

*Tesis
Doctoral*

*Nuria Esther
Hernández Barrera*

REDES DE NEURONAS

ARTIFICIALES EN

OCEANOGRAFÍA

*Departamento de Estadística e Investigación
Operativa.*

Facultad de Matemáticas. Universidad de Sevilla.

Noviembre 2.004

AGRADECIMIENTOS.

En primer lugar me gustaría expresar mi agradecimiento a mis directores de tesis, D. Joaquín Muñoz García y D. Rafael Pino Mejías, por su apoyo, sus innumerables consejos y su inestimable ayuda durante todo este tiempo.

En segundo lugar, a mis compañeros del Instituto Español de Oceanografía, de la Secretaría General de Pesca Marítima y del Buque de Investigación Oceanográfico Vizconde de Eza, por transmitirme los conocimientos adquiridos durante tantos años de experiencia, y también a D. Jesús Gálvez, del Instituto Hidrográfico de la Marina, que siempre me animó a continuar trabajando en el difícil mundo de la investigación.

Quisiera agradecer especialmente a mi familia y a mis grandes amigos, entre ellos Ángel, Rocío, Amilcar, Isabel, Diana, Elena, Gerardo,... por su incondicional apoyo y paciencia, sin las cuales no hubiera dispuesto de la fuerza y constancia necesarias durante todos estos años para poder realizar este trabajo, y también por la confianza que ellos siempre han demostrado en mi, incluso en los momentos de mayor desánimo.

Por último, hacer constar que los datos necesarios para la realización de esta tesis doctoral se han obtenido durante unas campañas de investigación oceanográfica realizadas por el Instituto Español de Oceanografía y la Secretaría General de Pesca Marítima, y que mi participación en dichas campañas y el posterior acceso a los datos ha sido posible gracias a una beca de formación como personal investigador que me fue concedida por el Instituto Español de Oceanografía.

ÍNDICE.

INTRODUCCIÓN	1
1. CONCEPTOS BÁSICOS DE OCEANOGRAFÍA FÍSICA	7
1.1. PRINCIPALES PROPIEDADES FÍSICAS DEL AGUA DEL MAR	7
1.1.1. TEMPERATURA	7
1.1.2. SALINIDAD	9
1.1.3. DENSIDAD	11
1.2. MASAS DE AGUA	13
1.3. CIRCULACIÓN GENERAL DE LOS OCÉANOS	15
1.3.1. INTERACCIÓN OCÉANO-ATMÓSFERA	15
1.3.2. CIRCULACIÓN TERMOHALINA	17
1.4. INSTRUMENTACIÓN MARINA	17
1.5. EL SONIDO EN EL AGUA DEL MAR	18
1.5.1. CÁLCULO DE LA VELOCIDAD DE SONIDO	21
2. BREVE DESCRIPCIÓN DEL MAR DE ALBORÁN	25
2.1. FISIOGRAFÍA DEL MAR DE ALBORÁN	25
2.1.1. PLATAFORMA CONTINENTAL	25
2.1.2. TALUD CONTINENTAL	25
2.1.3. LLANURA ABISAL	28
2.2. OCEANOGRAFÍA FÍSICA DEL MAR DE ALBORÁN	28
2.2.1. CONDICIONES ATMOSFÉRICAS	28
2.2.2. MASAS DE AGUA	29
2.2.3. CIRCULACIÓN OCEÁNICA	30
2.3. TIPOS DE PERFILES ACÚSTICOS DETECTADOS EN EL MAR DE ALBORÁN	30
2.3.1. TALUD CONTINENTAL DE LA COSTA ESPAÑOLA	32
2.3.2. TALUD CONTINENTAL DE LA COSTA AFRICANA	35
2.3.3. LLANURA ABISAL	37
3. REDES DE NEURONAS ARTIFICIALES	43
3.1. INTRODUCCIÓN HISTÓRICA	43
3.2. SISTEMAS DE NEURONAS BIOLÓGICAS	47
3.3. ESTRUCTURA DE LAS REDES DE NEURONAS ARTIFICIALES	49
3.3.1. DEFINICIÓN DE RED NEURONAL ARTIFICIAL	50
3.3.2. ELEMENTOS DE PROCESAMIENTO	51
3.3.3. ARQUITECTURA DE LAS REDES DE NEURONAS ARTIFICIALES	56
3.3.4. ALGORITMOS DE APRENDIZAJE	57
3.3.5. TRATAMIENTO DE LA INFORMACIÓN	60

3.4. VENTAJAS DE LAS REDES DE NEURONAS ARTIFICIALES	62
3.5. MODELOS DE REDES DE NEURONAS ARTIFICIALES Y APLICACIONES	63
3.6. ASPECTOS ESTADÍSTICOS DE LAS REDES DE NEURONAS ARTIFICIALES	68
4. ALGUNOS MODELOS DE REDES DE NEURONAS ARTIFICIALES	73
4.1. EL PERCEPTRÓN SIMPLE	73
4.1.1. DEFINICIÓN DEL PERCEPTRÓN SIMPLE	73
4.1.2. ALGORITMO DE APRENDIZAJE DEL PERCEPTRÓN SIMPLE	75
4.1.3. CARACTERÍSTICAS DEL PERCEPTRÓN SIMPLE	76
4.2. EL PERCEPTRÓN MULTINIVEL O MULTICAPA (MLP)	79
4.2.1. DEFINICIÓN DEL PERCEPTRÓN MULTINIVEL O MULTICAPA	80
4.2.2. ALGORITMO DE APRENDIZAJE DEL PERCEPTRÓN MULTINIVEL O MULTICAPA: RETROPROPAGACIÓN DE ERRORES (BP)	82
4.2.3. CARACTERÍSTICAS DEL PERCEPTRÓN MULTINIVEL O MULTICAPA	88
4.2.4. ASPECTOS PRÁCTICOS RELATIVOS AL PROCESO DE ENTRENAMIENTO DEL PERCEPTRÓN MULTINIVEL O MULTICAPA	91
5. CONSTRUCCIÓN DE REDES DE NEURONAS ARTIFICIALES	97
5.1. ALGORITMOS DE ENTRENAMIENTO	97
5.2. IMPLEMENTACIÓN Y DISEÑO EXPERIMENTAL	103
5.3. MLP DE UNA CAPA CON DOS PREDICTORES	104
5.3.1. RESUMEN DE LOS RESULTADOS	105
5.3.2. ANÁLISIS DE LOS RESULTADOS	119
5.4. MLP DE UNA CAPA CON ONCE PREDICTORES	139
5.4.1. RESUMEN DE LOS RESULTADOS	139
5.4.2. ANÁLISIS DE LOS RESULTADOS	154
5.5. COMPARACIÓN CON OTROS MODELOS DE PREDICCIÓN	180
6. CONCLUSIONES	193
ANEXO 1	199
ANEXO 2	217
ANEXO 3	239
BIBLIOGRAFÍA	261

INTRODUCCIÓN.

La preocupación de la humanidad por medir la profundidad del fondo del mar existe desde hace muchos siglos, por lo que han ido surgiendo numerosas técnicas para realizar dicha medición. Así, durante el siglo XIX se utilizaba el escandallo, que consistía en un peso o plomada que colgaba de una cuerda graduada, la cual se dejaba caer por la proa o el costado del barco hasta tocar el fondo y después se leía la profundidad de acuerdo con la marca correspondiente en la cuerda. Se inventaron diversos mecanismos para poder obtener con mayor facilidad una medida de la profundidad (ruedas mecánicas, aparatos de presión, etc.), pero no fue hasta principios del siglo XX cuando se utilizaron las ecosondas, instrumentos que emplean los principios físicos de la propagación del sonido en el agua de mar, con el fin de conseguir un registro continuo y medidas más fiables de la profundidad.

Las técnicas acústicas son métodos físicos que permiten conocer la morfología del fondo marino y la estructura del subfondo marino de un modo indirecto. Las ecosondas se basan en la medida del tiempo que tarda una onda acústica en recorrer la distancia existente entre el punto de partida y el fondo del mar, donde se refleja, y su retorno al punto de partida. La emisión y recepción acústica se realiza generalmente a través de un mismo transductor que convierte las variaciones mecánicas en pulsos eléctricos y viceversa, de forma que en la emisión, la energía eléctrica se convierte en acústica, y en la recepción la onda acústica se transforma en señal eléctrica. El haz puede comprender un único pulso, o bien una banda de pulsos que se distribuyen con un ángulo variable a babor y estribor del barco. Estos últimos, que representan los mayores avances en sistemas de ecosonda reciben el nombre de ecosondas multihaz.

Para la utilización de estas técnicas acústicas se precisa determinar la velocidad de propagación del sonido en el agua del mar, ya que éstas se basan en la medición del tiempo que tarda en propagarse una señal acústica desde su emisión hasta su recepción tras ser reflejada en el fondo marino. A partir de la velocidad de propagación del sonido en la columna de agua, el tiempo empleado proporciona una medida de la profundidad existente, por lo que es primordial disponer de un buen conocimiento de la distribución espacial de la velocidad de propagación del sonido.

La velocidad de propagación del sonido en el agua del mar es una variable oceanográfica que además permite determinar muchas de las peculiaridades

del medio en el que ésta se transmite, lo que provocó a principios del siglo XX un gran interés en su cálculo. En este sentido, a finales de los años 30 se elaboraron las primeras tablas de la velocidad de propagación del sonido en el agua del mar para determinadas zonas del mundo, y ya en los años 60 se obtuvieron los primeros algoritmos de cálculo basados en mediciones realizadas en laboratorios, a partir de los que se desarrollaron otra serie de algoritmos que actualmente constituyen el procedimiento estándar internacionalmente adoptado por la UNESCO.

Para la correcta utilización de estos algoritmos de cálculo de la velocidad de propagación del sonido en el agua del mar se precisa poseer información de determinadas variables oceanográficas, aunque en muchas ocasiones esto no es posible y por tanto no siempre se dispone de toda la información necesaria, lo que genera que la fiabilidad de estos algoritmos se reduzca de forma considerable, surgiendo la necesidad de buscar nuevas metodologías para proceder a un cálculo más preciso.

Hay que considerar que los océanos constituyen un medio en continuo movimiento y cambiante, por lo que no es posible la recogida de muestras bajo las mismas condiciones, resultando imposible conseguir unas suposiciones estadísticas adecuadas y homogéneas para las variables que se analicen. Así, resulta prácticamente imposible suponer determinados modelos distribucionales o imponer condiciones de independencia a las observaciones que se realizan, lo que conduce a recurrir a métodos y/o técnicas estadísticas que no se encuentran sometidas a hipótesis muy restrictivas y que pueden denominarse de "gran espectro".

En este sentido, con este trabajo se pretende abordar el problema desde una perspectiva totalmente novedosa e innovadora mediante la utilización de redes de neuronas artificiales. Este tipo de técnicas han sido aplicadas en numerosos campos de la ciencia con resultados satisfactorios, pero hasta ahora se han abordado muy pocos problemas relacionados con el campo de la oceanografía a través de redes de neuronas artificiales, por lo que se pretende así estudiar la posibilidad de tratar diversas cuestiones de carácter oceanográfico mediante ellas.

Uno de los objetivos principales que se plantean es la obtención de predicciones para la velocidad de propagación del sonido en el agua del mar mediante la construcción de redes de neuronas artificiales. Para ello, en función de la información procedente de las bases de datos obtenidas a través de diversas campañas de investigación oceanográficas realizadas por el Instituto Español

de Oceanografía y la Secretaría General de Pesca Marítima durante los meses de julio y agosto de los años 2.002 y 2.003 en la zona del Mar de Alborán, se han construido varias redes de neuronas artificiales.

Estas redes de neuronas artificiales han sido construidas mediante varios de los algoritmos de entrenamiento existentes y tomando diversas arquitecturas hasta encontrar aquella que optimiza los resultados. Este proceso de búsqueda ha sido realizado de forma exhaustiva, identificando así diversas características de los métodos de entrenamiento empleados.

Las predicciones obtenidas mediante los modelos de redes finalmente seleccionados han sido analizadas y comparadas con las proporcionadas por los algoritmos clásicos que se emplean usualmente para el cálculo de la velocidad de propagación del sonido en el agua del mar.

Una vez descritas las líneas generales del estudio realizado, se ha de indicar que en el primer capítulo se procede a describir algunos de los conceptos básicos y fundamentales de la oceanografía física, necesarios para comprender el problema que se pretende abordar, explicándose, de forma breve y resumida, las propiedades físicas más importantes del agua del mar, el concepto de masas de agua, la circulación general de los océanos, los instrumentos científicos diseñados para obtener datos oceanográficos y por último el proceso de propagación del sonido en el agua del mar.

En el capítulo segundo se analiza de forma descriptiva el Mar de Alborán, ya que de esta zona proceden los datos a modelizar, y por lo tanto, es preciso conocer la fisiografía y los procesos físicos que la caracterizan. Tras la descripción de estos conceptos, se comentan los distintos perfiles de sonido observados a partir de los datos recogidos en la zona de estudio durante las campañas de investigación oceanográfica.

En el tercer capítulo se introducen distintos conceptos relativos a las redes de neuronas artificiales. Se comienza con una breve evolución histórica, y se continúa describiendo el funcionamiento de las neuronas que existen en el cerebro humano. Tras proporcionar una definición genérica de las redes de neuronas artificiales, se explica cómo está estructurada una red de neuronas artificial, los elementos que la constituyen, su arquitectura, los diversos tipos de algoritmos de aprendizaje existentes y la forma en que la información es asociada en una red de neuronas artificial. Por último, se citan las ventajas que se obtienen con la utilización de las redes de neuronas artificiales, los diversos tipos existentes, sus múltiples y variadas aplicaciones, y la interacción entre las redes de neuronas artificiales y la estadística.

El capítulo cuarto recoge una descripción detallada de dos modelos de redes de neuronas artificiales: el perceptrón simple y el perceptrón multinivel o multicapa. Para ambos modelos se proporciona su definición, sus características y su algoritmo de aprendizaje. Para el perceptrón multinivel o multicapa, que es el modelo de red de neuronas artificial escogido para modelizar la variable de estudio, se analizan diversos aspectos relacionados con el proceso de aprendizaje, proporcionando distintos criterios y reglas prácticas.

En el quinto capítulo se muestran las redes de neuronas artificiales construidas para resolver el problema de la predicción de la velocidad de propagación del sonido en el agua del mar, describiendo los diversos algoritmos de entrenamiento utilizados y exponiendo los criterios seguidos en su implementación. Para cada una de las redes de neuronas artificiales que se han construido, se analizan sus resultados y se proporcionan las estimaciones óptimas. Por último, se comparan las predicciones de la velocidad de propagación del sonido en el agua del mar obtenidas a partir de las técnicas de redes de neuronas artificiales con las resultantes de los algoritmos clásicos de cálculo adoptados internacionalmente por la UNESCO.

El último capítulo recoge diferentes conclusiones y propone futuras líneas de investigación en este campo, el cual está muy abierto por la novedad que representa en sí misma la aplicación de las redes de neuronas artificiales a un problema oceanográfico.

1 CONCEPTOS BÁSICOS DE OCEANOGRAFÍA FÍSICA.

1.1 PRINCIPALES PROPIEDADES FÍSICAS DEL AGUA DEL MAR.

1.1.1 TEMPERATURA.

La temperatura de los océanos depende básicamente de la absorción de la energía solar y de la interacción que se produce entre los océanos y la atmósfera.

La distribución de la temperatura en la superficie de los océanos es zonal, con las líneas de temperatura constante (isotermas) discurriendo en sentido este-oeste, apareciendo los valores más elevados en las zonas tropicales y disminuyendo progresivamente hacia los polos. Las distorsiones de las isotermas, por lo general cerca de los continentes, se producen por las corrientes que transportan las aguas cálidas hacia latitudes altas y a la inversa, aguas frías hacia latitudes bajas. En las Figuras 1.1.1 y 1.1.2 se muestran los valores de la temperatura media en la superficie de los océanos durante los meses de febrero y agosto respectivamente.

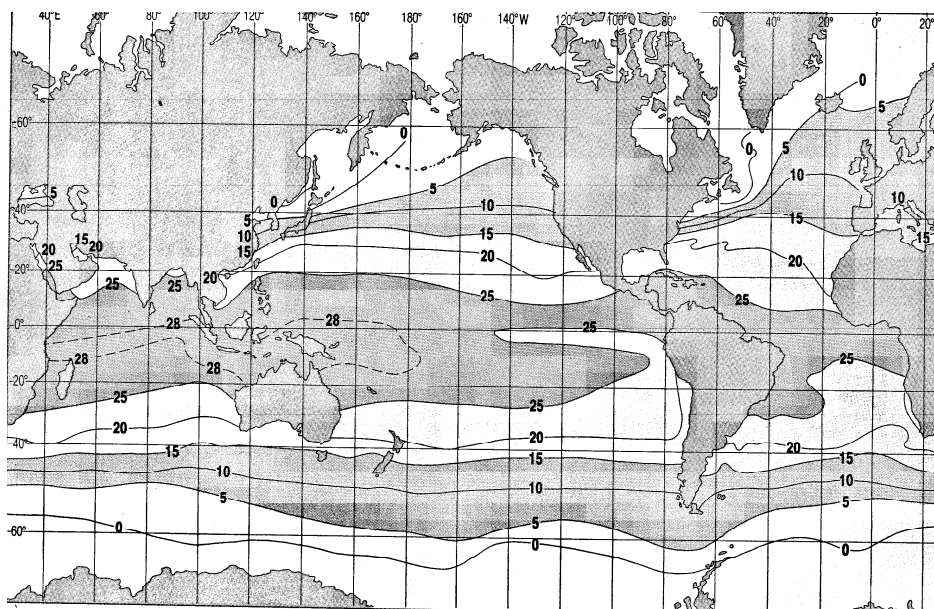


Figura 1.1.1. Temperatura superficial media (°C) en febrero (Open University).

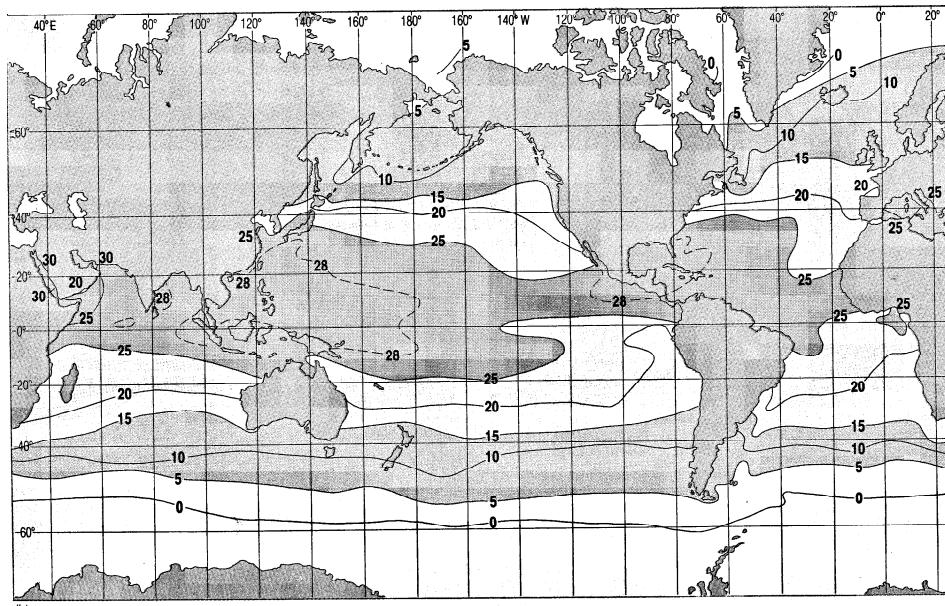


Figura 1.1.2. Temperatura superficial media ($^{\circ}\text{C}$) en agosto (Open University).

En cuanto a su distribución con respecto a la profundidad, cuyo perfil típico se muestra en la Figura 1.1.3, se distinguen tres zonas claramente diferenciadas. En la zona superior, desde la superficie hasta una profundidad que varía entre los 50 y los 200 metros, existe lo que se denomina capa de mezcla. En esta zona se absorbe la mayor parte de la energía solar, transmitiéndose el calor por la mezcla turbulenta generada por los vientos y las olas, lo que produce una capa donde las variaciones de temperatura son pequeñas. Entre los 200 y los 1000 metros, se extiende lo que se denomina zona intermedia y en la que se produce un descenso acusado de la temperatura. Y a partir de los 1000 metros comienza la zona profunda, en la que la temperatura decrece más lentamente.

La profundidad a la que el gradiente de la temperatura es máximo se conoce como termoclina, pero como es difícil determinar con exactitud esta profundidad, se recurre a lo que se conoce como "zona de termoclina", que es el rango de profundidades donde el gradiente de la temperatura es mayor que el existente por arriba y por abajo. En las latitudes bajas y medias existe una termoclina presente todo el tiempo entre los 200 y los 1000 metros, que se denomina termoclina permanente, y por debajo de la cual no existen variaciones estacionales de temperatura. En cambio, en latitudes altas no

existe termoclina permanente.

La temperatura en la capa de mezcla muestra variaciones estacionales, especialmente en latitudes medias. En invierno la temperatura en la superficie es baja y los vientos son más intensos, por lo que la capa de mezcla es más profunda y puede extenderse hasta la termoclina permanente, pero en verano, la temperatura aumenta y a menudo aparece en esta zona una termoclina estacional. La termoclina estacional también puede aparecer en latitudes altas, pero no en latitudes tropicales, ya que en estas zonas no se produce enfriamiento invernal.

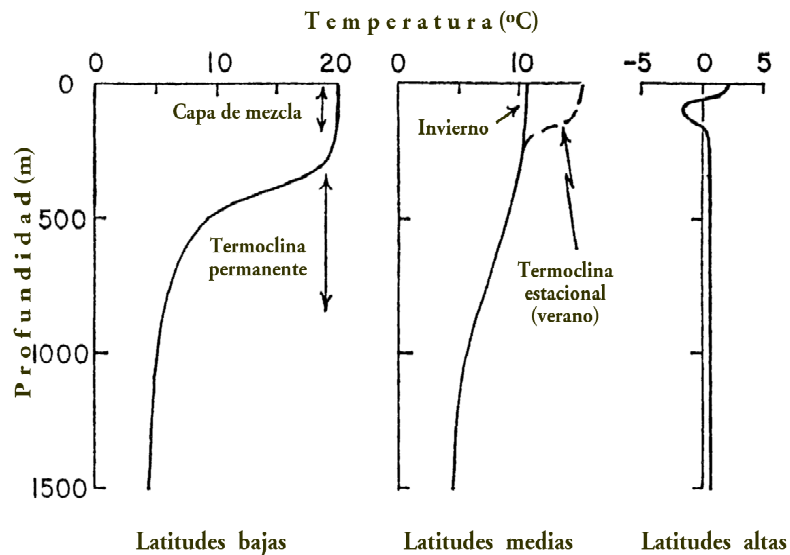


Figura 1.1.3. Perfil típico de temperatura (adaptado de Pickard).

