



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **EXTENSIÓN TEMPORAL DE HINDCAST DE OLEAJE UTILIZANDO MACHINE LEARNING O DEEP LEARNING**

**JAVIER IGNACIO HURTADO RAMÍREZ**

Trabajo de Título para optar al título de Ingeniero Civil de  
Industrias, Diploma en Ingeniería Hidráulica

Profesor Guía:  
**RODRIGO CIENFUEGOS CARRASCO**

Santiago de Chile, 2021



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA  
Departamento de Ingeniería Hidráulica y Ambiental

---

# **EXTENSIÓN TEMPORAL DE HINDCAST DE OLEAJE UTILIZANDO MACHINE LEARNING O DEEP LEARNING**

**JAVIER IGNACIO HURTADO RAMÍREZ**

Trabajo de Título presentada a la Comisión integrada por los profesores:

**RODRIGO CIENFUEGOS CARRASCO**

**HECTOR IVÁN JOAQUÍN JORQUERA GONZÁLEZ**

**BENJAMÍN CARRIÓN ARETXABALA**

Trabajo de Título para optar al título de Ingeniero Civil de Industrias,  
Diploma en Ingeniería Hidráulica

Santiago de Chile, 2021

(A mis Padres, hermanos y amigos,  
que me apoyaron.)

## **AGRADECIMIENTOS**

En primer lugar agradecer a mi familia por ayudarme a mantener un ambiente laboral óptimo durante el trabajo remoto desde la casa. Agradezco a Tomás Cuevas, Benjamín Carrión y Rodrigo Cienfuegos quienes han estado muy presentes y siempre disponibles para resolver todas mis dudas. Finalmente agradezco a José Del Piano por toda la ayuda en materia de redes neuronales.

## INDICE GENERAL

Pág.

DEDICATORIA .....	ii
AGRADECIMIENTOS .....	iii
INDICE DE TABLAS .....	1
INDICE DE FIGURAS .....	2
RESUMEN .....	4
ABSTRACT .....	5
I. INTRODUCCIÓN .....	6
1.1 Contextualización del problema .....	6
1.2 Objetivos de la empresa .....	6
1.3 Antecedentes del problema .....	7
II. MARCO TEÓRICO .....	7
2.1 Clima de oleaje .....	7
2.1.1 Caracterización del clima de oleaje .....	7
2.2 Regresor de machine learning .....	8
2.3 Redes de aprendizaje profundo .....	9
2.4 Indicadores de desempeño .....	13
III. METODOLOGÍA .....	14
3.1 Obtención y levantamiento de datos de oleaje y viento.....	14
3.1.1 Datos de oleaje .....	14
3.1.2 Datos de viento .....	15
3.2 Estudio de zonas de generación de oleaje y tiempos de respuesta .....	15
3.3 Implementación de modelos de <i>machine learning</i> y <i>deep learning</i> .....	16
3.3.1 Definición del modelo con mejores estimaciones del espectro.....	16
IV. RESULTADOS .....	18

4.1	Zonas de generación de oleaje .....	18
4.2	Modelos de <i>machine learning</i> .....	18
4.3	Desarrollo de modelo regresor Machine Learning .....	21
4.4	Desarrollo de modelo LSTM, Deep learning.....	22
V.	CONCLUSIONES .....	30
	 BIBLIOGRAFIA .....	31
	 A N E X O S .....	34
	Anexo 1 : Formación de oleaje a partir del viento.....	35
	Anexo 2 : Obtención de parámetros de oleaje a partir de un espectro direccional de energía de oleaje.....	37
	Anexo 3 : Estructura de aprendizaje de <i>machine learning</i> .....	40
	Anexo 4 : Zonas de generación de oleaje .....	42
	Anexo 5 : Desempeño del modelo LSTM - FCL frente alternativas existentes .....	44
	Anexo 6 : Desempeño del modelo LSTM - FCL frente distintas localidades .....	48
	Anexo 7 : Desempeño del modelo LSTM - FCL frente distintas estaciones del año.	50
	Anexo 8 : Desempeño del modelo LSTM - FCL para pronosticar espectros por componente de generación .....	51
	Anexo 9 : Series de tiempo y gráficos de dispersión .....	53
	Anexo 10 : Estimación de espectro de oleaje completo usando redes LSTM.....	55

## **INDICE DE TABLAS**

Pág.

Tabla 4.1: Comparación de resultado de los distintos modelos de estimación de periodo  
*peak* de oleaje ..... 21

Tabla 4.2: Calidad de ajuste de los parámetros de oleaje de regresor Ridge..... 22

## INDICE DE FIGURAS

Pág.

Figura 2.1: Espectro de oleaje completo zona de Valparaíso de Chile.....	8
Figura 2.2: Etapas de una red convolucional.....	10
Figura 3.1: Presión al nivel del mar ( <i>SLP</i> ) sin filtro (a) y con filtro (b). .....	15
Figura 3.2: Esquema de desiciones para mejorar el modelo.....	16
Figura 3.3: Diagrama de modelos a entrenar .....	17
Figura 4.1: Correlaciones entre SLP y oleaje en Valparaíso .....	18
Figura 4.2: Correlaciones entre variables de viento y oleaje en Valparaíso.....	19
Figura 4.3: Ajuste del modelo Ridge altura significativa para distintas memorias, resolución de mapa de presión y ajuste periodo pico para distintos $\alpha$ .....	20
Figura 4.4: Esquema red LSTM <i>many to many</i> por cada sub-serie de muestra de largo <i>timestep</i> .....	23
Figura 4.5: Esquema red LSTM con auto-regresión <i>many to many</i> por cada sub-serie de muestra de largo <i>timestep</i> precedido por un <i>traslape</i> .....	24
Figura 4.6: Esquema red LSTM <i>many to one</i> por cada sub-serie de largo <i>timestep</i> ..	24
Figura 4.7: Resultados arquitecturas LSTM.....	25
Figura 4.8: Ajuste de hiperparámetros del modelo LSTM .....	25
Figura 4.9: Resultados de regresor y LSTM con y sin la red convolucional.....	26
Figura 4.10: Comparación de resultados Lago Erie (H. Haoguo et al, 2021). .....	27
Figura 4.11: Ubicación espacial de los puntos con registro de oleaje disponibles en la costa de Chile, zona central.....	28

Figura 4.12: Serie de tiempo de oleaje, gráfico dispersión y Q - Q (rojo) del modelo LSTM  
– FCL *many to one* para parámetros resumen de oleaje usando dos puntos de oleajes cercanos y seas locales ..... 29

## RESUMEN

Para el diseño de proyectos de ingeniería de costas, es necesario contar con registros de oleaje hindcast (predicciones retrospectivas) que representen el estado del mar. Los registros en la costa de Chile dejaron de ser entregados por la NOAA (Oficina Nacional de Administración Oceánica) desde el año 2019, por lo que es necesario realizar modelamiento físico del océano Pacífico para obtener los registros de oleaje hindcast faltantes. En este proyecto se busca completar los registros de oleaje faltantes por medio de herramientas de inteligencia artificial. Para esto se prueban distintas arquitecturas exitosas en modelar oleaje en la costa de California ([Pirhooshyaran, M. and Lawrence, S., 2019](#)) y el Lago Erie ([Hu, Haoguo, 2021](#)) por medio de regresores con memoria de largo y corto plazo (LSTM).

Se prueban distintas arquitecturas de redes neuronales, donde se obtienen los mejores resultados por medio de la estructura *many-to-one*, la cual considera un entrenamiento de una semana utilizando la presión al nivel del mar en la zona sur del Pacífico para el pronóstico de cada registro de oleaje. Estos resultados mejoran levemente al incorporar redes convolucionales pre-entrenadas, debido a que obtiene características implícitas del mapa de presión. Se obtienen pronósticos 80% de correlación para la altura significativa y dirección de oleaje y 50% de correlación para el periodo pico, presentando un ajuste más pobre al logrado en el Lago Erie.

La principal causa de las diferencias entre los resultados de altura y dirección de oleaje frente al periodo, responden a la baja correlación entre el periodo y las variables de presión al nivel del mar y la alta auto-regresión que este tiene, demandando mayor capacidad de entrenamiento con memoria. Los resultados del periodo de oleaje también son inferiores a los logrados en el caso del Lago Erie, puesto que el oleaje pronosticado se encuentran en un sistema aislado, donde el periodo generado responde únicamente a las variables de viento presentes en el input del modelo. Este problema se soluciona al usar como variable predictor el oleaje en puntos cercanos, presentando de esta manera una correlación del 85% y 66% para la altura significativa y periodo pico de oleaje al usar el modelo LSTM.

## ABSTRACT

For the design of coastal engineering projects, it is necessary to have hindcast wave records that represent the sea. The records on the coast of Chile were no longer delivered by NOAA (National Oceanic and Atmospheric Administration) since 2019, so it is necessary to perform physical modeling of the Pacific Ocean to obtain the missing hindcast wave records. This project attempts to complete the missing hindcast wave records by using artificial intelligence tools. For this, different successful architectures are tested in modeling swell off the California coast ([Pirhooshyaran, M. and Lawrence, S., 2019](#)) and Lake Erie ([Hu, Haoguo, 2021](#)) by regressors with layers of long and short term memory (LSTM).

Different neural network architectures are tested, where the best results are obtained by the many-to-one structure, which considers a one-week training using sea level pressure in the South Pacific for forecasting each wave record for 10 years. These results are slightly improved by incorporating pre-trained convolutional networks, because it obtains spatial characteristics of the pressure map. Forecasts are 80% correlated for significant wave height and direction and 50% correlated for the peak period, presenting a poorer fit than that achieved for Lake Erie case.

The main cause of the differences between the results of wave height and direction versus period, respond to the low correlation between the period and sea level pressure parameters, the high auto-regression that this has, demanding greater training capacity with memory. The wave period results are also lower than those achieved in the case of Lake Erie, since the predicted waves are in an isolated system, where the generated period responds only to the wind variables that were given to the model. This problem is solved by using as a predictor variable the swell at nearby points, thus presenting a correlation of 85% and 66% for the significant height and peak period.

## I. INTRODUCCIÓN

### 1.1 Contextualización del problema

Es fundamental contar con hindcast (predicciones retrospectivas) de oleaje en el diseño de proyectos de ingeniería de costas. Se necesita principalmente el oleaje en aguas profundas como una condición de borde, el cual sirve para correr modelos hacia aguas poco profundas o someras. Típicamente corresponden a hindcast (cálculo matemático con registros del pasado) de oleaje distribuidos por el *National Centers of Environment Prediction (NCEP)*. Sin embargo, este dejó de actualizar los hindcast del espectro de oleaje en junio del 2019, por lo que se necesita modelar físicamente el océano para completar los registros faltantes.

### 1.2 Objetivos del trabajo

Determinar la factibilidad de llenar el oleaje en aguas profundas en la costa de Chile faltante a partir de campos de vientos resumidos en la presión al nivel del mar, mediante un regresor con memoria correspondiente a un modelo LSTM. El cual ha sido probado con éxito en la costa de California ([Pirhooshyaran, M. and Lawrence, S., 2019](#)) y el Lago Erie ([Hu, Haoguo, 2021](#)).

Finalmente, se busca comprender y aplicar principios de cantidad de movimiento y conservación de la energía, integrar principios físicos en el desarrollo de modelos conceptuales y cuantitativos, y diagnosticar y predecir el comportamiento de sistemas hidráulicos.

### 1.3 Antecedentes del problema

Es necesario realizar un estudio de las zonas de generación de oleaje más relevantes para cada parámetro de oleaje a pronosticar. Para esto es importante considerar tanto la existencia de oleaje local o *seas*, y también el llamado mar de fondo o *swell*, generado por tormentas ubicadas a miles de kilómetros del sitio de estudio. También es necesario estudiar los tiempos de llegada de los *swell*, e incluirlos dentro de una “memoria” para que el regresor y la red con memoria pueda considerarlos en el entrenamiento y la estimación. Finalmente, se diseñan distintas arquitecturas de redes recurrentes LSTM y se evalúa la incorporación de redes convolucionales para la abstracción de características. Este modelo debe ser validado contra casos de éxito para corroborar el buen funcionamiento del modelo.

## II. MARCO TEÓRICO

### 2.1 Clima de oleaje

La principal fuente de generación de oleaje es el intercambio de energía entre el viento y el mar producto de la fuerza de presión y de roce ([Anexo 1](#)). Principalmente la generación de oleaje responde a tres factores; la velocidad media del viento, la duración del tiempo en que actúa y la superficie sobre la que ha soplado (esta distancia de aplicación se llama *fetch*). Este intercambio de energía entre el viento y el mar puede darse de manera completa, generando un estado de mar desarrollado cuando el *fetch* es lo suficientemente largo.

#### 2.1.1 Caracterización del clima de oleaje

El oleaje se describe por distintos parámetros básicos que envuelven las principales características que tiene un clima de oleaje compuesto por una representación estadística.

##### a) Parámetros de ola

Los siguientes parámetros describen una ola como un modelo de onda simple ([Kamphuis, 2010](#)). En el [Anexo 2](#) se encuentra la información de base matemática para calcular los parámetros resumen del oleaje correspondiente a la representación estadística del estado de mar.

###### i) H, Altura de ola

Representa el tamaño de la ola y corresponde a la distancia entre la cresta y el valle.

###### ii) L, Longitud de onda

Es la distancia horizontal entre dos crestas de olas o valles sucesivas.

###### iii) T, Periodo

El tiempo entre dos ondas consecutivas.

###### iv) $\theta$ , Dirección de propagación

La dirección de propagación de una ola.

### b) Caracterización de un estado de oleaje

El estado de mar se puede caracterizar por medio de parámetros resumen del oleaje ([Anexo 2](#)). Una alternativa más completa para representar el clima de oleaje es el espectro direccional de energía. Este hindcast de oleaje se genera con un modelo matemático y presenta la densidad de energía del oleaje para cada periodo en una dirección de propagación como se presenta en el espectro de la localidad de Valparaíso de Chile en la Figura 2.1.

