra parte, a partir de la Figura 4.3 se aprecia que al evaluar el modelo sobre el conjunto de prueba el error absoluto medio es de  $15,42^\circ$ . Tal nivel de error da cuenta que al modelo le resulta de mayor dificultad el realizar estimaciones precisas frente a señales referidas a microsismicidad. Lo anterior también se ve reflejado en el valor de la desviación estándar de los errores de estimación ( $25,86^\circ$ ), los cuales se presentan en la Figura 4.10. Nuevamente, el aumento en el error observado puede atribuirse al menor SNR con que cuenta la base de datos de microsismicidad.

#### 4.5.2.3. Base de datos $M \geq 4$

A partir de la Figura 4.11 se aprecian las estimaciones promedio registradas sobre el conjunto de validación. Al comparar dichas estimaciones con su valor real, se obtiene un error absoluto medio de  $7,83^\circ$ . Al examinar el respectivo histograma de errores, el cual se presenta en la Figura 4.12, se observa una desviación estándar de 13,47. Esto refleja que la red se desempeña de manera exitosa sobre el conjunto de validación, considerando solo registros sísmicos asociados a eventos con magnitud mayor o igual a 4.

De la Figura 4.13 se aprecia que las estimaciones promedio obtenidas sobre el conjunto de prueba para la base de datos  $M \geq 4$  coinciden en gran medida con sus respectivos valores reales, lo cual se plasma en un error absoluto medio de  $8,51^\circ$ . Adicionalmente, la Figura 4.14

da cuenta que los errores resultan ser bajos en su gran mayoría. De hecho, un 75% de las predicciones registran un error absoluto que es menor a  $10^\circ$ , y sobre un 93% presenta un error absoluto menor a  $20^\circ$ , mientras que la desviación estándar del error de predicción es de  $18,08^\circ$ .

En base a lo visto, se destaca que el modelo CNN-MHAT presenta el mejor desempeño sobre la base de datos de eventos con magnitud  $M \geq 4$ . Este buen rendimiento se explica debido a que las señales cuentan con un alto valor de SNR, siendo en promedio igual a 42,90 dB. El bajo error observado, sobre todo en el conjunto de prueba, da cuenta que el modelo planteado podría ser de gran utilidad en el contexto de sistemas de alerta temprana, donde sismos de gran magnitud pueden representar un riesgo importante de generación de tsunamis.

## 4.6. Fortalezas y debilidades del modelo CNN-MHAT

A partir de los resultados obtenidos, se observa que el mejor desempeño del modelo de red CNN-MHAT se presenta cuando se usan señales provenientes de eventos cuya magnitud es mayor o igual a 4. Son este tipo de eventos los que revisten el mayor peligro para la generación de maremotos. Debido a lo anterior, una fortaleza del modelo planteado es que se comporta mejor en aquellos casos que pueden catalogarse como de mayor relevancia inmediata.

En adición a lo anterior, es de destacar que el modelo de red solo requiere de 3 segundos a partir del arribo de la onda P (y 1 segundo antes). Lo anterior se traduce en una gran reducción en el tiempo requerido para conocer la dirección de origen de un evento sísmico. A través de esto, es posible saber si el epicentro está en dirección a la costa, lo cual es un factor clave en materia de riesgo de tsunamis. De esta manera, el modelo desarrollado podría ser utilizado en el contexto de un sistema de alerta temprana.

Se destaca también que el modelo puede ser utilizado para identificar la localización (en términos de dirección) de sismos con magnitud pequeña y distancia cercana, lo cual es mencionado en este trabajo como microsismicidad. Sin embargo, debido a la dependencia con el SNR de las señales, se observa que esta labor reporta resultados con un grado de exactitud menor.

Una limitación del modelo desarrollado es que solamente puede ser empleado sobre avisamientos asociados a alguna de las 49 estaciones consideradas en el trabajo actual. Ello se debe a que el identificador ID station implementado solamente puede referirse a alguna de estas estaciones. Para superar esta limitante, el modelo puede ser re-entrenado considerando nuevos avisamientos pertenecientes a otras estaciones. Adicionalmente, es posible emplear una versión de la red que no haga uso del identificador ID station, lo cual disminuye el desempeño de la red pero no de manera crítica.