1.1.2 SALINIDAD.

El agua del mar es una compleja solución que contiene disueltos la mayoría de los elementos conocidos. A la concentración de sus iones disueltos se le denomina salinidad.

Aunque la concentración de las principales sales disueltas varían en diferentes regiones del océano, las proporciones entre ellas permanecen prácticamente constantes, lo que indica que existen una serie de procesos actuando continuamente para concentrar o diluir el agua del mar y mantener esas proporciones constantes. Esta característica se conoce como principio de

conservación de la sal, y permite además determinar la concentración media de las sales disueltas en el océano.

El agua del mar contiene en media unos 35 gramos de sal por litro de agua. Actualmente se utiliza la escala práctica de salinidad, que no posee unidades, ya que se emplea la relación entre la conductividad eléctrica de la muestra y la de una patrón.

La variación de la salinidad en el océano depende de la evaporación, las precipitaciones, la descarga de ríos, la congelación y fusión de hielos, y la mezcla entre las aguas superficiales y profundas. Así, la salinidad en la superficie de mares abiertos oscila generalmente entre 33 y 37, tomando sus valores máximos en latitudes próximas a 20° y decreciendo hacia el Ecuador y hacia altas latitudes. Modificaciones de este patrón de comportamiento tienen lugar de forma local, alcanzando valores inferiores en zonas próximas a la costa donde desembocan grandes ríos y en regiones polares donde se produce el deshielo, y alcanzando valores superiores en regiones donde la evaporación supera en gran medida a la cantidad de agua aportada por los ríos y por las precipitaciones, lo que ocurre en mares casi cerrados, como el Mar Mediterráneo y el Mar Rojo. La distribución de la salinidad en la superficie de los océanos se muestra en la Figura 1.1.4.

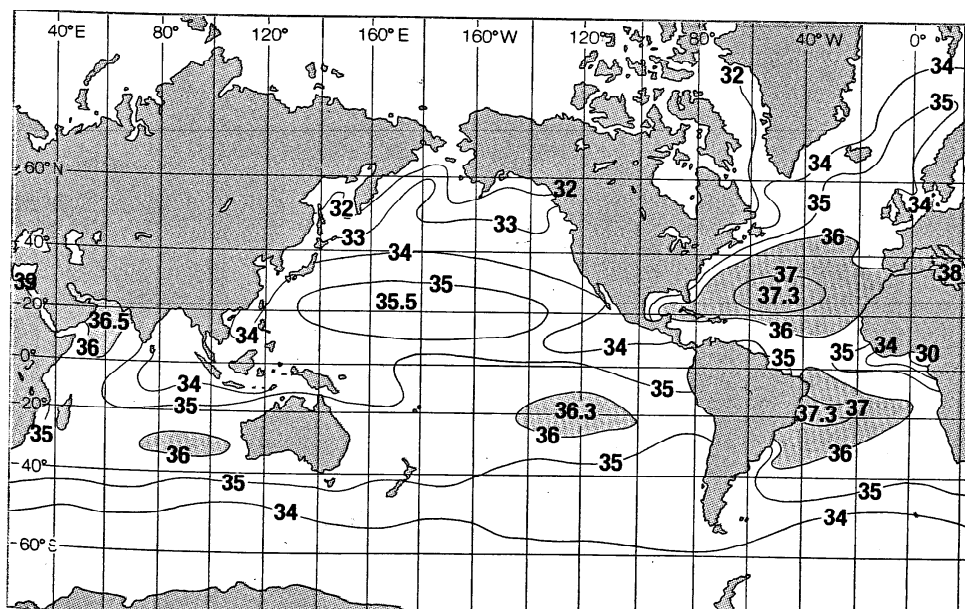


Figura 1.1.4. Salinidad superficial media (ups) anual (Open University).

En cuanto a la variación vertical de la salinidad, cuyo perfil típico se muestra en la Figura 1.1.5, se observan fluctuaciones con valores comprendidos entre 34 y 37, pero su distribución en los primeros metros es bastante compleja, ya que es en esta zona donde se ve más afectada de una forma directa y rápida por los procesos descritos anteriormente. En aguas profundas, la salinidad es relativamente uniforme, ya que ésta sólo fluctúa entre 34.5 y 35.

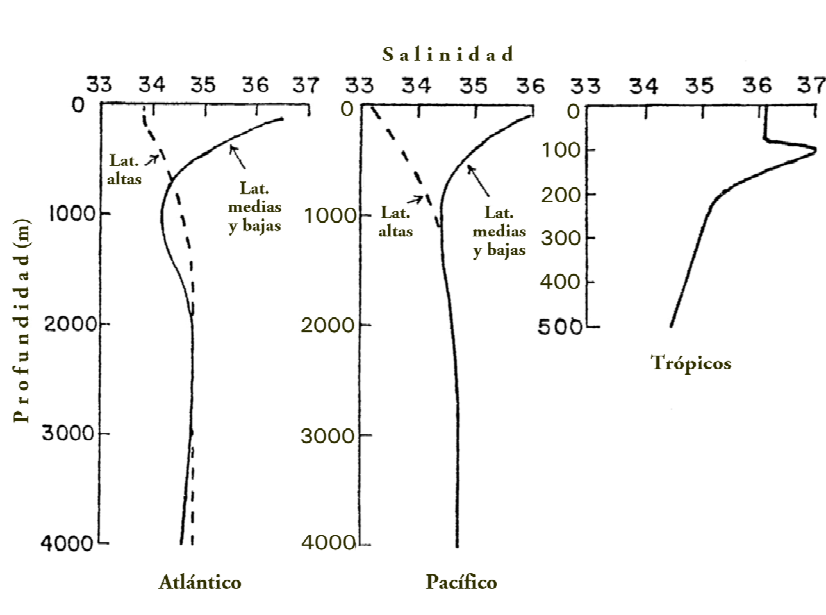


Figura 1.1.5. Perfil típico de salinidad (adaptado de Pickard).

1.1.3 DENSIDAD.

La temperatura junto con la salinidad y la presión determinan la densidad del agua del mar, lo que es importante porque ésta determina la profundidad a la que las masas de agua se encuentran en equilibrio, es decir, la densidad es el factor más importante en el movimiento vertical de las aguas de los océanos. La densidad se expresa en kg/m^3 y presenta un rango de variación en los océanos que oscila entre 1021 kg/m^3 en la superficie y 1070 kg/m^3 a una profundidad de 10000 metros, mientras que en aguas costeras se obtienen unos valores de densidad menores.

Por conveniencia, es usual emplear como medida de la densidad la canti-

dad $\sigma_{s,t,p}$ ¹, conocida como valor "in situ" y que viene dada por

$$\sigma_{s,t,p} = (\text{densidad} - 1000 \text{ kg/m}^3).$$

La distribución de la densidad del agua del mar en la superficie toma aproximadamente el valor 22 kg/m^3 en el Ecuador y va creciendo en ambas latitudes hasta alcanzar el valor de 26 ó 27 kg/m^3 en las latitudes próximas a 60° , comenzando a decrecer posteriormente de forma suave.

Aunque en general la densidad aumenta con la profundidad, este incremento no se produce uniformemente. La representación del perfil típico de la densidad con respecto a la profundidad aparece en la Figura 1.1.6.

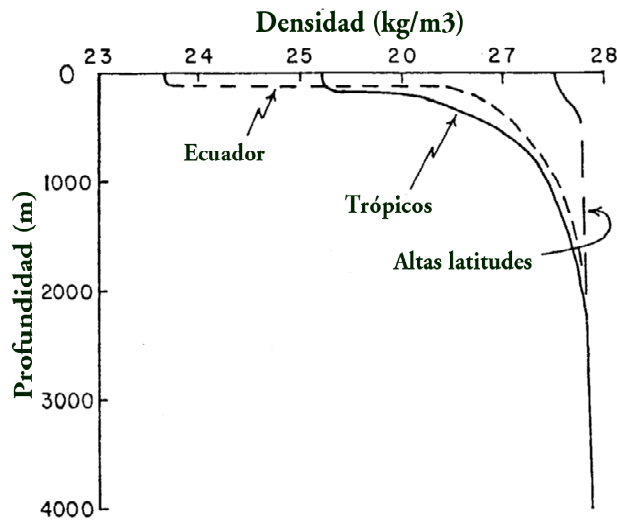


Figura 1.1.6. Perfil típico de densidad (adaptado de Pickard).

En las regiones tropicales y ecuatoriales existe una zona superficial donde la densidad es prácticamente uniforme, debajo una zona donde la densidad se incrementa rápidamente con la profundidad y por último una zona en la que la densidad aumenta de forma más suave con la profundidad.

El cambio de la densidad con respecto a la profundidad determina la estabilidad del agua, de forma que si la estabilidad es alta, los movimientos y mezclas verticales se minimizan.

¹No confundir esta sigma con la que se emplea habitualmente para designar el concepto estadístico de desviación típica.

Cuando la densidad no cambia con la profundidad, el agua posee una estabilidad neutral y puede mezclarse verticalmente sin ningún esfuerzo, pero si la densidad decrece con la profundidad, esto indica que las condiciones son inestables, lo que sólo puede ocurrir en la zona superficial (hasta unos 50-100 metros). En cambio, el agua en la zona donde la densidad aumenta rápidamente con la profundidad, que es conocida con el nombre de pycnoclina, es muy estable, lo que significa que se precisa mucha más energía para desplazarla hacia arriba o hacia abajo que en una región donde la densidad cambia lentamente con la profundidad. El resultado de esta característica, es que la turbulencia, que causa la mayoría de las mezclas entre los distintos volúmenes o masas de agua, es incapaz de traspasar esta capa estable, por lo que la pycnoclina ofrece una especie de barrera que impide o amortigua el paso del agua y de sus propiedades en la dirección vertical.

Obsérvese que en mares abiertos la mayoría de las variaciones de la densidad están asociadas a cambios en la temperatura, por lo que la pycnoclina se correspondería con la termoclina permanente, mientras que en las zonas costeras, fiordos y estuarios, es la salinidad el factor determinante de la densidad.

1.2 MASAS DE AGUA.

Las propiedades físicas y químicas del agua de los océanos no se combinan de cualquier manera, sino que sólo existe un número determinado de posibles combinaciones, lo que sirve para identificar lo que se conoce como masas de agua. Las propiedades que habitualmente se emplean para identificarlas son la temperatura y la salinidad, ya que se trata de propiedades conservativas, esto es, una vez que el agua deja de estar en contacto con la atmósfera, estas propiedades sólo pueden cambiar debido a la mezcla con aguas de diferentes características. Las masas de agua también pueden ser identificadas por sus características químicas o biológicas, tales como el contenido de oxígeno o de nutrientes disueltos, e incluso por la presencia de determinadas comunidades de organismos, pero estas características pueden cambiar por razones muy diferentes a la mezcla de masas de agua, tales como procesos químicos, físicos o biológicos, esto es, se trata de propiedades no conservativas.

Para identificar las masas de agua presentes en una localización determinada se emplea el diagrama salinidad-temperatura (diagrama S-T), en el que se representan los valores de las observaciones de salinidad y temperatura para sucesivas profundidades y se unen los puntos conforme se va incremen-

tando la profundidad, obteniendo así un gráfico con la temperatura en el eje vertical y la salinidad en el eje horizontal. Un ejemplo de un diagrama S-T se muestra en la Figura 1.2.1.

Si la masa de agua es completamente homogénea (caso ideal), ésta es representada en el diagrama S-T como un único punto, y se le conoce como "agua tipo". Así, si las observaciones aparecen agrupadas alrededor de un punto, eso indicaría la existencia de un agua tipo. Como las masas de agua se forman por la mezcla de dos o más aguas tipo, éstas serán representadas por una línea.

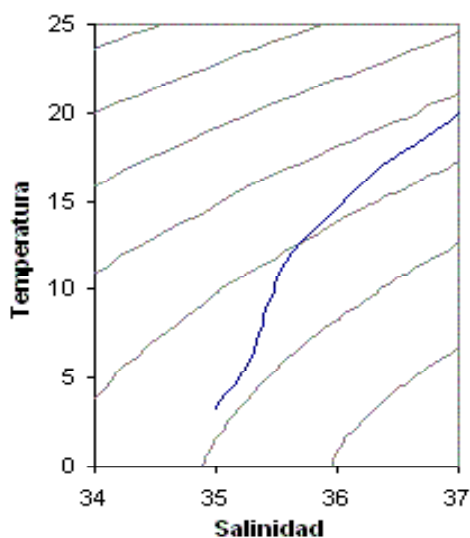


Figura 1.2.1. Ejemplo de diagrama S-T.

Existe un gran número de masas de agua, cada una caracterizada por sus valores de salinidad y temperatura. Generalmente se distingue entre masas de agua superiores, cuya distribución está fuertemente influenciada por las corrientes superficiales, y que abarcan toda la capa de mezcla y la parte superior de la termoclina permanente, por lo que presentan un grosor variable, las masas de agua intermedias y las masas de agua profundas. Si el agua más profunda que está en contacto con el fondo del mar presenta unas características diferentes a las de la masa de agua que la rodea, entonces se le denomina masa de agua de fondo.

Sin embargo, entre las masas de agua se produce mezcla, ya que el agua de los océanos está continuamente en movimiento y lo hace no de forma laminar, sino de una forma turbulenta, lo que puede llevar a que aguas de muy distintas características se encuentren en contacto. Estas turbulencias pueden estar asociadas a una gran cantidad de fenómenos, como pueden ser las olas, las variaciones en la velocidad debido a la profundidad, las corrientes producidas por las mareas, etc. Pero la mezcla de las aguas producida por estas turbulencias tiene una escala mucho mayor horizontalmente que verticalmente, ya que el efecto del incremento de la densidad con la profundidad inhibe o suprime la mezcla vertical por la estabilidad de dichas aguas. Sin embargo, aun cuando la turbulencia no produzca la mezcla verticalmente, las propias variaciones de la salinidad y la temperatura pueden provocarla, ya que se produciría un proceso de difusión del calor y de la sal a nivel molecular, lo que puede provocar pequeñas inestabilidades que contribuyen de forma significativa a la mezcla vertical de los océanos, lo que se reflejará a su vez en las características de las masas de agua.

Al igual que en la superficie del mar, en la superficie que separa dos capas de agua con distinta densidad, se pueden producir ondas llamadas internas que pueden propagar energía a través de los océanos.

1.3 CIRCULACIÓN GENERAL DE LOS OCÉANOS.

1.3.1 INTERACCIÓN OCÉANO-ATMÓSFERA:

La temperatura en la superficie de la Tierra varía según la latitud y las estaciones, ya que la temperatura depende de la intensidad de la radiación de los rayos solares que alcanzan su superficie, y esta intensidad no es la misma ni en todas las zonas de nuestro planeta ni a lo largo de las estaciones del año, lo que es debido a la inclinación del eje de la Tierra con respecto a su órbita alrededor del Sol. Pero la radiación recibida no es absorbida completamente, sino que una parte es reflejada de nuevo a la atmósfera. Esta radiación emitida por la Tierra a la atmósfera, en cambio, es más uniforme, ya que el océano y la atmósfera transportan calor desde el ecuador a los polos.

Cuando el viento se desliza sobre la superficie de los océanos transmite energía a las capas más superficiales de éstos, poniendo sus aguas en movimiento. Es evidente que cuanto mayor sea la velocidad del viento, mayor será la fuerza de rozamiento actuando sobre la superficie del mar y por lo

tanto, mayor será la corriente superficial que éste genera. Pero la dirección de las corrientes superficiales no es la misma que la del viento, ya que ésta se ve afectada por la rotación de la tierra, de tal manera que la fuerza de Coriolis desvía la dirección de las corrientes hacia la derecha (izquierda) de la dirección del viento en el Hemisferio Norte (Sur).

Por otra parte, el efecto del viento sobre la superficie de los océanos es transmitido hacia el interior, pero este efecto va disminuyendo con la profundidad debido a la fricción interna generada por el movimiento de las distintas capas de agua, cuya magnitud resulta mucho mayor en los océanos por el movimiento turbulento de sus aguas.

Así, de la combinación de los efectos producidos por la fuerza de Coriolis y por la fricción interna, Ekman dedujo que la velocidad de las corrientes producidas por el efecto del viento decrecen exponencialmente con la profundidad, y que su dirección se desvía 45° de la dirección del viento en la superficie, incrementándose la desviación de dicho ángulo conforme se produce un incremento de la profundidad. Los vectores de la dirección de las corrientes forman una espiral cuyos valores teóricos son conocidos como la espiral de Ekman, y a la capa de agua comprendida entre la superficie y la profundidad donde termina el efecto del viento, se le denomina capa de Ekman.

Lo más importante de esta teoría es el hecho de que el transporte medio de la corriente para la capa de Ekman se desvía 90° a la derecha (izquierda) de la dirección del viento en el Hemisferio Norte (Sur), y que este transporte contribuye significativamente a la circulación general de los océanos.

Pero el efecto que produce el viento sobre la superficie de los océanos no solo causa movimientos horizontales de las aguas, sino que también provoca movimientos verticales. Estos movimientos verticales pueden ser de divergencia de las aguas superficiales, en cuyo caso las aguas profundas suben a la superficie ocupando su lugar, y de convergencia de las aguas superficiales cuando son éstas las que se hunden en zonas más profundas.

La presión atmosférica produce sobre la superficie de los océanos un efecto conocido como barómetro invertido, esto es, que un sistema de alta presión deprime la superficie del mar y una baja la elevará. Esta diferencia de nivel produce un flujo, ya que el agua tenderá a moverse hacia los niveles inferiores.

En la superficie de los océanos se producen además dos procesos de transferencia de calor y agua con la atmósfera, que influyen en los movimientos de ambos fluidos geofísicos.

1.3.2 CIRCULACIÓN TERMOHALINA.

La circulación de los océanos está controlada por las variaciones en densidad de sus aguas debido a cambios producidos en su salinidad y temperatura, lo que se conoce con el nombre de circulación termohalina.

El agua superficial de la zonas próximas al Ecuador es calentada en exceso, lo que provoca que ésta sea desplazada hacia el norte y el sur, en dirección a los polos. Cuando estas aguas llegan a altas latitudes se han producido tanto una reducción considerable de su temperatura como un aumento de la salinidad, todo esto debido a los procesos de evaporación, radiación y conducción, por lo que se incrementa su densidad. Este aumento de la densidad de las aguas superficiales provoca movimientos de convergencia, hundiéndolas a las capas profundas, y siendo transportadas a través de todos los océanos mediante el flujo horizontal. Posteriormente, estas aguas frías se van calentando lentamente al irse mezclando con las aguas más someras y calientes, produciendo movimientos de divergencia que las llevan gradualmente hacia la superficie, reiniciándose de nuevo el ciclo completo.

1.4 INSTRUMENTACIÓN MARINA.

Existe una gran cantidad de instrumentos diseñados para la obtención de datos oceanográficos que permitan la comprensión de los procesos que ocurren en los océanos, pero aquí sólo se citarán aquellos relacionados con las características físicas del agua del mar, tanto en la capa superficial de interacción océano-atmósfera, como por debajo de ésta.

Los datos oceanográficos pueden ser recogidos in situ y también a partir de la información recibida por los satélites que orbitan alrededor de la Tierra, pero los sensores de los satélites sólo son capaces de obtener mediciones de la superficie de los océanos o de muy poca profundidad, y no son capaces de detectar el movimiento y las propiedades de las aguas situadas a más de un metro de profundidad, por lo que es necesario recurrir a los métodos más convencionales de recogida de datos in situ.

Los datos relativos a la superficie de los océanos, tales como temperatura, salinidad, oleaje,..., o a las condiciones meteorológicas, como pueden ser la temperatura del aire, la velocidad y dirección del viento, la presión atmosférica,..., pueden proceder de sensores instalados en barcos o de boyas colocadas sobre la superficie del mar cuyas localizaciones son controladas por satélite.

El conocimiento sobre la circulación oceánica se ha obtenido a partir de mediciones de temperatura y salinidad realizadas generalmente desde barcos de investigación mediante instrumentos que proporcionan perfiles verticales de dichas propiedades, lo que permite calcular la densidad de dichas aguas y la realización de diagramas S-T que servirían para detectar las masas de agua y sus movimientos.

Las primeras mediciones de las propiedades del agua de los océanos bajo la superficie fueron llevadas a cabo a principios del siglo XIX a bordo de barcos de investigación, y aun hoy día algunos siguen siendo empleados aunque mejorados y en combinación con otras técnicas desarrolladas con posterioridad. Inicialmente consistían en el lanzamiento de botellas que se cerraban a unas profundidades fijadas y recogían así muestras de agua para determinar su salinidad (botellas Nansen y Niskin), y de termómetros de mercurio calibrados para obtener mediciones precisas de la temperatura. Mediante el lanzamiento de botellas no sólo se obtienen datos sobre la salinidad para determinadas profundidades que van desde la superficie hasta el fondo del mar, sino que la recogida de muestras de agua permite estudiar además otras propiedades, como los nutrientes o la cantidad de oxígeno disuelto.

En los años cuarenta se desarrollaron los batitermógrafos mecánicos (MBT), instrumentos diseñados específicamente para medir la temperatura con respecto a la profundidad, pero que estaban limitados únicamente a los primeros 100 ó 200 metros. Con esa misma finalidad pero con una tecnología totalmente distinta, fueron desarrollados en los años sesenta los batitermógrafos extensibles (XBT), que pueden alcanzar mayores profundidades y cuya facilidad y bajo coste hacen que en la actualidad sean usados frecuentemente.

A mediados de los años sesenta también se desarrollaron instrumentos de carácter electrónico que medían de forma continua la salinidad, la temperatura y la profundidad (STD), y que a partir de los años setenta evolucionaron en instrumentos mucho más precisos que miden la salinidad a partir de la conductividad, la temperatura y la profundidad (CTD), y que en la actualidad siguen utilizándose.

1.5 EL SONIDO EN EL AGUA DEL MAR.

La velocidad de sonido en el agua del mar (c) es una variable oceanográfica de tipo escalar que determina muchas de las características de la transmisión del sonido en el medio en el que se propaga.