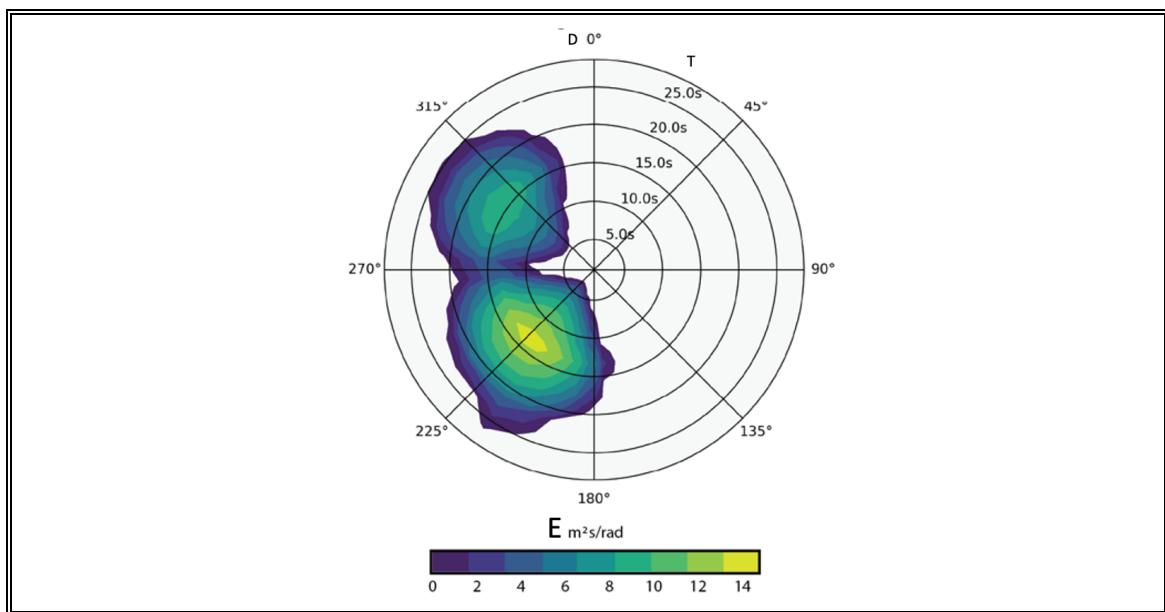


Figura 2.1: Espectro de oleaje completo zona de Valparaíso de Chile.

El espectro direccional de energía abarca toda la información que contiene el oleaje. Por otro lado, es posible resumir el espectro mediante parámetros estadísticos.

## 2.2 Regresores de *machine learning*

Como alternativa a los modelos numéricos a base física para estimar oleaje, se busca implementar regresores de *machine learning* y *deep learning* para

obtener un espectro direccional de energía del oleaje. Como primer acercamiento buscará un modelo que estime los parámetros resumen del espectro.

Las técnicas de *machine learning* permiten predicciones prácticas y eficientes por medio de regresores o clasificadores. Estas técnicas también conocidas como aprendizaje automático, son capaces de aprender por sí solas al incorporar su propia experiencia combinando datos de entrada y salidas. Existen dos subgrupos importantes; el aprendizaje supervisado y el no supervisado. En el aprendizaje supervisado, la atención se centra en un objeto esperado o valor buscado, mientras que en el no supervisado, se espera encontrar datos que describan o caractericen el objeto de entrada ([Shetty, B. 2018](#)). En el [Anexo 3](#) se encuentra mayor información de los algoritmos utilizados en *machine learning* y el funcionamiento de estos.

Por otro lado, el uso de *deep learning* es un subconjunto de la clase *machine learning* que ha sido muy exitoso para realizar tareas de manera supervisada, ubicándose en el estado de arte en muchas aplicaciones ([Valdés, L, et al. 2019](#)).

### 2.2.1 Regresor Ridge

Se selecciona el regresor Ridge como técnica de *machine learning* porque corresponde a un optimizador lineal preparado para un aprendizaje supervisado y continuo (Figura A.3 del [Anexo 3](#)). Está estructurado para minimizar la suma cuadrada de las diferencias entre el resultado y el valor real por medio del entrenamiento de los pesos  $w$  como se muestra en la ecuación 2.1. El factor  $\alpha$  corresponde a un hiperparámetro que controla el nivel de estrechamiento entre datos y estimaciones.

$$\min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2 \quad (2.1)$$

Donde se tiene X, como el vector de *input* (viento), vector de coeficientes de peso  $w$ , el vector  $y$  como *output* (oleaje) y  $\alpha$  que debe ser optimizado.

### 2.2.2 Redes de aprendizaje profundo

Las *deep neural networks* están entre las herramientas más exitosas en el mundo de la inteligencia artificial, donde se han obtenido grandes resultados y avances en el pronóstico supervisado ([K. Simonyan, 2015](#)). Existen varias estructuras de redes neuronales, entre ellas se encuentran las redes convolucionales y las recurrentes. Las

redes convolucionales reciben imágenes 2D como *input*, y sirven para entrenar el mapa de viento considerando la ubicación espacial de cada magnitud. Por otro lado, las redes recurrentes corresponden a un regresor con memoria especializado para series de tiempo.

### a) Redes neuronales convolucionales CNN

Las convoluciones corresponden a operaciones matriciales que permiten obtener características a partir de una imagen, considerando la espacialidad de los datos y así poder entrenar un modelo. Es de especial interés para el pronóstico de oleaje, debido a que nos interesa la extensión superficial del viento (*fetch*), por lo que es necesario incorporar convoluciones para obtener buenos pronósticos. Ya que las redes convolucionales tienen la capacidad de encontrar relaciones en imágenes 2D y de esta forma identificar la extensión superficial de los vientos en el Pacífico. En la Figura 2.2 se muestra un proceso de etapas de una red neuronal convolucional.

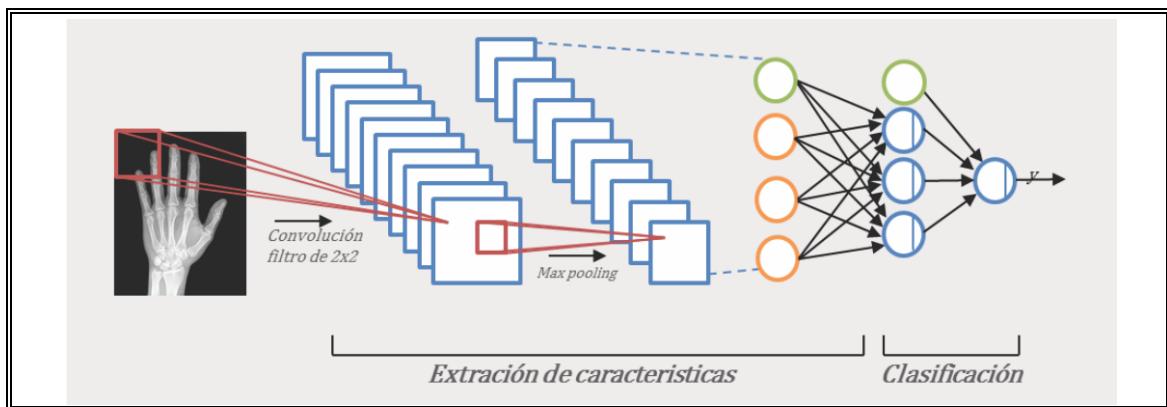


Figura 2.2: Etapas de una red convolucional. Fuente: [Arias, V. et al., 2019](#).

Dentro de las convoluciones, existen distintas alternativas. Entre ellas están las capas convolucionales que se pueden entrenar, poseen peso propio y estos son entrenados durante el aprendizaje. Las capas convolucionales *pooling* que promedian (*Average pooling*) o mantienen el máximo (*Max pooling*) de los datos en cada convolución, son útiles para disminuir la resolución de una imagen y poder

acelerar el aprendizaje. Existen diversos modelos que hoy realizan una buena clasificación de imágenes disponibles en librerías abiertas de [Keras](#). A estos se les puede extraer la capa de clasificación que se muestra en la Figura 2.2, para obtener las características del *input*.

### b) Redes neuronales recurrentes (*recurrent neural networks*)

Dentro del grupo de las redes neuronales recurrentes, es de especial interés la red LSTM (*Long short-term memory*) porque ha sido probada con éxito para estimaciones de series de oleaje usando registros de oleaje de zonas cercanas ([Pirhooshyaran, M. and Lawrence, S., 2019](#)). También han sido probadas para pronosticar hindcast de oleaje en sistemas aislados ([Haoguo, H. et al, 2021](#)). Es de especial utilidad, debido a que las redes recurrentes convencionales existe el problema llamado “*vanishing gradient*”, donde la red pierde capacidad de memoria a medida que avanza la estimación ([Hochreiter S. et al, 1997](#)). Debido a que el ajuste del gradiente “hacia atrás” en el proceso, da mayor importancia a los datos cercanos que los lejanos. Este problema es solucionado con las redes LSTM, debido a que incorpora un segundo vector oculto que memoriza la información de largo plazo ([Hochreiter S. et al, 1997](#)).

#### i) Función de pérdida y optimizador a utilizar

Se usa la función de pérdida como el error cuadrático medio (MSE) como la más conveniente para la estimación de oleaje ([Pirhooshyaran, M., et al. 2019](#)). Esta ecuación busca disminuir las distancias entre el valor estimado y el real, penalizando la distancia entre ambos al cuadrado.

$$mse = \frac{1}{M} \sum_{i=0}^M (y_{estimado} - y_{real})^2 \quad (2.2)$$

Por medio de la propagación “hacia atrás” o *backward*, se adaptan los pesos  $w$  del modelo según el error de estos en el calculo final del MSE. Esto se calcula mediante el gradiente de los pesos  $w$  por medio de un algoritmo de optimización de primero orden llamado *Stochastic gradient descent* (SGD). Este actualiza directamente los pesos  $w$  en dirección negativa (*backward*) a la secuencia de los datos por sub-lotes

$b_i$  llamados *batches*. Los pesos que definirán la importancia de cada valor del *input* durante la red neuronal son actualizados por medio de este algoritmo de optimización como se muestra en la fórmula 2.3. La medida del cambio ocurre en base al error medido por la función de pérdida definida como el *mse* (fórmula 2.4)

$$w_{i+1} = w_i - \alpha \nabla_w f_w(x) \quad (2.3)$$

$$f_w(x) = \frac{1}{M} \sum_{i=0}^M (y(x, w) - y_{real})^2 \quad (2.4)$$

Donde,  $\alpha$  es el *learning rate* fijado en 0.001 ([Pirhooshyaran, M., et al. 2019](#)),  $\nabla_w$  el gradiente de  $w$  y  $f_w(x)$  la función de pérdida definida como *mse*. Existen modificaciones del optimizador SGD, que mejoran el proceso de aprendizaje. Para la estimación de oleaje se usará el optimizador *Adaptive momentum estimation*, conocido como ADAM ([Diederik, P. et al. 2014](#)). Donde los pesos  $w$  se actualizan de manera adaptativa al tamaño, lo que produce una caída exponencial en valor de la función de perdida, en este caso el *mse*.

$$w_{i+1} = w_i - \frac{\alpha_i}{\sqrt{\hat{v}_i} + \epsilon} \hat{m}_i \quad (2.5)$$

$$\hat{m}_i = \frac{m_i}{1 - \beta_1^i} \quad (2.6)$$

$$\hat{v}_i = \frac{v_i}{1 - \beta_2^i} \quad (2.7)$$

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) \nabla_w f_w(x) \quad (2.8)$$

$$v_i = \beta_2 m_{i-1} + (1 - \beta_2) \nabla_w^2 f_w(x) \quad (2.9)$$

Donde  $\beta_1$  y  $\beta_2$  son parámetros reguladores y  $\epsilon$  sirve para reducir problemas numéricos como la división por cero. Se definen eficientemente como  $\beta_1 = 0.9$ ,  $\beta_2 = 0.99$  y  $\epsilon = 1e - 8$  ([Diederik, P. et al. 2014](#)). De esta manera se define el modelo LSTM - FCL, como una herramienta de *deep learning* para la estimación de oleaje. Existen más optimizadores que pueden ser usados para el entrenamiento y entregar distintos resultados.

## 2.3

### Indicadores de desempeño

Se usan como medidores de desempeño la correlación de los datos (fórmula 2.10) y el coeficiente de determinación (fórmula 2.11). Estas tienen sentido cuando se busca correlacionar dos variables que pueden tener distintas unidades, ya que proporciona información que ayuda a entender la proporcionalidad y el comportamiento lineal entre dos variables. Finalmente para comparar el rendimiento entre una serie de datos y su pronóstico, es conveniente usar la raíz del error cuadrático medio (fórmula 2.14). El cual muestra un error en la misma dimensión de las series que se comparan, entregando una idea de la variación de los resultados.

$$R = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2.10)$$

$$R^2 = \left( \frac{\sigma_{xy}}{\sigma_x \sigma_y} \right)^2 \quad (2.11)$$

$$\sigma_{xy} = \frac{\sum x_i y_i}{M} - \bar{x} \bar{y} \quad (2.12)$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{M} - \bar{x}^2} \quad (2.13)$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{M} - \bar{y}^2} \quad (2.14)$$

$$rmse = \sqrt{\frac{1}{M} \sum_{i=0}^M (x_i - y_i)^2} \quad (2.15)$$

### **III. METODOLOGÍA**

#### **3.1 Obtención y levantamiento de datos de oleaje y viento**

##### **3.1.1 Datos de oleaje**

Se disponen de registros hindcast de oleaje cada tres horas en la zona de Valparaíso de Chile entre los años 1979 al 2019. Estos modelos son distribuidos por el *National Center of Environmental Prediction* ([NCEP, 2021](#)). Para trabajar estos datos se usa las librerías de manejo de datos llamada Pandas para *dataframes* y Xarray para arreglos multidimensionales. En el proyecto se utilizan los registros entre el año 1999 y 2008 para el entrenamiento de los modelos, y los años 2008 y 2010 para la validación. También se cuenta con registros de oleaje en dos puntos cercanos ([Furgo Oceanor y Empresa Barid, 2021](#)) para la validación del modelo.

##### **3.1.2 Datos de viento**

El registro de datos a utilizar corresponden a los parámetros de magnitud de presión al nivel del mar en el Pacífico sur por cada  $0.25^{\circ}$  de latitud y  $0.25^{\circ}$  longitud, cada tres horas desde el año 1979 hasta el 2020. Estos registros se extraen del Centro Europeo de Mediciones Meteorológicas ([ECMWF, 2021](#)). Los datos se pueden descargar en formato netCDF, los cuales son manejados con la librería Xarray disponible en Python.

#### **3.2 Estudio de zonas de generación de oleaje y tiempos de respuestas**

Para generar estimaciones de oleaje, es relevante definir las zonas en el mar que son responsables de la generación de oleaje en el punto de interés y encontrar los tiempos que se demora en llegar el oleaje generado por cada zona al lugar que se busca estimar.

Para esto se promedia el viento por ventanas rectangulares para definir las distintas zonas de generación como se muestra en la Figura 3.1. Esto se realiza por medio de un filtro de imagen estructurado como una convolución *Average pooling*, de esta forma se filtra la presión en el nivel del mar.

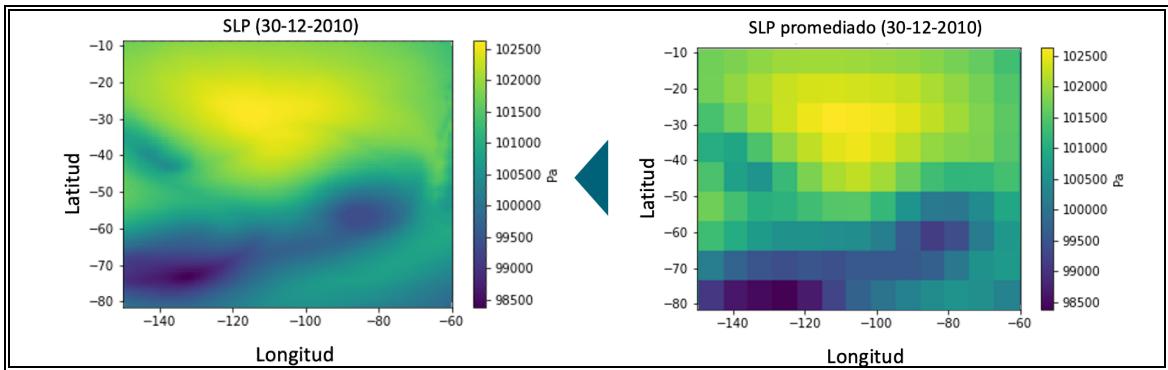


Figura 3.1: Presión al nivel del mar (*SLP*) en un instante de tiempo sin filtro (a) y con filtro (b).

