

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO  
FACULTAD DE INGENIERÍA  
ESCUELA DE INGENIERÍA INFORMÁTICA

**PRONÓSTICO DE CAPTURAS DE ALTA Y BAJA  
FRECUENCIA DE SARDINAS USANDO REDES  
NEURONALES DE BASE RADIAL CON EXTREME  
LEARNING MACHINE**

JORGE CARVAJAL PEÑA

INFORME FINAL DEL PROYECTO  
PARA OPTAR AL TÍTULO  
PROFESIONAL DE  
INGENIERO CIVIL EN INFORMÁTICA

**DICIEMBRE, 2013**

Pontificia Universidad Católica de Valparaíso  
Facultad de Ingeniería  
Escuela de Ingeniería Informática

**PRONÓSTICO DE CAPTURAS DE ALTA  
Y BAJA FRECUENCIA DE SARDINAS  
USANDO REDES NEURONALES DE BASE  
RADIAL CON EXTREME LEARNING  
MACHINE**

**JORGE EDUARDO CARVAJAL PEÑA**

Profesor Guía: **Nibaldo Rodríguez Agurto**

Profesor Co-referente: **Wenceslao Palma Muñoz**

Carrera: **Ingeniería Civil Informática**

**Diciembre, 2013**

### *Dedicatoria*

Dedicada a mi familia por haber dado siempre lo mejor de sí para darme lo mejor, mis amigos por estar siempre ahí y ser mi equilibrio y a los profesores que creyeron en mí.

## *Agradecimientos*

Dios, Familia, a mis amigos y profesor guía.

## Resumen

En el presente proyecto se presentan los objetivos, estado de arte y modelos de red desarrollados para pronosticar el nivel de captura mensual de sardina en la zona norte de Chile, este pez es una de las especies pelágicas más importantes de nuestro país, en 1983 debido a la caída en la población de esta especie, la extracción de la sardina está sujeta a regulaciones. Es por ello que el pronóstico de capturas toma un rol muy importante en el nivel de stocks, previo a la toma de decisiones. El modelo usado para resolver este problema, es una Red Neuronal Artificial junto a la técnica Extreme Learning Machine, la cual está basada en la Pseudoinversa de Moore-Penrose.

La mejor topología de la red que presenta el mejor desempeño fue con 12 nodos de entrada y 100 iteraciones, para la calibración del modelo.

Finalmente, los resultados indican que el modelo basado en ELM presenta un desempeño de un 98,62% de la varianza explicada y un 14,61% de MAPE, con tiempos de ejecución notoriamente más bajos en comparación a otras técnicas utilizadas con anterioridad como la Inteligencia de Enjambres o Máquina de Soporte Vectorial.

**Palabras claves:** Redes Neuronales Artificiales, Extreme Learning Machine, Pseudoinversa de Moore-Penrose.

## Abstract

On this project the objectives, state of the art and developed network model are presented to forecast the monthly capture level of sardine in the north area of Chile, this fish is one of the most important pelagic species of our country, in 1983 due to the fall in the population of this species, the extraction of sardine is subject to regulations. That is why the forecast of catches taken a very important role in the level of stocks, before to the decision making. The model used to solve this problem, is an Artificial Neural Network with Extreme Learning Machine technique, based on Moore-Penrose Pseudoinverse.

The best topology that presents the best performance was 12 input nodes and 100 iterations, for model calibration.

Finally, the results indicate that model based on ELM presents a performance of 98,62% of the explained variance and 14,61% of MAPE, with lower runtime than others techniques previously used like Swarm Intelligence or Support Vector Machine.

**Keywords:** Artificial Neural Network, Extreme Learning Machine, Moore-Penrose Pseudoinverse.

# ÍNDICE

Índice .....	vi
Índice de Imágenes .....	ix
1. Introducción.....	1
1.1    Objetivos Generales.....	3
1.2    Objetivos Específicos .....	3
1.3    Organización del texto.....	3
2.    Pronóstico con Redes Neuronales.....	4
2.1    Áreas de Aplicación.....	5
2.2    Redes con Función de Base Radial.....	7
2.3    Arquitectura de las RBF .....	7
2.4    Aprendizaje de una RNA.....	9
2.4.1    Aprendizaje Supervisado.....	10
2.4.1.1    Modelo de aprendizaje supervisado según valor deseado .....	10
2.4.2    Aprendizaje No Supervisado .....	12
2.4.3    Aprendizaje Híbrido.....	12
3.    Extreme Learning Machine .....	13
3.1    Selección Óptima de Neuronas .....	14
3.2    Pseudoinversa de Moore-Penrose.....	14
3.3    Tipos de Extreme learning machine. ....	14
3.3.1    Extreme Learning Machine autoadaptativa Evolucionaria.....	15
3.3.2    Extreme learning Machine en línea secuencial.....	15
3.3.3    Extreme Learning Machine ponderado para el aprendizaje en desequilibrio.....	15
3.3.4    Extreme Learning Machine Robusto. ....	16
3.4    Aplicaciones de las ELM.....	16
4.    Modelo Propuesto. ....	17
4.1    Métricas De Evaluación.....	18
4.2    Selección de Topología y Parámetros del modelo.....	19
4.3    Resultados de Baja frecuencia.....	21
4.4    Resultados de Alta frecuencia .....	23
4.5    Resultados de las componentes sumadas.....	24

5. Discusión de resultados .....	27
6. Conclusiones.....	28
7. Referencias .....	30

## Lista de Abreviaturas o Siglas

- ELM: Extreme Learning Machine
- RBF: Redes Neuronales con Función de Base Radial
- MAPE: Porcentaje de Error Medio Absoluto
- RNA: Red Neuronal Artificial
- $R$ : Coeficiente de Correlación de Pearson
- $R^2$ : Coeficiente de Determinación



## ÍNDICE DE IMÁGENES

<b>Ilustración 1: Función de Base Radial</b> .....	7
<b>Ilustración 2: neurona gaussiana</b> .....	8
<b>Ilustración 3: Entrada – Valor deseado</b> .....	11
<b>Ilustración 4: Entrada suavizada – Valor deseado suavizado</b> .....	11
<b>Ilustración 5: Bloque del Modelo</b> .....	17
<b>Ilustración 6: Modelo propuesto</b> .....	18
<b>Ilustración 7: (a) Gráfico de Pronóstico vs Data Deseada Testing mejor resultado baja frecuencia. (b) Gráfico Correlación (c) Error relativo</b> .....	21
<b>Ilustración 8: (a) Gráfico de Pronóstico vs Data Deseada Testing Peor resultado baja frecuencia (b) Gráfico de Correlación (c) Error relativo porcentual.</b> .....	22
<b>Ilustración 9: (a) Gráfico de Pronóstico vs Data Deseada Testing mejor resultado Alta frecuencia (b) Gráfico de correlación (c) Error relativo.</b> .....	23
<b>Ilustración 10: (a) Gráfico de Correlación Peor resultado Alta frecuencia Testing. (b) Gráfico de correlación (c) Error relativo</b> .....	24
<b>Ilustración 11: (a) Gráfico de Pronóstico vs Data Deseada Testing mejor resultado Frecuencias sumadas (b) Gráfico de Correlación (c) Error Relativo</b> .....	25
<b>Ilustración 12: (a) Gráfico de Pronóstico vs Data Deseada Testing Peor resultado frecuencia sumada. (b) Gráfico de Correlación (c) Error relativo.</b> .....	26

## Tablas

<b>Tabla 1 Iteraciones para búsqueda de mejores resultados de las componentes sumadas</b> .....	20
<b>Tabla 2 Mejor resultado Baja Frecuencia</b> .....	21
<b>Tabla 3 Peor resultado Baja Frecuencia</b> .....	22
<b>Tabla 4 Mejor resultado Alta Frecuencia</b> .....	23
<b>Tabla 5 Peor resultado Baja Frecuencia</b> .....	24
<b>Tabla 6 Mejor resultado Frecuencias sumadas</b> .....	25
<b>Tabla 7 Peor resultado Frecuencias sumadas</b> .....	26

# 1. INTRODUCCIÓN

El sureste del Océano Pacífico, en especial, el mar del norte de Chile y sur del Perú, es un ecosistema altamente productivo en cuanto a especies pelágicas de refiere, dentro de las cuales se encuentra la especie Sardina - *Sardinops Sagax*.

Esta especie tuvo un ciclo de abundancia en la zona norte de Chile entre 1978 y 1991, como resultado de condiciones ambientales favorables para su reproducción, coincidiendo con un periodo decadal más cálido en la región, con un máximo de capturas de Chile durante 1984 de 2,6 millones de toneladas. A partir de 1992, comienza una declinación de su abundancia, la cual se prolonga hasta nuestros días, donde nuevamente la anchoveta ha sido la especie predominante en la región. Durante el periodo de mayor intensidad de este evento El Niño, los desembarques de esta especie se incrementaron levemente, sin ser relevantes en las capturas totales. Algo similar ocurre a fines del 98, sin que ello medie el efecto del Niño. Esta especie, al igual que el jurel, se ve favorecida por periodos cálidos prolongados. Pero en respuesta a los calentamientos de corto periodo como los asociados a El Niño, esta especie presenta un incremento en sus capturas, ya que el frente cálido que se genera actúa como barrera concentrando la población en la zona costera. Ello ocurrió en los eventos de El Niño anteriores, cuando esta especie era dominante en el ecosistema de la zona norte [36]. Como conclusión de esa investigación, las condiciones actuales de su biomasa tornó los desembarques relativamente menos importantes.

En este escenario, el pronóstico de las capturas es un tópico básico, debido a que juegan un papel fundamental en el papel de los stocks, previo a la toma de decisiones. Para alcanzar esta meta, es necesario pronosticar eventos incontrolables como la posible abundancia o aumento de la biomasa. La variabilidad de su nivel de abundancia se ve afectada directamente por una serie de variables ambientales, entre las cuales se encuentran: La corriente del Niño, temperatura superficial del mar y concentración de alimento (plancton).

Chile tiene acceso, junto con Perú, a uno de los cinco "oasis" pesqueros más importantes del mundo, es por ello, que cuenta con una gran dotación de recursos marítimos a lo largo de sus costas.

Con el fin de desarrollar políticas sostenibles de explotación, el pronóstico de la existencia y capturas de especies pelágicas en el norte de Chile es uno de los objetivos principales de la industria pesquera y el gobierno. Sin embargo, fluctuaciones en las variables del entorno complican esta tarea. A consecuencia de lo expuesto anteriormente, las disminuciones en las capturas totales de sardina afecta directamente a la producción y exportación de la especie, por lo que obtener una cuota anual de desembarques que se ajuste a la realidad del sector pesquero es gravitante para el futuro de la actividad. Para ello, es que la Subsecretaría de Pesca implementó en el año 2003, a través de una ley transitoria, un sistema de Cuotas Individuales y Transferibles de Pesca (CIT) en las principales pesquerías pelágicas y demersales del país, como una herramienta orientada a la conservación de los recursos, así como también, a la sustentabilidad de la actividad, por lo tanto, su determinación es un elemento relevante para el sector, con el objetivo de evitar el colapso biológico y económico de las unidades pesqueras.

Debido a esto, el proyecto tiene como objetivo principal modelar las capturas mensuales de la sardina, con el fin de elaborar un modelo predictivo que permita, a través de datos históricos, apoyar a las entidades gubernamentales en la fijación de las cuotas anuales de desembarques para la especie pelágica de sardina en la zona Norte del país y de esta manera, permitir la disponibilidad de la especie en el futuro.