El sonido se propaga como ondas de presión y, aunque no se trata de ondas

sinusoidales, como el comportamiento de la presión acústica cuando pasa la onda sonora sí es de tipo sinusoidal, ésto permite caracterizar la velocidad de propagación de las ondas sonoras por su frecuencia (η) y su longitud de onda (λ) mediante la relación

$$c = \eta\lambda.$$

Evidentemente, la propagación del sonido va a depender tanto del medio por el que se propaga como de las características de éste. Así, la velocidad de sonido en un determinado medio viene dada por

$$c = \sqrt{B/\rho},$$

siendo B el módulo de la compresibilidad del medio donde se propaga el sonido y ρ su densidad.

Cuando el sonido se propaga en los océanos, tanto B como ρ van a depender de la temperatura, la salinidad y la presión, por lo que el valor de la velocidad de sonido en el agua del mar se puede obtener a partir de los valores de estas propiedades. Es evidente que un aumento de temperatura implica una disminución de la densidad, lo que hace que c también aumente. En cambio, si la salinidad se incrementa, ésto implica un incremento de la densidad, pero también implica que la compresibilidad del medio aumenta, por lo que el incremento de densidad es compensado y c vuelve a aumentar su valor, aunque en una proporción menor.

Por otro lado, el hecho de que la velocidad de sonido dependa de las propiedades del agua implica refracción, esto es, que la dirección de propagación del sonido puede ser alterada por los cambios producidos en estas propiedades. Así, en la mayoría de los océanos se ha comprobado la existencia de un mínimo de velocidad de sonido aproximadamente en torno a los 1000 metros de profundidad. Esto permite que las ondas sonoras que comienzan a propagarse en torno a esta profundidad permanezcan en esta zona, ya que la energía que emiten se ven afectadas por la refracción tanto por arriba como por debajo de esta profundidad, lo que las obliga a propagarse únicamente en sentido horizontal, generando lo que se conoce como canal de sonido, y a través de los cuáles el sonido puede propagarse a lo largo de miles de kilómetros debido a que la energía perdida es mucho menor que si se propagan verticalmente.

Tal y como ya se ha comentado en apartados anteriores, los gradientes horizontales de salinidad y temperatura son mucho más pequeños que los verticales, por lo que las variaciones horizontales de la velocidad de sonido

van a ser también menores que las producidas verticalmente. Con respecto a la distribución vertical de c , es posible establecer que ésta generalmente aumenta con un aumento de la presión, salvo en los canales de sonido ya mencionados. A la función que determina la variación de la velocidad de sonido con respecto a la profundidad se le denomina usualmente perfil de velocidad de sonido, y en la Figura 1.5.1 se muestra un perfil típico.

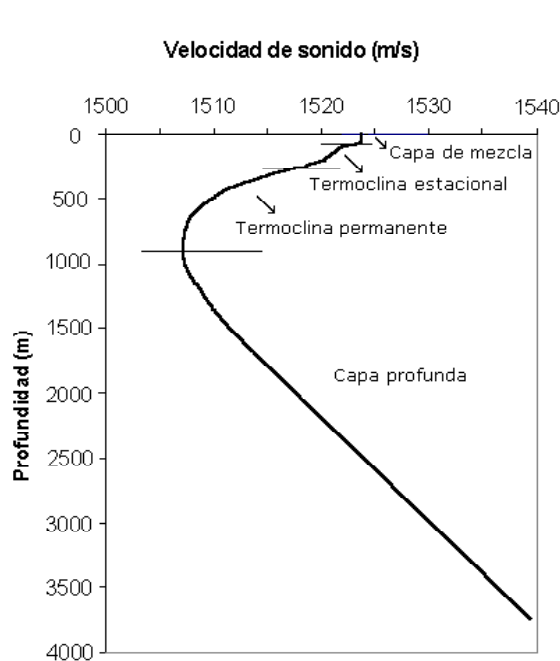


Figura 1.5.1. Perfil típico de velocidad de sonido en el agua del mar.

En un perfil de velocidad de sonido se pueden distinguir claramente diversas capas que presentan características y propiedades diferentes. Justo debajo de la superficie se encuentra la capa de mezcla, en la que la velocidad de sonido es susceptible a cambios diarios por la acción del viento, por las condiciones meteorológicas, por el intercambio de calor y agua entre atmósfera y océano, etc. El sonido tiende a ser atrapado en esta zona, por lo que presenta un valor prácticamente constante. Por debajo puede aparecer la termoclina estacional, caracterizada por un descenso pronunciado de la velocidad de sonido con respecto a la profundidad, y que, evidentemente, varía con las estaciones. A continuación se encuentra la termoclina permanente,

en la que la velocidad de sonido continúa disminuyendo con la profundidad pero de una forma más suave. Y por último, la capa profunda, en la que la velocidad de sonido incrementa con la profundidad, debido al efecto de la presión, ya que la temperatura es prácticamente constante en esta zona.

1.5.1 CÁLCULO DE LA VELOCIDAD DE SONIDO.

Las primeras tablas de velocidad de propagación del sonido para diversas zonas y océanos del mundo fueron elaboradas por D. J. Matthews (1939), siendo éstas adoptadas internacionalmente hasta que fueron sustituidas por las tablas de D. J. T. Carter (1980).

Existen numerosas ecuaciones para determinar el cálculo de la velocidad del sonido a partir de la temperatura, la salinidad y la presión del agua del mar. Los primeros trabajos adoptados internacionalmente fueron desarrollados por W. Wilson (1960), pero estas ecuaciones fueron sustituidas en los años setenta por las de V. A. Del Grosso (1974) y las de C. T. Chen y F. J. Millero (1977), debido a que se observaron resultados erróneos en sus medidas de la velocidad del sonido realizadas en agua pura, especialmente con la utilización de las modernas técnicas de medición de salinidad y profundidad. Posteriormente, G. S. K. Wong y S. Zhu (1995) adaptaron estos algoritmos a la nueva escala internacional de temperatura de 1990.

Hay que tener en cuenta que dichos algoritmos fueron derivados de ajustes realizados experimentalmente en laboratorios, por lo que poseen unos rangos de temperatura, salinidad y presión dentro de los cuáles únicamente pueden ser utilizados. En el caso de la ecuación de Del Grosso, ésta tiene validez para valores de temperatura que oscilan entre 0 y 30°C, para unos valores de salinidad comprendidos entre 30 y 40, y para una presión que varía entre 0 y 1000 kg/cm² (que equivale aproximadamente a 980.6652 bares). La ecuación de Chen y Millero presenta un rango de validez más amplio, oscilando los valores de temperatura entre 0 y 40°C, los de salinidad entre 0 y 40, y los de presión entre 0 y 1000 bares.

Evidentemente, estas fórmulas llevan asociadas unos errores de predicción, por lo que la elección del algoritmo a emplear va a depender de la precisión que se pretenda conseguir. En este sentido, el algoritmo de Del Grosso posee una desviación típica de 0.05 m/s, mientras que la desviación típica de la ecuación de Chen y Millero asciende a 0.19 m/s. Nótese que estos errores aumentarán si se emplean para su cálculo los datos procedentes de XBT, lo que puede ocurrir de forma frecuente. En estos casos, como no se dispone de

mediciones de la salinidad con respecto a la profundidad, el procedimiento habitual consiste en asignarle valores a la distribución de la salinidad, siendo determinante la elección de estos valores en el error de predicción cometido. Estos valores pueden proceder de otros sensores de los que se disponga en el barco o bien de estudios precedentes que se hayan realizado en la zona.

Aunque la ecuación de Chen y Millero es la adoptada en la actualidad internacionalmente por la UNESCO, existen numerosos estudios que abordan su precisión y su rango de aplicabilidad, comparándola con la de Del Grosso, e incluso algunos investigadores consideran esta última más consistente. Véase por ejemplo B. D. Dushaw et al. (1993) y J. L. Speisberger y K. Metzger (1991a, 1991b).

Por otro lado, como ambos algoritmos emplean como variable en su cálculo la presión, y habitualmente de lo que se dispone es de temperatura y salinidad respecto a la profundidad, se hace necesario el uso de un algoritmo apropiado para convertir la presión en profundidad y viceversa. Varias fórmulas han sido desarrolladas para realizar esta conversión, siendo el proporcionado por C. C. Leroy y F. Parthiot (1998) el empleado internacionalmente en la actualidad. En él se establece una fórmula general para todos los océanos y mares del mundo que viene acompañada por un conjunto de términos correctores específicos para determinadas áreas, como por ejemplo cuencas cerradas o aguas polares.

Los algoritmos para el cálculo de la velocidad del sonido en el agua del mar proporcionados por Del Grosso y por Chen y Millero adaptados a la escala internacional de temperatura de 1990 por Wong y Zhu, y los de conversión de presión a profundidad y de profundidad a presión por Leroy y Parthiot, han sido implementados en MATLAB, y sus códigos figuran en el Anexo 1.

2 BREVE DESCRIPCIÓN DEL MAR DE ALBORÁN.

2.1 FISIOGRAFÍA DEL MAR DE ALBORÁN.

El Mar de Alborán es la cuenca más occidental del Mar Mediterráneo y tiene un área aproximada de 54000 km². Conecta al oeste con el Océano Atlántico a través del Estrecho de Gibraltar y al Mediterráneo en su límite oriental. Según la *International Hydrographic Organization*, se encuentra limitado al norte por la costa española, al sur por Marruecos y Argelia, al oeste por el Estrecho de Gibraltar y al este por la línea que une el Cabo de Gata (Almería, España) con el Cabo Figalo (Argelia).

Como puede observarse en su mapa fisiográfico, que se muestra en la Figura 2.1.1, el Cabo de Tres Forcas, la Isla de Alborán y el final de las depresiones que existen tanto al norte como al sur de la isla, separan el Mar de Alborán alrededor de la longitud 3° W en dos partes, una occidental y una oriental, con unas profundidades máximas de 1500 metros y de más de 2000 metros en ambas partes respectivamente.

2.1.1 PLATAFORMA CONTINENTAL.

La plataforma continental puede dividirse en cuatro zonas en función de su anchura. En la costa española, la anchura oscila entre 10 km en Málaga y 2 km en Cabo Sacratif, teniendo una media de 5 km. A lo largo de la costa africana, la plataforma tiene una anchura de 6 km en media desde el Estrecho de Gibraltar hasta Punta Basicur, hasta el Cabo de Tres Forcas fluctúa entre 3 km y 18 km, y a partir de este punto hasta su límite oriental mantiene una anchura prácticamente uniforme de aproximadamente 15 km. La profundidad del límite de la plataforma se encuentra alrededor de los 100 metros, salvo en la parte de la costa africana que discurre entre el Cabo de Tres Forcas y el Cabo Figalo, cuya profundidad desciende a 150 metros.

2.1.2 TALUD CONTINENTAL.

El talud continental es una región extensa con una topografía muy variada, estando interrumpido por mesetas y bancos de diversos tamaños y profundidades. El gradiente del talud en la costa española es relativamente constante, con un valor representativo de 1:30, salvo para las zonas interrumpidas por

la existencia de dos mesetas: una pequeña meseta situada al suroeste del Cabo de Gata y delimitada al oeste por el Cañón de Almería, y otra de mayores dimensiones ($80 \text{ km} \times 60 \text{ km}$), que presenta numerosos picos volcánicos y se encuentra situada entre 3° W y 4° W . Su límite inferior está definido aproximadamente por la isobata de los 900 metros y presenta un talud continental muy acusado que conduce a una gran depresión conocida como la Depresión de Alborán. Los gradientes del talud a lo largo de la costa africana son suaves y varían desde aproximadamente 1:15 en Punta Basicur hasta 1:9 en Cabo Figalo, existiendo una pendiente anómala alrededor de la longitud 2° W que conduce a una pequeña depresión que alcanza una profundidad máxima próxima a los 2400 metros. El talud continental de la costa africana se encuentra interrumpido por tres mesetas, una cordillera y una pequeña depresión.

Entre Punta Almina y hasta aproximadamente la longitud $4^\circ 20' \text{ W}$, se extiende una alargada meseta caracterizada por suaves pendientes. Tiene unas dimensiones aproximadas de $80 \text{ km} \times 55 \text{ km}$, y su límite superior oscila alrededor de la isobata de los 600 metros. Entre aproximadamente $4^\circ 20' \text{ W}$ y el Cabo de Tres Forcas discurre una cordillera alargada de unos 160 km de longitud y orientada en la dirección noreste-suroeste, que contiene en su extremo suroeste a los bancos de Xauen y Tofiño, y en su extremo noreste a la Isla de Alborán. La zona suroeste de esta cordillera, denominada Cordillera de Alborán, está separada de la plataforma continental africana por una estrecha depresión de aproximadamente 400 metros de profundidad y que discurre inicialmente paralela a la costa y luego cambia para seguir la tendencia de la cordillera, y junto a esta depresión aparece una pequeña meseta. La volcánica Isla de Alborán está situada en el extremo más oriental de esta cordillera y tiene unas dimensiones de 600 metros \times 250 metros. Se encuentra rodeada por una meseta de unos 25 ó 30 metros de anchura, a partir de la cual el fondo desciende abruptamente. Por último, existe en esta zona otra depresión que discurre paralelamente al sur de la cordillera. Su parte más profunda, que es conocida como la Cuenca Oriental de Alborán, tiene unas dimensiones de $13 \text{ km} \times 6 \text{ km}$, y está limitada por la isobata de los 1200 metros, superando ligeramente esta profundidad. Entre el Cabo de Tres Forcas y el Cabo Figalo existe una extensa meseta de dimensiones aproximadas $135 \text{ km} \times 63 \text{ km}$ y limitada al norte por la isobata de los 1000 metros. Su extremo occidental está surcado por una serie de cordilleras y valles orientados en dirección norte-sur. En el resto de la meseta existen varios bancos, como los de Provençaux, Alidade y Cabliers.

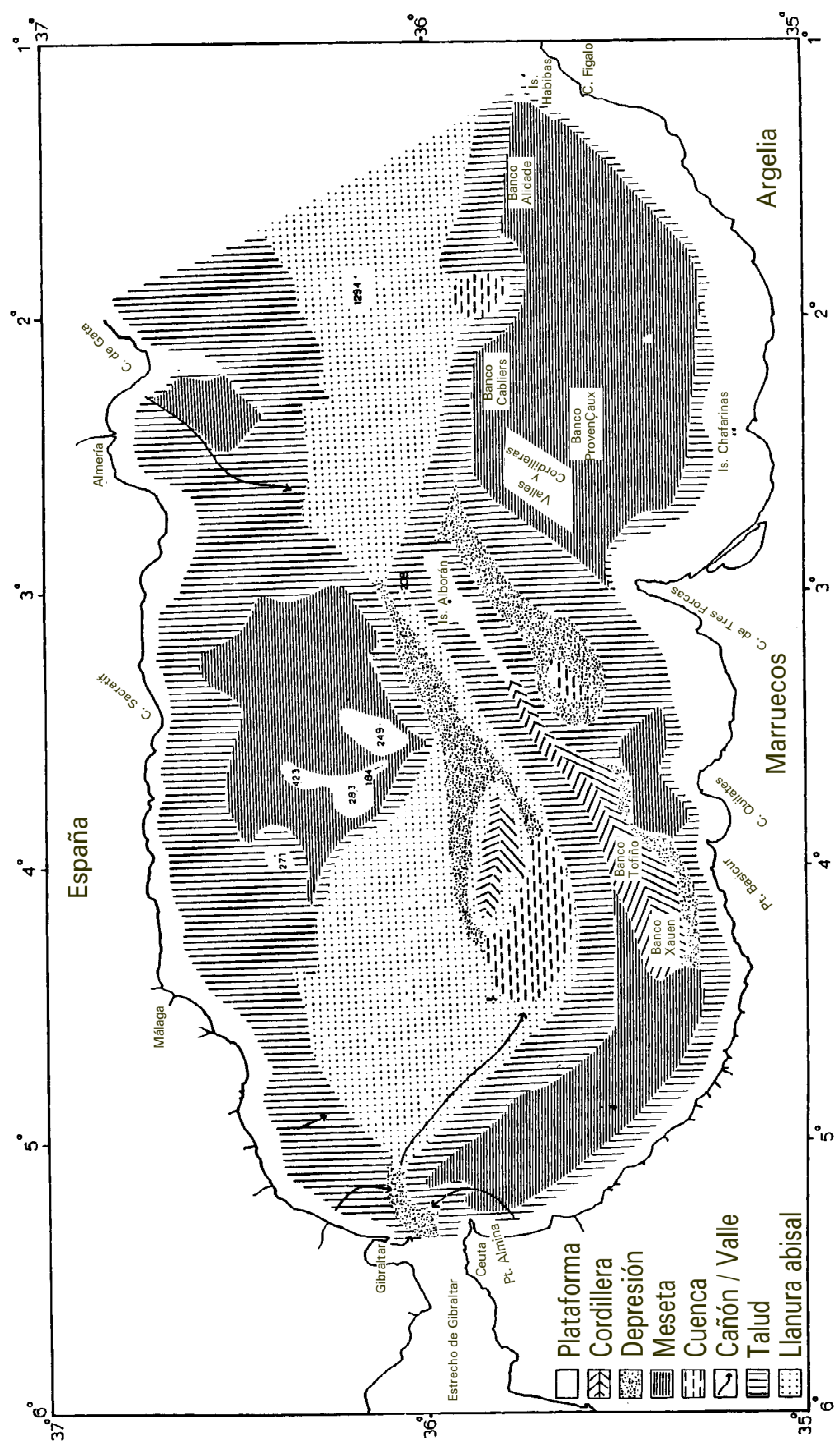


Figura 2.1.1. Mapa fisiográfico del Mar de Alborán (adaptado de Carter et al.).

2.1.3 LLANURA ABISAL.

Al norte de la Cordillera de Alborán discurre paralelamente a ésta una alargada depresión, cuya parte más profunda se conoce como la Cuenca Occidental de Alborán. Dicha cuenca está delimitada por la isobata de los 1400 metros y tiene unas dimensiones de 40 km × 22 km, alcanzando una profundidad máxima ligeramente superior a 1500 metros. A lo largo del borde noreste de dicha cuenca se extiende una pequeña cordillera de aproximadamente 32 km de longitud y que contiene varios picos.

2.2 OCEANOGRAFÍA FÍSICA DEL MAR DE ALBORÁN.

2.2.1 CONDICIONES ATMOSFÉRICAS.

Los movimientos del anticiclón semipermanente de las Azores son determinantes para la climatología del Mar de Alborán.

Durante el verano (junio-septiembre) este anticiclón se extiende hacia el noreste y como consecuencia el tiempo es seco y templado con vientos ligeros de levante, aunque el aire caliente, al moverse sobre el agua más fría, puede producir nieblas y estratos bajos en Gibraltar. Estos vientos de levante también pueden aparecer en otras épocas del año, y en ese caso suelen ir acompañados de nubes bajas y de grandes lluvias. A principios de verano pueden además aparecer vientos del sur o suroeste, aunque no son frecuentes. En estos casos, los cielos se nublan, las temperaturas son tibias y se producen fuertes lluvias. En ocasiones se producen los terrales calientes de verano cuando los vientos provienen del norte.

A lo largo del invierno (noviembre-febrero) el tiempo es inestable, húmedo y ventoso, ya que el anticiclón de las Azores disminuye y se desplaza hacia el sur. En esta época del año habitualmente soplan vientos húmedos y suaves de poniente, provocando que los cielos estén generalmente despejados con algunos cúmulos y que esporádicamente puedan aparecer chaparrones. También pueden soplar vientos del norte produciendo los ya mencionados terrales, aunque en este caso de invierno.

Respecto a las estaciones de transición, el otoño dura sólo el mes de octubre y el paso al invierno es muy brusco. En cuanto a la primavera (marzo-junio), ésta consta de períodos alternos de tiempo invernal y de tiempo veraniego, por lo que en lo referente a los vientos, pueden soplar de poniente

o del norte en sus períodos de invierno, y de levante en los de verano. Los vientos del sur o suroeste también pueden aparecer a finales de esta estación.

2.2.2 MASAS DE AGUA.

Las variaciones en temperatura y salinidad respecto a la profundidad muestran la existencia de tres importantes masas de agua para el Mar de Alborán: el Agua Atlántica (AA), el Agua Levantina Intermedia (ALI) y el Agua Profunda (AP).

La mayor parte del mar de Alborán está cubierto por una capa superior de AA, cuyo grosor varía desde 150 metros en el centro de la zona occidental de Alborán hasta menos de 50 metros cerca de la costa española. Esta capa procede de la mezcla del Agua Central Noratlántica y del Agua Atlántica Superficial que penetran en el Mar de Alborán a través del Estrecho de Gibraltar y que están caracterizadas por unos valores de salinidad relativamente inferiores y de temperatura diferentes a la de las aguas superficiales propias del Mar de Alborán.

La salinidad del AA varía desde 36.2 a 36.5, y su distribución vertical de la temperatura cambia a lo largo del año. En verano se observan unos gradientes acusados de 0.04 °C/m de la termoclina estacional y permanente combinadas hasta una profundidad de 150 metros, mientras que en invierno es casi isoterma.

Debajo del AA existe la capa del mínimo de temperaturas, zona de transición que separa las aguas atlánticas de las mediterráneas (ALI y AP). Esta capa fría pero moderadamente salina presenta unos gradientes elevados de temperatura en torno a 0.02 °C/m y de salinidad de 0.02/m, y alcanza su densidad de equilibrio por encima del ALI.

El estrato del ALI se origina durante el invierno en la zona de Rodas y Chipre, extendiéndose por el Mediterráneo hacia el oeste y penetrando en el Mar de Alborán por su parte oriental. Se encuentra normalmente entre 200 metros y 600 metros de profundidad, y se caracteriza por unos gradientes de salinidad y temperatura pequeños, alcanzando sus valores máximos de temperatura en torno a 13.2 °C y de salinidad entre 38.45 y 38.50, aunque durante los últimos años se ha observado un ALI más cálida y ligeramente más salina.

Por debajo de estos máximos, la temperatura y la salinidad decrecen más lentamente, indicando la zona ocupada por el AP, cuyos valores varían interanualmente entre 12.91 °C y 12.80 °C y entre 38.415 y 38.46, respectivamente.

Este AP se forma durante el invierno frente a la costa mediterránea francesa, hundiéndose y extendiéndose hacia el sur ocupando las capas profundas de los mares Balear, Tirreno y Alborán.

2.2.3 CIRCULACIÓN OCEÁNICA.

De la distribución de las masas de agua se puede deducir de forma general cómo es la circulación oceánica en el Mar de Alborán.