Se tiene en la Figura 3.1.b, el promedio del valor de la presión en cada coordenada latitud y longitud. Cada posición en el mar tiene un registro completo de viento, este registro se compara con cada parámetro de oleaje para obtener una correlación ( $R$ , fórmula 2.11) o coeficiente de determinación ( $R^2$ , fórmula 2.12). Ambos entregan un acercamiento al nivel de explicación de oleaje que se puede obtener de cada zona en particular. A partir de estas correlaciones se pueden generar gráficos de correlación de cada punto en el mar respecto a cada parámetro resumen del espectro de oleaje.

Finalmente, al definir la región total de datos a usar para estimar el oleaje, se debe estudiar la correlación en función del desfase temporal. Esto se realiza por medio de un gráfico que muestra la evolución de la correlación máxima a medida que se aumenta el desfase entre los datos de viento con los de oleaje.

Luego se divide el mapa completo en zonas de generación para ver la evolución de la correlación en cada zona respecto a los parámetros de oleaje. De esta forma se puede ver qué sector corresponde a la zona de generación más relevante de cada parámetro, y cual es el tiempo de respuesta entre la zona de generación y el parámetro de oleaje con mayor correlación.

### 3.3 Implementación de los modelos de *machine learning* y *deep learning*

La metodología para establecer un buen modelo debe abarcar tanto el manejo de la estructura de los *inputs* como el ajuste de los hiperparámetros que definen

cada modelo. Para esto se realizan iteraciones en torno al esquema de la Figura 3.2, donde se prueba con el modelo Ridge el cual se implementa usando la [librería scikit-learn](#) de Python. Luego se prueban distintas arquitecturas de modelos de redes neuronales recurrentes LSTM, se selecciona la mejor y finalmente se evalúa la incorporación de un modelo pre-entrenado con redes convolucionales.

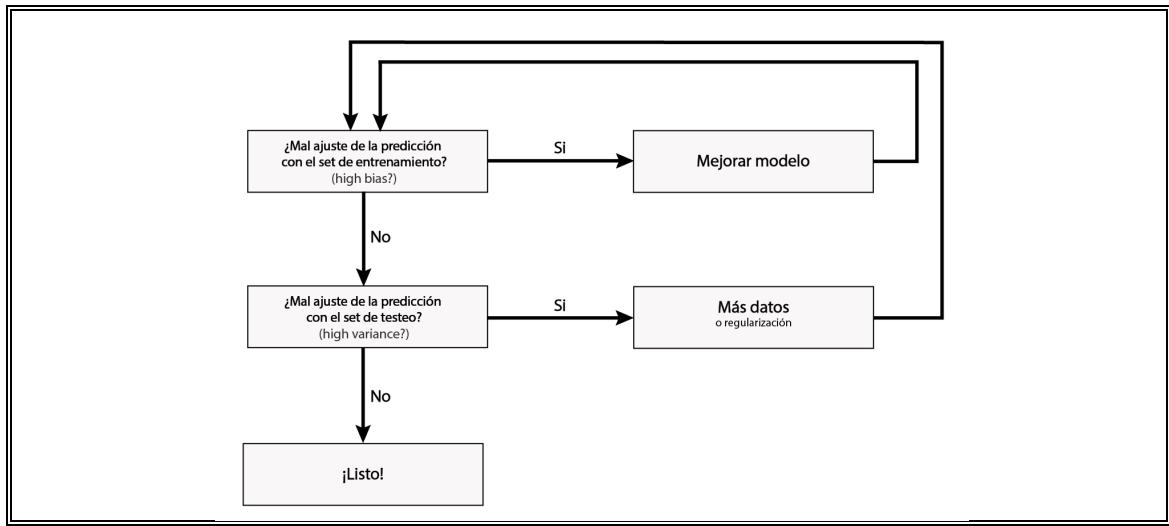


Figura 3.2: Esquema de decisiones para mejorar el modelo.

### 3.3.1 Definición del modelo y validación frente a la casos de éxito

Una vez definido el *input* necesario para poder maximizar las estimaciones usando el regresor, se prueba con un modelo de *deep learning* definido en la [sección 2.2.2](#). La implementación se realiza por medio de las herramientas disponibles de la librería Keras de Python. Y los parámetros se ajustan de manera inicial con los valores propuestos por Pirhooshyaran, M. (2019). A partir de estos resultados es necesario iterar sobre los hiperparámetros para llegar al mejor resultado, para esto se define un tamaño óptimo de lotes (*batch*) tal que, el computador pueda ser capaz de procesarlos y también logre converger a un resultado que sea un óptimo global. El set de entrenamiento y validación son divididos en una proporción 80% de entrenamiento y 20% de validación.

La red neuronal recursiva actúa de manera que estima una secuencia de valores a partir de una secuencia *input*. El regresor calcula un valor a partir de un vector de características que incorporaba los desfases de manera forzada (expansión del vector al agregar manualmente los vectores anteriores), en cambio el modelo LSTM – FCL se entrena usando esta misma memoria de manera lineal en el tiempo. Debido a su funcionamiento por compuertas que permiten recordar las características relevantes del pasado. Se espera que el modelo LSTM – FCL mejore las estimaciones al incorporar una manera más optimizada de abordar los desfases entre la generación y la llegada del oleaje al punto de interés ([Sección 2.2.2.b](#)).

Finalmente se visualiza los resultados de la calidad de ajuste del regresor y del modelo LSTM – FCL, con y sin el vector pre-entrenado para poder determinar cual de todas las alternativas es la que presenta mejor correlación y menor error cuadrático medio como se presenta en la Figura 3.3.

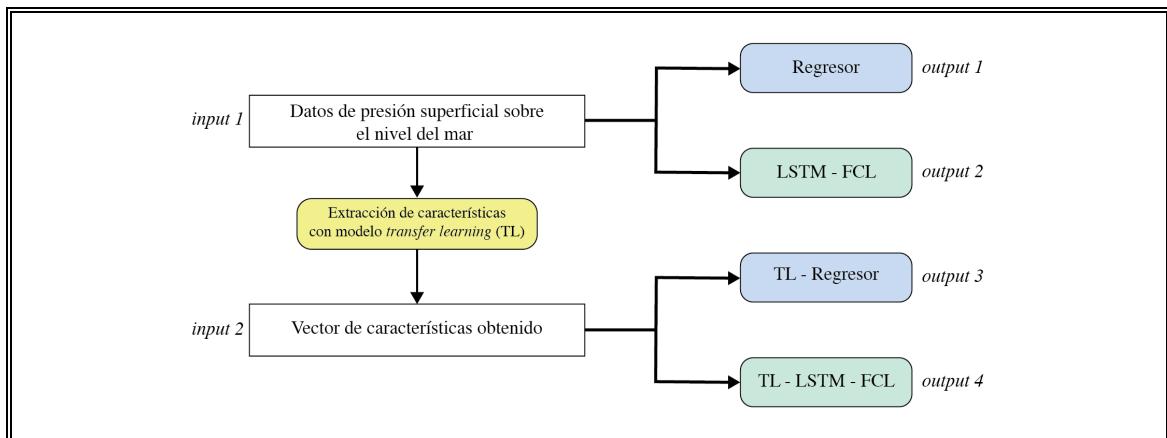


Figura 3.3: Diagrama de modelos a entrenar.

Finalmente se compara los resultados obtenidos frente a los presentados en el Lago Erie ([H. Haoguo et al, 2021](#)). Y luego se evalúa el funcionamiento del modelo en comparación a los resultados presentados en la costa de California ([Pirhooshyaran, M. and Lawrence, S. 2019](#)).

## 4

## RESULTADOS

### 4.1 Zona de generación de oleaje

El estado atmosférico del viento generador de oleaje puede ser bien caracterizado por la variable de presión superficial al nivel del mar (SLP) ([Camus, P. et al, 2014](#)). En la Figura 4.1 se muestra la correlación de cada variable de presión al nivel del mar en cada punto del espacio respecto a cada parámetro de oleaje registrado en Valparaíso. Luego de iteraciones, se define el área de influencia que muestra zonas con mayor correlación entre la zona de latitud -10N a -70N y longitud -130E a -70E.

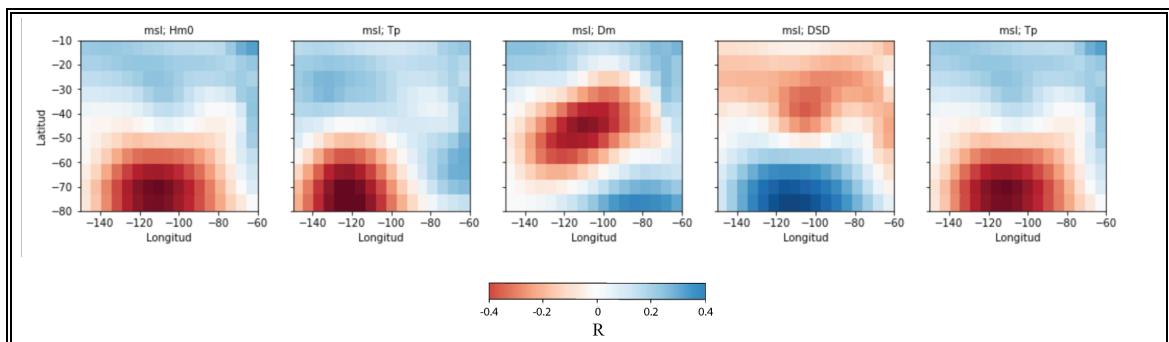


Figura 4.1: Correlaciones entre SLP y oleaje en Valparaíso.

Donde *SLP* corresponde a la presión al nivel del mar. Los parámetros resumen del espectro son *Hm0* (altura significativa en metros), *Tm* (periodo medio en segundos), *Tp* (periodo *peak* en segundos), *Dm* (dirección media de propagación en grados) y *DSD* (desviación estándar direccional o *spreading* en grados). La formulación matemática se presenta en el [Anexo 2](#). Para poder determinar los desfases y con esto el tamaño necesario de la memoria de los regresores, se realiza el cálculo de las correlaciones para distintas zonas de generación, y se registra el valor máximo de la correlación ( $R$ , fórmula 2.10) para distintos desfases de datos entre el *input* y el *output*. Los resultados se muestran en la Figura 4.2 para distintas zonas de generación.

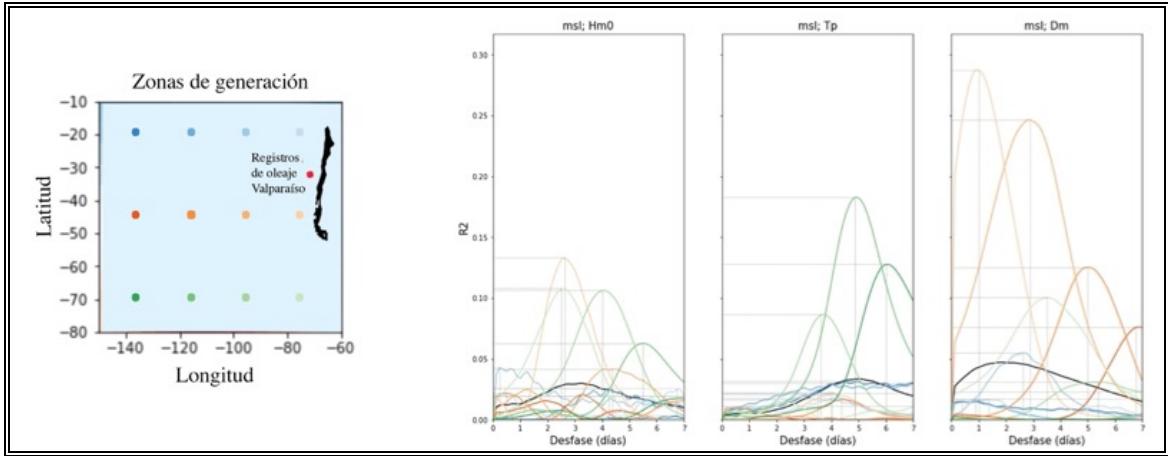


Figura 4.2: Correlación de los parámetros de oleaje en función de las variables de viento en el tiempo.

En la Figura 4.2 se observa que los parámetros de oleaje de altura significativa y dirección de oleaje tienen una respuesta más rápida y cercana a los registros de oleaje en Valparaíso. Por otro lado, el periodo pico, tiene un tiempo de respuesta más largo y respecto a una zona de generación más lejana (área verde).

Se excluye del estudio los registros correspondientes al oleaje generado en la zona del noroeste del Pacífico ([Anexo 4](#)). Esto con el fin de poder focalizar el estudio en la generación de oleajes con componente suroeste, el cual corresponde al oleaje con mayor incidencia en Chile ([Camus et al, 2017](#)).

## 4.2 Modelos de *machine learning*

### a) Ajuste de hiperparámetros de regresor Ridge

Usando un registro de 10 años entre las fechas de 2000 y 2010, se define el regresor Ridge con los mejores resultados para valores de  $\alpha = 1$ , tamaño de memoria mayor a una semana y una calidad de input de oleaje superior a 36 registros de oleaje como se muestran en la Figura 4.3.