En el presente proyecto se tomarán, como series temporales, los datos relacionados con los volúmenes de captura de sardinas en el norte de Chile y mediante el uso de Redes Neuronales (RBF) de Base Radial con aprendizaje basado en la técnica de Extreme Learning Machine.

Las Redes Neuronales Artificiales son modelos que intentan reproducir el comportamiento del cerebro humano [37]. Una red neuronal artificial consiste en un conjunto de elementos de procesamiento, llamados neuronas, los cuales se conectan entre sí. El modelo de la red neuronal artificial está basado una red neuronal biológica, su composición y el sistema se ven en mayor detalle en [3]

Existen distintas técnicas para entrenar una red neuronal, alguna implican un aprendizaje supervisado y otras uno no supervisado. En el presente caso de estudio, se realiza un aprendizaje supervisado de la red mediante la utilización de Extreme Learning Machine.

La utilización de esta técnica para definir el estado interno de la red se debe a sus grandes cualidades de rapidez, la posibilidad de no quedarse en un mínimo local, y la implementación del algoritmo LARS que selecciona las neuronas de una forma óptima, utilizando un modelo lineal simple.

Más adelante se interiorizará en el modelo propuesto en esta investigación, utilizando datos de alta y baja frecuencia previamente procesados a través de la transformada wavelet, en el que se expondrán los valores entregados a través de métricas para luego discutir los resultados entregados en la parte final del modelo, en el que las métricas surgirán a partir de la suma de los dos pronosticadores.

Finalmente, en el último capítulo, se revisarán las conclusiones de la investigación junto con las referencias utilizadas.

## 1.1 OBJETIVOS GENERALES

Desarrollar y evaluar un modelo de pronóstico para las capturas mensuales de Alta y Baja frecuencia de Sardina en la zona norte de Chile utilizando RBF NN con Máquina de aprendizaje Extremo (Extreme Learning Machine).

## 1.2 OBJETIVOS ESPECÍFICOS

- Explicar la estructura de RNA basada en Extreme Learning Machine.
- Diseñar y estimar la topología de una RNA usando Extreme Learning Machine.
- Evaluar el rendimiento del modelo de pronóstico propuesto usando las métricas residuales denominadas Raíz de Error Cuadrático medio, Coeficiente de Correlación de Pearson, Coeficiente de Determinación y Porcentaje del Error Medio Absoluto.

## 1.3 ORGANIZACIÓN DEL TEXTO

La presente memoria se divide en capítulos. En el capítulo 2 se destina a lo relacionado al Pronóstico con Redes Neuronales. Este capítulo está conformado por cuatro secciones, las que comprende a las Áreas de aplicación de las RNA, exponiendo investigaciones recientes en el área de los negocios, industria y ciencia, entre otros, Las RNA con función de base radial, donde se muestra la Arquitectura de las RBF que será utilizada en este trabajo, sus distintas formas y expresiones y finalmente el Aprendizaje de una RNA, en el que destacan el Aprendizaje Supervisado, Aprendizaje no Supervisado e híbrido.

En el Capítulo 3 se presenta lo relacionado al algoritmo Extreme Learning Machine. Este capítulo comienza con una sección introductoria de sus fundamentos. Sus cuatro secciones comprenden la Selección Óptima de Neuronas, la Pseudoinversa de Moore-Penrose, la que se utiliza en el modelo propuesto en este trabajo, algunos tipos de ELM que se encuentran en la literatura y Aplicaciones de las ELM.

En el Capítulo 4 se presenta el Modelo Propuesto, y en sus secciones los puntos relevantes de esta investigación: las métricas residuales utilizadas, la Selección de Topología y parámetros del modelo y los resultados entregados por el modelo.

En el Capítulo 5 se encuentra la discusión de los resultados entregados por el modelo, y finalmente en el Capítulo 6 las Conclusiones.

## 2. PRONÓSTICO CON REDES NEURONALES

Predecir es hacer declaraciones sobre algo que va a suceder, a menudo se basan en información del pasado y del estado actual. Por ejemplo, cada persona hace diversos pronósticos en su diario vivir, ya sea sobre el clima, el tiempo se demora en llegar a un cierto destino, etcétera.

En el aspecto técnico, los parámetros predecibles de un sistema generalmente pueden ser expresados y evaluados por medio de ecuaciones, por lo tanto, la predicción es simplemente la evaluación o solución de una posible ecuación. Sin embargo, existen problemas muy complejos y quizás imposibles de solucionar. Además, la solución computacional podría ser muy complicada y el resultado podría ser generado una vez que ya sucedió el evento a predecir.

Las redes neuronales artificiales corresponden a un modelo de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso en los animales. Consiste en un sistema de interconexión de neuronas artificiales formando una red que colabora para producir un estímulo de salida. Estas neuronas corresponden a unidades de procesamiento que se comunican unas a otras mediante el envío de señales por medio de una serie de conexiones las cuales tienen un peso asociado.

Una de las mayores aplicaciones de las RNAs es el pronóstico. Las RNAs proveen una atractiva herramienta alternativa para los investigadores ya que estas poseen varias características que las hacen valiosas para las tareas de predicción. Primero, al contrario de los métodos tradicionales estadísticos, las RNAs son métodos adaptativos basados en datos, donde a veces se tiene muy poco conocimiento del problema bajo estudio. Ellas aprenden en base a ejemplos capturando relaciones funcionales sutiles entre los datos aun cuando estas relaciones sean desconocidas o difíciles de describir. Además son muy adecuadas para aquellos problemas cuyas soluciones requieren un conocimiento difícil de especificar, pero para los cuales existen suficientes datos u observaciones.

Segundo, las RNAs pueden generalizar, luego de la fase de entrenamiento pueden correctamente inferir el futuro de la población aun cuando la data contenga información ruidosa. Como el pronóstico se realiza vía la predicción del comportamiento futuro en base al comportamiento pasado, el pronóstico es una aplicación ideal para las RNAs al menos en principio.

Tercero, las RNAs son aproximadores funcionales universales. La experiencia de trabajos previos ha demostrado que una red puede aproximar cualquier función continua a una precisión deseada. Las RNAs tienen una estructura más flexible que los métodos estadísticos tradicionales. Cualquier modelo de pronóstico asume que existe una relación fundamental entre las variables de entrada (valores pasados de una serie de tiempo en este caso u otras variables relevantes) y las salidas (valores futuros). Frecuentemente los métodos estadísticos tradicionales tienen limitaciones a la hora de realizar las estimaciones de estas relaciones fundamentales debido a la complejidad de los problemas reales.

Por último las RNAs no son lineales, por mucho tiempo el pronóstico ha estado dentro del dominio de la estadística lineal, los cuales asumen que las series de tiempo bajo estudio son generadas por procesos lineales. Los procesos lineales tienen la ventaja de que pueden ser entendidos y analizados en gran detalle, lo que los hace fáciles de explicar e implementar. Sin embargo pueden ser inapropiados si el proceso fundamental bajo estudio es no lineal. No es razonable a priori asumir que la realización de cierta serie de tiempo es generada por un proceso lineal. De hecho en la vida real los sistemas a menudo son no lineales.

Las redes neuronales convencionales (NN) y la máquina de vector de soportes (SVM) son dos de las mayores técnicas de inteligencia computacional, que ha sido popular en las pasadas dos o tres décadas. Especialmente en la pasada década, SVM ha jugado un rol dominante en aplicaciones de clasificación [26]. De hecho, SVM puede ser considerado como uno de los métodos de aprendizaje de redes neuronales feedforward. [27]. Como trabajo de tesis para optar por el título en esta casa de estudio, se han realizado investigaciones relacionadas con SVM [29,34,35]. En [28], utilizando Máquinas de Soporte Vectorial y PSO obtuvo un 93,91% de la varianza explicada. Investigaciones relacionadas con NN [31,32,33] en el ámbito de la predicción de especies pelágicas en Chile, como ejemplo en [30] Los resultados obtenidos mediante el modelo de red Neuro-Wavelet con Algoritmo Genético presenta una explicación de la varianza de hasta un 85 %, resultado que alcanza un 86% al aplicar pseudoinversa para determinar los pesos ocultos o lineales de la red, mismo pseudoinversa utilizada en esta investigación, y [34,36] en relación a pronóstico de otro tipo de especies, sin embargo la mayor parte de las investigaciones relacionadas al pronóstico han sido de anchovetas.

## 2.1 ÁREAS DE APLICACIÓN

Recientes actividades de investigación en Redes Neuronales Artificiales han mostrado que las RNAs tienen gran poder en la clasificación de patrones y capacidad de reconocimiento de patrones. Inspirados en sistemas biológicos, particularmente por la investigación del cerebro humano, las RNAs son capaces de aprender y generalizar desde la experiencia, dado a la intromisión al problema del reconocimiento de patrones que fue introducido en [4]. Actualmente, las RNAs están siendo utilizadas para una extensa variedad de tareas en muchos campos diferentes del área de los negocios, industria y ciencia [38].

Una de las primeras aplicaciones exitosas en el pronóstico con redes neuronales, es reportada por [46] Usando dos series de tiempo caótico determinista, generados por el mapa logístico y la ecuación Glass-Mackey, ellos diseñaron la red neuronal feed-forward, las que pueden imitar con precisión y predecir el comportamiento no lineal del sistema. Sus resultados muestran que las redes neuronales pueden ser usadas para modelar y pronosticar series de tiempo con alta precisión.

Por otro lado, las series de manchas solares han servido como punto de referencia y han sido bien estudiadas en la literatura estadística. Ya que los datos que son no lineales, no estacionarios y no gaussianos, a menudo se utilizan como un criterio para evaluar y comparar los nuevos métodos de predicción. Algunos autores se centran en el uso de redes neuronales para mejorar la precisión en la predicción de las actividades de las manchas solares a través de los métodos tradicionales, mientras que otros utilizan los datos para ilustrar un método.

Otro campo en que el pronóstico de las redes neuronales es usado, es la economía, ya que han sido utilizadas para prevenir quiebres de empresas, comportamiento del precio de acciones, para saber los posibles tipos de cambio. En [15], se persigue evaluar la rentabilidad que habría obtenido un inversionista si hubiese utilizado la información obtenida por las redes neuronales para la conformación de sus carteras de bonos, definiéndose que es una herramienta poderosa para lapsos anuales, y define a su vez, que todas las variables relacionadas con la rentabilidad (el monto transado, la tasa interbancaria, el benchmarck rezagado) son factores determinantes, señalando en que el uso de las redes neuronales tienen propiedades interesantes que pueden ser utilizadas en este ambiente.

La estrategia energética de un país es extremadamente importante para su desarrollo, sobre todo para aquellos países en vías de desarrollo, debido a que su economía va creciendo rápidamente la demanda por el consumo energético también aumenta, [40] afirman que una realizar un pronostico acertado acerca del consumo en el futuro permitiría a aquellos que hacen las políticas energéticas poder elaborar estrategias críticas para enfrentar esta problemática. En su experimento utilizaron una red neuronal de retropropagación obteniendo un MAPE de 3.8% para los datos de prueba.

En una rama mucho menos explorada varios investigadores han estado utilizando redes neuronales para poder predecir la magnitud de los terremotos, [39] proponen un modelo basado en redes neuronales recurrentes con capas ocultas múltiples, para pronosticar las magnitudes de futuros terremotos utilizando las magnitudes de eventos telúricos pasados en el Norte del Mar Rojo, península del Sinai, el golfo de Aqaba y golfo de Suez. En sus resultados afirman que su red neuronal obtuvo una ganancia del 32% respecto a otros métodos estadísticos tradicionales.