El AA penetra rápidamente en la zona occidental del Mar de Alborán por el Estrecho de Gibraltar, dirigiéndose inicialmente en la dirección del eje del Estrecho hasta que gira hacia el sur alrededor de la longitud 3° 30' W y se bifurca. Una rama vuelve hacia el oeste, generando un giro anticiclónico, y la otra rama, que es la parte principal de la corriente atlántica, continúa hacia el este, pasando al sur de la Isla de Alborán y penetrando en la zona oriental del Mar de Alborán. En esta zona forma meandros, originando remolinos que separan el flujo atlántico de la costa africana, hasta que tras pasar estos meandros, el flujo se aproxima a la costa argelina y discurre paralela a ésta hacia el este, ensanchándose y perdiendo velocidad conforme fluye en el Mediterráneo.

Sobre la circulación de las aguas mediterráneas, ALI y AP, no existe un gran conocimiento. Lo único que se puede afirmar es que fluyen hacia el oeste y que contribuyen en determinada medida al flujo mediterráneo que sale al Atlántico a través del Estrecho de Gibraltar.

Por último, respecto a las velocidades de los flujos, el análisis de los perfiles verticales de las corrientes medias indican una disminución brusca del flujo en los primeros 250 metros, obteniendo unos valores superiores a 60 cm/seg para la superficie y de 1 cm/seg para la capa intermedia, y posteriormente se produce una disminución suave hacia el fondo.

2.3 TIPOS DE PERFILES ACÚSTICOS DETECTADOS EN EL MAR DE ALBORÁN.

Los perfiles de velocidad de sonido analizados en este trabajo fueron recogidos a bordo del B/O Vizconde de Eza en las campañas oceanográficas "Alborán 2.002" y "Alborán 2.003", realizadas por el Instituto Español de Oceanografía y la Secretaría General de Pesca Marítima durante los meses de julio y agosto

de dichos años. El área de trabajo se encuentra comprendida entre las latitudes 35.7° N y 36.8° N, y las longitudes 2° W y 4.7° W. Sus localizaciones aparecen en la Figura 2.3.1.

Las mediciones fueron realizadas con el instrumento "Sound Velocity, Temperature and Depth Profiler Model SVPlus", aparato que forma parte de la instrumentación científica de la que se dispone a bordo del B/O Vizconde de Eza. Este instrumento consta de una serie de sensores diseñados para medir la presión (precisión de ± 2 metros), la temperatura (precisión de ± 0.005 $^{\circ}$ C) y la velocidad de sonido (precisión de ± 0.06 m/s).

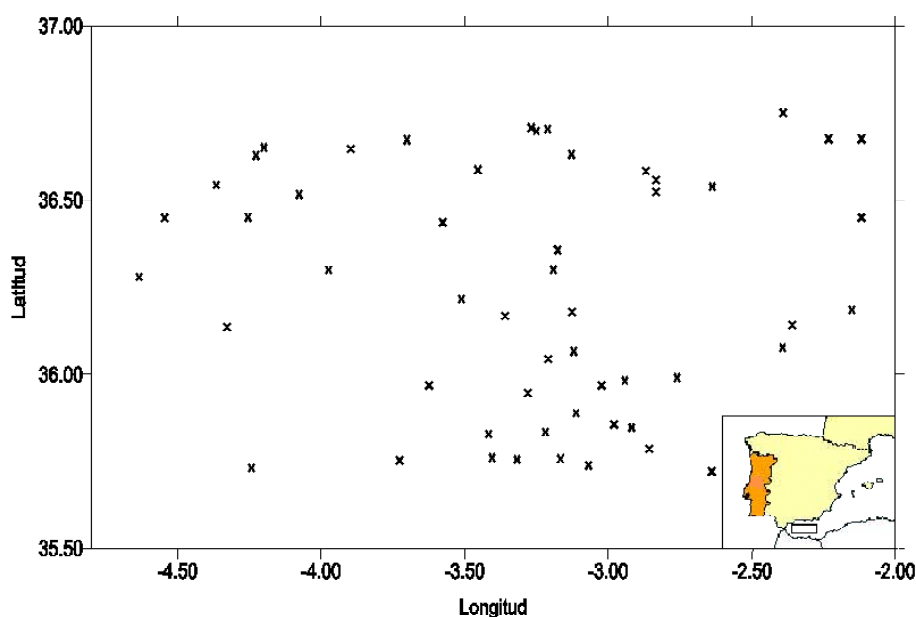


Figura 2.3.1. Localizaciones de los perfiles de velocidad de sonido analizados en este estudio.

Del análisis de los perfiles realizados, se puede deducir que el final de la termoclina permanente se encuentra entre 150 metros y 200 metros de profundidad, y a partir de esta profundidad, el comportamiento de la velocidad de sonido es lineal, creciente con la profundidad y similar para todas las zonas. Presenta un valor mínimo de aproximadamente 1508 m/s a la profundidad de 200 metros, y un máximo de aproximadamente 1537 m/s a 1900 metros de profundidad. En cambio, hasta esa profundidad de 200 metros, que se corresponde con la capa de mezcla y con la termoclina, se

han observado distintos perfiles acústicos según las unidades fisiográficas a las que pertenecen las localizaciones en las que estos perfiles fueron tomados, detectándose incluso subtipos dentro de una misma unidad fisiográfica.

2.3.1 TALUD CONTINENTAL DE LA COSTA ESPAÑOLA.

En el área de estudio que se encuentra en el talud continental de la costa española se han detectado cuatro subzonas con distintos perfiles acústicos, que serán denotados por TC1, TC2, TC3 y TC4 respectivamente.

El primer tipo de perfil acústico, TC1, corresponde a la pequeña meseta situada frente al Cabo de Gata y sus representaciones gráficas se muestran en la Figura 2.3.2. Este perfil se caracteriza por una capa de mezcla bastante fina, profundizando sólo hasta aproximadamente 5 metros, y cuyo valor de velocidad de sonido oscila alrededor de 1530 m/s, siendo éste el menor de los observados para la capa de mezcla de toda la zona del talud continental. La termoclina estacional en esta zona abarca hasta una profundidad de 25 metros aproximadamente, rondando un valor de 1510 m/s. El final de la termoclina permanente se encuentra en esta zona bastante más superficial, situándose alrededor de una profundidad de 100 metros.

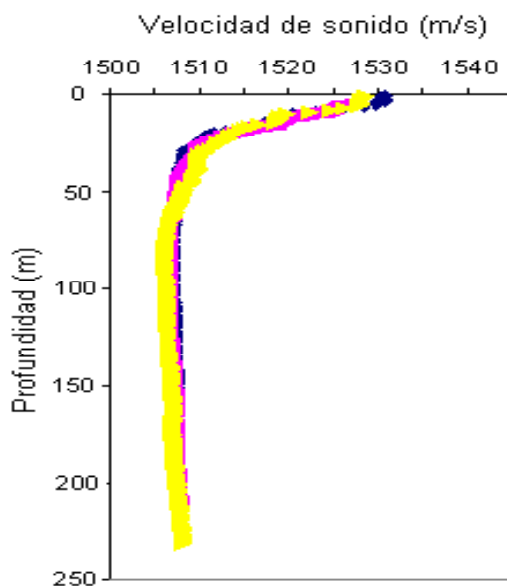


Figura 2.3.2. Perfil acústico TC1.

El segundo tipo de perfil acústico, TC2, corresponde a la meseta situada entre las longitudes 3° W y 4° W, y sus representaciones gráficas se muestran en la Figura 2.3.3. En estos gráficos se observan claramente dos tendencias diferentes hasta la profundidad de 80 metros aproximadamente. Dichas diferencias son debidas a que unos perfiles fueron realizados durante el día mientras que otros se realizaron por la noche. Para los perfiles diurnos la capa de mezcla abarca hasta una profundidad de 15 metros aproximadamente, rondando un valor de 1535 m/s, mientras que para los nocturnos se aprecia una capa de mezcla prácticamente inexistente y con una velocidad de sonido más elevada, en torno a 1538 m/s. En cuanto a la termoclina, se observa una pendiente de termoclina mucho más acusada en los perfiles nocturnos que en los diurnos, pero puede considerarse que en esta zona abarca hasta una profundidad de 140 metros.

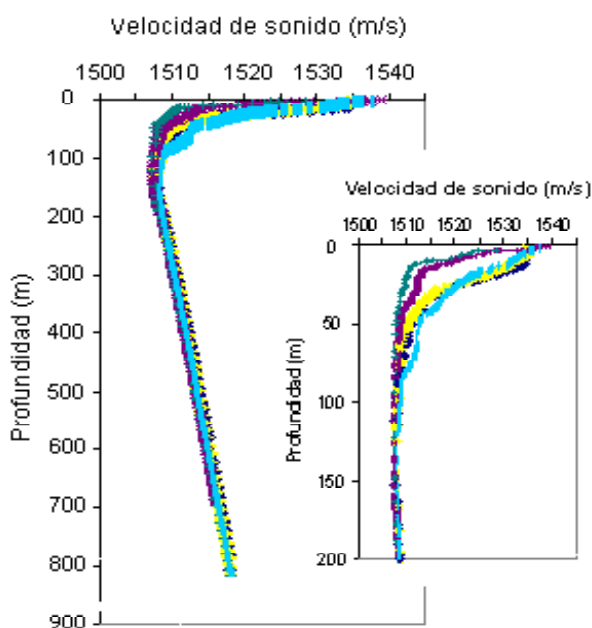


Figura 2.3.3. Perfil acústico TC2.

El tercer tipo de perfil acústico, TC3, corresponde a la zona del talud que se encuentra aproximadamente al este de la longitud 3.30° W y sus representaciones gráficas se muestran en la Figura 2.3.4. La capa de mezcla

presenta unos rangos muy variados, profundizando desde 7 metros hasta 15 metros y con valores oscilando entre 1532 m/s y 1538 m/s. El final de la termoclina estacional se encuentra en torno a la profundidad de 25 metros, con un rango de valores comprendido entre 1512 m/s y 1517 m/s. Con respecto a la termoclina permanente, puede considerarse que finaliza a una profundidad de 160 metros.

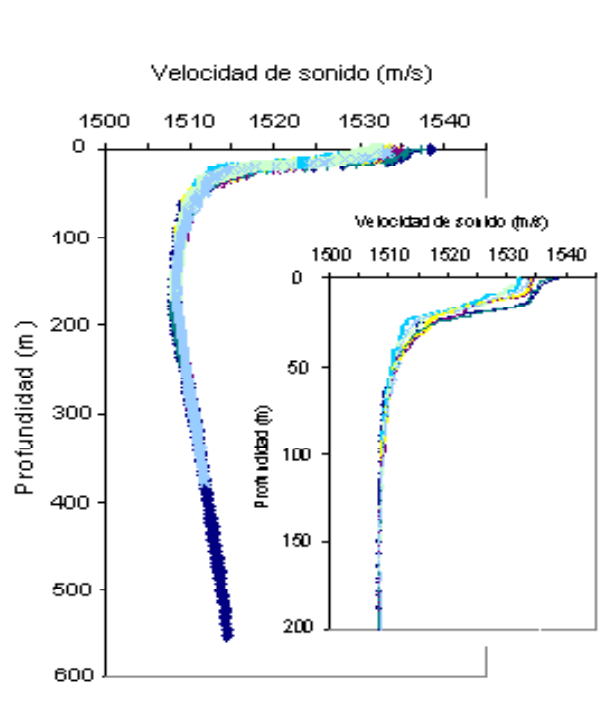


Figura 2.3.4. Perfil acústico TC3.

El cuarto tipo de perfil acústico, TC4, corresponde a la zona del talud que se encuentra aproximadamente al oeste de la longitud 3.30° W y sus representaciones gráficas se muestran en la Figura 2.3.5. La capa de mezcla es bastante fina, abarcando sólo los primeros 4 ó 5 metros y alcanzando un valor de 1538 m/s. En esta zona vuelve a quedar reflejada la diferencia diurna de la nocturna, obteniéndose que la termoclina estacional para los perfiles diurnos finaliza en torno a 10 metros de profundidad, presentando una pendiente muy pronunciada y con valores oscilando entre 1515 m/s y 1525 m/s, mientras que para los perfiles nocturnos la pendiente de la termoclina

es mucho más suave y puede considerarse que finaliza en torno a 30 metros de profundidad, presentando unos valores menores que varían entre 1512 m/s y 1517 m/s. El final de la termoclina permanente sin embargo está reflejada claramente en torno a una profundidad de 150 metros.

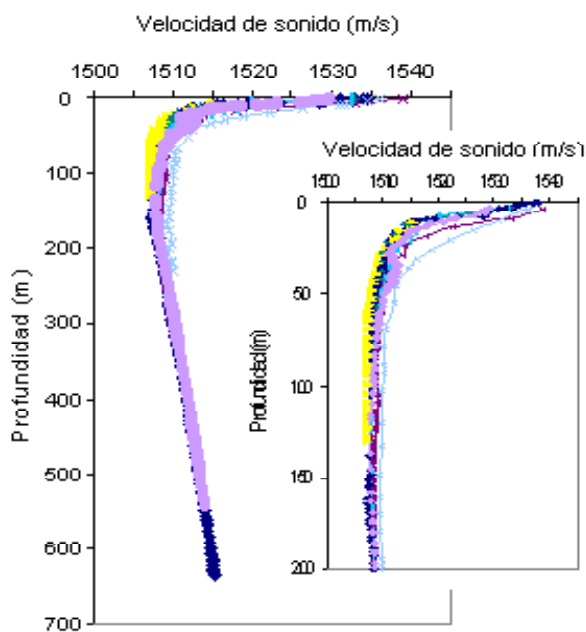


Figura 2.3.5. Perfil acústico TC4.

2.3.2 TALUD CONTINENTAL DE LA COSTA AFRICANA.

En el área de estudio que se encuentra en el talud continental de la costa africana se han detectado dos subzonas con distintos perfiles acústicos, que serán denotados por TC5 y TC6 respectivamente.

El perfil TC5 corresponde con la zona de la cordillera de Alborán y sus representaciones gráficas se muestran en la Figura 2.3.6. En estas gráficas se pueden distinguir dos tendencias diferentes en los primeros 50 metros, según las localizaciones se encuentren al norte o al sur de la Isla de Alborán.

Ambos subgrupos presentan un valor similar de la velocidad de sonido en la superficie, oscilando alrededor de 1533 m/s. La capa de mezcla para

las localizaciones situadas al sur de la Isla abarca hasta una profundidad de 10 ó 12 metros, mientras que para las que se encuentran al norte es más profunda, llegando hasta 17 ó 18 metros. Esta diferencia se traslada también a la profundidad de la termoclina estacional, llegando hasta 20 metros en los del sur y hasta 28 metros en los del norte. El rango de valores de velocidad de sonido que se alcanza al final de la termoclina estacional para las localizaciones que se encuentran al sur de la Isla es muy amplio, abarcando desde 1515 m/s hasta 1520 m/s, mientras que para los que se encuentran al norte ronda un valor de 1518 m/s. A partir de 50 metros de profundidad, el comportamiento es el mismo en ambas zonas, detectándose el final de la termoclina permanente aproximadamente a una profundidad de 175 metros. En cuanto a las pendientes de las termoclinas estacional y permanente, éstas son similares en ambas zonas.

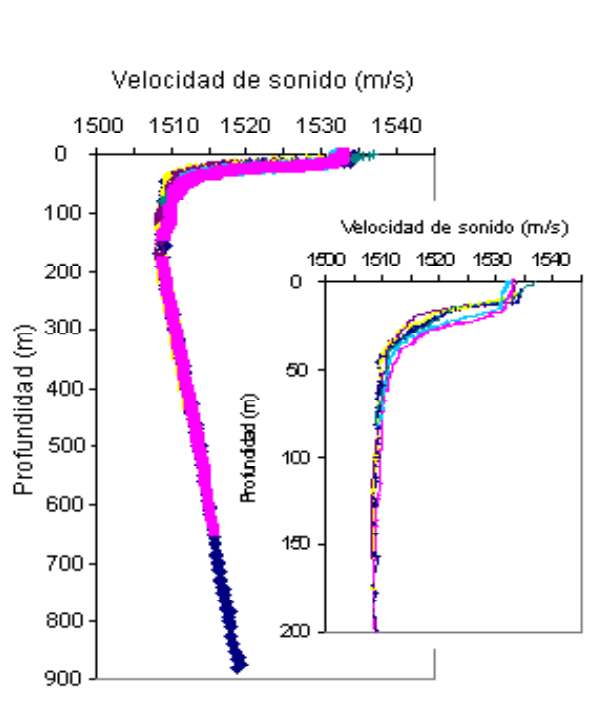


Figura 2.3.6. Perfil acústico TC5.

El perfil TC6 corresponde con la estrecha y profunda depresión que se encuentra al sur de la cordillera de Alborán y sus representaciones gráficas se

muestran en la Figura 2.3.7. En esta zona la capa de mezcla es estrecha, profundizando únicamente hasta 8 metros y presentando un valor de velocidad de sonido de aproximadamente 1534 m/s. La termoclina estacional abarca hasta una profundidad de 15 metros y sus valores en el límite oscilan entre 1517 m/s y 1521 m/s. De toda el área de estudio perteneciente al talud continental, tanto de la costa española como de la africana, es en esta zona donde la termoclina permanente se encuentra más profunda, abarcando hasta una profundidad de 200 metros.

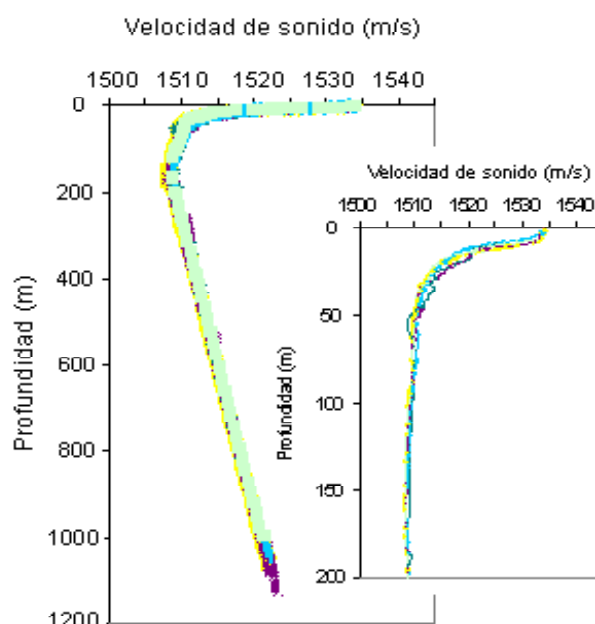


Figura 2.3.7. Perfil acústico TC6.

2.3.3 LLANURA ABISAL.

El área de estudio que se encuentra en la llanura abisal se ha dividido en dos zonas, según corresponda a la parte oriental o a la occidental, que serán denotados por LAOR y LAOC respectivamente.

Las representaciones gráficas del perfil acústico LAOR se muestran en la Figura 2.3.8. Este perfil se caracteriza por una capa de mezcla casi inexistente, considerándose que profundiza únicamente 2 ó 3 metros, y con un valor

de velocidad de sonido en torno a 1536 m/s. La termoclina estacional presenta una suave pendiente y se encuentra a bastante profundidad, abarcando hasta 50 metros y con un valor de aproximadamente 1513 m/s en su límite inferior. El final de la termoclina permanente en esta zona se encuentra a una profundidad de 200 metros.

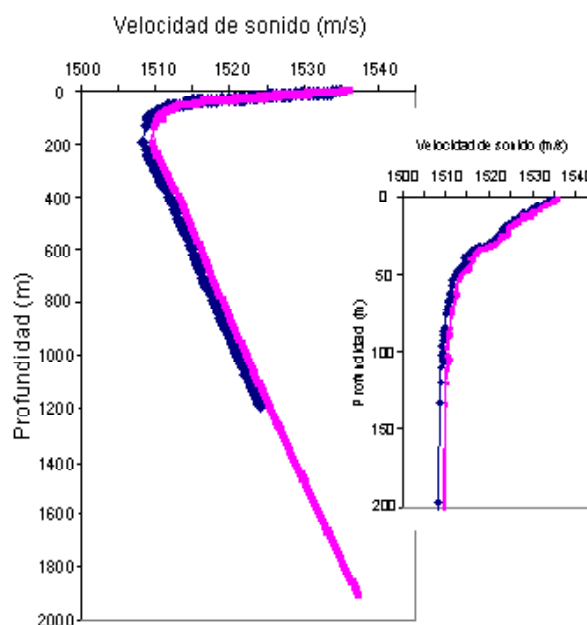


Figura 2.3.8. Perfil acústico LAOR.

En cuanto al perfil acústico LAOC, cuyas representaciones gráficas se muestran en la Figura 2.3.9, se observan dos subgrupos de perfiles, uno correspondiente a la zona que se encuentra al norte de la pequeña cordillera existente en este área, y otro al sur de ésta.

Los perfiles de la zona norte presentan gran variabilidad hasta aproximadamente una profundidad de 100 metros, pero todos tienen en común una capa de mezcla prácticamente inexistente y cuyo límite inferior no está muy claro. Los valores de la velocidad de sonido para la capa de mezcla presentan un rango muy amplio que abarca desde 1527 m/s hasta casi 1535 m/s. El final de la termoclina estacional se encuentra alrededor de 25 metros de profundidad, presentando un rango amplio de valores que

oscila entre 1512 m/s y 1517 m/s. El final de la termoclina permanente se encuentra aproximadamente a una profundidad de 200 metros.

Respecto a los perfiles de la zona sur, éstos presentan claramente una capa de mezcla hasta la profundidad de 10 metros y con valores más elevados que los de la zona norte, siendo éstos de aproximadamente 1538 m/s. Hasta el final de la termoclina, que en esta zona se encuentra aproximadamente a 210 metros de profundidad, se observan claramente cuatro cambios de pendiente que sucesivamente se van suavizando. El primero, cuya pendiente es muy pronunciada, prácticamente vertical, puede considerarse que corresponde a la termoclina estacional, abarcando hasta 15 metros y presentando un valor de velocidad de sonido de 1530 m/s en esta profundidad. El segundo abarca hasta 50 metros, con un valor de velocidad de sonido de 1520 m/s. El tercero profundiza hasta 150 metros y su valor de velocidad de sonido a esta profundidad es de 1509 m/s. El cuarto y último ya coincide con el final de la termoclina permanente.

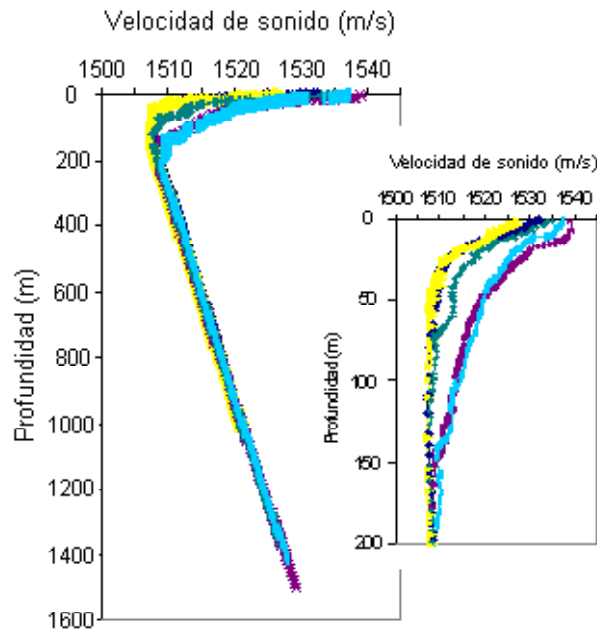


Figura 2.3.9. Perfil acústico LAOC.