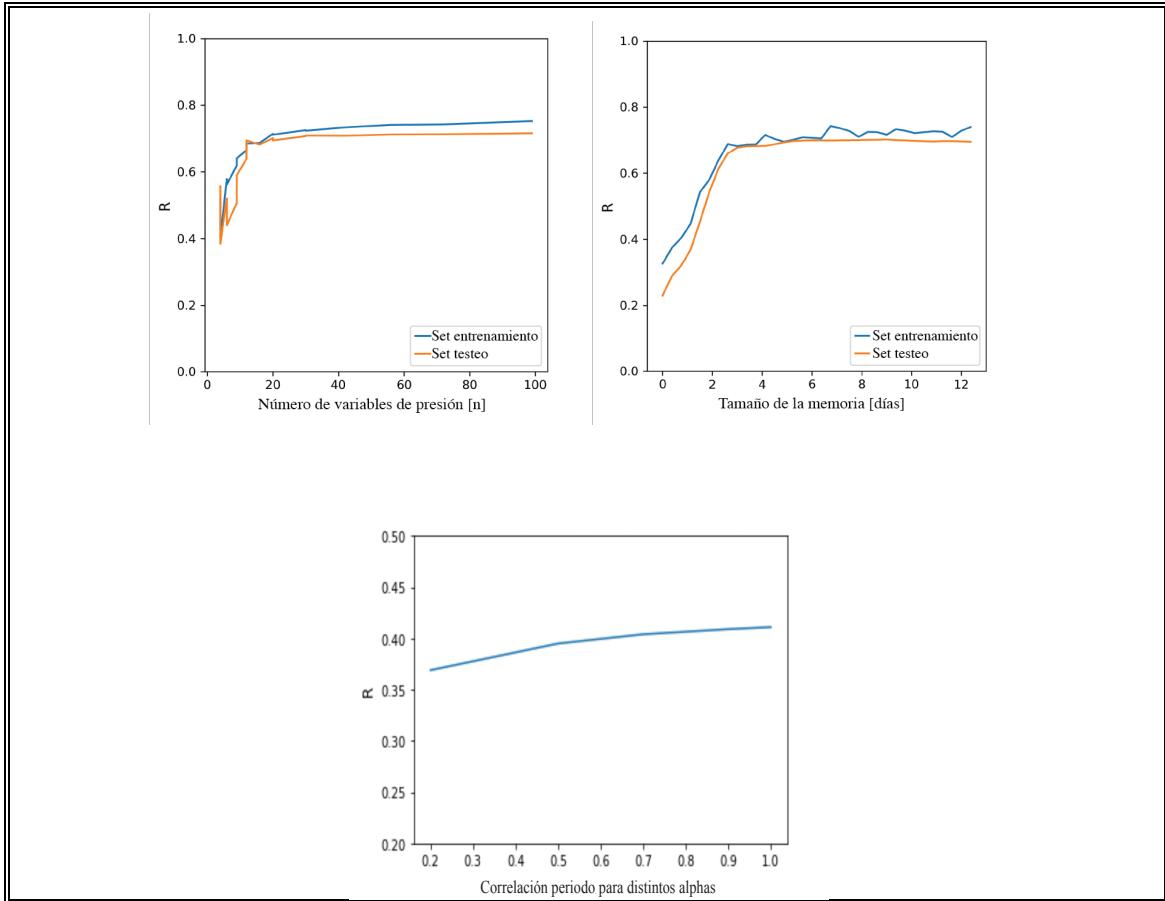


Figura 4.3: Ajuste del modelo Ridge altura significativa para distintas memorias, resolución de mapa de presión y ajuste periodo pico para distintos  $\alpha$ .

Se obtiene una correlación de 0.76 para la altura significativa y 0.44 para el periodo *peak* del oleaje. Es relevante destacar el bajo desempeño en la estimación del período *peak*. Donde la razón más directa de este bajo desempeño está relacionada con la dependencia del período con el *fetch* de la propagación del viento ([Deo M. C. et al, 2001](#)). El vector de características al incorporar los números en forma de vector, no permite entregar información de la localización espacial entre ellos, por lo que es más difícil para el regresor encontrar la relación para estimar el período del oleaje. Para esto es necesario incorporar modelos pre-entrenados que puedan extraer características de manera más inteligente del mapa de presiones, donde se espera que se incorpore implícitamente el gradiente de presiones y *fetch* por medio de las redes convolucionales.

b) Selección de modelo de redes convolucionales

Para mejorar el desempeño del modelo es necesario extraer características del mapa de presiones que define un estado atmosférico en cada registro de tiempo. La extracción de características se obtiene por medio de redes convolucionales, que logran extraer de manera inteligente patrones e identificar comportamientos que están relacionados con distintos fenómenos meteorológicos y tormentas. Por medio de la técnica *transfer learning* descrita en la [sección 2.2.2.a](#), se entrena distintos modelos pre-entrenados disponibles en la librería de Keras. Estos modelos se descargan, luego se ingresa el registro completo de los mapas de presiones con resolución completa, para poder extraer las características de cada mapa de presión. Para esto es necesario descargar cada modelo, descargar los pesos  $w$  entrenados y excluir las capas clasificadoras correspondiente a las últimas capas de cada modelo (Figura 3.3). Los resultados se presentan en la Tabla 4.1, donde se obtiene que el modelo que mejor logra estimar el período *peak* de oleaje es el modelo que incluye la capa VGG16.

Tabla 4.1: Comparación de resultado de los distintos modelos de estimación de período *peak* de oleaje.

Modelo pre-entrenado con regresor Ridge	Correlación entrenamiento	Correlación validación
Sin modelo	0.89	0.44
InceptionV3	0.91	0.23
ResNet50	0.75	0.46
ResNet152V2	0.84	0.25
Xception	0.95	0.27
VGG16	0.92	0.47

Si bien se obtiene un vector de características pre-entrenado por la red VGG-16, que mejora ligeramente los resultados obtenidos en la estimación del período de oleaje. No aumenta considerablemente el resultado por lo que no está reconociendo de manera óptima las tormentas en el mapa de presiones.

#### 4.3 Desarrollo de modelo regresor *Machine Learning*

Para estimar los parámetros resumen del espectro de oleaje, se comienza estimando los parámetros resumen por medio del regresor Ridge sin las redes convolucionales integradas en el modelo pre-entrenado VGG16, y luego se entrena el regresor con la red para ver si mejora el desempeño o empeora. Los resultados se presentan en la Tabla 4.2.

Tabla 4.2: Calidad de ajuste de los parámetros resumen de oleaje regresor Ridge.

	Hm0		Tm		Tp		Mean Dir		DSD	
	r	rmse (m)	r	rmse (s)	r	rmse (s)	r	rmse (grados)	r	rmse (grados)
Regresor	0.73	0.43	0.59	1.49	0.41	1.88	0.85	6.83	0.59	5.05
VGG16 - Regresor	0.77	0.39	0.55	1.58	0.45	1.88	0.76	8.71	0.51	5.54

Los resultados del regresor presentan buenos pronósticos para la altura significativa y dirección media, sin embargo la estimación del periodo medio y *peak* es muy baja. Las redes convolucionales del modelo VGG16, ayudan a mejorar las estimaciones del periodo *peak*. Por lo tanto la incorporación de redes convolucionales en el caso del regresor no aporta a mejores resultados a excepción del periodo *peak*.

La estimaciones del periodo *peak* y periodo medio son inferiores al 50% de correlación. Esto se debe a que el periodo es un parámetro de oleaje que presenta una alta auto-regresión, es decir, la relación del parámetro no se explica solamente por las variables de viento, sino que este necesita ser explicado con datos propios. Es por esto que se avanza con la implementación del modelo LSTM, el cual es un regresor con memoria que presenta una arquitectura que considera un entrenamiento sobre si misma.

#### 4.4 Desarrollo de modelo LSTM, *Deep Learning*

Se evalúan tres arquitecturas de redes LSTM con hiperparámetros fijos de *timestep* una semana, 128 capas ocultas dentro de la LSTM y un mapa de presión al nivel del mar de alta resolución con 500 valores de presión por registro de tiempo. Luego de seleccionar la arquitectura que tenga mejor desempeño, se optimizan los hiperparámetros.

**a)      LSTM – FCL (*Many to many*)**

Se define la arquitectura de la red LSTM many to many 1, donde el *input* entra como ejemplos de sub-series de largo *timestep* para estimar un *output* con ejemplos de sub-series del mismo largo *timestep* como se muestra en la Figura 4.4. También se considera una segunda arquitectura LSTM many to many 2 con desfase similar, donde el largo de la sub-serie de input es más largo que la serie del output, permitiendo entrenar los valores de oleajes iniciales.

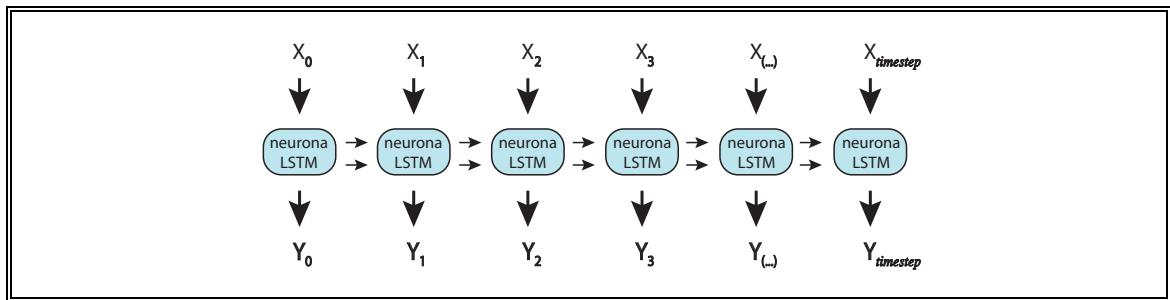


Figura 4.4: Esquema red LSTM *many to many* por cada sub-serie de muestra de largo *timestep*.

**b)      LSTM – FCL con entrenamiento previo (*Many to many 3*)**

Se entrena un modelo LSTM – FCL incorporando el *output* como se muestra en la Figura 4.5 también se considera un *traslape* de 7 días para entrenar los estados ocultos de la red LSTM antes de pronosticar el primer parámetro de oleaje con *timestep* de dos semanas. De esta forma se espera que la red recurrente logre entrenarse con una consideración más explícita de los valores pronosticados, buscando incorporar la auto-correlación que tienen las variables como el periodo *peak* y el periodo medio.

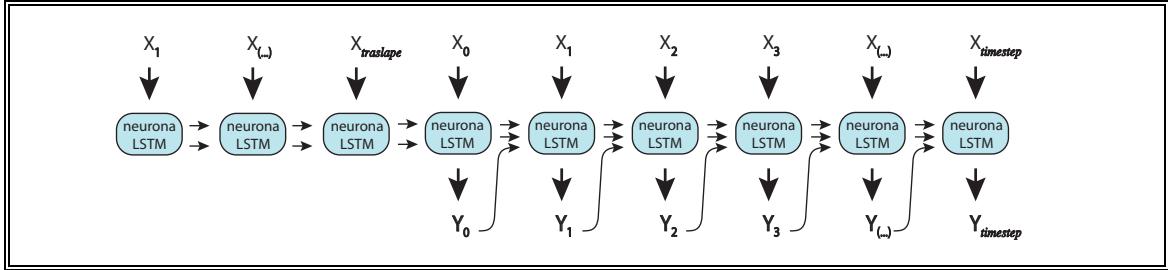


Figura 4.5: Esquema red LSTM con auto-regresión *many to many* por cada sub-serie de muestra de largo *timestep* precedido por un *traslape*.

### c) LSTM – FCL (*Many to one*)

Finalmente, se prueba entregándole sub-series de largo *timestep* de manera similar al entrenamiento del regresor. Donde se entrena una sub-serie de 7 días para pronosticar solo un instante de tiempo del *output* como se muestra en la Figura 4.6. Esta arquitectura es la más costosa computacionalmente, dado que implica entrenar el modelo usando 7 días para pronosticar un instante de tiempo.

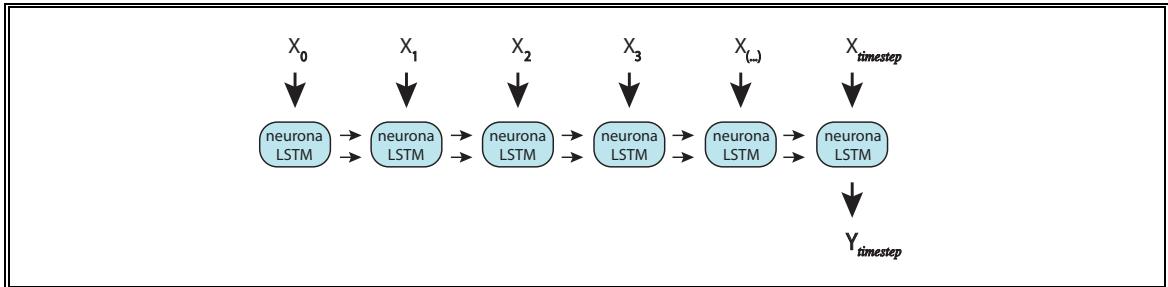


Figura 4.6: Esquema red LSTM *many to one* por cada sub-serie de largo *timestep*.

Frente a las cuatro arquitecturas, se obtiene que las mejores estimaciones se encuentran en la red LSTM – FCL many to one como se presenta en el gráfico 4.7

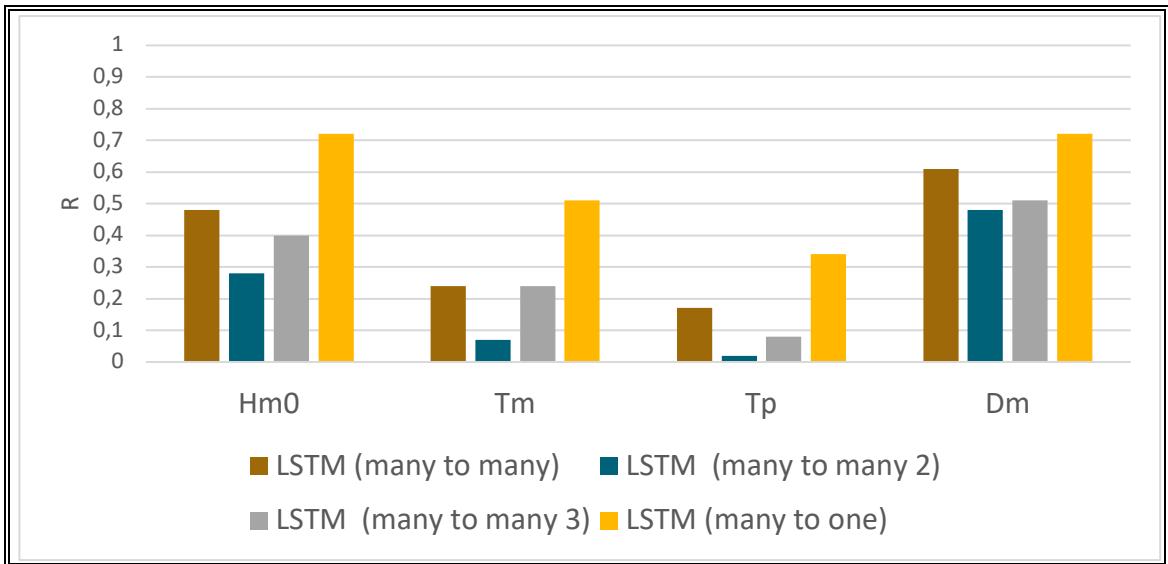


Figura 4.7: Resultados arquitecturas LSTM.

Se optimizan los hiperparámetros de la red many to many, para poder obtener mejores resultados. Los resultados del análisis de sensibilidad se muestran en la Figura 4.8. De esta forma se selecciona una resolución de mapa de 16 valores de SLP por registro en el tiempo, un número de capas ocultas dentro de la red LSTM de 36 y un Dropout de 0.5.