Una debilidad que debe ser considerada es que el modelo requiere a priori que se haya realizado un marcado de la llegada de la onda P, pues esto no está integrado en el modelo planteado. En el caso de este trabajo, dicho marcado se realizó de manera manual, sin embargo, al integrar el modelo en un módulo automatizado, puede suceder que el marcado de la llegada de la onda P no sea siempre acertado.

Otra debilidad a considerar es que el modelo requiere de la información registrada en los tres canales (Z, N y E). Si por algún motivo, existe un problema con alguno de los canales de la estación considerada, entonces es esperable que el modelo no se desempeñe de manera correcta. Más aún, si alguno de los canales no está disponible, entonces no es posible emplear el método propuesto.

# Capítulo 5

## Conclusión

A partir del trabajo realizado y los resultados obtenidos, se destaca que se pudo desarrollar una arquitectura de red capaz de estimar el ángulo back-azimut con un bajo error absoluto en la mayoría de los casos. En base a lo anterior, se concluye que el objetivo principal planteado en este trabajo se pudo cumplir de manera satisfactoria.

Respecto a la arquitectura de red planteada, se concluye que tanto el añadir información respecto a la estación de registro (ID station) como el uso del mecanismo de auto-atención Multi-Head Attention, permiten una mejora en el desempeño de la red. Adicionalmente, los resultados obtenidos dan cuenta que el problema de estimación de back-azimut es dependiente del SNR de las señales.

De esta manera, el modelo desarrollado se desempeña con mayor éxito sobre aquellos eventos que presentan una magnitud considerable, los cuales pueden ser considerados de mayor riesgo de generación de tsunamis. Ello, junto al hecho de que el modelo solo requiere de tres segundos a partir de la llegada de la onda P, son grandes fortalezas a considerar.

### 5.1. Trabajo futuro

Es importante mencionar que los resultados obtenidos se encuentran enmarcados en un contexto en que los datos utilizados para entrenamiento y evaluación son limitados. Como todo modelo de *Deep Learning*, el contar con una mayor cantidad y variedad de datos de entrenamiento es recomendable para aumentar su generalización. De esta manera, se propone como trabajo a futuro el aumentar el tamaño de la base de datos.

Una ventaja del enfoque correspondiente a estimar la localización de un evento sísmico a partir de una sola estación sismológica, es que un solo evento puede ser registrado por múltiples estaciones, lo cual se traduce en una cantidad similar de avistamientos que pueden ser empleados para entrenar o evaluar la red. Esto puede ser aprovechado para incrementar la base de datos actual.

Adicionalmente, no solo se pueden descargar avistamientos de la zona norte, en las 49 estaciones iniciales consideradas, sino que nuevas estaciones pueden ser utilizadas. Es posible incluso añadir avistamientos sísmicos referidos a la zona sur del país, lo cual también implica el uso de nuevas estaciones sismológicas.



Otra manera de aumentar la cantidad de datos disponibles es el uso de técnicas de aumentación de datos. Para el caso de las señales sísmicas, son diversas las técnicas que han sido planteadas, como lo puede ser por ejemplo el superponer ruido a las trazas [39]. A través de ello es posible no solo aumentar la cantidad de datos para los rangos de ángulos con menor representación, sino que también se podría mejorar el desempeño de la red en condiciones de menor valor SNR.

Otro aspecto relevante a tratar a futuro es el evaluar la integración del modelo con métodos de estimación automáticos de la onda P. Esto es un aspecto sumamente relevante, pues es probable que en un escenario realista, el picado de la onda P de la mayor parte de los avistamientos se realice usando métodos de estimación automáticos, como lo puede ser el algoritmo STA-LTA<sup>16</sup> u otros.

---

<sup>16</sup> Este algoritmo compara el promedio de las amplitudes en una ventana antes y otra ventana después del instante en cuestión.