3 REDES DE NEURONAS ARTIFICIALES.

3.1 INTRODUCCIÓN HISTÓRICA.

Los primeros trabajos realizados sobre Inteligencia Artificial fueron enfocados desde dos esquemas genéricos bien diferenciados. Por un lado el modelado de la actividad racional por medio de sistemas formales de reglas y manipulación simbólica, que es la rama más conocida y que se denomina simbólico-deductiva, y por otro, el desarrollo de modelos inspirados en las redes de neuronas biológicas, denominados inductivos o también conexionistas.

Si bien cada vez es mayor el conocimiento que se tiene del comportamiento del cerebro, de su funcionamiento así como de su estructura física y lógica, existe un problema de fondo que limita los resultados del segundo enfoque: estos modelos se implementan en ordenadores que están basados en la arquitectura Von Neumann, que en definitiva se apoyan en una descripción secuencial del proceso de tratamiento de la información, es decir, en algoritmos.

Las redes de neuronas artificiales, en adelante RNAS, simulan el comportamiento de los sistemas nerviosos biológicos mediante la combinación de muchos elementos simples de computación en sistemas altamente interconectados, con la esperanza de que fenómenos complejos, como la inteligencia, emergerán como resultado de la auto-organización o aprendizaje del sistema. Se trata pues de una aproximación al comportamiento de las redes de neuronas biológicas, inspirado en lo que se conoce de ellas y de su comportamiento.

Según J. Hertz, A. Krogh y R. Palmer (1991) la aproximación al procesamiento de información que se realiza en el cerebro humano y en un ordenador digital convencional puede ser analizada según los siguientes aspectos:

- **Velocidad de procesamiento.**

El tiempo de ciclo, es decir, el tiempo que se tarda en procesar una unidad elemental de información desde la entrada hasta la salida, se mide, para un ordenador de los más potentes, en nanosegundos (10^{-9} segs.), mientras que para una neurona del cerebro, se mide en milisegundos (10^{-3} segs.). Por tanto, un ordenador procesa la información un millón de veces (10^6) más rápido.

- **Orden de procesamiento.**

A pesar del hecho anterior, el cerebro es muy superior en el procesamiento de información humana debido a que, en contraste con los ordenadores seriales, el cerebro se estructura en una arquitectura de procesadores masivamente paralela, cada uno de ellos relativamente pequeño.

- **Abundancia y complejidad.**

Se estima que el número de neuronas presentes en cualquier momento en el cerebro oscila entre 10^{11} y 10^{14} , cada una de ellas con entre 10^4 y 10^5 conexiones. En un ordenador convencional solo existe un nodo de computación, la unidad central de proceso (CPU), y en los ordenadores de arquitectura paralela el número de procesadores se mide a lo sumo en miles.

- **Almacenamiento de la información.**

En cualquier ordenador, una copia estática del conocimiento se almacena en una determinada dirección de memoria, y la llegada de nueva información destruye a la anterior. En cambio, en el cerebro la información no sólo es almacenada en las interconexiones, sino que la información nueva es añadida al cerebro ajustando la intensidad de las conexiones entre neuronas. Esta adaptación proporciona una posible explicación de la habilidad del cerebro para generalizar. Por tanto, mientras que la información en el cerebro es adaptable, en el ordenador es reemplazable.

- **Tolerancia a fallos.**

Los fallos (daños) en neuronas individuales no producen una degradación importante del rendimiento del cerebro, es más, todos los días mueren en el cerebro células nerviosas sin afectar su rendimiento. Esto indica que cada idea o concepto no es almacenada en el cerebro en una única neurona, sino a través de muchas y de sus interconexiones. Por el contrario, la mayoría de los ordenadores convencionales no son tolerantes a fallos, esto es, la eliminación de cualquier componente de procesamiento o cualquier fallo de la memoria del ordenador produce una máquina inefectiva.

- **Control de procesamiento.**

El cerebro no cuenta con un centro de control del procesamiento de información, sino que la actividad de una neurona solo depende de sus conexiones locales. En cambio, la CPU de un ordenador controla todas las actividades y tiene acceso a información global, por lo que ésta constituye un punto crítico en caso de fallo.

La evolución histórica del desarrollo de modelos inspirados en las redes de neuronas biológicas comenzó en los años 40 con W. McCulloch y W. Pitts (1943), quienes propusieron una teoría general del procesado de la información basada en redes de elementos binarios de decisión, que podrían ser consideradas como neuronas, aunque en un estado primitivo.

En los comienzos de la década de los años 50, D. Hebb (1949) introdujo dos ideas fundamentales: la de que una percepción o concepto se representa en el cerebro por un conjunto de neuronas activas simultáneamente, y la de que la memoria se localiza en las conexiones entre las neuronas, llamadas sinapsis. Asimismo, proporcionó una regla de aprendizaje basada en el refuerzo de las conexiones físicas entre neuronas cuando éstas se activan (por ejemplo, al realizar una tarea de forma repetitiva), y que es conocida como la regla de Hebb. Esta regla constituyó la inspiración de muchos algoritmos de aprendizaje, aunque actualmente se sabe que existen conexiones inhibitorias y no sólo excitatorias.

Alrededor de 1960, F. Rosenblatt (1957, 1958) y sus colaboradores presentan el concepto de Perceptrón, enfocado a la Percepción Automática, de ahí el nombre. Éste consiste en un modelo de naturaleza lineal que simplifica los mecanismos biológicos del procesamiento de la información sensorial, y como tal, posee una serie de limitaciones para solucionar problemas no lineales, como por ejemplo la función XOR (OR-Exclusivo). Esta limitación provocó con el tiempo que disminuyera el interés por las redes de neuronas, sobre todo después de la crítica, al final de los años 60, que M. Minsky y S. Papert (1969) realizaron del Perceptrón, donde demostraban que aun conectando un gran número de ellos, no eran capaces de resolver el problema de la linealidad.

Paralelamente al concepto de Perceptrón, a principios de los años 60 se presentaron varios modelos, tales como el Adaline (Adaptative Linear Element) o el Madaline (Multiple Adaline) donde se utilizaba la regla de aprendizaje de mínimos cuadrados (LMS), también denominada regla delta, y que fue introducida en estos modelos por B. Widrow y M. E. Hoff (1960).

Estos modelos de redes permitieron usar por primera vez una red de neuronas artificial para solucionar un problema importante del mundo real: se usaron como filtros adaptativos para eliminar el ruido en las líneas telefónicas.

El resurgir del interés por el Perceptrón se produjo en la década de los 80 con la aparición del algoritmo de aprendizaje de retropropagación (algoritmo BP), el cual es fundamental en las redes de neuronas denominadas Perceptrón Multicapa, caracterizadas por estar formadas por más de dos capas. Presentado por D. E. Rumelhart, J. L. McClelland, G. E. Hinton y R. J. Williams (1986), este algoritmo de aprendizaje permitió entrenar modelos basados en el Perceptrón Multicapa que solventaban las limitaciones del Perceptrón Simple como las expuestas por M. Minsky y S. Papert. A pesar de las propias limitaciones de este algoritmo de aprendizaje, debido como se verá más adelante, a la posibilidad de quedar atrapado en un mínimo local en el proceso de optimización en la selección de los parámetros óptimos, el conjunto de aplicaciones de las redes de neuronas se amplió de forma considerable.

Otras de las razones del resurgir del Perceptrón, y junto con él, de las redes de neuronas artificiales, se debió según D. E. Rumelhart, B. Widrow y M. A. Lehr (1994) a los siguientes aspectos:

1. Los ordenadores disponibles en los años 80 eran mucho más rápidos que los existentes en las décadas de los 50 y los 60, e incluso los ordenadores de tipo serial ya permitían simular y experimentar con modelos intrínsecamente paralelos. De hecho, gracias a los protocolos de comunicaciones locales y remotas, es posible distribuir los cálculos entre varios ordenadores seriales, tipo ordenador personal o estación de trabajo, consiguiendo así un entorno distribuido y aproximadamente paralelo.
2. El modelo de computación paralelo es uno de los más referenciados como mecanismo para lograr un mayor aumento de velocidad en los cálculos científicos, sin embargo, si bien existen muchos ordenadores paralelos, incluso con diversos tipos de arquitectura, no es tan fácil el desarrollo de algoritmos eficientes para tales tipos de equipos. Cabe esperar que el desarrollo de algoritmos que se muestren eficientes en las redes de neuronas puedan constituir un paradigma para la computación paralela.

Simultáneamente, en la década de los años 80, T. Kohonen (1982a, 1982b, 1984) propuso un mecanismo a través del cual una red de neuronas puede organizarse por sí misma de forma que las neuronas se disponen en una región determinada en función de su actividad, dando lugar a lo que se conoce como Mapa de Kohonen o Mapas Autoorganizados, de gran utilidad en el reconocimiento de patrones y en el análisis de conglomerados.

Seguidamente, a finales de la década de los 80 fueron introducidas las redes de neuronas basadas en las Funciones de Base Radial (RBF) por D. S. Broomhead y D. Lowe (1988), que surgieron como alternativa al Perceptrón Multicapa. Y también a finales de los años 80 y principios de los 90, G. A. Carpenter y S. Grossberg (1987a, 1987b, 1990, 1994) establecieron la Teoría de la Resonancia Adaptativa.

3.2 SISTEMAS DE NEURONAS BIOLÓGICAS.

Como ya se ha comentado, los modelos conexionistas están basados en el comportamiento real del cerebro, en cómo está estructurado y en el funcionamiento de los sistemas nerviosos. En éstos, la neurona es el elemento fundamental, ya que el cerebro se compone de billones de estas células vivas, interconectadas entre sí formando redes con unas funcionalidades específicas.

Desde el punto de vista fisiológico, todas las neuronas están constituidas por los mismos elementos, independientemente de su tamaño y su forma. Así, una neurona está compuesta por una parte central denominada cuerpo celular o soma, del que surge una rama principal llamada axón, y otras ramas más cortas denominadas dendritas, que conforman en su conjunto una delicada estructura que rodea el cuerpo celular. El axón, por su parte, se escinde en su extremo en millares de ramificaciones que alcanzan las dendritas de otras neuronas, en una estructura de conexión entre ellas llamada sinapsis. Esta estructura esquemática de una neurona aparece en la Figura 3.2.1.

Esta topología de neuronas manifiesta la capacidad funcional de los sistemas de neuronas de comunicación entre sus unidades. Las dendritas sirven como emisores-receptores de las señales procedentes de las neuronas adyacentes, mientras que a través del axón se transmite la actividad neural generada a otras células nerviosas, a las fibras musculares o a los órganos sensitivos, y a la inversa.

Este complejo mecanismo de comunicación entre neuronas fue estudiado por nuestro premio Nobel Ramón y Cajal, que demostró por primera vez en 1888, que los axones procedentes de las células nerviosas terminan libre-

mente, relacionándose con las dendritas y con los cuerpos de otras neuronas por contacto y no por continuidad. Más aún, Ramón y Cajal dio a conocer en 1891 su teoría de la polarización dinámica: "La transmisión del impulso nervioso se produce siempre desde las dendritas y el cuerpo celular en dirección al axón. Toda neurona posee, pues, un aparato de recepción (el soma y las dendritas), un aparato de emisión (el axón) y un aparato de distribución (la arborización nerviosa terminal)".

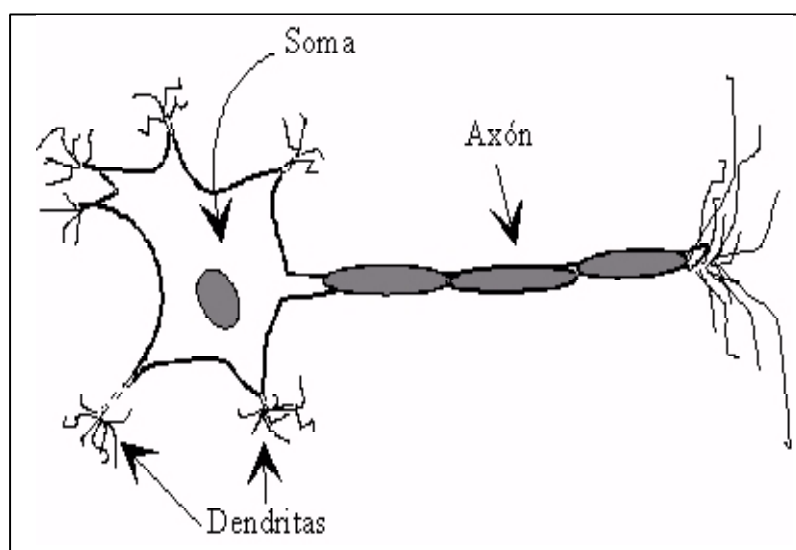


Figura 3.2.1. Estructura de una neurona biológica típica.

Básicamente, el paso de información de una neurona a otra se realiza por medio de compuestos químicos llamados neurotransmisores, que son liberados por los Botones Terminales situados en las ramificaciones del axón, y fluyen en el espacio sináptico hacia las dendritas de la neurona siguiente. Todo este movimiento químico está provocado por, y a su vez genera, un impulso eléctrico que se transmite a lo largo del axón, llamado potencial de acción, que modifica una diferencia de potencial estable que existe entre el exterior de la neurona y su interior a través de su membrana, llamado potencial de reposo. El impulso es provocado por la acumulación de neurotransmisores en las dendritas, que modifica la permeabilidad de la membrana, permitiendo la entrada masiva de iones de signo contrario a los que se encuentran en su interior. Esta entrada masiva de iones, además de cambiar la polaridad de la diferencia de potencial que hay en la membrana, cierra el paso de

iones del exterior, y permite el de iones del interior de la membrana, que llevan a restablecer el potencial de reposo, produciéndose este mecanismo a lo largo de toda la membrana. Al transmitirse el impulso por todo el axón y llegar a sus extremidades, es este mismo quien provoca la liberación de los neurotransmisores que llegan a la siguiente neurona, siguiendo así la transmisión de la información de una neurona a otra a través de toda la red de neuronas.

En otro nivel, los neurotransmisores pueden ser excitatorios o inhibitorios, y además su efecto es variable, por lo que se puede hablar de una fuerza o efectividad de la sinapsis. En definitiva, todas las señales recibidas por una neurona son combinadas, y en función de la estimulación total que esta combinación provoca, la neurona alcanza un cierto nivel de activación, que si llega a estimular a la misma, provocará su activación, provocando con ello un tren de impulsos, y determinando a qué velocidad se transmite éste a lo largo de la neurona.

Se puede afirmar, por tanto, que el aprendizaje en las redes de neuronas biológicas se produce variando la efectividad de la sinapsis, esto es, modificando cómo unas neuronas influyen sobre otras. De ahí la afirmación de que la memoria, o dicho de otro modo, el conocimiento implícito en la red, queda definido en las conexiones, es decir, en las sinapsis.

3.3 ESTRUCTURA DE LAS REDES DE NEURONAS ARTIFICIALES.

Las RNAS modelizan el comportamiento de las redes de neuronas biológicas, aunque evidentemente, esta modelización implica cierta simplificación del sistema.

Al igual que en el cerebro, la neurona es el elemento fundamental de las RNAS, que queda definida como la unidad de proceso. Más adelante se verá cómo se organizan estos elementos y se introducirán otros conceptos, como la Regla de Aprendizaje, que caracterizan la red y definen su comportamiento.

3.3.1 DEFINICIÓN DE RED NEURONAL ARTIFICIAL.

Entre las diversas definiciones de RNAS existentes, tal vez la más completa sea la que formalizan en su libro J. Hertz, A. Krogh y R. Palmer (1991), capaz de acomodar modelos tanto físicos como lógicos.

Definición 3.3.1.

Una Red Neuronal Artificial es un modelo de procesamiento distribuido y en paralelo de la información, en forma de grafo dirigido, con la siguiente estructura interna:

1. Los nodos del grafo se llaman también elementos de procesamiento, neuronas artificiales ó neuronas.
2. Los arcos del grafo se llaman conexiones sinápticas. Cada conexión sináptica funciona como un canal unidireccional instantáneo de transmisión de información.
3. Cada elemento de procesamiento puede recibir cualquier número de conexiones entrantes.
4. Cada elemento de procesamiento puede tener cualquier número de conexiones salientes, portando todas ellas la misma información.
5. Los elementos de procesamiento pueden tener memoria local.
6. Cada elemento de procesamiento posee una función, llamada función de activación o función de transferencia, que se aplica a las informaciones proporcionadas por las conexiones de entrada, lo que determina la información a transmitir por las conexiones de salida. La función de transferencia puede depender de los valores almacenados en la memoria local.
7. Las funciones de activación pueden operar de forma continuada o bien a intervalos de tiempo de longitudes no necesariamente iguales. En este último caso, debe haber una entrada que active la función de activación. Esta entrada le llega al nodo desde un elemento de secuenciación, que es parte también de la red.

3.3.2 ELEMENTOS DE PROCESAMIENTO.

El elemento fundamental de una RNAS lo constituyen las neuronas artificiales. De acuerdo con la *definición 3.3.1*, una neurona artificial es una unidad de procesamiento de información que, a partir de m valores de entrada x_1, \dots, x_m , produce un valor de salida, y , que a su vez puede ser parte de la entrada a otras neuronas.

Se pueden considerar así tres tipos de neuronas artificiales: aquellas que reciben señales del exterior, llamadas neuronas de entrada; las denominadas neuronas ocultas, que recogen la información de las anteriores y la procesan, componiendo así capas internas, puesto que no tienen relación directa ni con la información de entrada ni con la de salida; y las que proporcionan la respuesta del sistema, recibiendo la información de las capas ocultas y transmitiéndola al exterior, que se denominan neuronas de salida.

La red quedaría así conformada por tres tipos de capas: una de entrada, una o varias ocultas y una última de salida.

Se verán ahora detalladamente cada uno de los elementos de la neurona artificial, con los que ésta quedará definida.

- **Conexiones sinápticas.**

La comunicación con el exterior de cada neurona se establece por medio de una serie de conexiones denominadas de entrada y por una conexión de salida. Las entradas, según el tipo de neurona, vendrán del exterior o bien de una capa anterior. La salida se podrá conectar a cuantas neuronas de la capa siguiente se desee, o bien formará parte de la salida global de la red.

Como se observa, la unión de la salida de la capa i con la entrada de la capa $i + 1$ es lo equivalente a la sinapsis biológica, si bien, como se ha mencionado, éstas no son meras transmisoras de información, sino que la refuerzan o la inhiben. Matemáticamente, esto se consigue ponderando las entradas por medio de una serie de pesos que se denominarán pesos sinápticos.

Así, sea un conjunto de m conexiones que proporcionan los valores de entrada x_1, \dots, x_m a la neurona y sean w_1, \dots, w_m los pesos sinápticos asociados a dichas conexiones.

Habitualmente se añade al conjunto de pesos sinápticos de la neurona un parámetro adicional u , que será denominado umbral, y que se introduce a través de una conexión de entrada ficticia x_0 de valor constante

1 y con un peso asociado w_0 cuyo valor será el umbral pero con el signo opuesto.

- **Potencial sináptico.**

En las neuronas biológicas las señales de las conexiones de entrada son combinadas de manera que provocan una estimulación global sobre la neurona, lo que se trasladará a la neurona artificial a través del concepto denominado potencial sináptico de la neurona.

Definición 3.3.2.

Se define el potencial sináptico de la neurona s , como la combinación lineal de los valores de entrada x_1, \dots, x_m , cada uno ponderado por sus pesos sinápticos asociados w_1, \dots, w_m , esto es,

$$s = \sum_{i=1}^m w_i x_i - u.$$

Introduciendo la entrada ficticia x_0 idénticamente igual a 1 y con un peso asociado $w_0 = -u$, esta expresión es equivalente a

$$s = \sum_{i=0}^m w_i x_i.$$

- **Estado de activación.**

El potencial sináptico de la neurona combina las entradas, de manera que éstas puedan condicionar el siguiente estado de la neurona, de una manera análoga a como una neurona es "excitada" por las informaciones que recibe por medio de las sinapsis.

Definición 3.3.3.

Se define como estado de activación de la neurona al valor numérico que tiene asociado la neurona, modificada por las entradas y transmitido hacia la salida.

Este estado de activación puede ser un valor continuo o discreto, así como limitado o bien ilimitado. Un ejemplo de valores discretos sería el conjunto de valores $\{0,1\}$ o el conjunto $\{-1,1\}$, y ejemplos de valores continuos y limitados serían los intervalos $[0,1]$ y $[-1,1]$, todos ellos muy utilizados.

- **Función de activación.**

Definición 3.3.4.

Una función de activación o de transferencia f asociada a una neurona, es aquella función que proporciona su estado de activación, y dependerá del potencial sináptico de la neurona y del valor del umbral.

Las funciones de transferencia de las neuronas se suelen considerar deterministas, y en la mayor parte de los modelos es monótona creciente y continua.

Definición 3.3.5.

Se dice que una función $f : \mathbb{R} \rightarrow \mathbb{R}$ es sigmoïdal si verifica

$$-\infty < \lim_{x \rightarrow -\infty} f(x) < \lim_{x \rightarrow +\infty} f(x) < +\infty.$$

En la Figura 3.3.1 se muestran algunas de las funciones de transferencia más habituales.

Como se verá en el capítulo siguiente, diversos resultados teóricos sugieren el empleo de funciones de activación de tipo sigmoïdal. Como ejemplos de funciones sigmoïdales se encuentran la función paso, la función signo y la función logística.

- **Función de salida.**

Definición 3.3.6.

Una función se dirá de salida de una neurona cuando proporciona el valor de la salida global y de ésta.