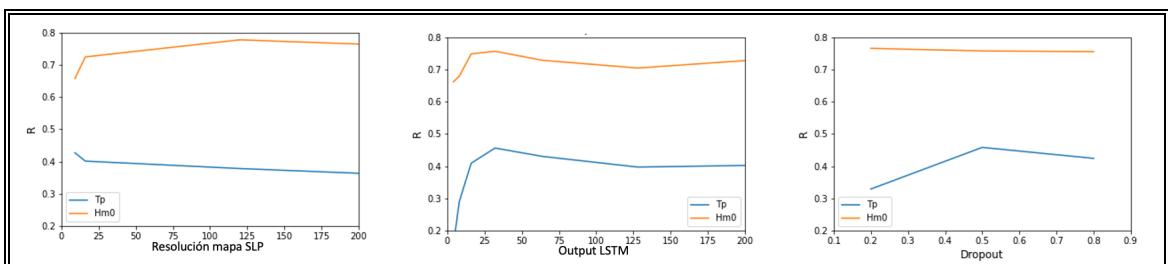


Figura 4.8: Ajuste de hiperparámetros del modelo LSTM.

De esta manera se obtienen resultados de pronósticos de oleaje en base a registros de oleaje con y sin red VGG16 como se presentan en el gráfico de la Figura 4.9.

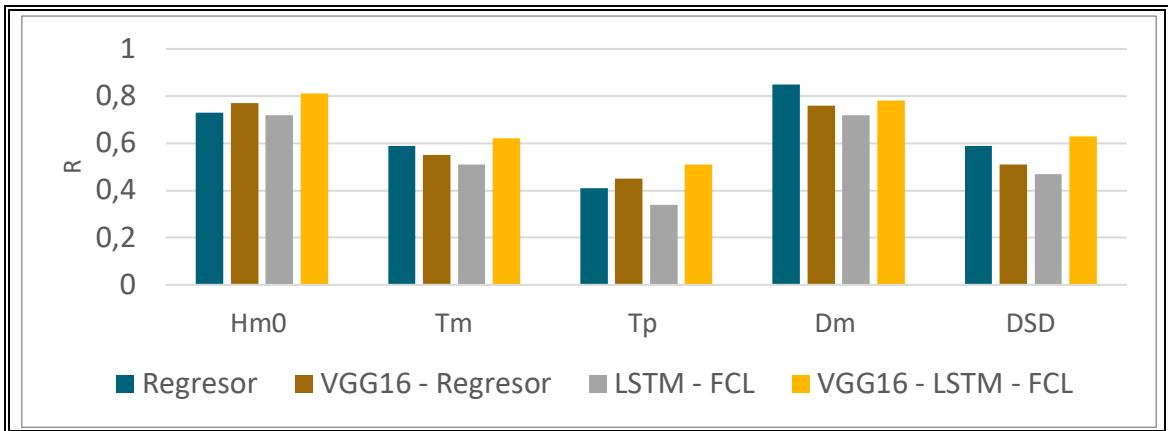


Figura 4.9: Resultados de regresor y LSTM con y sin la red convolucional.

Los mejores resultados se obtienen por medio de la incorporación de redes convolucionales y el uso de la red LSTM many to one a excepción de la dirección de oleaje, la cual está mejor representada por el regresor. Ya que este parámetro de oleaje está más correlacionado a las magnitudes de la presión al nivel del mar de puntos cercanos.

#### 4.5 Análisis de resultados

En la Figura 4.10 se pueden ver las diferencias entre los ajustes del periodo pico pronosticados para el sector de Valparaíso contra los resultados en el Lago Erie. Estos resultados son inferiores a los obtenidos por [H. Haoguo et al. \(2021\)](#) en el pronóstico de oleaje del lago Erie. La principal razón de esta diferencia de resultados corresponde a que el oleaje del lago Erie corresponde a un sistema aislado, donde la zona de generación de oleaje está determinada por los dos puntos de vientos y la variabilidad es menor. Como resultado se obtiene que no se puede pronosticar el periodo de oleaje usando únicamente el registro de presión al nivel del mar.

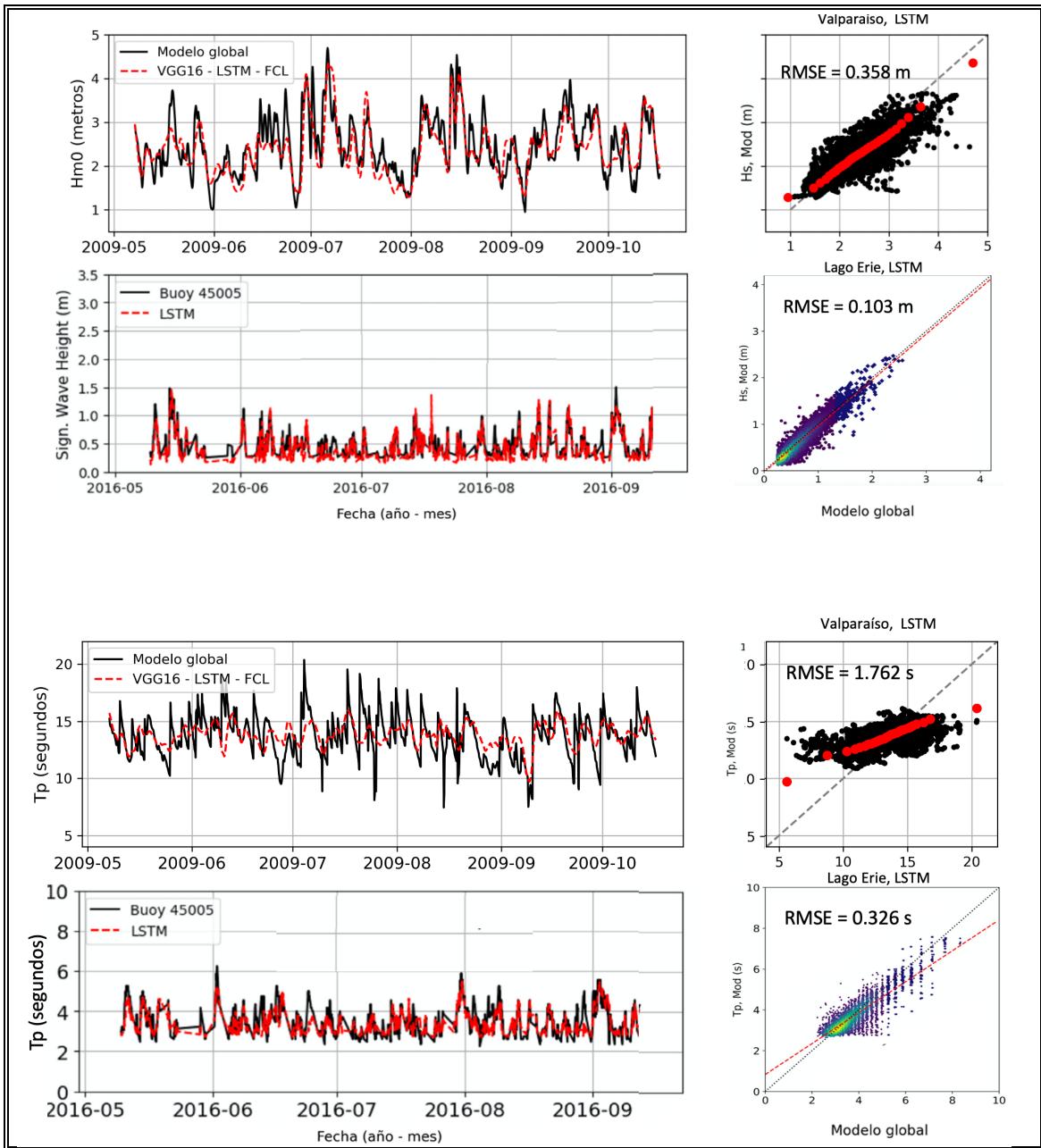


Figura 4.10: Comparación de resultados Lago Erie ([H. Haoguo et al, 2021](#)) contra resultados en Valparaíso. Grafico de dispersión de Valparaíso presenta la dispersión por quintil en rojo.

Para validar los resultados anteriores y la inviabilidad de pronosticar periodo en el mar abierto usando solo registros de presión al nivel del mar con LSTM, es necesario validar el correcto funcionamiento de la red neuronal LSTM entrenada y su

arquitectura, para esto se replica la metodología propuesta en [Pirhooshyaran, M. and Lawrence, S. \(2019\)](#), donde usan dos puntos de oleaje cercano y los seas locales para entrenar un modelo LSTM y pronosticar un tercer punto de oleaje como se muestra en la Figura 4.11. Los oleajes del punto 2 y 3 son hindcast de oleaje distribuidos por [Furgo Oceanor](#) y la empresa [Barid](#). Se entrena el modelo LSTM many to one usando registros entre el año 2000 y 2010.



Figura 4.11: Ubicación espacial de los puntos con registro de oleaje disponibles en la costa de Chile, zona central.

Los resultados obtenidos en la costa de Chile con el modelo LSTM many to one entrega resultados similares a los obtenidos por en [Pirhooshyaran, M. and Lawrence, S. \(2019\)](#) en la costa de California. Estos se presentan en la Figura 4.12, donde se obtiene una correlación de 85% frente al 89% documentado en el *paper* para la altura significativa. El periodo de oleaje logra ser estimado con un 66% de correlación presentando un satisfactorio ajuste frente a la gran variabilidad y alta auto-regresividad de este parámetro. De esta forma la red LSTM con arquitectura many to one, tiene un buen funcionamiento y necesita incorporar registros de oleaje de puntos cercanos como variable predictor como en el caso de la costa de California ([Pirhooshyaran, M. and Lawrence, S. 2019](#)).

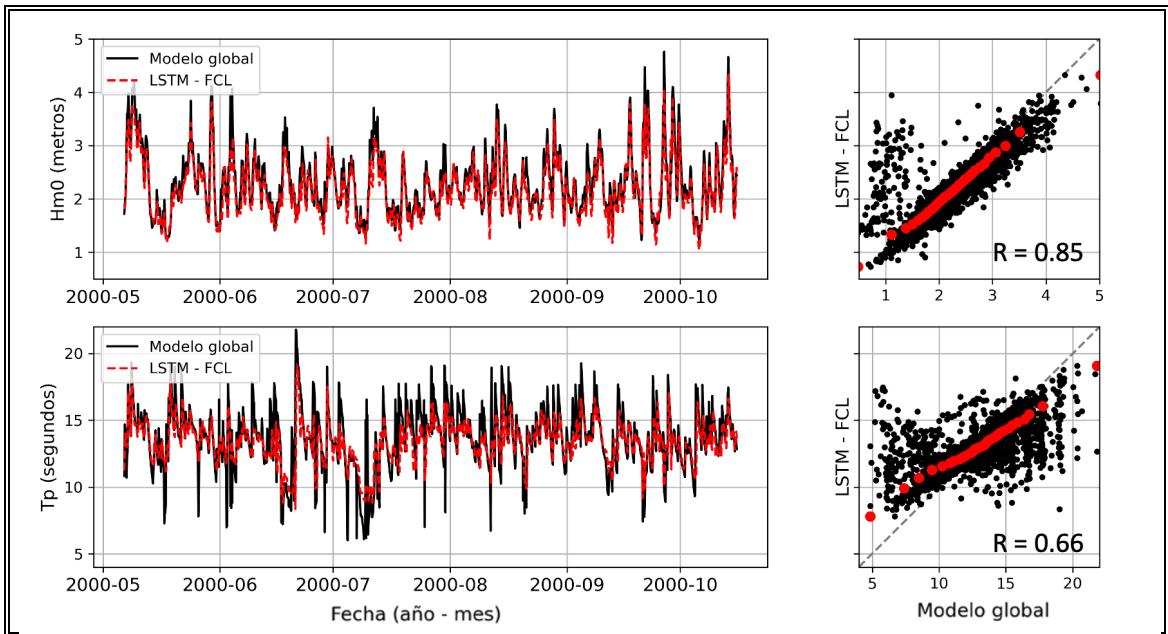


Figura 4.12: Serie de tiempo de oleaje, gráfico dispersión y Q - Q (rojo) del modelo LSTM – FCL *many to one* para parámetros resumen de oleaje usando dos puntos de oleajes cercanos y seas locales.

## 5 CONCLUSIONES

- Es posible llenar exitosamente los hindcast de parámetros resumen de oleaje faltantes. Esto se logra por medio del entrenamiento de redes neuronales LSTM con arquitectura *many to one* usando como predictor hindcast de oleajes en puntos cercanos y *seas* locales.
- Si no se cuenta con registros de oleaje cercanos en un sistema abierto al mar, la alternativa con mejores resultados se obtiene usando redes convolucionales pre-entrenadas (modelo VGG16). La cual extrae características del viento sobre el mar, para entrenar la red LSTM. Sin embargo, esta alternativa solo logra pronosticar altura significativa y dirección de oleaje. En el caso de un sistema cerrado como un lago, basta utilizar los *inputs* del mapa de viento para entrenar la red LSTM exitosamente ([H. Haoguo et al, 2021](#)).
- A lo largo del proyecto se comprenden y aplican principios de cantidad de movimiento y conservación de la energía en los fenómenos de la generación de oleaje producto del viento. Se integran los principios físicos del oleaje y el desarrollo de modelos conceptuales para usar modelos hindcast de oleajes. También se diagnostican y predicen parámetros resumen y parámetros resumen de oleajes por medio de redes neuronales cumpliendo con las competencias del perfil de egreso declaradas.

## BIBLIOGRAFIA

- Arias, V., Salazar, J., Garicano, C., et al. (2019) **An introduction to artificial intelligence applications in medicine: Historical aspects.** Latin American Society of Hypertension.
- Baird, Empresa consultora ingeniería costera y fluvial.  
<https://www.baird.com/?lang=es>
- Camus, P. et al (2017) **Statistical wave climate projections for coastal impact assessments.** Environmental Hydraulics Institute, Universidad de Cantabria, Santander, España.
- Camus, P. et al (2014) **A weather-type statistical downscaling framework for ocean wave climate.** J. Geophys. Res.: Oceans, 119.
- Cienfuegos, R. (2019) **Ingeniería de Costas ICH324,** Pontificia Universidad de Chile. Clase 3a.
- Deo, M.C, Jha, A, Chaphekar, A.S, & Ravikant, K. (2001). **Neural networks for wave forecasting.** *Ocean Engineering*, 28(7), 889-898.
- Diederik, P. Kingma, and Jimmy B. (2014) **A method for stochastic optimization.** arXiv preprint, 1412.6980.
- ECMWF (2021) **ERA5 hourly data on single levels from 1979 to present,** recuperado de:  
<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>
- Furgo Oceanor As, **Environmental Monitoring System.** www.oceanor.no.
- Hochreiter, S. and Schmidhuber, J. (1997) **Long short-term memory.** Neural computation, 9(8): 1735-1780.