Dado a que nos encontramos en un país, en el cual sus recursos naturales son vitales para el crecimiento del país, es donde el pronóstico del precio de estos mismos los vuelve una herramienta poderosa, ya que pronosticando el precio se pueden realizar estimaciones de elaboraciones de presupuesto nacional, en función de un precio futuro, es así como en Pronóstico del precio del cobre mediante redes neuronales [47] se realiza un estudio en función del impacto del precio del cobre en la actividad minera y en la economía chilena, en el cual, como resultados, le entregan un MAPE de un 15 %.

En [41-44] se encuentran investigaciones aplicadas utilizando redes neuronales para la predicción en distintos ámbitos.

## 2.2 REDES CON FUNCIÓN DE BASE RADIAL

Las redes RBF son redes multicapa feed-forward formadas por una única capa oculta, donde cada una de las neuronas presentan un carácter local, es decir, se activa en una región diferentes del espacio de patrones de entrada. Este carácter se debe al uso de funciones de base radial como por ejemplo: la función gaussiana. Las neuronas de la capa de salida realizan una combinación lineal de las activaciones de las capas ocultas.

Estas redes nacen tras la necesidad de construir una red de neuronas que requiere un menor tiempo de aprendizaje que el perceptrón multicapa de manera que sea apropiada para aplicaciones en tiempo real. Esto se consiguió pues mediante las activaciones locales pocas neuronas ocultas tendrían que ser procesadas para nuevos patrones de entrada.

La función de base radial define hiper esferas que dividen el espacio de entrada de tal manera que cada neurona de la capa oculta construye una aproximación local y no lineal en una determinada región de dicho espacio. Dado que la salida es una combinación lineal de las funciones de base radial, las aproximaciones que construyen las redes RBF son combinaciones lineales de múltiples funciones locales y no lineales. De esta manera, las redes RBF aproximan relaciones completas por medio de una colección de aproximaciones locales menos complejas, dividiendo el problema en varios sub-problemas menos complejos.

## 2.3 ARQUITECTURA DE LAS RBF

Las redes RBF se encuentran formadas por 3 capas:

- **Capa de Entrada:** reciben las señales desde el exterior y las transmiten a la siguiente capa sin realizar ningún procesamiento sobre estas señales, además no presentan ningún peso asociado.
- **Capa Oculta:** reciben las señales de la capa de entrada y realizan la transformación local y no lineal de la señal. Esta capa es la única que contiene componentes no lineales.
- **Capa de Salida:** actúa como salida de la red realizando una combinación lineal de las activaciones de las neuronas ocultas. Solo las neuronas de salida poseen umbral.

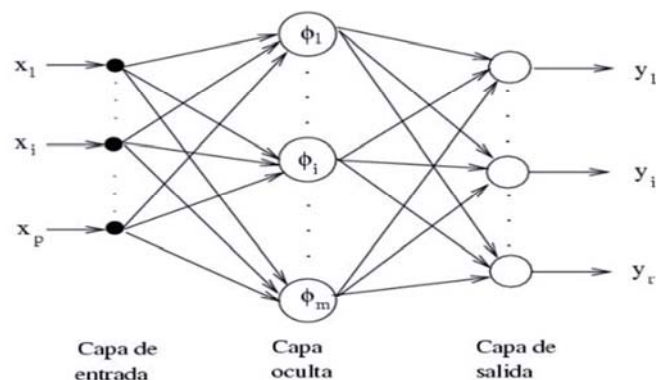


Ilustración 1: Función de Base Radial



En una red RBF, como la figura, con  $p$  neuronas en la capa de entrada,  $m$  en la capa oculta y  $r$  en la de salida, las activaciones de las neuronas de salida para el patrón de entrada  $n$ ,  $X(n) = (x_1(n), x_2(n), \dots, x_p(n))$  denotadas como  $y_k(n)$ , vienen dadas por la siguiente expresión:

$$y_k(n) = \sum_{i=1}^m w_{ik} \varphi_i(n) + b_k \text{ para } k = 1, 2, \dots, r \quad (2.1)$$

Donde  $w_{ik}$  representa a los pesos de la conexión de la neurona oculta  $i$  a la neurona de salida  $k$   $b_k$  corresponde al umbral de la neurona de salida  $k$  y  $\varphi_i(n)$  son las activaciones de las neuronas ocultas para el patrón de entrada  $X(n)$ .  $\varphi_i$  viene dada por la siguiente expresión.

$$\varphi_i(n) = \varphi\left(\frac{\|X(n) - C_i\|}{d_i}\right) \text{ para } i = 1, 2, \dots, m \quad (2.2)$$

Donde  $\varphi$  es la función de base radial,  $C_i = (C_{i1}, \dots, C_{ip})$  son vectores que representan los centros,  $d_i$  son números reales que representan desviación, anchura o dilatación y  $\| \cdot \|$  corresponden a la distancia euclídea del vector de entrada  $X(n)$  al centro  $C_i$  definida como

$$\|X(n) - C_i\| = \left(\sum_{j=1}^p (x_j(n) - c_{ij})^2\right)^{\frac{1}{2}} \quad (2.3)$$

La función de base radial puede adoptar distintas formas y expresiones, entre ellas se encuentran:

- Función Gaussiana:

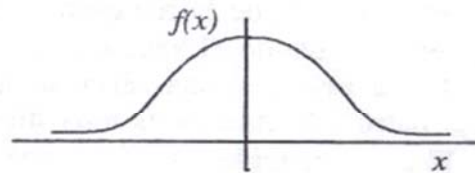
$$\varphi(r) = e^{\left(\frac{-r^2}{2}\right)} \quad (2.4)$$

- Función inversa cuadrática:

$$\varphi(r) = \frac{1}{1+r^2} \quad (2.5)$$

- Función inversa multicuadrática:

$$\varphi(r) = \frac{1}{\sqrt{1+r^2}} \quad (2.6)$$



*Capa oculta: neurona gaussiana*

**Ilustración 2: neurona gaussiana**

## 2.4 APRENDIZAJE DE UNA RNA.

En [1] el aprendizaje se define como un proceso en el cual los parámetros libres de una red neuronal son adaptados a través de un proceso de estimulación por el entorno en el cual se encuentra inserta la red. El tipo de aprendizaje es determinado por la manera en la cual los parámetros alterados toman lugar. Esta definición del proceso de aprendizaje implica la siguiente secuencia de eventos:

- La red neuronal es estimulada por el entorno
- La red neuronal realiza cambios en sus parámetros libres como resultado de la estimulación.
- La red neuronal responde de una nueva forma al entorno debido a los cambios que han existido en su estructura interna.

La interpretación de estos puntos, es que la red neuronal cambia su estructura interna en respuesta a una información de entrada, y estos cambios producen los tipos de conexiones de las neuronas.

Antes de comenzar el proceso de entrenamiento se debe determinar un estado inicial, lo cual significa, escoger un conjunto inicial de pesos para las diversas conexiones entre las neuronas de la red neuronal. Uno de los criterios a emplear consiste en otorgar valores aleatorios a las conexiones, encontrándose estos dentro de un intervalo  $[-n, n]$ , donde  $n$  es un número natural positivo. Cabe mencionar que durante el proceso de entrenamiento los pesos no se encontraran restringidos a éste intervalo.

Por otro lado, para determinar cuándo termina el proceso de aprendizaje es necesario establecer una condición de detención. Típicamente se distinguen 3 tipos de criterios:

- Verificación del error de los resultados: Se define una función o métrica que sentencie que los resultados son positivos, y se entrena hasta llegar a un rango definido.
- Resultado de entrenamiento irrelevante: Resulta cuando los resultados entre iteración son muy parecidos sino iguales.
- Alcance de un número fijo de ciclos: Se define un número determinado de iteraciones independiente de la verificación del error.

Un aspecto importante respecto al aprendizaje de las redes neuronales es el conocer cómo se modifican los valores de los pesos, es decir, cuáles son los criterios que se siguen para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información.

Existen 3 métodos de aprendizaje importantes que pueden distinguirse [45]:

- Aprendizaje Supervisado
- Aprendizaje No Supervisado
- Aprendizaje Híbrido

## 2.4.1 APRENDIZAJE SUPERVISADO

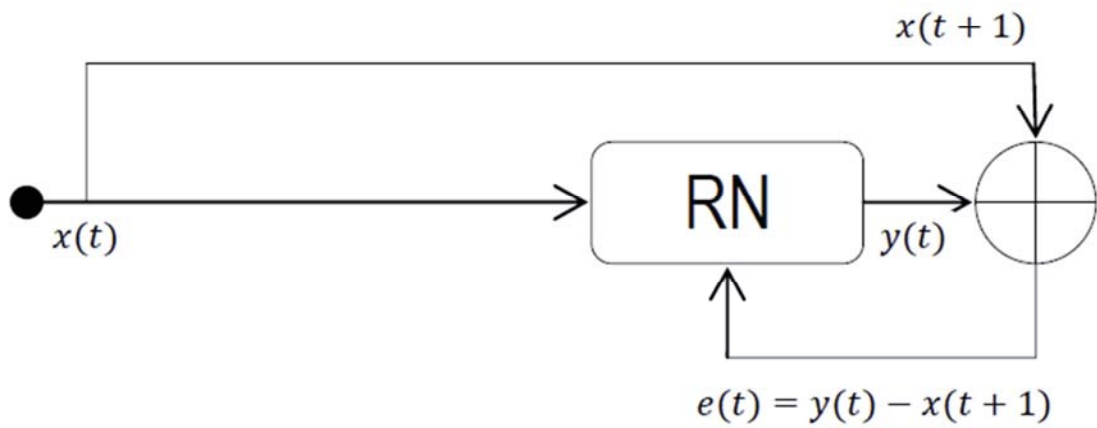
Se caracteriza porque el proceso de aprendizaje se encuentra controlado por un agente externo (supervisor, maestro) el cual determinan cuales debieran ser las respuestas entregadas por la red dada una entrada determinada. Este agente externo controla la salida y en caso de no corresponder a la deseada, se procederá a modificar los pesos de las conexiones, con el fin de que la salida obtenida se aproxime a la deseada.

Se suelen considerar 3 formas de llevar a cabo este tipo de aprendizaje, dando lugar a los siguientes aprendizajes supervisados.

- Aprendizaje por corrección de error: Ajustar los pesos de la red en función de la regla del mínimo error cuadrado (LMS error) o regla delta, que además de utilizar la desviación a la salida objetivo toma en consideración a todas las neuronas predecesoras que tienen la neurona de salida, cuantificando el error global y acelerando el aprendizaje (debido a que se tiene más información). También, se debe mencionar a la regla de aprendizaje hacia atrás o de backpropagation, la cual consiste en una generalización de la regla Delta y fue la primera que permitió realizar cambios sobre los pesos en las conexiones de la capa oculta [5].
- Aprendizaje por esfuerzo: Tipo de aprendizaje más lento que el antes mencionado. Se ajustan los pesos de la red basándose en probabilidades.
- Aprendizaje estocástico: Asocia a la red con un sólido físico que tiene cierto estado energético, de tal forma que un estado de mínima energía correspondería a una situación en que los pesos de las conexiones consiguen que su funcionamiento sea el que más se ajusta al objetivo deseado. De esta manera, el aprendizaje consistiría en asignar valores aleatorios a los pesos y determinar la energía de la red. Si el valor energético es menor después del cambio (comportamiento se acerca al deseado) se acepta el cambio, de lo contrario se aceptaría el cambio en función de una determinada y preestablecida función de probabilidades.