Normalmente se utiliza la función identidad, por lo que la salida quedaría definida por la función de transferencia en sí,

$$y = f(s).$$

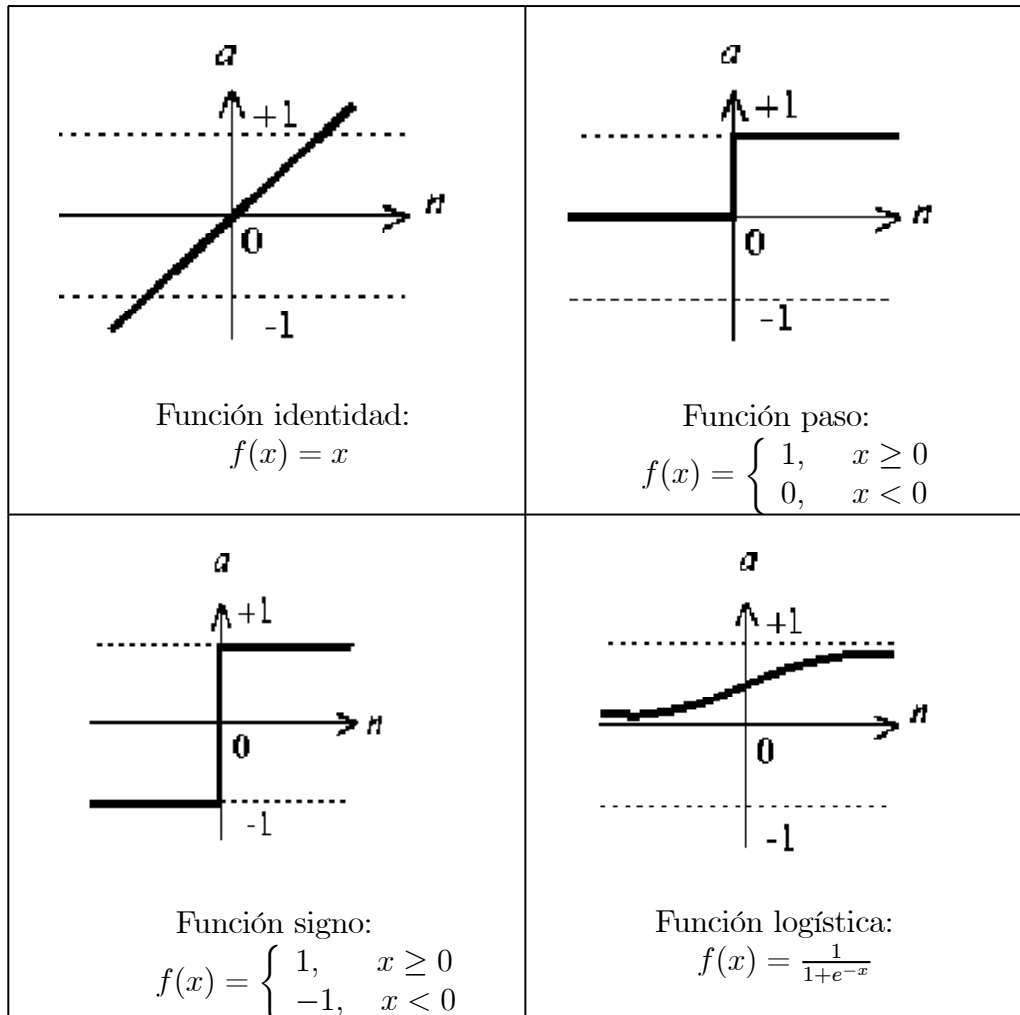


Figura 3.3.1. Algunas funciones de activación o transferencia.

A partir de este modelo general de neurona expuesto, se puede definir el concepto de neurona artificial estándar, cuyo esquema gráfico aparece en la Figura 3.3.2, como aquella neurona cuya regla de propagación es la suma ponderada de las entradas con sus pesos sinápticos asociados y cuya función de salida es la identidad.

Definición 3.3.7.

Se define la neurona artificial estándar como aquella que consta de los siguientes elementos:

1. Un conjunto de entradas x_1, \dots, x_m con sus pesos sinápticos asociados w_1, \dots, w_m .
2. El potencial sináptico s , obtenido a través de la función de propagación suma ponderada y al que se le sustrae el valor del umbral u ,

$$s = \sum_{i=1}^m w_i x_i - u,$$

lo que introduciendo la entrada ficticia x_0 con valor constante 1 y peso asociado $w_0 = -u$, es equivalente a

$$s = \sum_{i=0}^m w_i x_i.$$

3. Una función de activación f .
4. La salida y de la neurona viene dada por

$$y = f(s).$$

Y con todo esto, quedaría definido lo que es la microestructura de la red, es decir, la caracterización de cada nodo de la red, de cada neurona artificial que la compone. Con ello, sin embargo, no se define el comportamiento de la red como tal, para ello se debe especificar cómo está organizada dicha red.

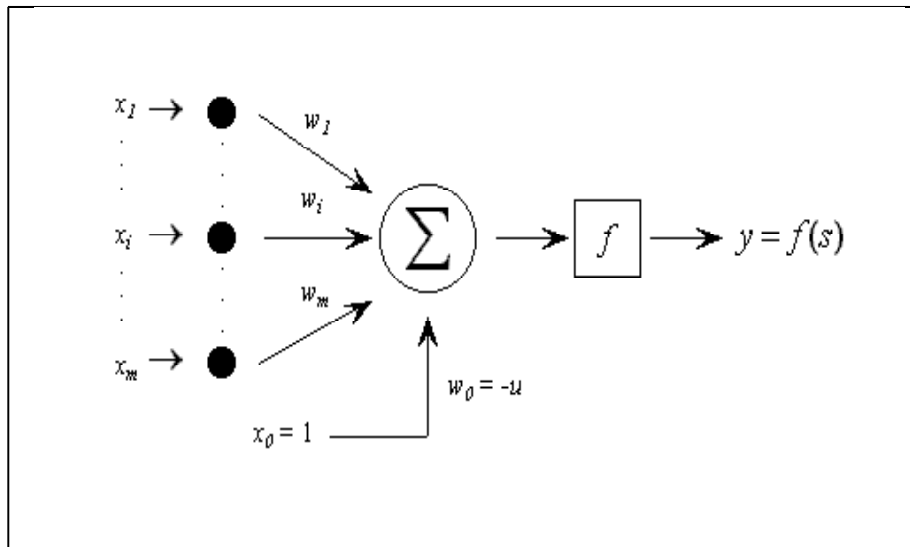


Figura 3.3.2. Modelo de neurona artificial estándar.

3.3.3 ARQUITECTURA DE LAS REDES DE NEURONAS ARTIFICIALES.

El concepto de arquitectura se refiere a la topología de las RNAS, es decir, a la organización o estructura de las conexiones de las neuronas entre sí. Al igual que las neuronas biológicas, las RNAS se conectan por medio de las sinapsis, cuya característica principal es que son direccionales, esto es, que la información únicamente se propaga en un solo sentido, lo que determina el comportamiento de la red.

Como ya se ha mencionado, según de dónde provenga la información que reciben las RNAS y hacia dónde éstas la dirigen, existen tres tipos de neuronas, que a su vez se agrupan en unidades estructurales llamadas capas. Así, en función del tipo de neuronas que conforman un capa, aparecen tres tipos de capas en una red de neuronas artificial: de entrada, oculta y de salida.

En general, atendiendo a la estructura en capas de las RNAS, existen redes monocapa, que están compuestas por una única capa de neuronas, y redes multicapa, que son aquellas cuyas neuronas se organizan en varias capas.

Se tiene que considerar además el tipo de conexiones que hay dentro de una red, en relación a si se producen dentro de una misma capa, en cuyo caso

se denominan conexiones intracapa o laterales, si se producen entre neuronas de capas distintas, denominadas conexiones intercapa, o bien las conexiones llamadas realimentadas, en el que el flujo de información discurre en sentido contrario al de entrada-salida, e incluso en algunas de éstas, la salida de una neurona se conecta a su propia entrada.

Por otro lado, el signo del peso sináptico asociado a una conexión entre neuronas determina si esa conexión es excitatoria o inhibitoria, de forma que un peso sináptico negativo define una conexión inhibitoria, mientras que un peso sináptico positivo indica una conexión excitatoria.

De esta forma, atendiendo al flujo de datos, que vendrá determinado por el sentido de las conexiones entre neuronas, existen redes unidireccionales o alimentadas hacia adelante y redes recurrentes.

Definición 3.3.8.

Una red de neuronas artificiales unidireccional o alimentada hacia adelante (feedforward) es una red de neuronas artificiales cuyos elementos de procesamiento están organizados en capas o niveles sucesivos, y de forma que, una vez ordenadas las capas de izquierda a derecha, solo existen conexiones entre nodos de niveles sucesivos, en el sentido izquierda-derecha.

Definición 3.3.9.

Una red de neuronas artificiales recurrente (feedback) es una red de neuronas artificiales cuyos elementos de procesamiento están organizados en capas o niveles sucesivos, y donde es posible establecer conexiones entre niveles no adyacentes, o bien unir nodos en sentido contrario al determinado por el orden de izquierda a derecha, o bien conectar nodos consigo mismo.

De todo esto se puede deducir que, según el número de capas en una red, éstas pueden ser monocapa o multicapa, mientras que según el flujo de la información por la red (que queda definido en cómo se conectan las neuronas entre sí), las hay unidireccionales o alimentadas hacia adelante y recurrentes.

3.3.4 ALGORITMOS DE APRENDIZAJE.

Como ya se ha comentado, el conocimiento en las redes biológicas se encuentra en las conexiones entre neuronas, es decir, en las sinapsis. Del mismo modo, se puede decir que el conocimiento en las RNAS está presente en el conjunto de pesos sinápticos de toda la red, por lo que habrá que elegir el

valor de cada uno de los pesos sinápticos de acuerdo a determinado criterio. Por ello, puede afirmarse que el mecanismo de aprendizaje es aquel proceso por el que una red de neuronas artificiales modifica sus pesos sinápticos en respuesta a unos datos de entrada durante una fase que puede llamarse de entrenamiento.

Definición 3.3.10.

Dada una red de neuronas artificiales, se llama algoritmo, método o regla de aprendizaje a cualquier algoritmo que permite obtener una asignación de valores para cada uno de los coeficientes sinápticos.

En definitiva, cada tipo de aprendizaje queda definido por el criterio que se seguirá para modificar el valor de los pesos sinápticos, y básicamente pueden distinguirse dos grandes grupos, que se conocen como aprendizaje supervisado y aprendizaje no supervisado. La diferencia fundamental entre ambos tipos es que en el supervisado existe un agente externo, también llamado supervisor, que controla todo el proceso de aprendizaje, mientras que en el no supervisado dicho agente no está presente.

Aprendizaje supervisado.

Mediante este tipo de regla de aprendizaje, lo que hace el supervisor, básicamente, es observar la salida de la red para un conjunto de entrada dado, y compararla con la que debería haberse obtenido de la red si ya estuviera entrenada. En base a una serie de criterios, el supervisor procederá a modificar los pesos sinápticos de manera que la salida obtenida se aproxime a la deseada.

De esta manera, según la forma en que actúe el supervisor, se tendrá:

- Aprendizaje por corrección de error, donde la modificación de los pesos sinápticos será proporcional al error encontrado en la salida de cada neurona. Como ejemplos de este tipo de regla de aprendizaje, está el algoritmo de error perceptrón, y el denominado regla delta o regla de mínimos cuadrados.
- Aprendizaje por refuerzo, donde la función del supervisor se asemeja más a la de un crítico. Se trata de un proceso de aprendizaje más lento que el anterior, ya que no se dispone de un ejemplo completo del comportamiento deseado de la red, es decir, no se indica cual debe ser

la salida exacta, sino solo el comportamiento de la red de una manera general. El supervisor va indicando si la salida para el conjunto de entrenamiento es correcta o no, es decir, si se ajusta a la deseada, dando su respuesta mediante una serie de valores indicativos de éxito o fracaso (que por ejemplo pueden estar representados con 1 y -1 respectivamente). Estas respuestas las va generando el supervisor en modo on-line al ejecutarse el aprendizaje, y se denominan señales de refuerzo. Si hay éxito, se dejan los pesos sinápticos como están, y si hay fracaso (la salida obtenida no se ajusta a la deseada) se modificarán los pesos sinápticos siguiendo una distribución de probabilidades ya establecida.

- Aprendizaje estocástico, donde el supervisor va realizando modificaciones a los pesos sinápticos de manera aleatoria, sin seguir en principio ningún patrón establecido, ni observar las salidas obtenidas con las entradas del conjunto de entrenamiento y comparándolas con las deseadas. A continuación observa las salidas obtenidas y las evalúa en función de su ajuste con las deseadas. Si las salidas obtenidas se encuentran muy próximas a las deseadas, se aceptarán los cambios aleatorios realizados a los pesos sinápticos. Y en caso contrario, el supervisor optará por usar una determinada y preestablecida distribución de probabilidades para aceptar o no el cambio aleatorio realizado a los pesos sinápticos.

Aprendizaje no supervisado.

Las redes con aprendizaje no supervisado no necesitan influencia externa para ajustar los pesos sinápticos de las conexiones entre neuronas. Esto significa que el ajuste de los pesos sinápticos, que se realiza durante el algoritmo de aprendizaje y por tanto depende de él, no acudirá a comparaciones de la salida obtenida durante el entrenamiento con un conjunto de salidas deseadas, ni con señales de refuerzo que indique la calidad de estas salidas obtenidas, ni se evalúen de ninguna forma. Estas técnicas se utilizan sobre todo para obtener estructuras, relaciones o clasificaciones del conjunto de datos de entrada.

Un ejemplo de algoritmo de aprendizaje no supervisado, es el aprendizaje Hebbiano, que refuerza las conexiones entre neuronas con la misma salida, e inhibe aquellas entre neuronas con salidas distintas. El resultado es que la

red lleva a cabo una separación de los valores de las conexiones de entrada en una serie de patrones, según la semejanza entre ellos.

3.3.5 TRATAMIENTO DE LA INFORMACIÓN.

Ya se ha dicho que en las redes de neuronas biológicas el conocimiento se almacena en las sinapsis. En concreto, y de forma análoga, en las redes de neuronas artificiales, ese conocimiento está almacenado en el valor de todos los pesos sinápticos de la red, que se van modificando mediante el proceso de aprendizaje, con el objetivo de que la salida de la red ante un conjunto de entrada sea de cierta manera, como puede ser la solución de una determinada función o bien la clasificación de las entradas en una serie de patrones de comportamiento.

Esta idea de almacenamiento de información hace que una red de neuronas pueda ser considerada de otra manera muy distinta a través del concepto de memoria.

El acceso a los datos se realiza mediante la introducción de cierta entrada en la red, que tras el proceso de aprendizaje dará como consecuencia una respuesta, esto es, una salida de la red, que constituye el dato que se está buscando. De esta forma, los datos se graban en la red por medio del aprendizaje. Funciona de manera análoga a una memoria de computador, la cual necesita que se le dé una dirección (el dato de entrada) y ésta responde con la información almacenada en la misma (el dato de salida).

Este comportamiento de la red de asociar dato buscado o de salida, con dato origen de la búsqueda o de entrada, se conoce habitualmente como perteneciente a una memoria de tipo asociativo, y existen dos formas de realizar esta asociación entre datos de entrada y de salida, según la naturaleza de la información almacenada.

Definición 3.3.11.

Se dirá que una red de neuronas artificiales es heteroasociativa cuando asocia parejas de datos cuyos elementos son de distinto tipo.

Así, sea A un patrón de entrada de la red y B un patrón de salida asociado al de entrada A y diferente de él. Estas redes son entrenadas para que ante la presentación del patrón de entrada A , devuelva como salida un patrón de tipo B , es decir, la red asocia datos de distinta naturaleza.

Definición 3.3.12.

Una red de neuronas artificiales se dirá que es autoasociativa cuando asocia parejas de datos en las que el tipo de información de salida es lo más parecido posible al de la información de entrada, dentro de los almacenados en la red, es decir, de los conocidos por ella.

Así, si A es un patrón de entrada de la red, estas redes son entrenadas para que asocie el patrón de entrada A consigo mismo. En definitiva, la red asocia datos del mismo tipo.

Una última forma de clasificar las redes de neuronas es mediante la forma en que la información se representa en ellas, entendiendo por información al conjunto de valores de las entradas posibles y al conjunto de valores de las respuestas de la red o datos de salida.

Desde un punto de vista sencillo, los datos pueden ser de naturaleza analógica o discreta.

- Analógica: se trata de valores reales continuos, que por regla general están normalizados, presentando un valor absoluto menor que 1.
- Discreta: se trata de valores definidos dentro de un conjunto finito, y generalmente aparecen en forma binaria, como por ejemplo los conjuntos $\{0,1\}$ y $\{-1,1\}$.

En función del tipo de datos de entrada y de salida de la red, éstas pueden clasificarse en redes continuas, cuando ambos conjuntos de datos son analógicos, redes híbridas, cuando el conjunto de datos de entrada es analógico y el de salida binario, y discretas, cuando ambos conjuntos de datos son binarios.

En la Tabla 3.3.1 se presentan unos ejemplos de cómo se representa la información en algunos de los modelos más utilizados de redes de neuronas.

Redes continuas	Redes híbridas	Redes discretas
Grossberg	Perceptrón	Máquina de Cauchy
Brain-State-in-Box	Adaline / Madaline	Máquina de Boltzmann
Memoria Asociativa Lineal		Teoría de la Resonancia Adaptativa
Hopfield Continuo		Hopfield Discreto

Tabla 3.3.1. Clasificación de algunas redes de neuronas según la representación de la información.

3.4 VENTAJAS DE LAS REDES DE NEURONAS ARTIFICIALES.

Antes de continuar viendo ejemplos específicos de redes de neuronas, así como el empleado en este trabajo, es conveniente recordar los motivos que hacen interesante el uso de las RNAS para la resolución de problemas complejos.

Al tratarse de un modelo del comportamiento del cerebro humano, parece lógico pensar que todas estas ventajas están relacionadas con la habilidad propia de estos sistemas biológicos.

En general, son sistemas capaces de aprender de la experiencia, como lo demuestra el hecho de necesitar un periodo de aprendizaje, generalizan nuevos casos que no han visto con anterioridad a partir de casos anteriores, son capaces de extraer información relevante de entre una maraña de información redundante, etc.

En particular, las ventajas incluyen:

1. Aprendizaje Adaptativo. Como característica más interesante aparece su capacidad para aprender en base a un entrenamiento. No necesita ningún algoritmo, y el diseñador solo ha de concentrarse en elegir la arquitectura adecuada, y los elementos idóneos. Incluso existen redes que siguen aprendiendo durante toda su vida útil, más allá de la fase de entrenamiento.

2. Autoorganización. Con este concepto se quiere expresar la capacidad de la red de organizar el conocimiento de manera propia y apropiada, durante el proceso de aprendizaje. Gracias a esta facultad, las redes de neuronas pueden responder satisfactoriamente en situaciones a las que antes no han sido expuestas, como por ejemplo al introducir conjuntos de entradas nuevos.
3. Tolerancia a Fallos. Se refiere a dos aspectos de tolerancia: en primer lugar a la posibilidad de dar respuesta satisfactoria al sistema, aun cuando los datos de entrada se presenten distorsionados o con errores, y en segundo lugar, a que pueden continuar funcionando con cierto nivel de fiabilidad aun cuando se degrade el sistema por la caída de algunas neuronas. Ésto último sucede a diferencia de los sistemas computacionales convencionales, los cuales se veían totalmente interrumpidos en su funcionamiento cuando sufrían algún pequeño error de memoria.
4. Operación en Paralelo. Aunque esta ventaja solo aparece en implementaciones de redes de neuronas que permitan procesos en paralelo y control eficaz de memoria del sistema, existen y supone un salto de calidad respecto a los sistemas tradicionales.
5. Fácil inserción dentro de la tecnología existente. No es necesaria la aparición de nueva tecnología, sino que con la actual todas las fases correspondientes a la implementación, al aprendizaje, a la verificación y al traslado a un hardware de bajo coste, se realizan de forma sencilla, lo que facilita su inserción dentro de sistemas ya existentes, dotándolos de una mejora incremental.

3.5 MODELOS DE REDES DE NEURONAS ARTIFICIALES Y APLICACIONES.

Existen numerosos modelos de redes de neuronas artificiales que han ido apareciendo a lo largo de la historia. En la Tabla 3.3.2 aparece la descripción de algunos de los más conocidos y utilizados modelos de RNAS. Los modelos que aparecen en esta tabla tienen como característica común que presentan una arquitectura en niveles y unas conexiones exclusivamente hacia adelante entre las neuronas.

MODELO DE RED	TOPOLOGÍA	APRENDIZAJE			TIPO DE ASOCIACIÓN
		ON/OFF LINE	TIPO	REGLA	
ADALINE/ MADALINE (WIDROW/HOFF 1960)	2 CAPAS FEEDFORWARD	OFF	SUPERVISADO	CORRECCIÓN DEL ERROR	HETEROASOCIATIVO
BRAIN -STATE -IN -A -BOX (ANDERSON 1977)	1 CAPA CONEXIONES LATERALES	OFF	SUPERVISADO	CORRECCIÓN DEL ERROR	AUTOASOCIATIVO
GROSSBERG (1968)	1 CAPA CONEXIONES LATERALES	ON	NO SUPERVISADO	HEBBIANO O COMPETITIVO	AUTOASOCIATIVO
HOPFIELD CONTINUO (1984)	1 CAPA CONEXIONES LATERALES	OFF	NO SUPERVISADO	HEBBIANO	AUTOASOCIATIVO
HOPFIELD DISCRETO (1982)	1 CAPA CONEXIONES LATERALES	OFF	NO SUPERVISADO	HEBBIANO	AUTOASOCIATIVO

Tabla 3.3.2. Características de diversos modelos de redes de neuronas.

MODELO DE RED	TOPOLOGÍA	APRENDIZAJE			TIPO DE ASOCIACIÓN
		ON/OFF LINE	TIPO	REGLA	
MÁQUINA DE CAUCHY (SZU 1986)	1 CAPA CONEXIONES LATERALES 3 CAPAS FEEDFORWARD	OFF	NO SUPERVISADO	ESTOCÁSTICO	HETEROASOCIATIVO
MÁQUINA DE BOLTZMANN (HILTON Y OTROS 1984)	1 CAPA CONEXIONES LATERALES 3 CAPAS FEEDFORWARD	OFF	SUPERVISADO	ESTOCÁSTICO + HEBBIANO	HETEROASOCIATIVO
MEMORIA ASOCIATIVA LINEAL (ANDERSON 1968)	2 CAPAS FEEDFORWARD	OFF	NO SUPERVISADO	HEBBIANO	HETEROASOCIATIVO
PERCEPTRÓN (ROSENBLAT 1958)	2 CAPAS FEEDFORWARD	OFF	SUPERVISADO	CORRECCIÓN DEL ERROR	HETEROASOCIATIVO
TEORÍA DE LA RESONANCIA ADAPTATIVA (CARPENTER 1986)	2 CAPAS FEEDFORWARD FEEDBACK CONEXIONES LATERALES	ON	NO SUPERVISADO	COMPETITIVO	HETEROASOCIATIVO

Tabla 3.3.2-Cont. Características de diversos modelos de redes de neuronas.