- Hu, Haoguo, Van der Westhuysen, André J, Chu, Philip, & Fujisaki-Manome, Ayumi. (2021). **Predicting Lake Erie wave heights and periods using XGBoost and LSTM.** *Ocean Modelling (Oxford)*, 164, 101832.
- Kamphuis, J. (2010). **Introduction to coastal engineering and management** (2nd ed., Advanced series on ocean engineering ; v. 30).
- Krizhevsky, Sutskever, and Hilton, E. (2012) **Imagenet classification with deep convolutional neural networks**, Nikkosoft consultores.
- Miles, J. W. (1957) **On the generation of surface waves by shear flows.** Journal of Fluids Mechanics, 7, pp. 469-478
- NCEP (2021) **National Centers for Environmental Prediction**
- NOAA (2021) **National and Oceanic Atmospheric Administration**
- Pirhooshyaran, M. and Lawrence, S. (2019) **Multivariate, multistep forecasting, reconstruction and feature selection of ocean waves via recurrent and sequence-to-sequence networks.** 1906.00195v1
- Shetty, B. (2018) **Supervised machine learning: Classification.** Toward Data Science.
- Valdés, L. Baquero, J. (2019) **Deep learning aplicado a imágenes satelitales como herramienta de detección de viviendas sin servicio de energía en el caserío Media Luna Uribia Guajira.** Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.

## **A N E X O S**

## ANEXO 1 : FORMACIÓN DE OLEAJE POR EL VIENTO

Para entender la formación de oleaje en aguas profundas es necesario revisar la interacción inicial entre el viento y la superficie plana del mar. En la Figura A.1.a se muestra la interacción entre la presión del viento sobre la superficie del mar, donde Miles, J. W. (1957) demostró que esta interacción provoca perturbaciones que crecen linealmente, luego estas perturbaciones se retroalimentan de manera que forman las ondas de gravedad. Estas ondas se mueven y van creciendo debido a esta retroalimentación que se fundamenta en la diferencia de presiones entre el viento que sube y el que baja la onda como se muestra en la Figura A.1.b.

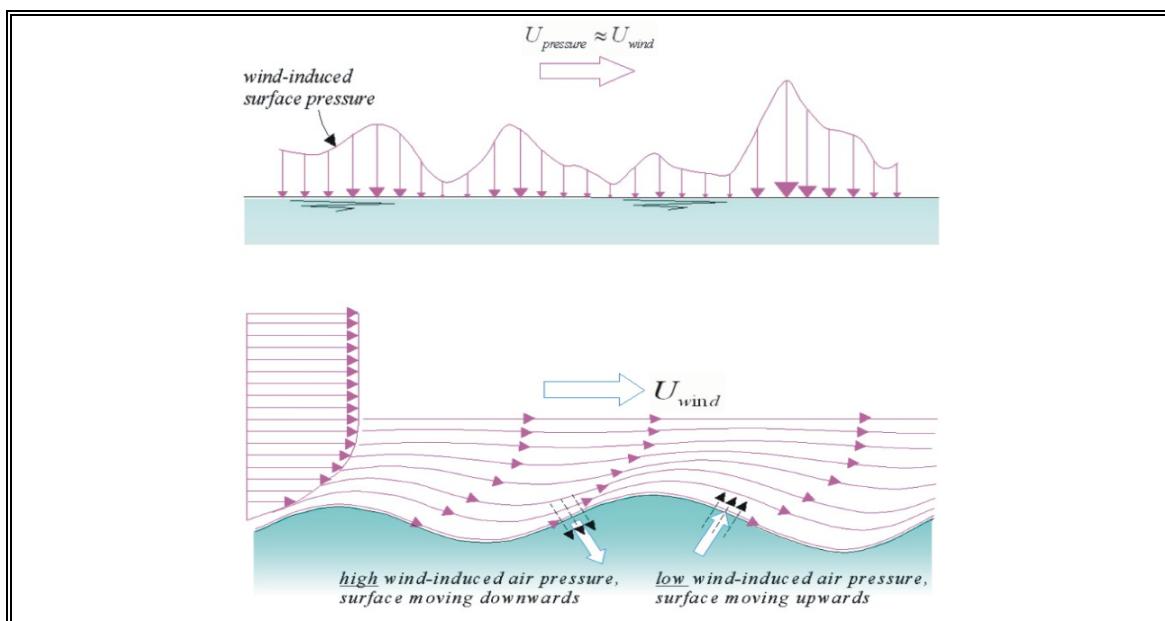


Figura A.1: Generación de olas por viento. (a) perturbación de la superficie debido a diferencias en la presión del viento sobre la superficie. (b) esquema de retroalimentación de la propagación de ondas producto del viento. Fuente:

Miles, J. W, 1957.

Las ondas de gravedad que tienen su motor en el viento, son la principal fuente de oleaje de aguas profundas. Este se puede pronosticar a partir de el viento como un valor idealizado correspondiente al valor promedio en el espacio, usando la

distancia de aplicación de este viento sobre la superficie del mar (*fetch*) y la prolongación de este viento sobre la superficie. De esta manera a medida que el *fetch* aumenta para un viento determinado en un periodo de tiempo prolongado, el período de la ola aumenta hasta alcanzar un equilibrio entre el intercambio de energía del viento y el mar. Este punto de equilibrio se denomina *swell*, el cual es un estado de oleaje completamente desarrollado, donde el mar tiene concentrada la altura de ola en un solo período o frecuencia de oleaje.

Es común la existencia de *swell* que interfieren con el estado del mar común producto de tormentas en el mar. Las tormentas y distintos vientos a lo largo del mar componen un mapa de zonas de generación de oleaje que determinan el estado de mar en un punto. Todo el oleaje en un cierto punto puede ser graficado y determinado en un espectro direccional de energía, el cual contiene la energía específica para cada dirección y período o frecuencia de oleaje como se muestra en la Figura 2.A. Es muy importante el espectro direccional de energía debido a que nos entrega toda la información del oleaje, por lo tanto se tiene conocimiento completo del comportamiento del oleaje y permite diseñar obras de ingeniería de cosas con mayor seguridad.

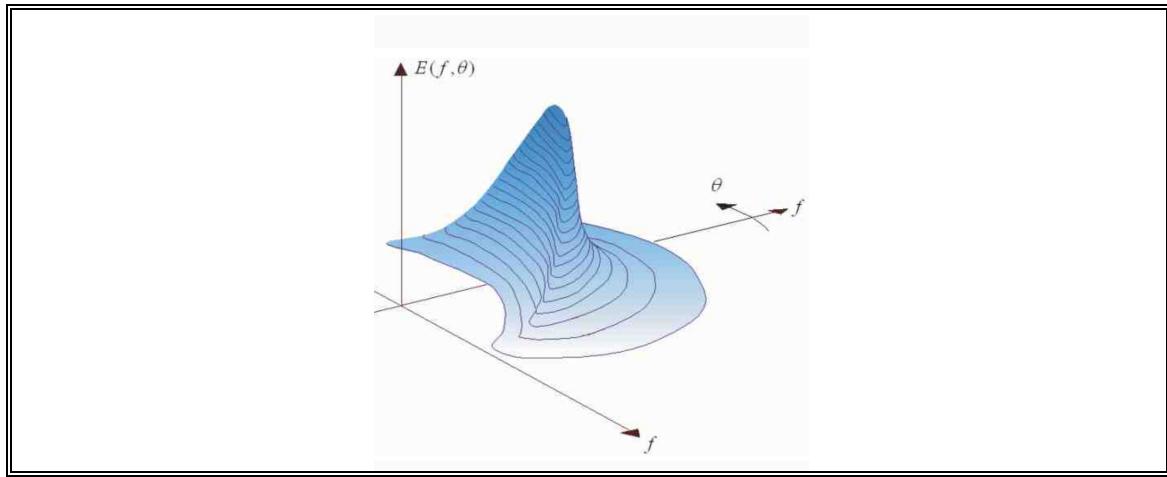


Figura A.2: Espectro direccional de energía del oleaje. Fuente: Leo H. Holthuijsen, 2018.

## ANEXO 2 : OBTENCIÓN DE PARÁMETROS DE OLEAJE A PARTIR DE UN ESPECTRO DIRECCIONAL DE ENERGÍA DE OLEAJE

El estado de mar se puede sintetizar de manera completa en un espectro de oleaje direccional de energía Figura A.2. Este espectro contiene la energía del oleaje para cada frecuencia (o período) y dirección de propagación. Este entrega la información completa como condición de borde para poder realizar modelamiento de oleajes en proyectos de ingeniería de costas.

A partir del espectro direccional de energía se busca obtener los parámetros de oleaje para poder comparar el pronostico con los registros existentes. También es de utilidad para poder comparar resultados con pronósticos de modelos matemáticos. Las ecuaciones provienen del clases de Ingeniería de costas Cienfuegos, R, 2018 y extraídas del manual de usuario MIKE 21.

- Altura de oleaje significativa,  $Hm_0$ .

$$m_0 = \iint E(f, \theta) df d\theta \quad (\text{A.1})$$

$$Hm_0 = 4\sqrt{m_0} \quad (\text{A.2})$$

- Periodo medio,  $Tm$ .

$$m_1 = \iint E(f, \theta) * f df d\theta \quad (\text{A.3})$$

$$Tm = \frac{m_0}{m_1} \quad (\text{A.4})$$

- Periodo peak,  $Tp$ .

$$Tp = T_i \text{ donde } E(f, \theta)_i \text{ es máximo} \quad (\text{A.5})$$

- Dirección media,  $Dm$ .

$$Md = \tan^{-1} \left( \frac{\iint \sin(\theta) E(f, \theta) df d\theta}{\iint \cos(\theta) E(f, \theta) df d\theta} \right) \quad (\text{A.6})$$

- Dirección peak,  $Dp$ .

$$Dp = \theta_i \text{ donde } E(f, \theta)_i \text{ es máximo} \quad (\text{A.7})$$

- Desviación estándar direccional, DSD.

$$DSD = \sqrt{2(1 - \sqrt{a^2 + b^2})} \quad (\text{A.8})$$

$$a = \frac{1}{m_0} \iint \cos(\theta) E(f, \theta) df d\theta \quad (\text{A.9})$$

$$b = \frac{1}{m_0} \iint \sin(\theta) E(f, \theta) df d\theta \quad (\text{A.10})$$

### ANEXO 3 : ESTRUCTURA DEL APRENDIZAJE DE MACHINE LEARNING

El aprendizaje automático es cada vez más habitual, donde los programas aprenden a medida que combinan la experiencia. En términos generales existen dos sub grupos de aprendizaje; el supervisado y el no supervisado. En el primero la atención está centrada en un objetivo que se busca estimar o clasificar, en el segundo no se posee este objetivo y se busca obtener características propias de los datos que se tienen. En la Figura A.3 se muestra un diagrama de flujo del funcionamiento del *machine learning*. Y en la Figura A.4 las distintas aplicaciones que estos tienen.

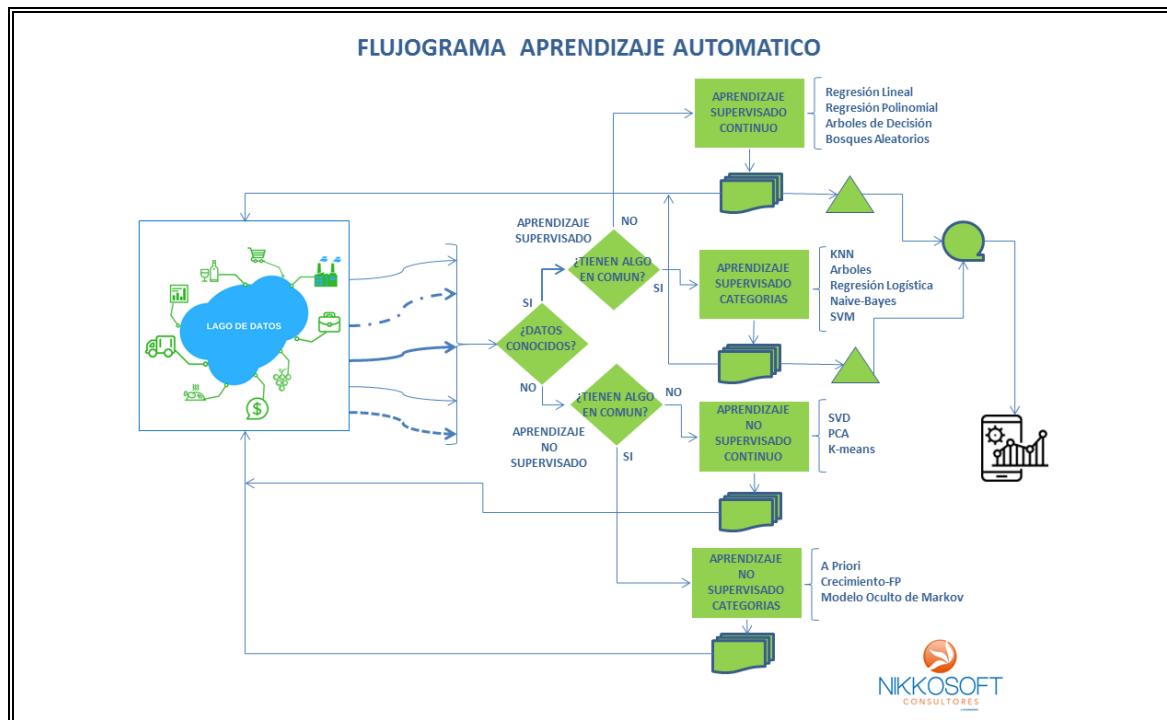


Figura A.3: Diagrama de flujo de *machine learning*. Fuente: Krizhevsky et al., 2012.

	<u>Supervisado</u>	<u>No supervisado</u>
<u>Continuo</u>	<ul style="list-style-type: none"> <li>• Regresión           <ul style="list-style-type: none"> <li>• lineal</li> <li>• polinomial</li> </ul> </li> <li>• Árboles de decisión</li> <li>• Bosque aleatorio</li> </ul>	<ul style="list-style-type: none"> <li>• Agrupaciones y reducciones de dimensionalidad:           <ul style="list-style-type: none"> <li>• SVD</li> <li>• PCA</li> <li>• k-media</li> </ul> </li> </ul>
<u>Categórico</u>	<ul style="list-style-type: none"> <li>• Clasificación           <ul style="list-style-type: none"> <li>• CNN</li> <li>• Árboles</li> <li>• Regresión lógica</li> <li>• Naive-bayes</li> <li>• SVM</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Análisis de asociación           <ul style="list-style-type: none"> <li>• A priori</li> <li>• Crecimiento de FP</li> <li>• Modelo oculto de Markov</li> </ul> </li> </ul>

Figura A.4: Algoritmos utilizados en *machine learning*. Fuente: Shetty B., 2018.

El aprendizaje a utilizar es supervisado y continuo por lo que es conveniente usar algoritmos de regresores lineales, árboles de decisión o bosque aleatorio. Por otro lado, también es una alternativa entrenar el modelo por medio del *deep learning*, que es una de las herramientas más populares, se obtienen aprendizajes supervisado por medio de redes neuronales profundas, que son entrenadas a medida que iteran al realizar los movimientos *forward*, cálculo del error por medio de un criterio de evaluación y luego un *backward* para ajustar las neuronas y disminuir el error del pronóstico, cada una de estas iteraciones es llamada época.