# Bibliografía

- [1] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning”, *Nature*, vol. 521, no. 7553, pp. 436-444, 2015. Available: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [2] K. Hornik, M. Stinchcombe and H. White, “Multilayer feedforward networks are universal approximators”, *Neural Networks*, vol. 2, no. 5, pp. 359-366, 1989. Available: [10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- [3] D. Zhou, “Universality of deep convolutional neural networks”, *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 787-794, 2020. Available: [10.1016/j.acha.2019.06.004](https://doi.org/10.1016/j.acha.2019.06.004).
- [4] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory”, *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. Available: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [5] G. Chevalier, “LARNN: Linear Attention Recurrent Neural Network”. *arXiv*, 2018. doi: [10.48550/ARXIV.1808.05578](https://doi.org/10.48550/ARXIV.1808.05578).
- [6] A. SCHÄFER and H. ZIMMERMANN, “RECURRENT NEURAL NETWORKS ARE UNIVERSAL APPROXIMATORS”, *International Journal of Neural Systems*, vol. 17, no. 04, pp. 253-263, 2007. Available: [10.1142/s0129065707001111](https://doi.org/10.1142/s0129065707001111).
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need”, in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [8] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer Normalization”. *arXiv*, 2016. doi: [10.48550/ARXIV.1607.06450](https://doi.org/10.48550/ARXIV.1607.06450).
- [9] I. Sucholutsky, A. Narayan, M. Schonlau, and S. Fischmeister, “Pay attention and you won’t lose it: A deep learning approach to sequence imputation,” *PeerJ Computer Science*, vol. 5, 2019.
- [10] Q. Wen et al., “Transformers in Time Series: A Survey”. *arXiv*, 2022. doi: [10.48550/ARXIV.2202.07125](https://doi.org/10.48550/ARXIV.2202.07125).
- [11] Y. Feng and Y. Cheng, “Short text sentiment analysis based on multi-channel CNN with multi-head attention mechanism,” *IEEE Access*, vol. 9, pp. 19854–19863, 2021.
- [12] R. A. Hamad, M. Kimura, L. Yang, W. L. Woo, and B. Wei, “Dilated causal convolution with multi-head self attention for sensor human activity recognition,” *Neural Computing and Applications*, vol. 33, no. 20, pp. 13705–13722, 2021.
- [13] “SISMOS - ONEMI: Ministerio del interior y seguridad pública -,” ONEMI. [Online]. Available: <https://www.onemi.gov.cl/terremoto/>.
- [14] M. Saez “Correlación cruzada de ruido sísmico para la obtención de perfiles profun-