Las redes de neuronas artificiales pueden utilizarse en un gran número y variedad de aplicaciones, tanto comerciales como militares.

Se pueden desarrollar redes de neuronas artificiales en un periodo de tiempo razonable, con la capacidad de realizar tareas concretas mejor que otras tecnologías. Cuando se implementan mediante hardware (redes de neuronas artificiales en chips VLSI), presentan una alta tolerancia a fallos del sistema y proporcionan un alto grado de paralelismo en el procesamiento de datos. Esto posibilita la inserción de redes de neuronas artificiales de bajo coste en sistemas existentes y recientemente desarrollados.

Tras el amplio desarrollo que en estos últimos años están teniendo las RNAS, éstas se utilizan básicamente en tres situaciones genéricas:

1. Como modelos de sistemas nerviosos biológicos, intentando representar y comprender mejor el mecanismo de funcionamiento de la inteligencia.
2. Como procesadores o controladores adaptativos de señales en tiempo real, implementados en hardware para aplicaciones tales como la robótica ó el diseño de circuitos integrados.
3. Como métodos de análisis estadístico de datos.

A continuación se citan algunas de las muchas aplicaciones particulares de las RNAS, ya existentes hoy en día:

- Diseño de chips de alta integración VLSI.
- Robots automatizados y sistemas de control (visión artificial y sensores de presión, temperatura, gas, ...).
- Analizadores del habla para ayudar en la audición de sordos profundos.
- Diagnóstico y tratamiento a partir de síntomas y/o de datos analíticos (electrocardiograma, encefalogramas, análisis sanguíneo, ...).
- Reconocimiento de aberraciones estructurales en cromosomas.
- Monitorización en cirugías.
- Predicción de reacciones adversas en los medicamentos.
- Obtención de modelos de la retina.

- Clasificación de las señales de radar.
- Detección de minas.
- Creación de armas inteligentes.
- Optimización del uso de recursos escasos.
- Los módems para transmisión en alta velocidad por líneas telefónicas utilizan dispositivos hardware que integran una red de neuronas artificial dentro de los dispositivos ecualizadores y canceladores de eco.
- Detección de transacciones fraudulentas con tarjetas de crédito.
- Previsión de la evolución de los precios.
- Identificación de falsificaciones.
- Interpretación de firmas.
- Concesión de créditos.
- Control de calidad.
- Control de procesos.
- Procesamiento del lenguaje natural.
- Reconocimiento de escritura, tanto a máquina como a mano.
- Reconocimiento del sexo de rostros humanos.
- Problemas de optimización, como el del viajante, plazas y horarios en líneas de vuelo, el flujo del tránsito controlando convenientemente la temporización de los semáforos, ...
- Localización de servicios.
- Detección de sucesos interesantes en aceleradores de partículas.
- Reconocimiento de la partida de droga de la que proviene una muestra incautada.
- Compresión/Descompresión de datos, etc.

- Extracción del conocimiento en grandes bases de datos (Knowledge Data Discovery).
- Soporte de la base de datos de reglas de un sistema basado en el conocimiento, G. Towell y J. Shavlik (1994).
- Evaluación de probabilidad de formaciones geológicas y petrolíferas.
- Identificación de candidatos para posiciones específicas.
- Previsión del tiempo.
- Obtención de parámetros atmosféricos y oceánicos a partir de datos procedentes de satélites.
- Determinación de la ecuación de estado del agua del mar para obtener un modelo de circulación oceánica.

3.6 ASPECTOS ESTADÍSTICOS DE LAS REDES DE NEURONAS ARTIFICIALES.

Las redes de neuronas artificiales pueden expresarse como modelos matemáticos que proporcionan un mecanismo bastante flexible para afrontar diversos problemas, en especial los de predicción y clasificación, los cuales pueden ser abordados también mediante técnicas estadísticas. Ambas metodologías, tal y como expresan B. Cheng y D. M. Titterington (1994), G. Hatabian (1994), B. D. Ripley (1993, 1994, 1996) o C. M. Bishop (1995), no son competidoras para el análisis de datos, sino que se trata de dos campos con muchos aspectos en común, y que pueden beneficiarse mutuamente.

La estadística puede aportar bastante al desarrollo de las RNAS, en aspectos como los siguientes:

- Diseño estadístico de experimentos a fin de asegurar la obtención de conjuntos de datos apropiados para ajustar los modelos estadísticos basados en RNAS.
- Aplicación de técnicas de muestreo para seleccionar una parte del conjunto de datos, algo especialmente útil también cuando el investigador se encuentra con información ya recogida en forma de un almacén

masivo de datos (Data Warehouse). Así los módulos implementados por productos comerciales de software dedicados a la extracción del conocimiento en grandes bases de datos, que emplean las RNAS entre otras técnicas, cuentan con módulos de muestreo. Es el caso del paquete estadístico SAS con su módulo SAS/Enterprise Miner.

- Recolección y análisis exploratorio de los datos, incluyendo técnicas de identificación y tratamiento de outliers.
- Análisis de datos con valores perdidos. Algunos autores incluso incorporan el algoritmo EM de estimación de máxima verosimilitud bajo datos incompletos dentro de algoritmos de aprendizaje de redes de neuronas artificiales, como S. Amari (1995).
- Aplicación de técnicas de optimización empleadas habitualmente en diversas áreas de la estadística computacional como algoritmos alternativos de aprendizaje. Así el módulo Neural Connection de SPSS dispone del método de gradientes conjugados, además del algoritmo de retropropagación.
- Técnicas de reducción de la dimensionalidad, necesarias en el tratamiento de imágenes y en general en el contexto de las grandes bases de datos. Encaja aquí la realización de un análisis de componentes principales para reducir el número de variables o bien aplicar el algoritmo de los centros móviles para obtener el conjunto de centros de una red de base radial.
- Aplicación de técnicas de selección de modelos, como ocurre con la construcción de criterios de información como el de Akaike o el Criterio de Información Bayesiano.
- Cálculo de intervalos de confianza para las predicciones proporcionadas por una red de neuronas artificiales. Dado el carácter no paramétrico de los modelos de predicción basados en RNAS, resulta especialmente útil emplear métodos de remuestreo con este fin. R. Tibshirani (1995) y A. Weigend y A. Le-Baron (1994) ilustran la aplicación de las técnicas bootstrap en este contexto, el primero en problemas generales de predicción y el segundo en el análisis de series temporales.

- Mecanismos para contemplar el llamado dilema sesgo-varianza, relacionado con el problema de la generalización, es decir, la capacidad de que la red proporcione respuestas adecuadas ante patrones no utilizados en el algoritmo de aprendizaje. S. Geman, E. Bienenstock y R. Doursat (1992) formalizan este problema.

En el sentido contrario, las RNAS también aportan ventajas y valores añadidos a la estadística, como son:

- Nuevas técnicas apropiadas para resolver el siguiente problema general: Dada una realización x de un vector aleatorio p -dimensional X , se pretende predecir el valor que tomará un vector aleatorio q -dimensional Y . Así, dado el carácter no lineal de la función de salida proporcionada por la red, el perceptrón multinivel se puede incluir dentro de los modelos de regresión no lineal, con la particularidad de incluir en general un número elevado de parámetros.
- Técnicas de análisis discriminante, utilizando para ello redes con una capa de salida con tantas neuronas como clases consideradas, y de forma que ante un vector de entrada una sola de las neuronas de salida vale 1, y las restantes 0, proporcionando así la clase donde se clasifica el patrón de entrada. Esta práctica es muy habitual en problemas de reconocimiento de patrones, como ilustra C. G. Looney (1997). Otras redes utilizan en la capa de salida funciones de activación logística, permitiendo así calcular estimaciones de la probabilidad de pertenencia a cada clase.
- En relación con el punto anterior se puede incluso implementar un perceptrón logístico, que difiere del perceptrón simple en la sustitución de la función paso o función signo por la función logística, modelo que puede ser ampliado incluyendo capas ocultas, lo que conduce al perceptrón multinivel con una sola neurona de salida con función de activación logística. M. Schumacher, R. Roßner y W. Vach (1996) comparan el modelo estadístico de regresión logística con este tipo de redes de neuronas artificiales.
- Edición de registros en bases de datos mediante redes de neuronas artificiales, incluyendo la imputación de valores perdidos o erróneos. S. Nordbotten (1995, 1996) lo lleva a cabo sobre el Censo de Noruega.

4 ALGUNOS MODELOS DE REDES DE NEURONAS ARTIFICIALES.

4.1 EL PERCEPTRÓN SIMPLE.

Este modelo fue desarrollado por F. Rosenblatt (1957, 1958), y era capaz de reconocer patrones sencillos, por lo que despertó gran interés con su aparición.

El perceptrón simple consiste en un modelo unidireccional, formado por dos capas, una de entrada y otra de salida. En este modelo, las neuronas de la capa de entrada no realizan ningún tratamiento de la información que reciben, sino que tan solo la traspasan a la capa de salida, que originariamente, en la definición proporcionada por F. Rosenblatt (1962) está formada por un único nodo.

4.1.1 DEFINICIÓN DEL PERCEPTRÓN SIMPLE.

Definición 4.1.1.

Sean x_1, \dots, x_m , los valores de entrada con los pesos sinápticos w_1, \dots, w_m asociados a sus conexiones con la neurona de salida, y sea y el valor de salida que proporciona la red.

Entonces, el modelo de perceptrón simple viene determinado por los siguientes elementos:

1. El potencial sináptico s es la ya conocida suma ponderada de los valores de entrada, que introduciendo el valor umbral u , puede ser expresada como

$$s = \sum_{i=0}^m w_i x_i,$$

siendo $x_0 = 1$ y su peso asociado $w_0 = -u$.

2. La función de transferencia de la neurona de la capa de salida en este modelo es una función paso del tipo

$$f(a) = \begin{cases} 0, & a < 0 \\ 1, & a \geq 0 \end{cases},$$

o bien, la función signo,

$$f(a) = \begin{cases} -1, & a < 0 \\ 1, & a \geq 0 \end{cases}.$$

3. El resultado de la neurona de la capa de salida de un perceptrón simple viene dada por la expresión

$$y = f \left(\sum_{i=0}^m w_i x_i \right).$$

Esta es la definición originaria de perceptrón simple que proporcionó F. Rosenblatt en 1962, pero ésta puede hacerse extensiva al caso en que la capa de salida esté formada por más de un único nodo. En la Figura 4.1.1 puede verse un esquema de un perceptrón simple formado por m neuronas en la capa de entrada y n neuronas en la capa de salida.

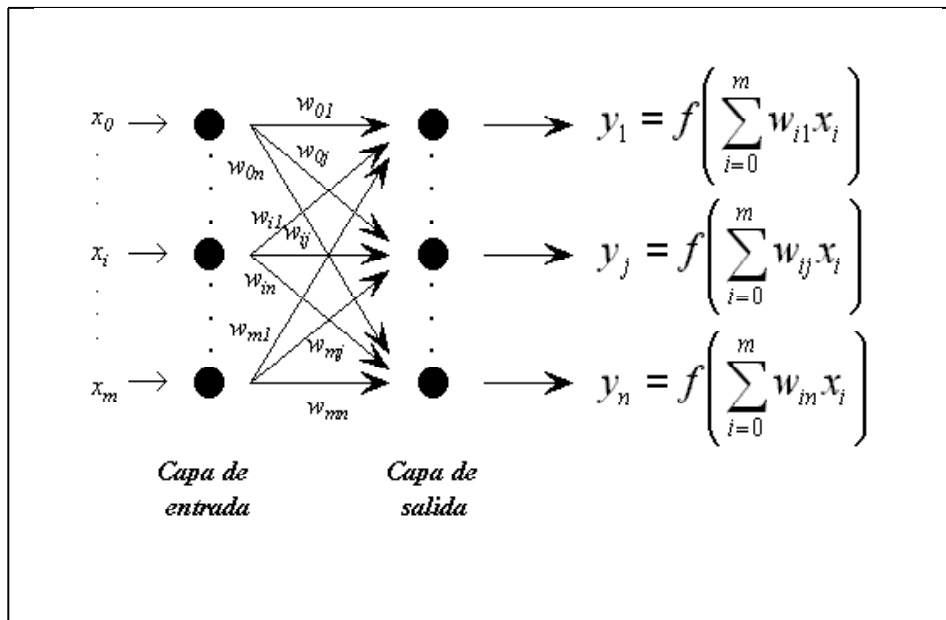


Figura 4.1.1. Esquema de un perceptrón simple formado por m neuronas en la capa de entrada y n neuronas en la capa de salida.

La importancia histórica del perceptrón simple estriba en su algoritmo de aprendizaje, que permite determinar de forma automática los pesos sinápticos para resolver problemas de clasificación de patrones a partir de un conjunto de datos de entrada etiquetados. En la próxima sección se analiza dicho algoritmo de aprendizaje.

4.1.2 ALGORITMO DE APRENDIZAJE DEL PERCEPTRÓN SIMPLE.

Se trata de un algoritmo de aprendizaje supervisado por corrección de errores. Como ya se ha mencionado, este tipo de algoritmos durante la fase de aprendizaje, ajusta los pesos de forma proporcional a la diferencia que existe entre la salida obtenida en la red y la que se desea obtener, con el objetivo de minimizar el error cometido por la red.

La actualización de los pesos se realiza mediante la regla de Hebb (D. Hebb, (1949)), cuya idea principal consiste en que si dos neuronas i y j tienen similares estados de activación, esto es, que ambos sean del mismo signo, entonces la conexión entre ellas debe hacerse más fuerte, lo que se traduce en que el peso sináptico asociado w_{ij} debe incrementarse. Por el contrario, si los estados de activación poseen signos contrarios, lo que indicaría una ausencia de similaridad, entonces la conexión entre ellas debe ser debilitada, lo que se traduce en que el peso sináptico asociado w_{ij} debe reducirse.

Definición 4.1.2.

Para una red de neuronas formada por m neuronas en la capa de entrada y n neuronas en la capa de salida, sea

$$X^r = (x_1, \dots, x_m)^r, \quad r = 1, \dots, p,$$

un conjunto de datos de entrada junto con los valores de las variables objetivo correspondientes

$$T^r = (t_1, \dots, t_n)^r, \quad r = 1, \dots, p.$$

Una vez fijada la condición de terminación C , la secuencia del algoritmo de aprendizaje del perceptrón simple es la siguiente:

1. Se determina la constante $\eta > 0$.
2. Se inicializan aleatoriamente los pesos sinápticos de las conexiones, $w_{11}, \dots, w_{1n}, w_{21}, \dots, w_{2n}, \dots, w_{m1}, \dots, w_{mn}$, y los valores umbrales, $w_{01} = -u_1, \dots, w_{0n} = -u_n$.
3. Mientras que no se verifique la condición de terminación C , repetir:
 - Se introduce en la red uno de los datos de entrada junto con los valores de las variables dependientes correspondientes.

- Se calcula la salida que proporciona la red,

$$y_j = f \left(\sum_{i=0}^m w_{ij} x_i \right), \quad j = 1, \dots, n,$$

siendo f una función paso de las descritas anteriormente.

- Se adaptan los pesos sinápticos según la siguiente expresión, conocida como la regla de Hebb:

$$\Delta w_{ij}^r = \begin{cases} 2\eta y_j^r x_i^r, & \text{si } y_j^r \neq t_j^r \\ 0, & \text{si } y_j^r = t_j^r \end{cases}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

que también se puede escribir como,

$$\Delta w_{ij}^r = \eta(t_j^r - y_j^r)x_i^r, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

En este algoritmo el peso sináptico no se modifica si la salida obtenida es igual que la deseada, y en caso contrario, esto es, en el caso de que la salida de la red sea incorrecta, el peso sináptico se actualiza mediante la regla de Hebb.

En cuanto al factor η , que recibe el nombre de tasa de aprendizaje, éste se utiliza para ajustar el ritmo de aprendizaje. Un valor pequeño implicaría un aprendizaje lento, mientras que uno excesivamente grande podría conducir a la inestabilidad de las estimaciones de los pesos sinápticos, por lo que η toma valores comprendidos entre 0 y 1.

Hay que tener en cuenta que, a pesar de la sencillez y facilidad de cálculo de la regla de Hebb, ésta únicamente es efectiva cuando existe una correlación fuerte entre los valores de entrada a la red y los de salida de la red.

4.1.3 CARACTERÍSTICAS DEL PERCEPTRÓN SIMPLE.

Definición 4.1.3.

Dado un conjunto de r vectores m -dimensionales $x^{(1)}, x^{(2)}, \dots, x^{(r)}$, donde cada uno de dichos vectores pertenece a una de entre dos clases posibles, C_1 y C_2 , se dice que los conjuntos de vectores correspondientes a cada una de las clases son separables linealmente si existe un conjunto de coeficientes

$w_0, w_1, w_2, \dots, w_m$, tal que

$$\begin{cases} \sum_{i=1}^m w_i x_i + w_0 \geq 0, & \forall (x_1, \dots, x_m) \in C_1 \\ \sum_{i=1}^m w_i x_i + w_0 < 0, & \forall (x_1, \dots, x_m) \in C_2 \end{cases} .$$

A continuación se muestra cómo el perceptrón simple únicamente permite realizar tareas de clasificación entre clases linealmente separables, lo que constituye una gran limitación para este tipo de red de neuronas y supuso el desarrollo de otros modelos de redes.

Proposición 4.1.1.

El perceptrón simple definido anteriormente constituye un separador lineal.

Demostración:

Estableciendo la suposición de que cada neurona del perceptrón representa una determinada clase, por lo que los valores de salida de cada neurona serán 0 en el caso de que no pertenezca a esa clase que representa, y 1 si pertenece.

Sea un perceptrón simple formado por dos neuronas en la capa de entrada, x_1 y x_2 , y una única neurona en la capa de salida, cuya salida y viene dada por

$$y = f \left(\sum_{i=0}^2 w_i x_i \right) = w_1 x_1 + w_2 x_2 - u,$$

lo que es equivalente a

$$y = \begin{cases} 1, & \text{si } w_1 x_1 + w_2 x_2 \geq u \\ 0, & \text{si } w_1 x_1 + w_2 x_2 < u \end{cases} .$$

Si se lleva a cabo su representación gráfica situando a las entradas x_1 y x_2 sobre los ejes de abscisas y coordenadas respectivamente, la siguiente condición,

$$w_1 x_1 + w_2 x_2 - u = 0,$$

o lo que es lo mismo,

$$x_2 = -\frac{w_1}{w_2} x_1 + \frac{u}{w_2},$$

representa la ecuación de una recta que divide el plano en dos regiones que corresponden con las determinadas por la salida de las neuronas. Este mismo resultado se obtendría en el caso de tener n entradas, ya que resultaría un hiperplano que delimitaría los distintos espacios.

Por tanto, un perceptrón simple representa un discriminador lineal, ya que implementa una condición lineal o hiperplano que separa las diversas regiones de decisión.

La limitación del perceptrón simple aparece claramente al considerar el problema de la función XOR, que devuelve un 0 lógico si las variables de entrada son iguales y un 1 lógico si son distintas. M. Minsky y S. Papert (1969) realizaron una importante crítica del perceptrón simple, y demostraron que aun conectando un gran número de ellos, no eran capaces de resolver el problema de la linealidad. Esto hizo decaer el interés por las RNAS, hasta la presentación del perceptrón multinivel, modelo que salvó las limitaciones del perceptrón simple.

Sea nuevamente considerado un perceptrón simple formado por dos neuronas en la capa de entrada, x_1 y x_2 , y una única neurona en la capa de salida, cuya salida y viene dada por

$$y = f\left(\sum_{i=0}^2 w_i x_i\right) = w_1 x_1 + w_2 x_2 - u.$$

Para la función XOR, la salida del perceptrón simple es equivalente a

x_1	x_2	y	Condición
0	0	0	$w_1 x_1 + w_2 x_2 < u$
0	1	1	$w_1 x_1 + w_2 x_2 > u$
1	0	1	$w_1 x_1 + w_2 x_2 > u$
1	1	0	$w_1 x_1 + w_2 x_2 < u$

La representación gráfica de estas condiciones en el plano se muestra en la Figura 4.1.2, y como puede apreciarse en ella, no es posible encontrar una única condición lineal para separar el conjunto de soluciones del problema, por lo que se dice que la función XOR no es separable linealmente.

De todo esto puede concluirse que la clase de funciones no separables linealmente no puede ser representada por medio de un perceptrón simple.

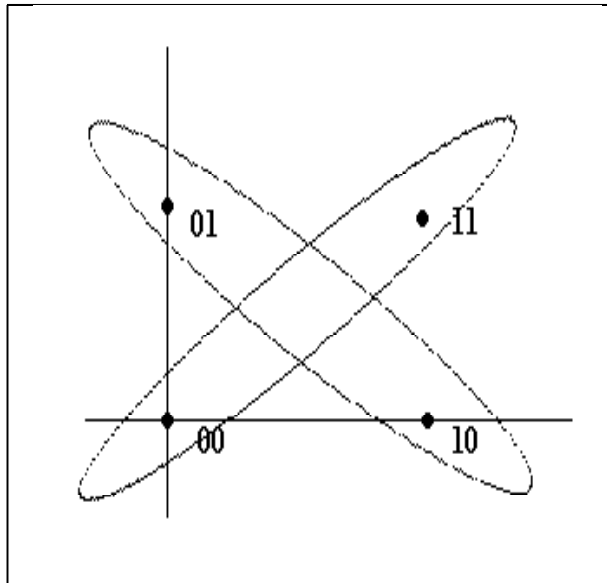


Figura 4.1.2. Representación gráfica de las condiciones del problema XOR.

Una solución a esta limitación consiste en la introducción de más capas en la arquitectura de la red, con lo que aparece el concepto de perceptrón multinivel o multicapa.

4.2 EL PERCEPTRÓN MULTINIVEL O MULTICAPA (MLP).

Ya a finales de los años 60 se señaló como solución a las limitaciones del perceptrón la introducción de capas ocultas en la red, pero el problema estribaba en que no se disponía de ningún algoritmo de aprendizaje que permitiera obtener los pesos sinápticos en una red multicapa con neuronas ocultas, lo que no se consiguió hasta mediados de los años 80, cuando D. E. Rumelhart y otros presentaron el algoritmo de retropropagación de errores o backpropagation (BP).

4.2.1 DEFINICIÓN DEL PERCEPTRÓN MULTINIVEL O MULTICAPA.

Definición 4.2.1.

El perceptrón multinivel o multicapa (MLP) es una red de tipo hacia adelante (feedforward) compuesto por una o varias capas de neuronas ocultas entre las de entrada y salida.