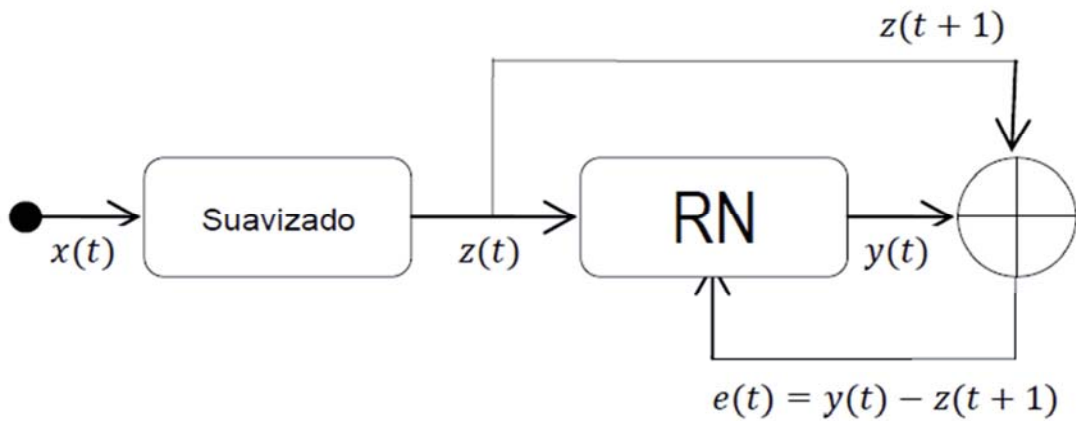
### 2.4.1.1 MODELO DE APRENDIZAJE SUPERVISADO SEGÚN VALOR DESEADO

En pronósticos se suele usar técnicas de suavizado de datos para reducir el efecto de los valores atípicos (outliers). La implementación de modelos como el de la Figura 3.4, en que tanto para la entrada como para el valor deseado se utilizan datos crudos, suele ser de difícil pronóstico dada la existencia de valores de comportamiento imprevisible.



**Ilustración 3: Entrada – Valor deseado**

Para reducir este efecto, suelen usarse técnicas de suavizado que se aplican a la señal, acercándola a la tendencia. Esto da lugar a modelos como el de la Figura 4.4, en que tanto la entrada como el valor deseado corresponden a data suavizada.



**Ilustración 4: Entrada suavizada – Valor deseado suavizado**

Existen modelos mixtos en que las entradas son suavizadas pero para el valor deseado se utilizan datos crudos.

## 2.4.2 APRENDIZAJE NO SUPERVISADO

Este tipo de aprendizaje se caracteriza por que la red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta de una determinada entrada es o no es correcta.

Estas redes deben encontrar características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presentan en su entrada. Existen varias posibilidades en cuanto a la interpretación de la salida de estas redes, que depende de su estructura y algoritmo de aprendizaje utilizado.

En cuanto a los algoritmos de este tipo de aprendizaje, en general se suelen utilizar 2 tipos que dan lugar a los siguientes aprendizajes:

- Aprendizaje Hebbiano. [2] Corresponde a una regla de aprendizaje que pretende medir la familiaridad o extraer características de los datos de entrada. Consiste en que si 2 neuronas toman el mismo estado simultáneamente (ambas activas o ambas inactivas, dado que el origen de la regla es en base a la neurona biológica clásica), el peso de la conexión entre ambas se incrementa.
- Aprendizaje Competitivo y Comparativo. Si un patrón nuevo se determina que pertenece a una clase previamente reconocida, entonces la inclusión de este nuevo patrón a esta clase matizará la representación de la misma. En caso contrario, la estructura y los pesos de la red serán ajustados para reconocer la nueva clase.

En el caso de las funciones de base radial, usualmente se utilizan algoritmos no supervisados para la obtención de los centros y las anchuras de las funciones de base radial.

## 2.4.3 APRENDIZAJE HÍBRIDO

Como su nombre lo indica, corresponde a la mezcla de los dos tipos de aprendizaje. Un claro ejemplo se observa en las redes de base radial, las cuales utilizan algoritmos que ayudan a determinar su configuración óptima de centros y anchura, independiente del algoritmo de aprendizaje. A su vez, ELM utiliza un algoritmo para determinar la selección óptima de neuronas (LARS) para la determinación de los pesos ocultos aleatorios.

En [50] se muestran investigaciones relacionadas en que mezclan la inteligencia de partículas (PSO) con ELM, en que utiliza las ventajas del PSO para enfrentar de mejor manera los inconvenientes de las NN. Se utiliza PSO para los pesos de entrada y centros ocultos y la Pseudoinversa para determinar los pesos ocultos. En este se observa que optimiza los resultados y tiene una mejor generalización que la ELM tradicional.

### 3. EXTREME LEARNING MACHINE

La necesidad de buscar nuevas técnicas de aprendizaje surge de los inconvenientes que muestran este tipo de técnicas antes mencionadas:

- Baja rapidez de aprendizaje.
- Parámetros ajustados por humanos triviales.
- Variante de aprendizajes triviales para diferentes aplicaciones de regresión, binaria, y clasificación multi-clase.

Extreme Learning Machine (ELM) [8,12,25] intenta superar estos inconvenientes. Trabajos recientes [6,7] han mostrado que supera a las técnicas mencionadas en términos de velocidad de aprendizaje y rendimiento de generalización. La esencia de ELM es que los parámetros de la capa oculta pueden ser independientes del entrenamiento, y la capa oculta de la generalización no necesita ser ajustada.

El algoritmo Extreme Learning Machine (ELM) está fundamentado en que un MLP compuesto por  $H$  neuronas, cuyos pesos de entrada están inicializados aleatoriamente, pueden “aprender”  $N$  distintos casos de entrenamiento produciendo un error cero, siendo  $N \geq H$ , y aproximar cualquier tipo de función continua. Tras inicializar de manera aleatoria los pesos de entrada, un MLP puede ser considerado como un sistema lineal y los pesos de salida pueden obtenerse de manera analítica mediante un simple cálculo de la pseudoinversa de la matriz de las salidas de las  $H$  neuronas ocultas para un determinado conjunto de entrenamiento. Así, dado un conjunto de  $N$  vectores de entrada, un MLP con  $H$  neuronas ocultas puede aproximar estos  $N$  casos con error nulo:

$$\sum_{i=1}^n \|y_i - t_i\| = 0 \quad (3.1)$$

Por ejemplo existen  $\beta_j, w_j$  y  $b_j$  tal que :

$$\sum_{j=1}^n \beta_j f(w_j \cdot x_i + b_j) = t_i, i = 1, \dots, N \quad (3.2)$$

Las anteriores  $N$  ecuaciones se pueden expresar  $H\beta = T$  donde  $H \in R^{N \times H}$  es la matriz de salidas de la capa oculta de neuronas de MLP,  $\beta \in R^{H \times m}$  es la matriz de pesos de salida, y  $T \in R^{N \times m}$  es la matriz de “targets” de los  $N$  casos de entrenamiento.

De esta forma, el entrenamiento del MLP viene dada la solución del problema de mínimos cuadrados establecido en (3), es decir, los pesos óptimos de la capa de salida son  $\hat{\beta} = H \dagger T$ , donde  $H \dagger$  es la pseudoinversa de Moore-Penrose. ELM proporciona un entrenamiento rápido y eficiente para MLPs, aunque es necesario preestablecer el número de neuronas ocultas. A continuación se describe un procedimiento para la selección automática del número de neuronas.

### 3.1 SELECCIÓN ÓPTIMA DE NEURONAS

El método Optimal Pruned ELM (OP-ELM) establece un número inicial muy elevado de neuronas ocultas ( $H \gg N$ ), y mediante el algoritmo Least Angle Regression (LARS) [10] elimina/poda aquellas variables (neuronas) que no son útiles para resolver el problema de mínimos cuadrados. Para ello ordena el conjunto posible de neuronas, conforme a su importancia para resolver. La solución obtenida por LARS es única cuando el problema es lineal.

La poda de neuronas es realizada mediante validación cruzada del tipo Leave-One-Out (LOO), escogiendo aquella combinación de neuronas (que han sido previamente ordenadas mediante el algoritmo LARS) que proporciona menor error de validación [9]. Por tanto únicamente son seleccionadas las neuronas más importantes según LARS, y junto con el entrenamiento ELM, se obtiene una solución única para el diseño de un MLP.

### 3.2 PSEUDOINVERSA DE MOORE-PENROSE.

Dada una matriz  $A \in M_{m \times n}(K)$ , una inversa generalizada de Moore-Penrose de  $A$  es una matriz  $X$  de orden  $n \times m$  de forma que:

$$AXA = A. \quad (3.3)$$

$$XAX = X. \quad (3.4)$$

$$AX \text{ y } XA \text{ son simétricas.} \quad (3.5)$$

#### **Proposición:**

Para cada matriz  $A$ , si existe inversa de Moore-Penrose de  $A$ , esta es única. A tal matriz se le nota por  $A^+$ . Observar que si  $A$  es regular, entonces  $A^+ = A^{-1}$ , si  $A$  es de rango pleno por filas, entonces  $A^+ = A^R$  y si  $A$  es de rango pleno por columnas, entonces  $A^+ = A^L$ .

Dada una matriz  $A$  de orden  $m \times n$ , llamaremos factorización de rango pleno de  $A$  a cada descomposición de  $A$  en producto de una matriz  $E$  de rango pleno por columnas y una matriz  $F$  de rango pleno por filas. Toda matriz posee una factorización de rango pleno,  $A = EF$ .

### 3.3 TIPOS DE EXTREME LEARNING MACHINE.

En [9,13,14,18,16,22] se encuentran investigaciones y modelos propuestos en la última década. Con la intención de mostrar las características fundamentales y a qué problema atacan en particular, se destacan los siguientes:

### 3.3.1 EXTREME LEARNING MACHINE AUTOADAPTATIVA EVOLUCIONARIA

En [18] se propone SaE-ELM (Self adaptative Evolutionary ELM) con el fin de superar las limitaciones de elegir manualmente una estrategia de generación de vectores de prueba y sus parámetros asociados en las versiones básicas. Se incluye este algoritmo con el fin de optimizar las ponderaciones de entrada a la red y la descentralización de los nodos ocultos, y a su vez se aplica la ELM para obtener los pesos de salida de la red.

### 3.3.2 EXTREME LEARNING MACHINE EN LÍNEA SECUENCIAL

OS-ELM (Online Sequential ELM) [14], es una variación de ELM que puede aprender la data de entrenamiento no solo uno-por-uno o parte-por-parte y descartar la data que ya se haya hecho. Se trata de un algoritmo de aprendizaje secuencial versátil con las siguientes características:

- Los datos observados de entrenamiento son secuenciales (uno-por-uno o parte-por-parte con variaciones o partes de largo fijo).
- En cualquier momento, solo el recién llegado o trozo de observaciones (en lugar de la totalidad de los datos pasados, en este caso puede ser la data de testing) se ven y son aprendidos.
- Un único o una parte de las observaciones de entrenamiento se descarta tan pronto como el procedimiento de aprendizaje para esas mismas partes está completo.
- El algoritmo de aprendizaje no tiene conocimiento previo en cuanto a cómo se presentarán muchas observaciones de entrenamiento.

En la investigación realizada por [17] el propósito de este algoritmo es la evaluación por comparación con otros tipos de algoritmos de aprendizaje secuencial.