- dos de velocidad de onda de corte en la cuenca de Santiago”, MSc thesis, Department of Civil Engineering, University of Chile, Santiago de Chile, Chile, 2016, doi: 10.13140/RG.2.2.22808.24325.
- [15] R. Hohensinn, “Detection of Hazardous Ground Movements with Instantaneous Velocity Estimates by GNSS,” phdthesis, 2019. doi: 10.3929/ethz-b-000348594.
- [16] T. Vincenty, “Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations,” *Survey Review*, vol. 23, no. 176, pp. 88–93, 1975.
- [17] J. Montalbetti and E. Kanasewich, “Enhancement of Teleseismic Body Phases with a Polarization Filter”, *Geophysical Journal International*, vol. 21, no. 2, pp. 119-129, 1970. Available: 10.1111/j.1365-246x.1970.tb01771.x.
- [18] E. Flinn, “Signal analysis using rectilinearity and direction of particle motion”, *Proceedings of the IEEE*, vol. 53, no. 12, pp. 1874-1876, 1965. Available: 10.1109/proc.1965.4462
- [19] C. Frohlich and J. Pulliam, “Single-station location of seismic events: A review and a plea for more research,” *Physics of the Earth and Planetary Interiors*, vol. 113, no. 1-4, pp. 277–291, 1999.
- [20] N. Magotra, N. Ahmed and E. Chael, “Seismic event detection and source location using single-station (three-component) data”, *Bulletin of the Seismological Society of America*, vol. 77, no. 3, pp. 958-971, 1987. Available: 10.1785/bssa0770030958.
- [21] N. Magotra, N. Ahmed and E. Chael, “Single-station seismic event detection and location”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 27, no. 1, pp. 15-23, 1989. Available: 10.1109/36.20270.
- [22] Noda, S. Yamamoto, S. Sato, N. Iwata, M. Korenaga and K. Ashiya, “Improvement of back-azimuth estimation in real-time by using a single station record”, *Earth, Planets and Space*, vol. 64, no. 3, pp. 305-308, 2012. Available: 10.5047/eps.2011.10.005.
- [23] R. Roberts, A. Christoffersson and F. Cassidy, “Real-Time Event Detection, Phase Identification and Source Location Estimation Using Single Station Three-Component Seismic Data”, *Geophysical Journal International*, vol. 97, no. 3, pp. 471-480, 1989. Available: 10.1111/j.1365-246x.1989.tb00517.x.
- [24] A. Eisermann, A. Ziv and G. Wust-Bloch, “Real-Time Back Azimuth for Earthquake Early Warning”, *Bulletin of the Seismological Society of America*, vol. 105, no. 4, pp. 2274-2285, 2015. Available: 10.1785/0120140298.
- [25] L. Ochoa Gutierrez, C. Vargas Jiménez and L. Niño Vásquez, “Fast estimation of earthquake arrival azimuth using a single seismological station and machine learning techniques”, *Earth Sciences Research Journal*, vol. 23, no. 2, pp. 103-109, 2019. Available: 10.15446/esrj.v23n2.70581.
- [26] S. Chanda and S. Somala, “Single-Component/Single-Station–Based Machine Learning for Estimating Magnitude and Location of an Earthquake: A Support Vector Machine Approach”, *Pure and Applied Geophysics*, vol. 178, no. 6, pp. 1959-1976, 2021. Available: 10.1007/s00024-021-02745-8.
- [27] T. Perol, M. Gharbi and M. Denolle, “Convolutional neural network for earthquake detection and location”, *Science Advances*, vol. 4, no. 2, 2018. Available: 10.1126/sciadv.1700578.
- [28] T. Bai and P. Tahmasebi, “Attention-based LSTM-FCN for earthquake detection and

- location”, *Geophysical Journal International*, vol. 228, no. 3, pp. 1568-1576, 2021. Available: 10.1093/gji/ggab401.
- [29] D. Bahdanau, K. Cho, Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate.” (9 2014). URL <http://arxiv.org/abs/1409.0473>
- [30] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks.” *arXiv*, 2017. doi: 10.48550/ARXIV.1709.01507.
- [31] A. Lomax, A. Micheline and D. Jozinović, “An Investigation of Rapid Earthquake Characterization Using Single-Station Waveforms and a Convolutional Neural Network”, *Seismological Research Letters*, vol. 90, no. 2, pp. 517-529, 2019. Available: 10.1785/0220180311.
- [32] S. Mousavi and G. Beroza, “Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 8211-8224, 2020. Available: 10.1109/tgrs.2020.2988770.
- [33] S. M. Mousavi, Y. Sheng, W. Zhu, and G. Beroza, “STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI,” *IEEE Access*. [Online]. Available: 10.1109/ACCESS.2019.2947848
- [34] J. Münchmeyer, D. Bindi, U. Leser and F. Tilmann, “Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network”, *Geophysical Journal International*, vol. 226, no. 2, pp. 1086-1104, 2021. Available: 10.1093/gji/ggab139.
- [35] A. Cofre et al., “End-to-End LSTM-Based Earthquake Magnitude Estimation With a Single Station”, *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022. Available: 10.1109/lgrs.2022.3175108.
- [36] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting.,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [37] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”. *arXiv*, 2014. doi: 10.48550/ARXIV.1412.6980.
- [38] I. Kandel and M. Castelli, “The effect of batch size on the generalizability of the Convolutional Neural Networks on a histopathology dataset,” *ICT Express*, vol. 6, no. 4, pp. 312–315, 2020.
- [39] W. Zhu, S. M. Mousavi, and G. C. Beroza, “Seismic signal augmentation to improve generalization of deep neural networks,” *Machine Learning in Geosciences*, pp. 151–177, 2020.