## ANEXO 4 : ZONAS DE GENERACIÓN DE OLEAJE

En la Figura A.5 se presenta el comportamiento en el tiempo (a medida que se desfasan los datos) de cada parámetro de oleaje según distintas zonas de generación.

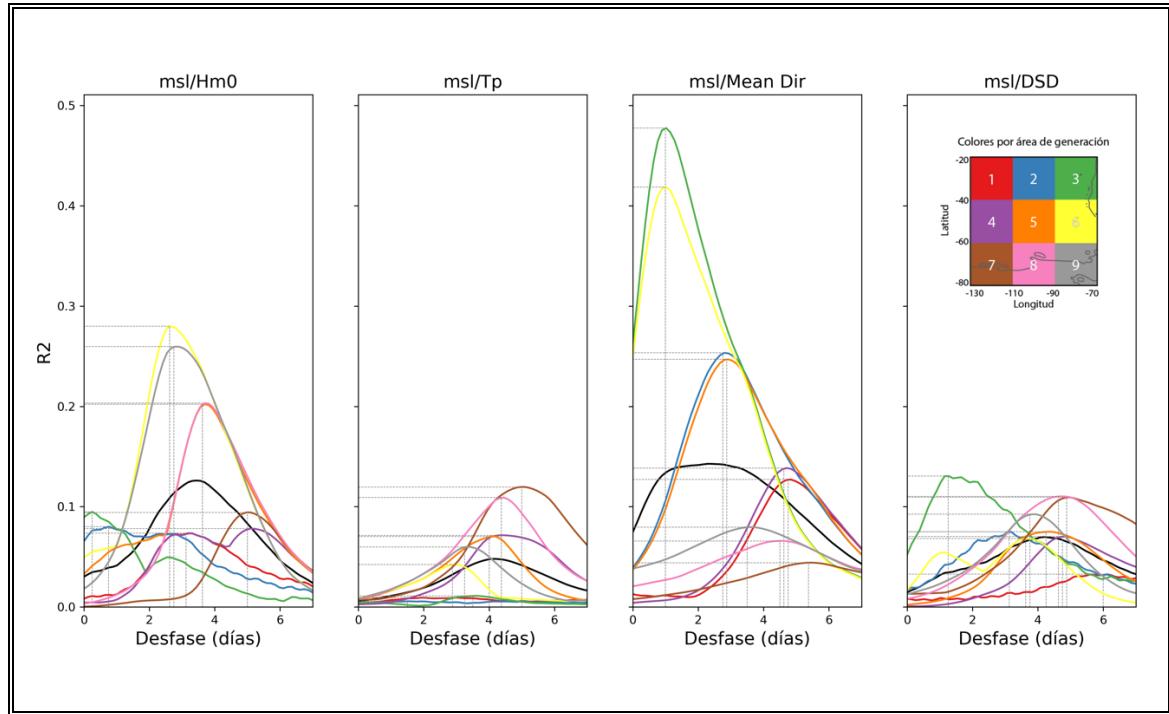


Figura A.5: Coeficiente de determinación de los parámetros de oleaje para distintos desfases en función para distintas zonas de generación.

Las zonas más relevantes para la estimación de los parámetros resumen de oleaje, se obtienen por medio de la presión al nivel del mar entre las zonas de generación 3, 6, 7, 8 y 9 (verde, amarillo, café, rosado y gris) de la Figura A.5. Las cuales corresponden a las áreas entre la sección de latitud -20N a -80N y longitud -90E a -70E (áreas verde, amarilla y gris) las cuales inciden producto de la cercanía al

punto de interés y la zona de latitud -60N a -80N y longitud -130E a -70E (áreas café, rosada y gris) las cuales corresponden a vientos aguas adentro vinculados al anticiclón del pacífico con dirección a la costa chilena.

En la figura A.5, se muestra en el centro el océano completo del mar, segmentado en distintas áreas. Cada área tiene un oleaje que proviene de distintas zonas de generación, las cuales se muestran en sub-mapas del mismo color y número al área de interés. En particular para la zona de Chile, interesa tener pronósticos en el área 8 (color rojo), y este área del océano tiene oleaje proveniente del océano sur completo bajo la latitud -30 N y también tiene una componente proveniente del Noroeste.

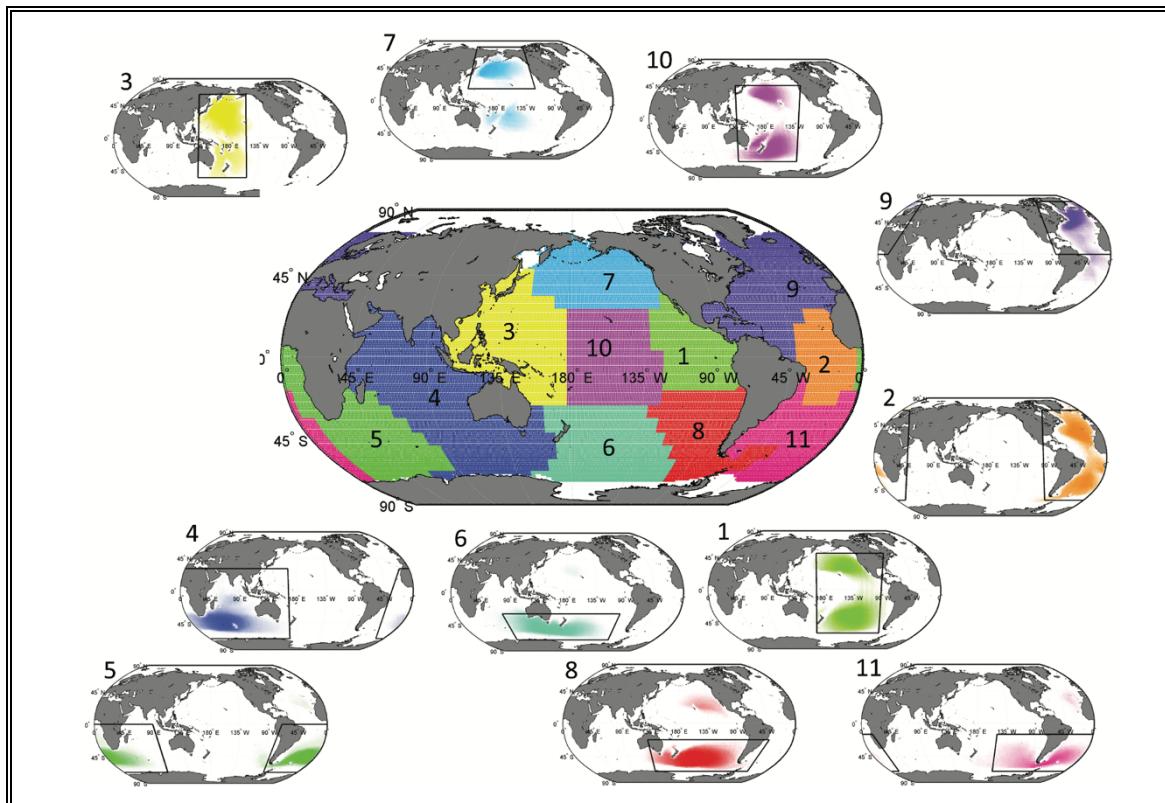


Figura A.6: Zonas de generación de oleaje. Fuente: Camus et al, 2017.

## ANEXO 5 : DESEMPEÑO DEL MODELO VGG16 – LSTM – FCL FRENTE ALTERNATIVAS EXISTENTES

Para evaluar los resultados, finalmente es importante comparar el desempeño de las estimaciones frente a distintas alternativas que existen hoy para obtener información del oleaje. En la Tabla A.7 se muestran los resultados de los ajustes entre los registros de oleaje con los pronósticos del NOAA (se tiene solo pronóstico de la altura significativa, periodo *peak* y dirección de oleaje *peak*). En otra columna, se muestra el ajuste entre los registros y los parámetros del espectro direccional de energía obtenido por la red VGG16 – LSTM – FCL. Donde los pronósticos de la red neuronal son inferiores en la estimación del Hm0 y Tp y superiores para el pronóstico de la dirección *peak* (Dp).

Tabla A.7: Desempeño de las estimaciones NOAA y VGG16 – LSTM – FCL.

Fuente: Elaboración propia.

Parámetros	Pronóstico NOAA		Pronóstico LSTM-FCL		Diferencia entre pronósticos	
	r	rmse	r	rmse	r	rmse
Hm0_Arica	0.97	0.05 m	0.80	0.69 m	-0.17	0.64 m
Hm0_Concepción	0.93	0.44 m	0.73	0.62 m	-0.20	0.18 m
Hm0_Valparaíso	0.94	0.35 m	0.74	0.45 m	-0.20	0.10 m
Tp_Arica	0.56	2.46 s	0.34	2.81 s	-0.22	0.35 s
Tp_Concepción	0.52	2.80 s	0.25	2.90 s	-0.27	0.10 s
Tp_Valparaíso	0.68	1.88 s	0.15	2.31 s	-0.53	0.43 s
Dm_Arica	0.72	30.1 °	0.32	7.2 °	-0.40	-22.9 °
Dm_Concepción	0.36	30.3 °	0.55	15.7 °	0.19	-14.6 °
Dm_Valparaíso	0.63	16.7 °	0.38	12.8 °	-0.25	- 3.9 °

## **ANEXO 6: DESEMPEÑO DEL MODELO VGG16 – LSTM – FCL PARA DISTINTAS LOCALIDADES**

Es importante que el modelo pueda ser validado en distintas localidades. Por lo que se realizan estimaciones usando datos de oleaje en aguas profundas en la zona de Arica y Concepción, localidades al norte y sur respectivamente de Valparaíso de Chile. Los resultados se muestran en la Tabla A.8, donde se obtienen resultados similares a los obtenidos en Valparaíso para las componentes del suroeste (SO). Se puede afirmar que el modelo presenta generalidad dentro de Chile. Sin embargo es necesario probar el modelo en distintas localidades, como Estados Unidos o Australia para poder concluir una generalidad total del modelo. Para esto sería necesario realizar nuevos estudios de zonas de generación de oleaje.

Tabla A.8: Calidad de ajuste del espectro para distintas localidades usando modelo VGG16 – LSTM – FCL. Fuente: Elaboración propia.

Espectro	Hm0		Tm		Tp		Dm		DSD	
	r	rmse (m)	r	rmse (s)	r	rmse (s)	r	rmse (grados)	r	rmse (grados)
Arica – SO	0.80	0.69	0.11	0.92	0.34	2.81	0.68	3.85	0.03	2.87
Concepción-SO	0.73	0.62	0.46	1.78	0.25	2.90	0.75	9.17	0.34	4.17
Valparaíso -SO	0.80	0.39	0.58	1.56	0.37	2.08	0.74	5.34	0.44	2.68

## ANEXO 7 : DESEMPEÑO DEL MODELO VGG16 – LSTM – FCL PARA DISTINTAS ESTACIONES DEL AÑO

En la Tabla A.9 se muestran los resultados de las estimaciones por el modelo VGG16 – LSTM – FCL, para distintas estaciones del año. Donde se concluye que la mejor estimación ocurre en las estaciones de verano y primavera, a excepción del rendimiento de la altura significativa, donde es superior en las estaciones de otoño e invierno.

Tabla A.9: Calidad de ajuste del espectro del suroeste según estaciones del año.

Espectro Suroeste	Hm0		Tm		Tp		Dm		DSD	
	r	rmse (m)	r	rmse (s)	r	rmse (s)	r	rmse (grados)	r	rmse (grados)
Verano	0.76	0.36	0.61	1.33	0.30	2.31	0.75	5.01	0.45	2.81
Otoño	0.82	0.40	0.48	1.68	0.24	1.89	0.68	7.96	0.31	3.03
Invierno	0.82	0.41	0.55	1.57	0.28	1.97	0.82	7.25	0.35	3.38
Primavera	0.78	0.37	0.69	1.37	0.41	2.09	0.79	5.66	0.39	2.82

## ANEXO 8 : DESEMPEÑO DEL MODELO VGG16 – LSTM – FCL PARA PRONOSTICAR ESPECTROS DE MANERA PARCIAL

Se aplica un filtro a los espectros direccionales de energía de oleaje para determinar la componente de oleaje y para el rango de períodos de oleaje que se deseen estudiar como se muestra en la Figura A.10. De esta forma se puede estudiar componentes del error en los resultados. También se realizará un estudio por estaciones del año.

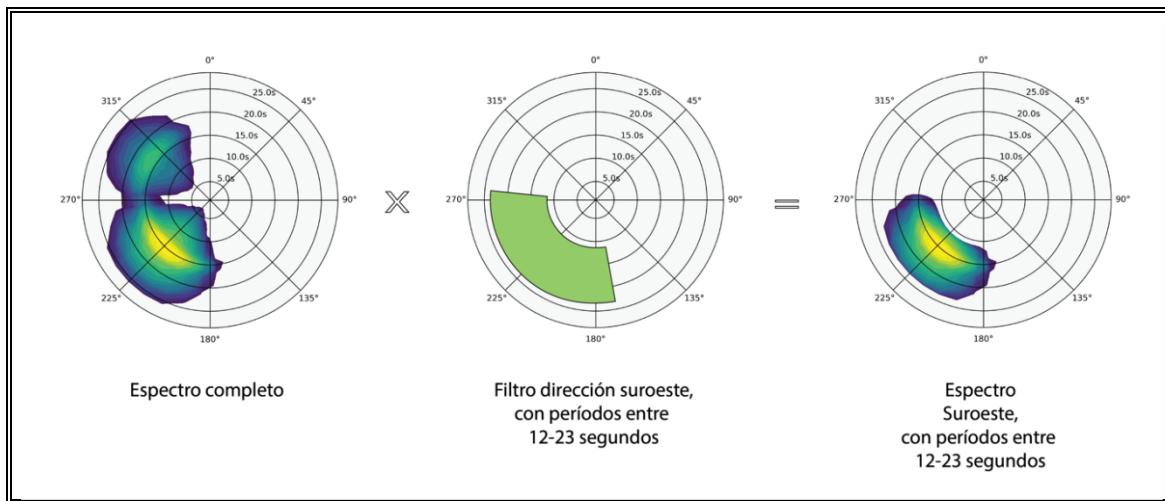


Figura A.10: Aplicación de filtro al espectro de oleaje.

Se aplican distintos filtros para calcular los parámetros resumen de cada espectro y de esta forma comparar los resultados. La Tabla A.11 muestra los resultados de distintos ajustes según la componente suroeste o noroeste, para períodos largos (mayor a 12 segundos) y cortos (menor a 12 segundos).

Tabla A.11: Calidad de ajuste del espectro estimado con distintos filtros.