Este algoritmo propuesto se utiliza en experimentos de problemas de benchmark de regresión, clasificación y series temporales, los cuales producen mejores rendimientos de las generalizaciones y un aprendizaje rápido.

### 3.3.3 EXTREME LEARNING MACHINE PONDERADO PARA EL APRENDIZAJE EN DESEQUILIBRIO

En comparación a la ELM mostrada en la primera parte de la investigación, en la cual es una ELM no ponderada, la ELM ponderada, está enfocada a ser un clasificador binario, en donde, dado un conjunto de datos de entrenamiento  $[x_j, t_j], i = 1, \dots, N$  pertenecientes a dos clases, donde  $T_i$  es +1 o -1 para indicar la clase positiva o negativa de la clase, definimos un  $N \times N$  matriz diagonal  $W$  asociados con cada muestra de entrenamiento  $x_i$ . Por lo general, si  $x_i$  proviene de una clase minoritaria (que se supone ser de clase positivo), el correspondiente peso  $W_{ii}$  es relativamente mayor que otras. Para maximizar la distancia marginal y para minimizar el error acumulativo ponderado con respecto a cada muestra, se tiene un problema



de optimización matemáticamente como minimizar:  $\|H\beta - T\|^2$  y  $\|\beta\|$  donde  $T = [t_1, \dots, t_N]$  y más precisamente minimizar:

$$Lp_{ELM} = \frac{1}{2} \|\beta\|^2 + CW \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2 \quad (3.6)$$

Sujeto a:

$$h(x_i)\beta = t_i^T - \xi_i^T, i = 1, \dots, N \quad (3.7)$$

Para recordar,  $h(x_i)$  es la función de mapeo de vectores en la capa oculta con respecto a  $x_i$  y  $\beta$  representa el vector de pesos de salida que conecta la capa oculta y la capa de salida. Dado a que es un clasificador binario, sólo hay un nodo en la capa de salida.  $\xi_i$ , el error de entrenamiento de la muestra  $x_i$  es causado por la diferencia de la salida deseada  $t_i$  y la salida real  $h(x_i)\beta$ .

Finalmente, aplicando pseudoinversa, como se muestra en más detalle en [16], se obtiene la función de salida para una matriz de tamaño  $L \times L$ .

$$f(x)_{N \times N} = \text{sign } h(x) H^T \left( \frac{1}{C} + WHH^T \right)^{-1} WT \quad (3.8)$$

$$f(x)_{L \times L} = \text{sign } h(x) \left( \frac{1}{C} + H^TWH \right)^{-1} H^TWT \quad (3.9)$$

### 3.3.4 EXTREME LEARNING MACHINE ROBUSTO.

La investigación realizada por [23] identifica puntos importantes a destacar, en cuanto ataca los problemas más comunes de ELM, los problemas computacionales que aparecen cuando la matriz de salida de la capa oculta no tiene su rango de columnas completo de la matriz o una matriz condicionada por la generación aleatoria de los pesos de entrada y centros y los problemas de robustez con outliers, que pueden ocurrir cuando la data de entrenamiento se contamina con outliers lo que hace que el radio de precisión se vea extremadamente afectado.

## 3.4 APLICACIONES DE LAS ELM

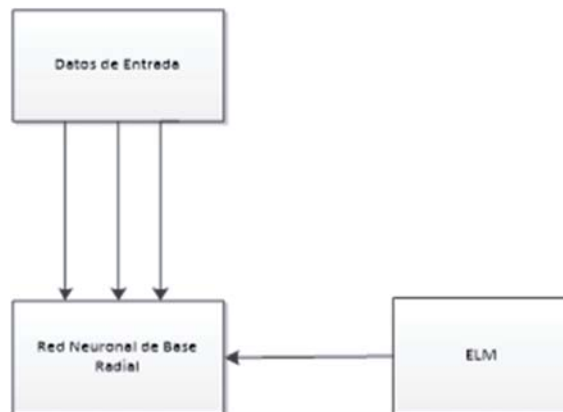
Al igual que las redes neuronales, las aplicaciones de las ELM son incontables. En [24] se realiza una investigación en relación a la predicción de energías de excitación de colorantes fluorescentes. [19] crean una herramienta para el análisis de Proteínas y Genoma humano. [20] presentaron un sistema automático de reconocimiento de objetivos basados en ELM wavelet, utilizando señales de eco de radares de rango de alta resolución, en el que se utiliza las capacidades de la ELM para mejorar iterativamente todos los parámetros de la red utilizando estos algoritmos de aprendizaje.

A su vez, sus características de clasificación se aplican en reconocimiento facial y clasificación de imágenes.

[49] exploró en la apreciación de la calidad de imagen. Los mecanismos de control de calidad son importantes en orden de proveer una experiencia visual satisfactoria en dispositivos electrónicos multimedia.

## 4. MODELO PROPUESTO.

El modelo propuesto considera dos fases para su elaboración, el algoritmo ELM que se comunicará con la red neuronal de base radial para que, a través de un número de iteraciones por definir por el usuario a modo de calibración, calcula los datos pronosticados estimados para los datos de Alta y Baja frecuencia. Como fitness, se utilizó la métrica del MAPE escalonado para buscar la mejor data estimada.



**Ilustración 5: Bloque del Modelo**

Una vez que los datos se desfazan (en el caso de esta investigación con un desfase de 12 meses), se aplicará la ELM para los datos de alta y baja frecuencia. En ese proceso, la ELM determinará los pesos de salida, o valores estimados por la red de base radial, realizándose pruebas con la función de activación gaussiana, inversa cuadrática e inversa multicuadrática. Más adelante se optará solo por la función de activación gaussiana, al observarse que los resultados entregados son relativamente similares.

La estructura de la ELM, se basa en la investigación de [48], donde se utilizan aproximadores universales como aproximadores de funciones.

Con los números de entradas definidos se inicializan los centros de la función de base radial de forma aleatoria, los cuales serán del mismo largo que los datos de entrada (en el caso de training 298 y 74 en el caso de testing).

Luego de inicializar los centros aleatorios, se inicializan la matriz de pesos con el largo de los centros, para secuencialmente simular la red de base radial. Se contabilizan los números de puntos y se inicializa Phi, las que serán la salida de la red de base radial. Enseguida se calcula la distancia entre los puntos de los centros aleatorios y se calcula Phi con la función de activación. Después de la simulación de la red, se calcula la Pseudoinversa de Phi para hallar la matriz de pesos. Luego del proceso de iteraciones, los valores estimado tanto de alta como de baja serán sumados, y una vez que se sumen se compararán con los valores reales de captura de sardinas, para efecto de las métricas y discusión de resultados.

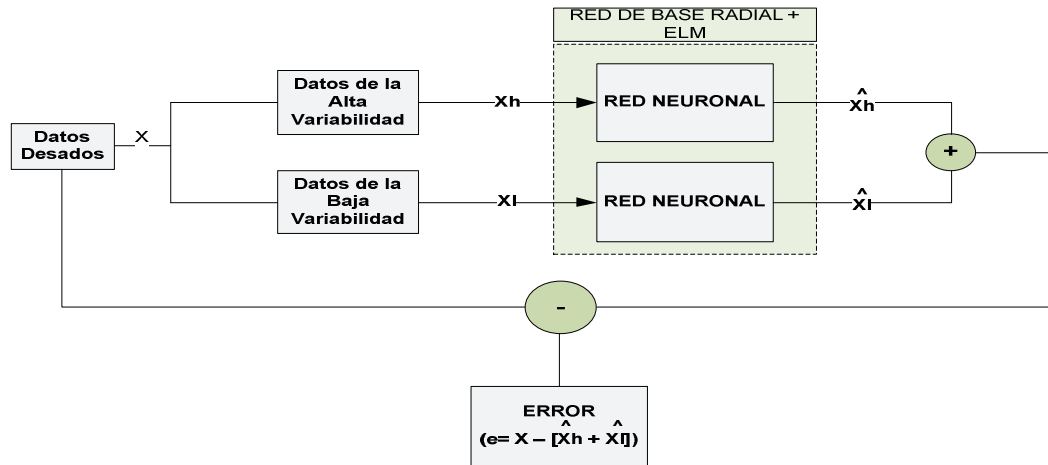


Ilustración 6: Modelo propuesto

## 4.1 MÉTRICAS DE EVALUACIÓN

En el presente estudio, para la selección de las topologías y, por consiguiente, de los modelos que presentan mejores resultados se utilizaron un conjunto de métricas de exactitud calculadas entre los datos observados (valores reales) y los datos pronosticados (entregados por la red). Estas métricas se nombran a continuación.

- Raíz Error Cuadrático Medio (RMSE). Consiste en la suma de las diferencias entre los datos observados y los datos proyectados por el modelo.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (d_i - y_i)^2}{N}} \quad (4.1)$$

- Coeficiente de Correlación de Pearson (R). El coeficiente de correlación de Pearson, R, es un índice que mide la relación lineal entre dos variables aleatorias cuantitativas. A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables

$$R = \frac{n \sum y_i \hat{y}_i - \sum y_i \sum \hat{y}_i}{\sqrt{n \sum y_i^2 - (\sum y_i)^2} \sqrt{n \sum \hat{y}_i^2 - (\sum \hat{y}_i)^2}} \quad (4.2)$$

- Coeficiente de Determinación (R<sup>2</sup>). Mide la dependencia entre los datos reales y los pronosticados. El 0 muestra independencia y el 1 lo contrario.

$$R^2 = 1 - \frac{\sum_{i=1}^N (d_i - y_i)^2}{\sum_{i=1}^N (d_i - \bar{d})^2} \quad (4.3)$$

- Porcentaje de Error Medio Absoluto (MAPE) Escalonado. Proporciona una indicación de que tan grandes son los errores de pronóstico comparados con los valores reales de la serie. También, correspondiente a la operación interna de la sumatoria se encuentra el Porcentaje de Error Absoluto (APE).

$$MAPE = \frac{\sum_{i=1}^N \left| \frac{d_i - y_i}{d_i} \right|}{N} \times 100 \quad d_i \neq 0 \quad (4.4)$$

Se define para las formulas presentadas a  $d_i$  al valor observado en el mes  $i$ ,  $y_i$  al valor pronosticado en el mes  $i$ ,  $\bar{d}$  como la media de la data observada y  $N$  como el número total de meses computados.

También dentro del análisis de los resultados obtenidos se utilizan dos métricas para hacer referencia al costo computacional que conlleva el entrenamiento de los modelos. El primero de ellos corresponde al tiempo promedio de entrenamiento que presenta cuánto tarda en promedio un modelo en especial en realizar el entrenamiento, esto, teniendo en cuenta que para todos los modelos se utilizará la misma cantidad de número de iteraciones y solo cambiarán las topologías. Debido a esto último, se agregó otra medida de referencia correspondiente al número de neuronas de manera que se pueda visualizar como influye la topología en el tiempo de entrenamiento.

## 4.2 SELECCIÓN DE TOPOLOGÍA Y PARÁMETROS DEL MODELO

Para los datos de alta y baja frecuencia, se consideraron ciertos criterios para definir una topología que encontrara los mejores valores pronosticados:

- Nodos de Entradas: Para obtener el número óptimo de entradas se realizó una calibración probando con 1, 6 y 12 nodos de entrada, y escogiendo la que entregara mejor MAPE escalonado.
- Se definió un límite inferior y superior para el número de iteraciones para el modelo, ya que a medida que se asignaban un número de iteraciones de esta forma encontraba mejores resultados de MAPE escalonado.

- Cabe destacar que el algoritmo ELM utiliza pesos ocultos aleatorios.
- Se utilizó un desfase de 12 meses para la data, con un 80% de entrenamiento y un 20% de testing, y los resultados entregados fueron para los datos desfasados de testing.
- El código se separó en testing de alta frecuencia, testing de baja frecuencia, y los pronosticadores finalmente se sumaron para efectos de cálculo de las métricas.
- En la tabla siguiente, se muestra los resultados presentados en relación al número de iteraciones. Se observa que al llegar a 100 iteraciones el modelo presenta buenos resultados, y que al aumentar de ese número de iteraciones no presenta mejores resultados, por tanto en 100 iteraciones presenta los mejores resultados en ambos casos (alta y baja frecuencia). Cabe destacar que se guarda la salida que tiene los mejores resultados de MAPE escalonado para caso por separado, para luego sumar las componentes.
- Finalmente, cabe destacar que el equipo en el cual se realizarán todos los entrenamientos y pruebas corresponde a un Intel Core i3 de 1.8 GHz y 4 GB de RAM y los modelos están hechos en MATLAB.

Iteraciones	MAPE	R	R <sup>2</sup>	Error medio
10	0,2596	0,9355	0,8752	0,001
50	0,2459	0,9795	0,9594	0,0005
100	0,1461	0,9931	0,9862	0,0041
200	0,2144	0,9798	0,9600	0,00053

**Tabla 1 Iteraciones para búsqueda de mejores resultados de las componentes sumadas.**

### 4.3 RESULTADOS DE BAJA FRECUENCIA

En la siguiente tabla se muestran los resultados del Mejor Caso de Baja frecuencia, con sus respectivos parámetros, escogidos de una serie de pruebas

Mejor Resultado Baja Frecuencia	
Nº Iteraciones	100
Nodos de Entrada	12
MAPE	0,009870
Error medio	0,0230
Coefficiente de Correlación	0,9919
Coefficiente de Determinación	0,9838

Tabla 2 Mejor resultado Baja Frecuencia

En la siguiente ilustración, se observa el gráfico de Pronóstico vs Data deseada con datos de baja frecuencia. Se observa una buena estimación casi en línea con los valores deseados. El gráfico de correlación muestra los datos muy cercano a la recta.

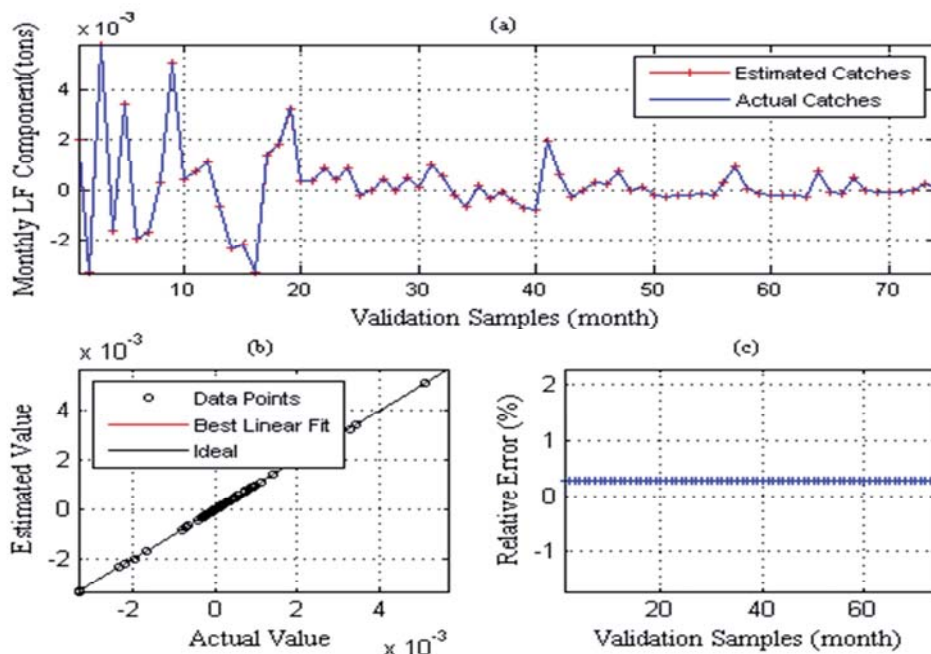


Ilustración 7: (a) Gráfico de Pronóstico vs Data Deseada Testing mejor resultado baja frecuencia. (b) Gráfico Correlación (c) Error relativo

En la siguiente tabla se entregan los valores del peor resultado para baja frecuencia, cabe destacar que fue con un número pequeño de iteraciones, junto a la ilustración del gráfico de pronóstico vs datos deseados del peor resultado.

Peor Resultado Baja Frecuencia	
N° Iteraciones	10
Nodos de Entrada	12
MAPE	5,2532
Error medio	0,2300
Coefficiente de Correlación	0,9235
Coefficiente de Determinación	0,8529

Tabla 3 Peor resultado Baja Frecuencia

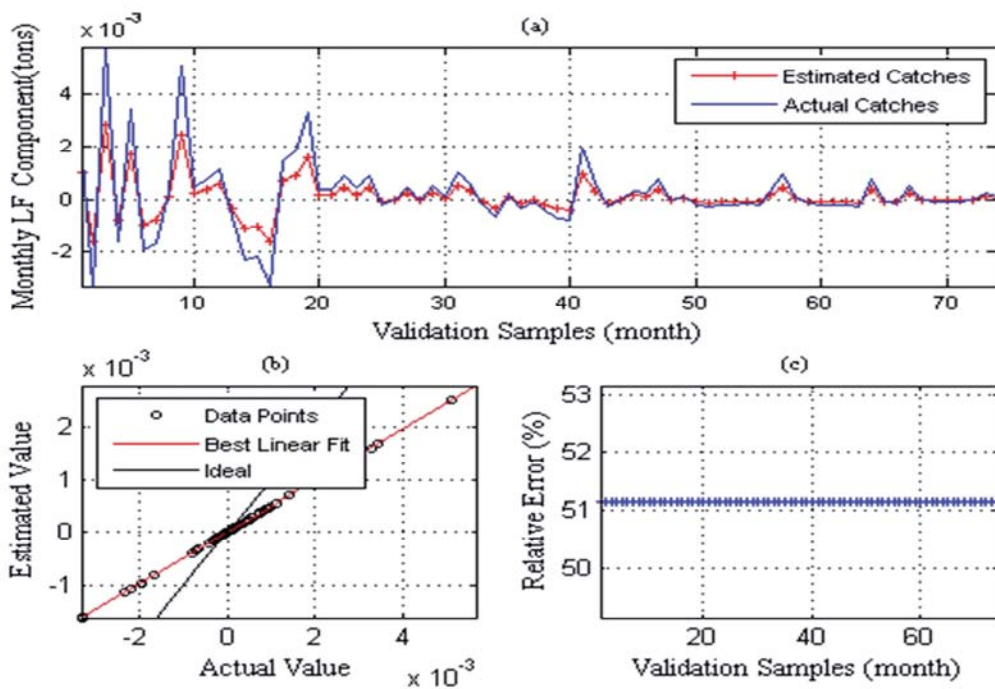


Ilustración 8: (a) Gráfico de Pronóstico vs Data Deseada Testing Peor resultado baja frecuencia (b) Gráfico de Correlación (c) Error relativo porcentual.

## 4.4 RESULTADOS DE ALTA FRECUENCIA

En la siguiente tabla se muestran los resultados del mejor caso de Alta frecuencia con sus respectivos parámetros, con 100 iteraciones definidas en la estructuración de las pruebas que encontraban mejores resultados.

Se puede observar que la data pronosticada se asemeja a los datos deseados.

Mejor Caso HF	
N° Iteraciones	100
Nodos de Entrada	12
MAPE	0,0079327
Error medio	0,0001
Coefficiente de Correlación	0,9937
Coefficiente de Determinación	0,98767

Tabla 4 Mejor resultado Alta Frecuencia

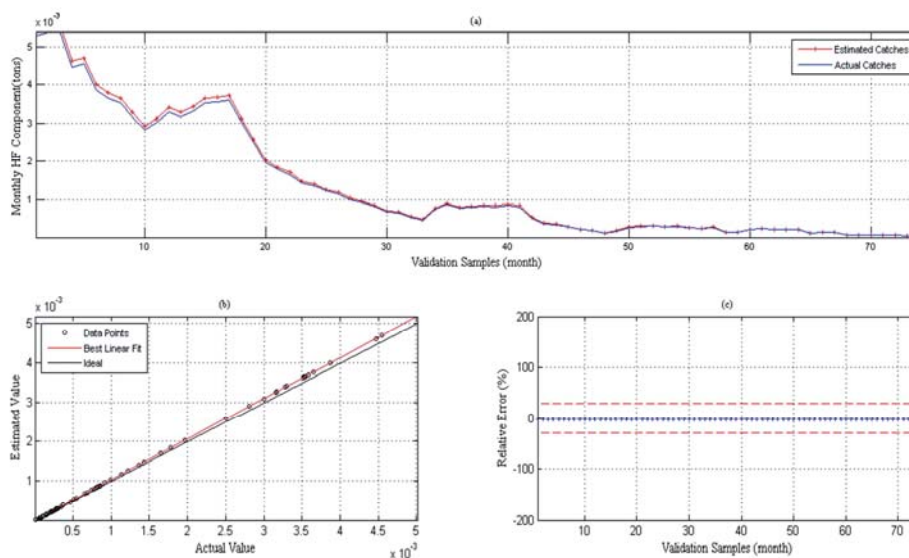


Ilustración 9: (a) Gráfico de Pronóstico vs Data Deseada Testing mejor resultado Alta frecuencia (b) Gráfico de correlación (c) Error relativo.

En la siguiente tabla se entregan los valores del peor resultado para Alta frecuencia, cabe destacar que fue con un número pequeño de iteraciones, junto a la ilustración del gráfico de pronóstico vs datos deseados del peor resultado



Peor Resultado Alta Frecuencia	
N° Iteraciones	10
Nodos de Entrada	12
MAPE	163,22
Error medio	12,24
Coefficiente de Correlación	0,9000
Coefficiente de Determinación	0,8100

Tabla 5 Peor resultado Baja Frecuencia

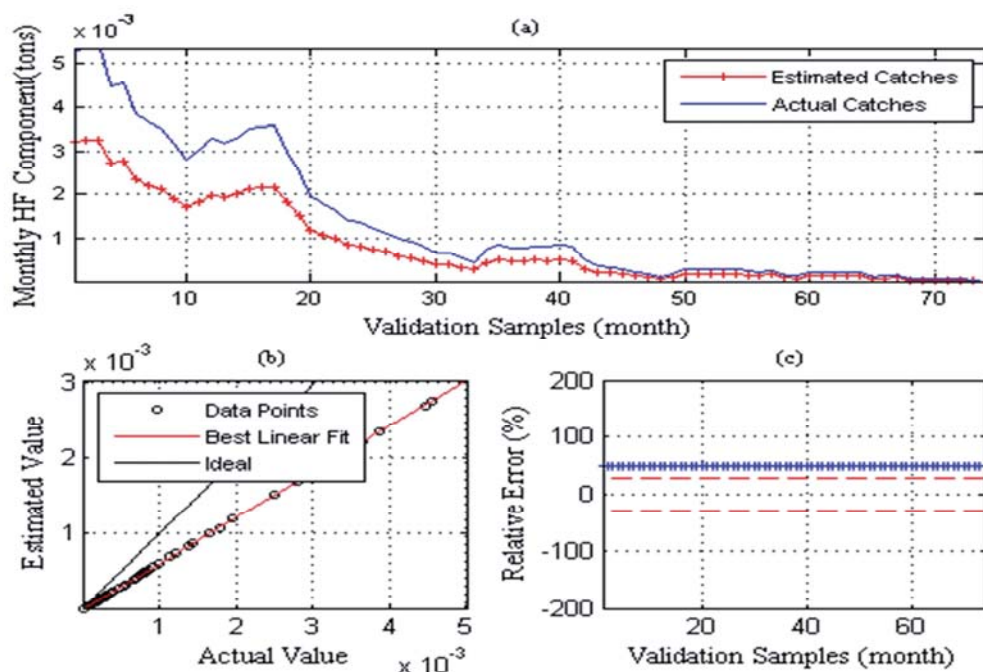


Ilustración 10: (a) Gráfico de Correlación Peor resultado Alta frecuencia Testing. (b) Gráfico de correlación (c) Error relativo

#### 4.5 RESULTADOS DE LAS COMPONENTES SUMADAS.

En la siguiente tabla se muestran los resultados del mejor caso de las frecuencias sumadas con sus respectivos parámetros, escogidos de una serie de pruebas, a su vez, se ve la ilustración del gráfico de pronóstico con relación a estas variables sumadas.