Espectro	Hm0		Tm		Tp		Mean Dir		DSD	
	r	rmse (m)	r	rmse (s)	r	rmse (s)	r	rmse (grados)	r	rmse (grados)
Completo	0.80	0.38	0.62	1.38	0.32	2.14	0.74	8.82	0.54	6.80
Noroeste	0.76	0.30	0.33	2.51	0.16	3.92	0.46	11.62	0.07	5.47

Suroeste	0.80	0.39	0.58	1.56	0.37	2.08	0.74	5.34	0.44	2.68
Suroeste Período cortos (menor a 12s)	0.72	0.32	0.56	0.77	0.36	0.92	0.83	5.10	0.59	2.36
Suroeste Período largos (mayor a 12s)	0.75	0.46	0.00	64.78	0.21	1.57	0.28	19.66	0.25	4.36

Es importante destacar que los resultados son muy positivos para la estimación del espectro suroeste, y esto se debe a que los datos de presión del noroeste del pacífico no están *dentro* del *inputs*. La estimación mejora considerablemente para los periodos cortos y es peor para periodos largos. Esto se debe a que hay una detección deficiente de tormentas con *fetch* largo por parte del modelo.

## ANEXO 9 : SERIES DE TIEMPO Y GRÁFICOS DE DISPERSIÓN.

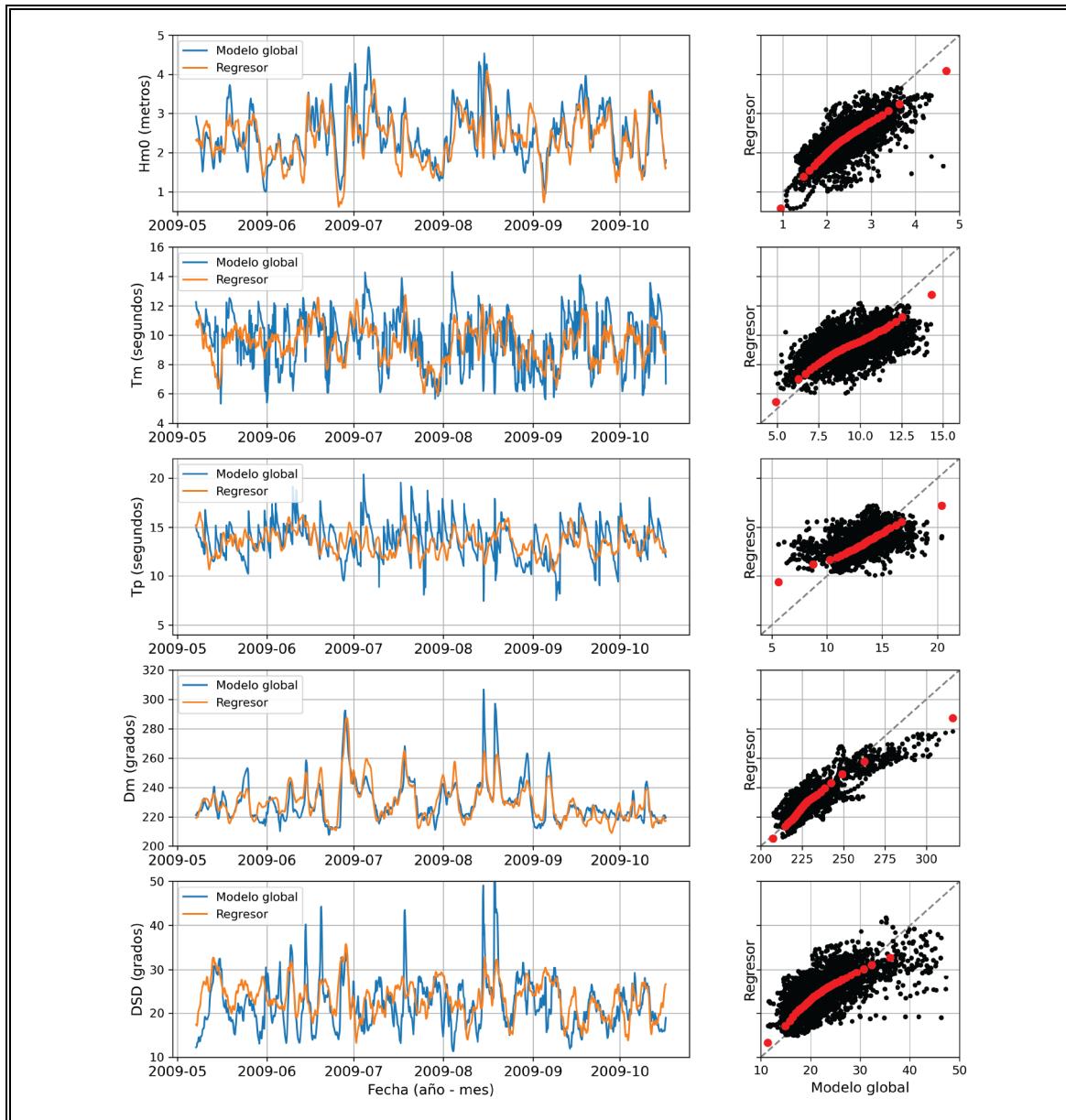


Figura A.12: Serie de tiempo de oleaje, gráfico dispersión y Q - Q (rojo) del modelo global contra estimaciones del regresor Ridge para cada parámetro resumen.

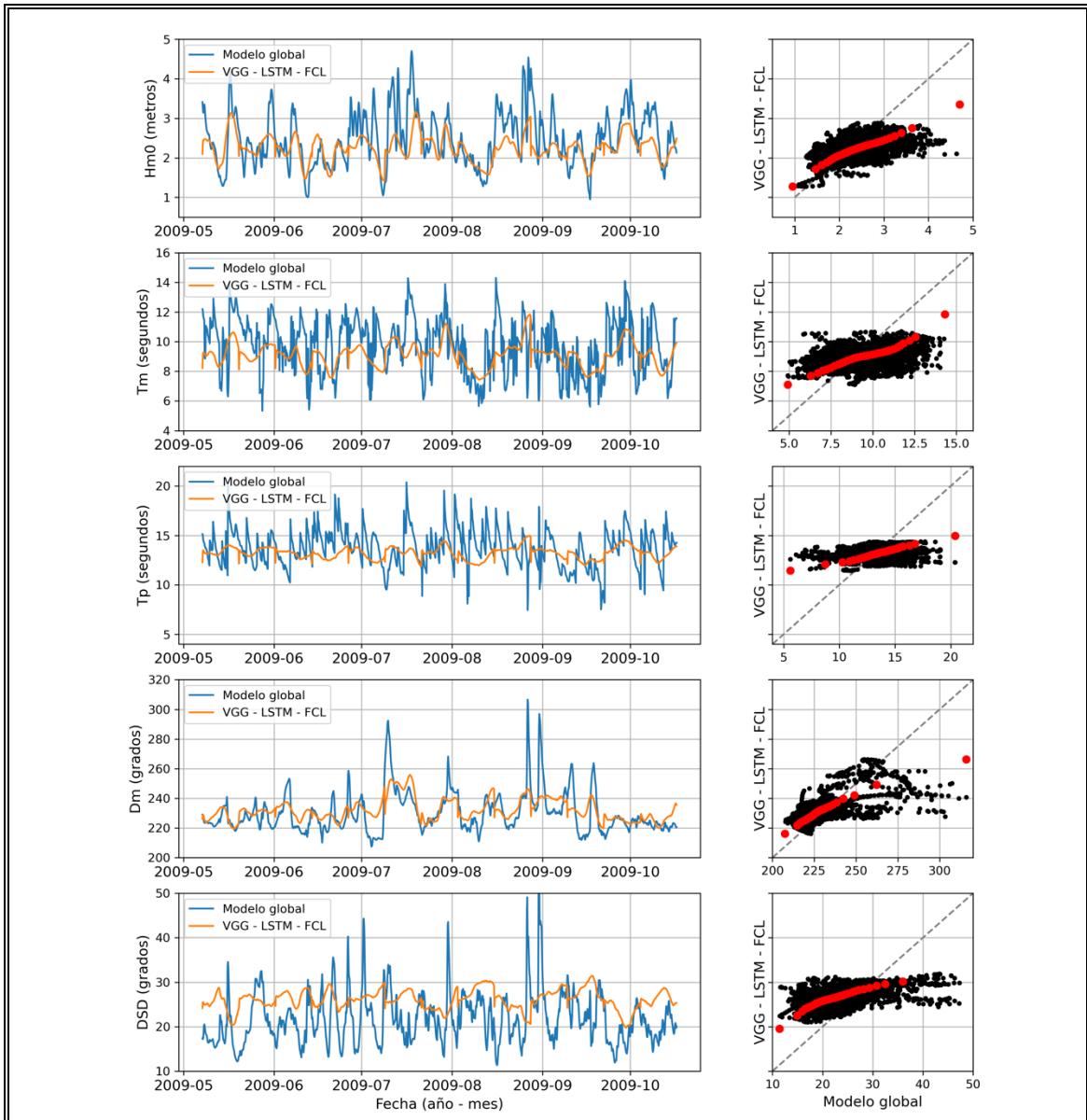


Figura A.13: Serie de tiempo de oleaje, gráfico dispersión y Q - Q (rojo) del modelo global contra estimaciones del VGG16 – LSTM – FCL many to many para cada parámetro resumen

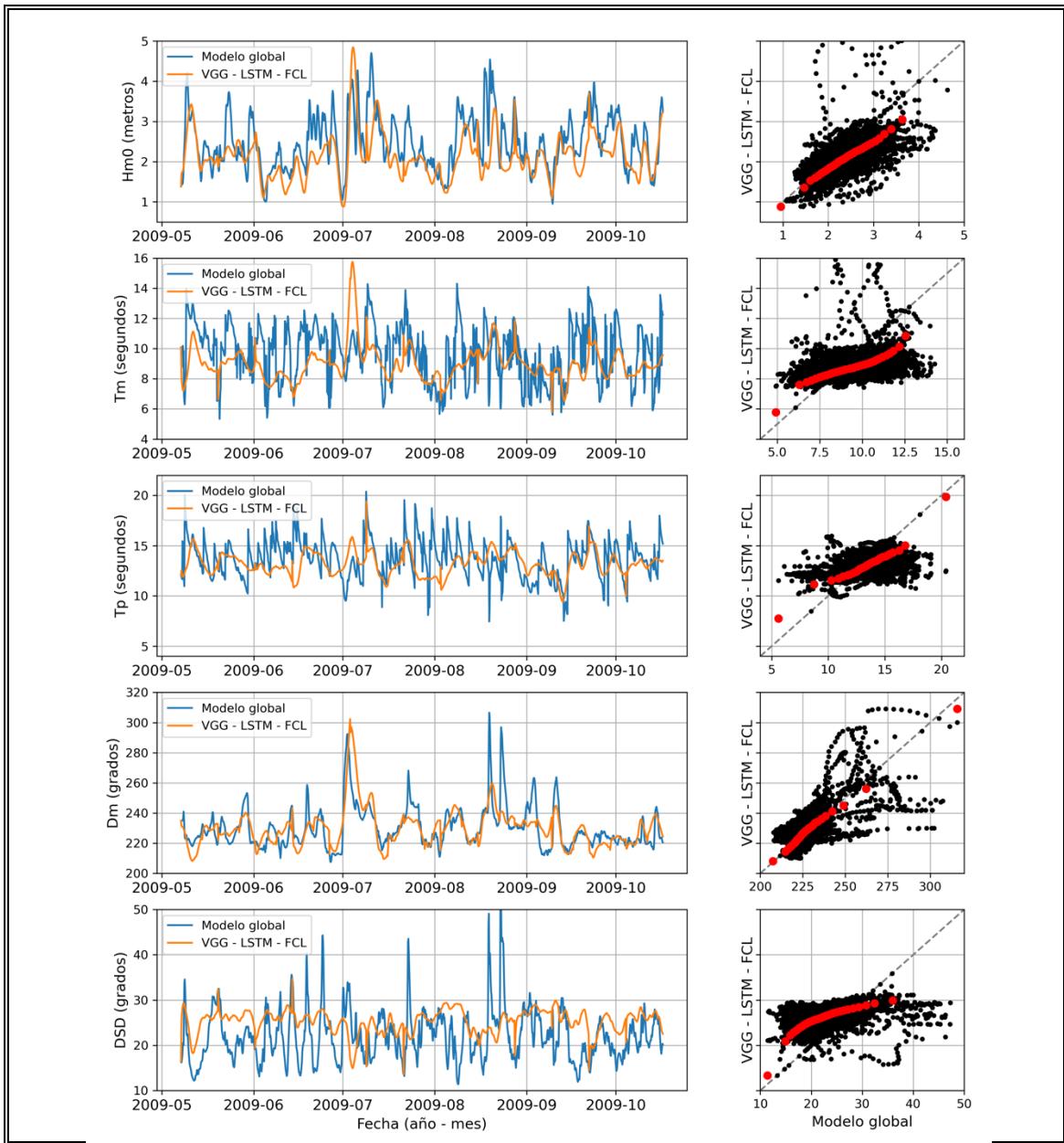


Figura A.14: Serie de tiempo de oleaje, gráfico dispersión y Q - Q (rojo) del modelo global contra estimaciones del VGG16 – LSTM – FCL *many to many* para cada parámetro resumen usando el *output* dentro del entrenamiento.

## ANEXO 10 : ESTIMACIÓN DE ESPECTRO DE OLEAJE COMPLETO USANDO REDES LSTM -FCL MANY TO ONE.

Se usan las redes LSTM – FCL *many to one* para estimar el espectro direccional de oleaje completo en lugar de estimar por separado cada parámetro resumen. Los resultados obtenidos, son comparados con los espectros registrados reales que se disponen. Los espectros se comparan por medio del cálculo de los parámetros que lo resumen (Anexo 2). De esta forma se construye la Tabla A.14, que corresponde a los parámetros resumen de oleaje calculados a partir del espectro direccional de energía estimado por los modelos LSTM – FCL y VGG16 – LSTM – FCL.

Tabla A.14: Calidad de ajuste del espectro completo de oleaje para distintos modelos.

	Hm0		Tm		Tp		Mean Dir		DSD	
	r	rmse (m)	r	rmse (s)	r	rmse (s)	r	rmse (grados)	r	rmse (grados)
LSTM - FCL	0.39	0.59	0.31	1.79	0.01	2.55	0.59	12.38	0.19	7.51
VGG16 - LSTM - FCL	0.80	0.38	0.62	1.38	0.32	2.14	0.74	8.82	0.54	6.80

Los resultados obtenidos de los parámetros calculados a partir del espectro estimado, son muy inferiores a los obtenidos al estimar cada parámetro resumen por separado. Es importante notar que el modelo VGG16 – LSTM – FCL muestra un rendimiento muy superior al modelo sin la incorporación del modelo VGG16. Por lo que se concluye que este modelo es el mejor para estimar el espectro direccional de energía de oleaje. Los resultados de la Tabla 4.3 muestran que la red logra extraer parámetros resumen a la hora de estimar todos los parámetros completos por medio de la estimación directa del espectro direccional de energía de oleaje. La incorporación de un modelo pre-entrenado VGG16, permite mejorar la estimación del período y las demás variables con una raíz de error cuadrático medio más baja. Sin embargo, no se alcanza a obtener estimaciones con correlación cercanas al 90%, por lo que el indicador de la correlación sigue estando alejado de un resultado satisfactorio.