Mejor Caso HF+LF	
N° Iteraciones	100
MAPE	0,1461
Error medio	0,0001
Coefficiente de Correlación	0,9931
Coefficiente de Determinación	0,9862

Tabla 6 Mejor resultado Frecuencias sumadas

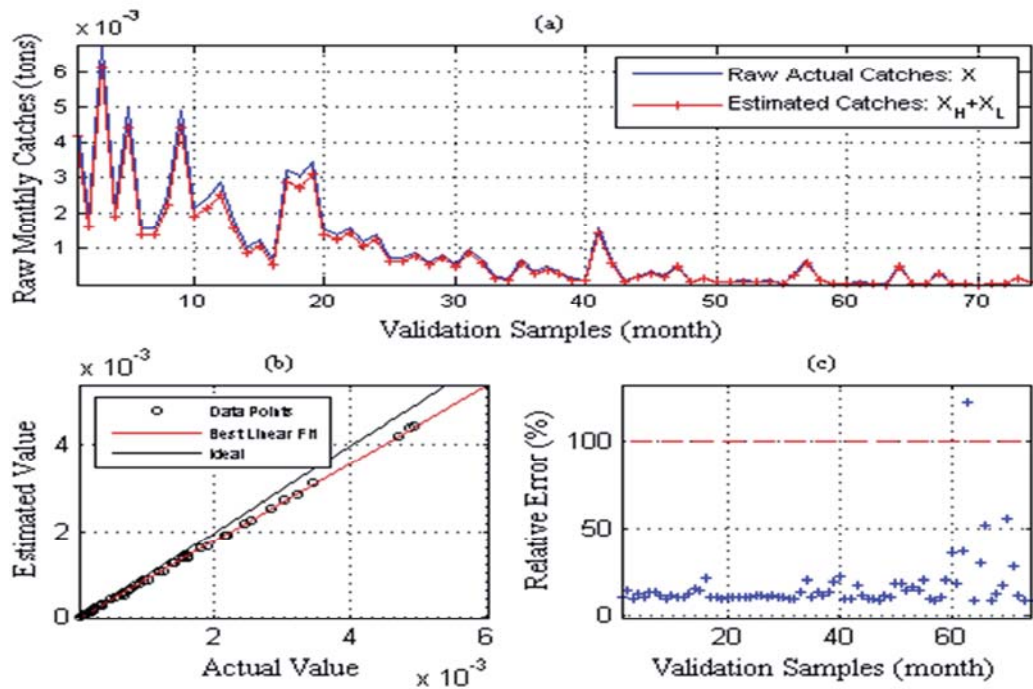


Ilustración 11: (a) Gráfico de Pronóstico vs Data Deseada Testing mejor resultado Frecuencias sumadas (b) Gráfico de Correlación (c) Error Relativo

En la siguiente tabla se entregan los valores del peor resultado para Alta frecuencia, cabe destacar que fue con un número pequeño de iteraciones, junto a la ilustración del gráfico de pronóstico vs datos deseados del peor resultado.

Peor resultado HF+LF	
N° Iteraciones	10
Nodos de Entrada	12
MAPE	114,23
Error medio	0,2220
Coefficiente de Correlación	0,9558
Coefficiente de Determinación	0,9136

Tabla 7 Peor resultado Frecuencias sumadas

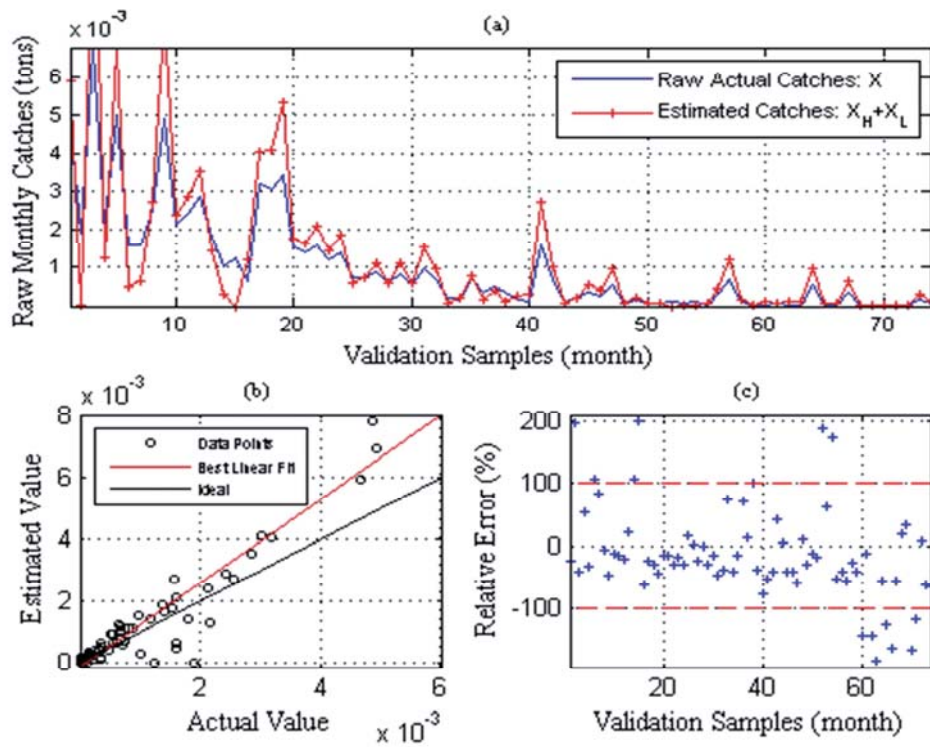


Ilustración 12: (a) Gráfico de Pronóstico vs Data Deseada Testing Peor resultado frecuencia sumada. (b) Gráfico de Correlación (c) Error relativo.

## 5. DISCUSIÓN DE RESULTADOS

En primer lugar, se debe recordar que el proceso de Aprendizaje extremo (ELM) utiliza la pseudoinversa, que emula una inversa generalizada, similar a lo que ocurre con el proceso de regresión lineal. Por tanto se puede obtener un resultado de una iteración. Al utilizar pesos ocultos aleatorios, eso produce que el valor que entregue el algoritmo sea muy bueno o muy malo, por tanto se vuelve necesario iterarlo una cantidad de veces limitada para que, mediante alguna métrica como lo es el MAPE escalonado, se pueda encontrar la mejor data estimada en este caso. Por tanto si se compara el peor caso, en el caso de alta y baja frecuencia, se concluye que es determinante realizar el proceso de calibración mediante iteraciones para obtener un valor de la métrica que se utilice.

Los gráficos de dispersión muestran los datos muy cercanos a la recta, lo cual lo vuelve con una dispersión pequeña pero concentrándose mucho en distancias cercanas. Esto podría atacarse utilizando algún tipo de metaheurística, como se utilizan en algoritmos como PSO, para que los valores converjan a los esperados. Los resultados obtenidos por la variable de Baja frecuencia fueron mejores a los de alta frecuencia, como se puede observar tanto en los gráficos como en la tabla, lo que se puede deducir que los eventos que ocurren a una baja frecuencia afectan de menor manera su correlación.

En la ilustración 5, se puede observar en (a) que la captura mensual de baja frecuencia deseada se encuentra en línea con la captura pronosticada, y al observar el valor del MAPE de 0,98 % la estimación se asemeja a la deseada, al igual como en (b). (c) muestra que el error relativo se encuentra siempre tendiendo a 0, lo que también es un buen indicador. Al igual que en baja frecuencia, ocurre algo similar con los datos de Alta frecuencia, ya que, en la Ilustración 7, se puede observar que los gráficos entregados contienen similitudes en la forma.

En la ilustración 9, se observa que, a pesar de que el MAPE entregado es de un 14,61%, se observa en (a) un gráfico de estimación aceptable en comparación con las investigaciones anteriores, a su vez mejoró las expectativas iniciales que entregaban un 21,34%. El  $R^2$  entregado por las componentes sumadas de 98,62% muestra que el grado de la varianza explicada es alto. También llama la atención en (c) que si bien los errores relativos porcentuales se acercan a 0, existen valores que también se alejan de la tendencia central.

En general, en la Ilustración 6,8 y 10 se observan malos resultados provocados principalmente por el poco número de iteraciones que se le hace al modelo, esto nos indica que es necesaria la calibración de los resultados de la Pseudoinversa, que si bien puede entregar valores sin necesidad de iterar ni comparar algún fitness, se vuelve necesario entregar una calibración para que los valores se ajusten.

## 6. CONCLUSIONES.

En el presente caso de estudio se desarrolló el estado del arte de los componentes fundamentales que se utilizaron en la confección del modelo de predicción propuesto en el presente trabajo. A su vez se muestran casos de aplicaciones reales tanto de las RNA como las de ELM, lo que evidencia su poder e importancia en las distintas áreas de las ciencias y la tecnología.

La investigación resultó beneficiosa ya que se encontró una nueva técnica que no había sido trabajada con anterioridad en la Escuela de Ingeniería Informática de la PUCV, marcando un precedente para futuras comparaciones de estimaciones que puedan realizar futuros postulantes a obtener el título, a su vez se genera una nueva opción para que se vuelva a utilizar por sus buenos resultados.

La técnica ELM obtiene prestaciones similares al entrenamiento estándar de un MLP en término de error, y teniendo como mayor ventaja la gran velocidad de convergencia que presenta.

Los datos de alta frecuencia y baja frecuencia se caracterizan por representar fenómenos que ocurren en un determinado lapso de tiempo, que si bien no es materia de estudio de este caso de estudio identificar cuales son, permiten entender de mejor manera la problemática, ya que se intenta predecir de qué forma ciertos fenómenos afectan o no afectan al sistema.

El algoritmo ELM, y el modelo propuesto para este caso de estudio son innovadores en el sentido de que, en estudios previos, se ingresaba la estructura a la red neuronal luego de la calibración del modelo. En este caso, el modelo puede entregar un valor estimado rápidamente. Por una parte es interesante observar que, al analizar datos de alta y baja variabilidad por separado, ayuda a que la muestra no se afecte, ya que al mezclarlos, se perdería la representatividad de los datos, dejándose fuera las variables exógenas que afectan a cierta frecuencia en estudio.

Para los resultados obtenidos, cabe destacar que los tiempos necesarios para generar los resultados para pronóstico fueron muy breves, por tanto utilizar ELM reduce significativamente el gasto computacional y facilita el encontrar un estimado adecuado con buenos valores, que con otras técnicas existentes como PSO.

En las discusiones de resultados se pudo observar que ELM entrega valores buenos considerando el bajo número de iteraciones (que en este caso funcionan a modo de calibrador) lo que la convierte en una herramienta poderosa de análisis en muestras de datos grandes. A su vez, MATLAB a través de sus funcionalidades y junto a una programación modular permite que el tiempo de procesamiento disminuya considerablemente en comparación a otras técnicas utilizadas para pronóstico (en la mayoría de los casos PSO).

El resultado del modelo propuesto fue el de un 14,61% de MAPE, lo que considera un buen modelo, el que puede ser revisado para mejorar el resultado, y un 98,62% de varianza explicada.

Por otra parte, se observa que, las cualidades por las que se optó realizar esta investigación utilizando ELM (mayor rapidez de aprendizaje y la no necesidad de ajustar los parámetros manualmente), cumplen las expectativas dado a que los resultados entregados por el modelo de pronóstico son relativamente buenos.

En comparación con otras técnicas utilizadas en otros estudios como la Inteligencia de Partículas en [28], [30] y [31] y Vector de Soporte Auto regresivo [29] con los mismos datos, cabe destacar que el tiempo es considerablemente bajo (al momento de la ejecución el tiempo de respuesta está alrededor de los 6 segundos). Las otras técnicas utilizadas (Inteligencia de partículas) junto con la calibración del modelo (buscando la topología que entrega la mejor estimación), en ocasiones, dura horas, por tanto más allá del tiempo computacional es importante destacar que a través de este modelo se puede encontrar una buena estimación en casos en que se requiere un apoyo a la toma de decisiones de forma rápida.

Finalmente, este trabajo de título desarrollado durante el año 2013 fue financiado económicamente por el proyecto FONDECYT Regular #1131105 "HARMONIC COEFFICIENTS FORECASTING OF NON-STATIONARY TIME SERIES TO SUPPORT THE MANAGEMENT OF THE SMALL PELAGIC FISHING IN CHILE".

## 7. REFERENCIAS

- [1] Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*.
- [2] Heeb, D. (1949). *Organization of Behavior*. John Wiley & Sons, New York.
- [3] Isasi, V., Galvan, L., *Redes de neuronas artificiales, un enfoque práctico*. Pearson Prentice Hall. 2004.
- [4] Rosenblatt, F. (1958). The perceptron: A theory of statistical separability in cognitive systems. Technical Report VG-1196-G-1, Cornell Aeronautical Laboratory.
- [5] Rumelhart, D., Hinton, G., and Williams, R. (1986b). *Parallel Distributed Processing, chapter Learning representations by back-propagation errors*. MIT Press. 1986.
- [6] Huang, Zhou, Ding, Zhang, “Extreme Learning Machine for Regression and Multiclass Classification,”(2012) *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 42, no. 2, pp. 513-529.
- [7] Huang, Wang, Lan, (2011) “Extreme Learning Machines: A Survey,” *International Journal of Machine Learning and Cybernetics*, pp. 107-122, vol. 2, no. 2.
- [8] Huang, Zhu, (2006) “Extreme Learning Machine: Theory and Applications”, *Neurocomputing*, vol. 70, pp. 489-501, 2006.
- [9] Míche, Bias, Jutten, Simula, Lendasse. (2008) “A methodology for Building Regression Models using Extreme Learning Machine”, *16th European Symposium in Artificial Neural Networks*, pp. 247-252, Bruges, Belgium.
- [10] Efron, Hastie, Tibshirani, (2004) “Least Angle Regression”, *Annual of Statistics*, vol. 32, pp. 407-499.
- [11] Weisberg, S. (1980). *Applied Linear Regression*. Wiley, New York. MR591462.
- [12] Huang, Zhu, Siew (2004). Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks, *International Joint Conference on Neural Networks*, Vol. 2, pp: 985-990, 2004.
- [13] Huang, Siew (2005). Extreme Learning Machine with Randomly Assigned RBF Kernels, *International Journal of Information Technology*, Vol. 11, No. 1, pp: 16-24.
- [14] Guang-Bin Huang, Nan-Ying Liang, Hai-Jun Rong, P. Saratchandran, N. Sundararajan, *On-Line Sequential Extreme Learning Machine*, *The IASTED International Conference on Computational Intelligence*, pp: 232-237, (2005).
- [15] Daniel Robles (2012), “Predicción de signo diario y semanal del Benchmark UF-5 mediante el uso de redes neuronales artificiales”. Universidad de Chile.

- [16] Weiwei Zong, Guang-Bin Huang, Yiqiang Chen, Weighted extreme learning machine for imbalance learning, *Neurocomputing*, Volume 101, 4 February 2013, Pages 229-242
- [17] Nan-Ying Liang, Guang-Bin Huang. A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks, *IEEE Transactions on neural networks*, Volume 17 n° 6, Noviembre 2006.
- [18] Jiuwen Cao, Zhiping Lin, Guang-Bin Huang, Self-Adaptative Evolutionary Extreme Learning Machine. Springer Science 18 July 2012.
- [19] C. Savojardo, P. Fariselli, and R. Casadio, "BETAWARE: a machine-learning tool to detect and predict transmembrane beta barrel proteins in Prokaryotes," *Bioinformatics*, Jan 13 2013.
- [20] Engin Avci, Resul Coteli A new automatic target recognition system base on wavelet extreme learning machine. *Expert System with applications* 39 (2012).
- [21] José M. Martínez-Martínez, Pablo Escandell-Montero, Emilio Soria-Olivas, José D. Martín-Guerrero, Rafael Magdalena-Benedito, Juan Gómez-Sanchis. (2011). Regularized extreme learning machine for regression problems. Elsevier B.V.
- [22] Hai-Jun Rong, Yew-Soon Ong, Ah-Hwee Tan, Zexuan Zhu. A fast pruned-extreme learning machine for classification problem. *Neurocomputing* 72 (2008).
- [23] Punyaphol Horata, Sirapat Chiewchanwattana, Khamron Sunat. Robust Extreme Learning Machine. *Neurocomputing* 102, 2013.
- [24] J.-N. Wang, J.-L. Jin, Y. Geng, S.-L. Sun, H.-L. Xu, Y.-H. Lu and Z.-M. Su, "An accurate and efficient method to predict the electronic excitation energies of BODIPY fluorescent dyes," *Journal of Computational Chemistry*, vol. 34, no. 7, pp. 566-575, 2013.
- [25] Pedro José García Laencina, Rafael Verdú Monedero, Jorge Larrey Ruiz, Juan Morales Sánchez, José Luis Sancho Gómez. Universidad Politécnica de Cartagena. Nuevas tendencias en Redes Neuronales Artificiales: Extreme Learning Machine. (2006).
- [26] Guang-Bin Huang, Dianhui Wang. Editorial. *Advances in Extreme Learning Machine (ELM 2011)*. Elsevier *Neurocomputing* 102. 19 de Agosto 2012.
- [27] C.Cortes, V.Vapnik, Support vector networks, *Mach. Learn.* 20 (1995) 273–297.
- [28] Juan Osorio Opazo. Vector de Soporte autoregresivo para pronosticar captura de anchovetas en la zona centro-sur de Chile. Biblioteca Pontificia Universidad Católica de Valparaíso. Mayo 2012.
- [29] Carlos Cosming González. Pronóstico de captura de anchovetas de la zona norte de Chile usando vector de soporte autoregresivo.



- Biblioteca Pontificia Universidad Católica de Valparaíso. Abril 2012.
- [30] Alejandro Trigo Galdames. Regresiones funcionales neurowavelet aplicadas al pronóstico de anchovetas. Biblioteca Pontificia Universidad Católica de Valparaíso. Septiembre 2011.
- [31] Francesco Villarroel Wells. Redes neuronales aplicativas para pronóstico de captura de anchoveta en la zona norte de Chile. Biblioteca Pontificia Universidad Católica de Valparaíso. Junio 2012.
- [32] Jorge Vergara Saavedra. Red sigmoideal recurrente con aprendizaje híbrido para el pronóstico del volumen de captura de anchovetas. Biblioteca Pontificia Universidad Católica de Valparaíso. Diciembre 2009
- [33] Pablo Quezada Román. Redes RRBf con aprendizaje híbrido para la predicción de captura de anchovetas. Biblioteca Pontificia Universidad Católica de Valparaíso. Diciembre 2009.
- [34] Diego Vera Bórquez. Pronóstico de marea roja para la Región de Magallanes utilizando máquinas de soporte vectorial con optimización por enjambre de partículas. Biblioteca Pontificia Universidad Católica de Valparaíso. Marzo 2012.
- [35] Nicolás Caselli Benavente. Pronóstico de capturas de anchovetas con algoritmos genéticos y máquinas de vector de soporte. Biblioteca Pontificia Universidad Católica de Valparaíso. Octubre 2011.
- [36] José Cañón. El Niño 1997-1998: Sus efectos en el sector pesquero industrial de la zona norte de Chile. Corpesca.
- [37] Hilera J. y Martínez V. (1995) Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones. RA-MA, Madrid.
- [38] B. Widrow, D. Rumelhart, and M.A. Lehr, "Neural Networks: Applications in Industry, Business, and Science," Communications of the ACM, 37(3):93-105, March 1994.
- [39] Abdulrahman S.N. Alarifi, Nassir S.N. Alarifi , Saad Al-Humidan. Earthquakes magnitude predication using artificial neural network in northern Red Sea área. Journal of King Saud University – Science (2012) 24, 301-313.
- [40] Ogcü, Demirel, Zaim. (2012). Forecasting Electricity Consumption with Neural Networks and Support Vector Regression. Procedia – Social and Behavior Sciences 58 1576-1585.
- [41] Gustavo Ovando, Monica Bocco, Silvina Sayago. Redes Neuronales para modelar la predicción de heladas.
- [42] Georgios Sermpinis, Konstantinos Theofilatos. Forecasting foreign exchange rates with adaptative neural networks using radial-basis functions and PSO. European Journal of Operational Research 225 (2013) 528-540.

- [43] Luis Zapata, Hugo Diaz. Predicción del tipo de cambio peso-dólar utilizando redes neuronales artificiales. Universidad del norte Barranquilla, Colombia pp. 29-42.
- [44] Fernando Villada, Nicolás Muñoz, Edwin García. Aplicación de las Redes Neuronales al Pronóstico de Precios en el mercado de valores.
- [45] Damián Jorge Matich. Redes Neuronales Conceptos Básicos y Aplicaciones. Rosario.
- [46] Alan Lapedes, Robert Farber. Nonlinear signal processing using NN: prediction and system modelling. TR-UR-87-2662.
- [47] Cristian Foix, Richard Weber. Pronóstico del precio del cobre mediante Redes Neuronales. Revista Ingeniería de Sistemas. Volumen XXI, 2007. Universidad de Chile.
- [48] Guang-Bin Huang, Lei Chen, Chee-Kheong Siew. Universal Approximation Using Incremental Constructive Feedforward Networks With Random Hidden Nodes. IEEE transactions on neural networks, Vol 17. No. 4. July 2006.
- [49] S. Decherchi, et al., Circular-ELM for the reduced-reference assessment of perceived image quality, Neurocomputing (2012).
- [50] Fei Han, Hai-Fen Yao, Qing -hua Ling. An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing 116 (2013) 87-93.