En la Figura 4.2.1 se muestra la estructura general de un perceptrón multicapa, en el que la capa de entrada está constituida por m neuronas, cada una de las cuales toma uno de los valores de entrada x_1, \dots, x_m , y la capa de salida está formada por n neuronas que proporcionan los valores de salida de la red y_1, \dots, y_n .

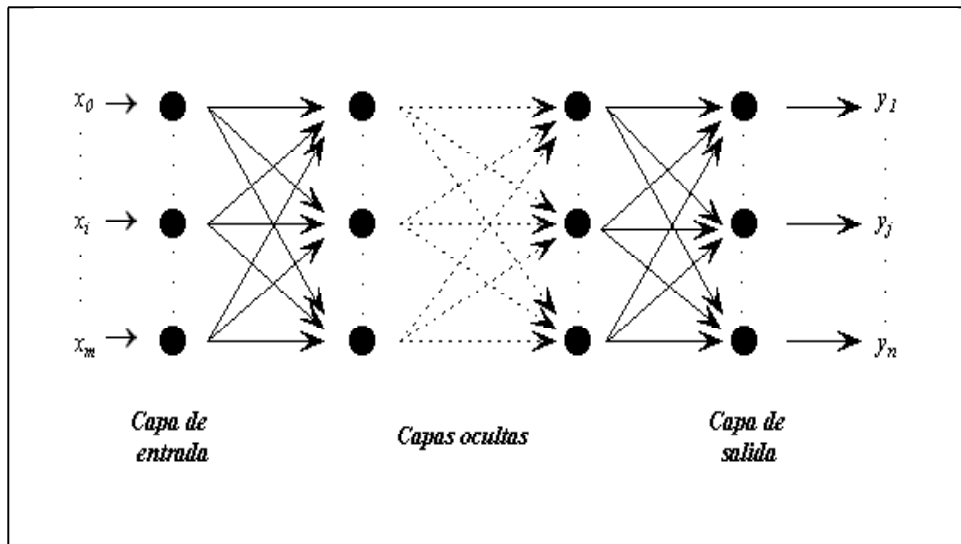


Figura 4.2.1. Estructura general de un perceptrón multinivel.

Por regla general, esta arquitectura de red se suele entrenar mediante el algoritmo de aprendizaje BP, o de algunas de sus variantes, por lo que al conjunto perceptrón multicapa + aprendizaje BP se le suele denominar red backpropagation o simplemente BP.

A continuación se describirá la estructura de un perceptrón multinivel o multicapa constituido por tres capas, tal y como puede observarse en la Figura 4.2.2.

Sea un MLP con una única capa oculta, en el que la capa de entrada está constituida por m neuronas, la capa oculta está formada por h neuronas, y n es el número de neuronas que constituyen la capa de salida. Entonces, un perceptrón multinivel formado por tres capas viene determinado por los siguientes elementos:

1. Los valores de entrada x_1, \dots, x_m con los pesos sinápticos, $w_{11}, \dots, w_{1h}, w_{21}, \dots, w_{2h}, \dots, w_{m1}, \dots, w_{mh}$, asociados a sus conexiones con las h neuronas de la capa oculta, y $v_{11}, \dots, v_{1n}, v_{21}, \dots, v_{2n}, \dots, v_{h1}, \dots, v_{hn}$, los pesos sinápticos asociados a las conexiones entre las neuronas de la capa oculta con las n neuronas de salida.
2. Los valores z_1, \dots, z_h que proporcionan las neuronas de la capa oculta vienen dados por

$$z_k = g \left(\sum_{i=0}^m w_{ik} x_i \right), \quad k = 1, \dots, h,$$

donde g es la función de activación de dicha capa y donde se introduce una neurona ficticia de entrada, x_0 , con valor idénticamente igual a 1, denotándose los valores umbrales de las neuronas de esta capa por $-w_{0k}$.

3. Y por último, los valores de salida y_1, \dots, y_n que proporciona la red vienen dados por

$$y_j = f \left(\sum_{k=0}^h v_{kj} z_k \right) = f \left(\sum_{k=0}^h v_{kj} g \left(\sum_{i=0}^m w_{ik} x_i \right) \right), \quad j = 1, \dots, n,$$

donde f es la función de activación de la capa de salida y donde se introduce una neurona oculta ficticia, z_0 , con valor idénticamente igual a 1, denotándose los valores umbrales de las neuronas de esta capa por $-v_{0j}$.

De aquí se deduce que el número de parámetros de un MLP con una capa oculta es $h(m + n + 1) + n$.

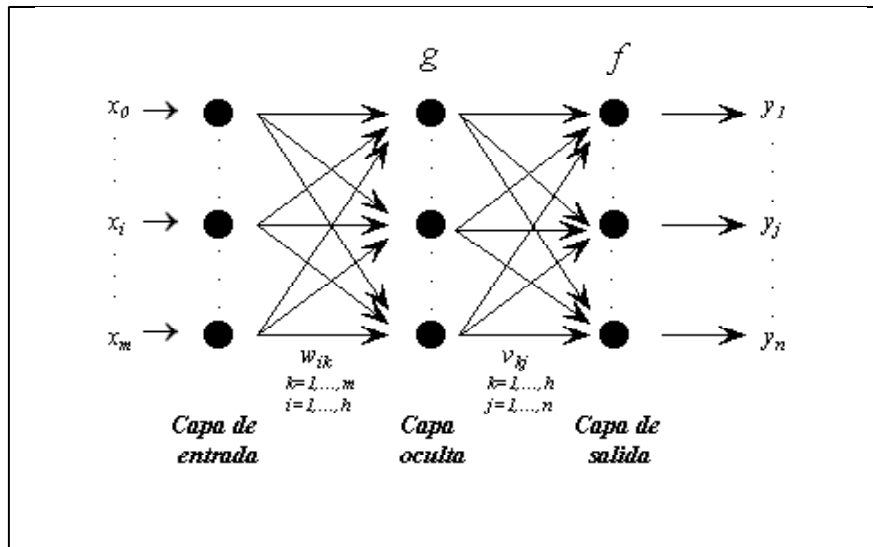


Figura 4.2.2. Estructura de un perceptrón multinivel constituido por tres capas.

4.2.2 ALGORITMO DE APRENDIZAJE DEL PERCEPTRÓN MULTINIVEL O MULTICAPA: RETROPROPAGACIÓN DE ERRORES (BP).

El método de aprendizaje por retropropagación de errores o backpropagation (BP), fue presentado por D. E. Rumelhart, J. L. McClelland, G. E. Hinton y R. J. Williams (1986). Este algoritmo está basado en la generalización de la regla delta (B. Widrow y M. E. Hoff, (1960)), y su aparición ha ampliado de forma considerable el rango de aplicación de las redes de neuronas.

Se trata de un algoritmo de aprendizaje supervisado por corrección de errores que emplea una función de error en la salida durante la fase de entrenamiento, e irá corrigiendo los pesos, primero los de la capa de salida, y luego progresivamente hacia atrás, los de las demás capas, de ahí su nombre. La función de error es similar a la del problema de mínimos cuadrados, pero además exige que las funciones de transferencia de las neuronas sean diferenciables. De esta forma, el BP aparece como consecuencia natural de extender a una red multicapa la regla del mínimo error cuadrático medio.

Sea un perceptrón multicapa formado por m neuronas en la capa de entrada, al menos una capa oculta de neuronas internas, y n neuronas en la capa de salida, siendo

$$X^r = (x_1, \dots, x_m)^r, \quad r = 1, \dots, p,$$

un conjunto de ejemplos de datos de entrada junto con los valores de las variables objetivo correspondientes

$$T^r = (t_1, \dots, t_n)^r, \quad r = 1, \dots, p,$$

y sea

$$Y^r = (y_1, \dots, y_n)^r, \quad r = 1, \dots, p,$$

las salidas que proporciona la red.

Definición 4.2.2.

Se define el error cuadrático sobre un patrón determinado r , E^r , como

$$E^r = \frac{1}{2} \sum_{j=1}^n (y_j^r - t_j^r)^2 = \frac{1}{2} \sum_{j=1}^n (e^r)^2, \quad r = 1, \dots, p,$$

siendo $e^r = (y_j^r - t_j^r)$ el vector de los errores de la red para el patrón r , $r = 1, \dots, p$.

Definición 4.2.3.

El error cuadrático total, E , vendrá definido por

$$E = \frac{1}{2} \sum_{r=1}^p \sum_{j=1}^n (y_j^r - t_j^r)^2 = \frac{1}{2} \sum_{r=1}^p \sum_{j=1}^n (e^r)^2 = \frac{1}{2} \|e\|^2,$$

siendo e el vector de los errores de la red.

Para forzar la similitud entre la salida que proporciona la red, Y^r , y los valores de las variables dependientes que se pretenden predecir, T^r , el vector de parámetros de pesos sinápticos es elegido de forma que se minimice el error cuadrático total E . Para ello se recurre al cálculo del vector gradiente de E , cuya dirección proporciona el mayor incremento en el valor del error cuadrático total, de tal forma que los pesos sinápticos serán iterativamente adaptados según la dirección contraria al vector gradiente.

Definición 4.2.4.

Sea un perceptrón multicapa con un total de q pesos sinápticos, incluidos los valores umbrales, y con funciones de activación diferenciables, y sea W el vector formado por esos q coeficientes sinápticos.

Entonces, se define el vector gradiente de E para un vector de pesos sinápticos determinado, W^* , con respecto a los pesos sinápticos como

$$\nabla E(W^*) = \frac{\partial E}{\partial W} \Big|_{W=W^*} = \left(\frac{\partial E}{\partial W_1} \Big|_{W_1=W_1^*}, \dots, \frac{\partial E}{\partial W_q} \Big|_{W_q=W_q^*} \right).$$

Mediante la regla de aprendizaje de retropropagación de errores (BP) o regla delta generalizada, los pesos sinápticos son actualizados a través de la fórmula

$$W = W^* - \eta \left(\frac{\partial E}{\partial W} \Big|_{W=W^*} \right),$$

donde al parámetro η se le conoce como tasa de aprendizaje.

A continuación se describe de forma detallada el algoritmo de aprendizaje BP.

Sea un perceptrón multicapa formado por m neuronas en la capa de entrada, al menos una capa oculta de neuronas internas, y n neuronas en la capa de salida, siendo

$$X^r = (x_1, \dots, x_m)^r, \quad r = 1, \dots, p,$$

un conjunto de ejemplos de datos de entrada junto con los valores de las variables objetivo correspondientes

$$T^r = (t_1, \dots, t_n)^r, \quad r = 1, \dots, p,$$

e

$$Y^r = (y_1, \dots, y_n)^r, \quad r = 1, \dots, p,$$

las salidas que proporciona la red.

Sea además q el total de pesos sinápticos, incluidos los valores umbrales, y con funciones de activación diferenciables, y sea W el vector formado por esos q coeficientes sinápticos.

Entonces, fijados una tasa de aprendizaje η y una condición de terminación C , el algoritmo de aprendizaje BP se desarrolla en las siguientes etapas:

- Paso 1: Establecer aleatoriamente los pesos y umbrales iniciales. No se debe empezar con valores nulos porque el aprendizaje no progresaría, ya que las salidas de las neuronas serían nulas, y el incremento de los pesos también.
- Paso 2: En la etapa ξ , para cada patrón r del conjunto de entrenamiento,
 1. Ejecutar la red para obtener la respuesta Y^r ante este patrón.
 2. Calcular el error cuadrático sobre el patrón r ,

$$E^r = \frac{1}{2} \sum_{j=1}^n (y_j^r - t_j^r)^2.$$

3. Hallar los incrementos de los pesos y umbrales de todas las capas mediante la regla

$$W^{(\xi)r} = W^{(\xi-1)r} - \eta^r \left(\frac{\partial E^r}{\partial W} \Big|_{W=W^{(\xi-1)r}} \right).$$

- Paso 3: Se obtienen los incrementos totales una vez entrenada la red con todos los patrones, $r = 1, \dots, p$, para todos los pesos y umbrales.
- Paso 4: Se actualizan los pesos y umbrales,

$$W^\xi = \sum_{r=1}^p W^{(\xi)r}.$$

- Paso 5: Se calcula el error cuadrático total,

$$E = \frac{1}{2} \sum_{r=1}^p \sum_{j=1}^n (y_j^r - t_j^r)^2 = \sum_{r=1}^p E^r,$$

y se vuelve al Paso 2 en el caso de que todavía no se haya satisfecho la condición de terminación C .

Para el caso de un perceptrón multinivel con tres capas, el algoritmo BP se expresaría del siguiente modo.

Sea una estructura de perceptrón multinivel como la que aparecía en la Figura 4.2.2 y sean las entradas, salidas, pesos y umbrales como se han definido anteriormente para el algoritmo BP con un perceptrón multicapa general.

Sean además g la función de activación de la capa oculta de tipo sigmoidal y f la función de activación de la capa de salida.

Entonces, para un MLP con tres capas, el algoritmo BP presenta las siguientes expresiones.

Para un patrón r del conjunto de entrenamiento, la salida de la red vendrá dada por la ecuación,

$$y_j^r = f \left(\sum_{k=0}^h v_{kj} z_k^r \right) = f \left(\sum_{k=0}^h v_{kj} g \left(\sum_{i=0}^m w_{ik} x_i^r \right) \right), \quad j = 1, \dots, n,$$

y el error cuadrático sobre el patrón r , $r = 1, \dots, p$, se expresará como

$$E^r = \frac{1}{2} \sum_{j=1}^n (y_j^r - t_j^r)^2 = \frac{1}{2} \sum_{j=1}^n \left(f \left(\sum_{k=0}^h v_{kj} g \left(\sum_{i=0}^m w_{ik} x_i^r \right) \right) - t_j^r \right)^2.$$

El método de aprendizaje trata de minimizar este error, lo que se lleva a cabo mediante descenso por el gradiente. Habrá por tanto dos gradientes, uno respecto de los pesos de la capa de salida y otro respecto de los de la capa oculta.

Para la etapa ξ se tendrá que

$$\begin{aligned} w_{ik}^{(\xi)r} &= w_{ik}^{(\xi-1)r} - \eta^r \left(\frac{\partial E^r}{\partial w_{ik}} \Big|_{w_{ik}=w_{ik}^{(\xi-1)r}} \right). \\ v_{kj}^{(\xi)r} &= v_{kj}^{(\xi-1)r} - \eta^r \left(\frac{\partial E^r}{\partial v_{kj}} \Big|_{v_{kj}=v_{kj}^{(\xi-1)r}} \right). \end{aligned}$$

Teniendo en cuenta las dependencias funcionales y aplicando adecuadamente la regla de la cadena, derivando se obtiene para una etapa del algoritmo ξ fijada, que las actualizaciones de los pesos vienen determinadas por

las siguientes expresiones:

$$\Delta w_{ik}^r = \eta^r \sum_{j=1}^n \left[(y_j^r - t_j^r) x_i^r g' \left(\sum_{i=0}^m w_{ik} x_i^r \right) \left(\sum_{k=0}^h v_{kj} \right) f'(y_j^r) \right].$$

$$\begin{aligned} \Delta v_{kj}^r &= \eta^r \sum_{j=1}^n \left[(y_j^r - t_j^r) g \left(\sum_{i=0}^m w_{ik} x_i^r \right) f' \left(\sum_{k=0}^h v_{kj} g \left(\sum_{i=0}^m w_{ik} x_i^r \right) \right) \right] \\ &= \eta^r \sum_{j=1}^n [(y_j^r - t_j^r) z_k^r f'(y_j^r)]. \end{aligned}$$

Una vez aplicado el entrenamiento para los p patrones, se obtendrán los incrementos totales y el error cuadrático total

$$\Delta w_{ik} = \sum_{r=1}^p \Delta w_{ik}^r.$$

$$\Delta v_{kj} = \sum_{r=1}^p \Delta v_{kj}^r.$$

$$E = \sum_{r=1}^p E^r.$$

Puede observarse que en la ecuación de actualización de los pesos sinápticos de las neuronas de la capa oculta están presentes los valores de los pesos de las neuronas de la capa de salida, por lo que se ve claramente el concepto de la propagación hacia atrás de los errores que da nombre al algoritmo.

También hay que indicar que en este proceso de aprendizaje se ejecuta la red para todos y cada uno de los patrones del conjunto de entrenamiento, acumulando la variación que supone este método en los pesos sinápticos y al final se procede a la efectiva actualización de los pesos. Es un proceso que se conoce como aprendizaje por lotes. Una variación de este proceso sería ir actualizando los pesos en cada iteración con cada patrón, lo que supondría un aprendizaje en serie, que se suele utilizar cuando se dispone de un conjunto enorme de patrones de entrenamiento, donde la redundancia de datos sea

grande, y la necesidad de almacenamiento parcial de los incrementos sea alta.

Obsérvese que el algoritmo BP no es más que la aplicación del método de optimización conocido como método del descenso más pronunciado, por lo que, al igual que este método de optimización, presenta el problema de no asegurar la convergencia a óptimos globales, pudiendo quedar la solución atrapada en óptimos locales.

4.2.3 CARACTERÍSTICAS DEL PERCEPTRÓN MULTINIVEL O MULTICAPA.

Tras observarse las limitaciones del perceptrón simple, se llegó a una arquitectura donde se introdujeron capas intermedias entre las de entrada y salida, de manera que la intersección de los planos definidos por las capas ocultas proporcionaran soluciones a problemas de clasificación de gran envergadura. Se comprobó experimentalmente que el MLP resolvía estos problemas de una manera eficaz y relativamente simple, por lo que, a partir de esta comprobación experimental, se intentó demostrar teóricamente cómo se comportaban este tipo de redes.

W. S. McCulloch y W. Pitts (1943) fueron los primeros que mostraron que su modelo de neurona, que consistía esencialmente en un dispositivo de umbral, podía representar cualquier función booleana.

Mucho más tarde, J. S. Denker y otros (1987) demostraron que toda función booleana podía ser representada por una red unidireccional multicapa de una sola capa oculta.

Por esas mismas fechas, R. P. Lippmann (1987) demostró que un perceptrón con dos capas ocultas bastaba para representar regiones de decisión arbitrariamente complejas.

Por otro lado, A. Lapedes y R. Farber (1988) demostraron que un perceptrón de dos capas ocultas es suficiente para representar cualquier función arbitraria (y no necesariamente booleana).

Más tarde, R. Hecht-Nielsen (1987, 1990), aplicando el teorema de Kolmogorov, demostró que una arquitectura de características similares al perceptrón multicapa, con una única capa oculta, resultaba ser un aproximador universal de funciones.

Y ya a finales de la década, surgieron diversos grupos que demostraron matemáticamente que el perceptrón multicapa de una única capa oculta era

en efecto un aproximador universal de funciones, como por ejemplo K. Funahashi (1989), K. Hornik, M. Stinchcombe y H. White (1989), y G. Cybenko (1989). Algunos de estos resultados se enumeran a continuación, y sus demostraciones pueden encontrarse en las referencias arriba citadas.

Teorema 4.2.1. (Teorema de Kolmogorov).

(Se presenta aquí una versión simplificada y adaptada al contexto de las redes de neuronas artificiales).

Dada una función $\varphi : K \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ con K compacto, existe un perceptrón multinivel con dos capas ocultas, la primera formada por $m(2m + 1)$ nodos y la segunda por $2m + 1$ nodos, que reproduce exáctamente la función φ .

Teorema 4.2.2. (Teorema de Funahashi).

Sea $g(x)$ una función no constante, acotada y monótona creciente. Sea K un subconjunto compacto (acotado y cerrado) de \mathbb{R}^m . Sea un número real $\varepsilon \in \mathbb{R}$, y sea un entero $k \in \mathbb{Z}$, fijado tal que $k \geq 3$. En estas condiciones, se tiene que:

Cualquier función

$$\varphi : x \in K \longrightarrow (\varphi_1(x), \varphi_2(x), \dots, \varphi_m(x)) \in \mathbb{R}^n,$$

con $\varphi_i(x)$ sumables en K , puede ser aproximado en el sentido de la topología L_2 en K por la función representada por una red de neuronas unidireccional de k capas ($k-2$ capas ocultas), con $g(x)$ la función de activación en la capa oculta, y con funciones de activación lineales para las de las capas de entrada y salida.

Esto es, $\forall \varepsilon > 0$, \exists un MLP de las características anteriores, que implementa

$$\varphi' : x \in K \longrightarrow (\varphi'_1(x), \varphi'_2(x), \dots, \varphi'_m(x)) \in \mathbb{R}^n,$$

de manera que

$$d_{L_2(K)}(\varphi, \varphi') = \left(\sum_{j=1}^n \int_K |\varphi_j(x_1, \dots, x_m) - \varphi'_j(x_1, \dots, x_m)|^2 dx \right)^{1/2} < \varepsilon.$$

Es fácil observar que las funciones sigmoideas empleadas habitualmente como función de transferencia en el MLP cumplen las condiciones exigidas a $g(x)$, lo que se recoge en el siguiente teorema.

Teorema 4.2.3. (Teorema de Cybenko).

Toda función absolutamente integrable Lebesgue $\varphi : K \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ con K compacto, puede ser aproximada uniformemente en L_1 mediante un MLP con una capa oculta, función de activación acotada, medible y sigmoideal en la capa oculta, y función de activación identidad en la capa de salida.

Esto es, $\forall \varepsilon > 0, \exists$ un MLP de las características anteriores tal que

$$\|\varphi - \varphi'\| = \int_K |\varphi(x_1, \dots, x_m) - \varphi'(x_1, \dots, x_m)| dx < \varepsilon.$$

Teorema 4.2.4. (Teorema de Hornik-Stinchcombe-White).

Toda función continua $\varphi : K \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^n$ con K compacto, puede ser aproximada mediante un MLP con una capa oculta y función de activación no decreciente en la capa oculta, siempre que sean utilizadas un número suficiente de neuronas en la capa oculta.

Como consecuencia de todos estos resultados enunciados, un perceptrón multicapa de una única capa oculta puede aproximar hasta el nivel deseado dentro de un conjunto compacto cualquier función continua, lo que demuestra que las redes multicapa unidireccionales pueden considerarse aproximadores funcionales universales.

Estos teoremas que demuestran el funcionamiento de este modelo de redes de neuronas fueron de gran importancia, ya que supusieron un resurgir en el campo de la neurocomputación. Sin embargo, todavía quedan muchas cuestiones por resolver, como por ejemplo, el número de nodos ocultos necesarios para aproximar una función determinada, ya que en el Teorema de Hornik-Stinchcombe-White solo se afirma que hay que emplear el número de neuronas en la capa oculta necesario para lograr el nivel de aproximación requerido. Podría ocurrir que, para un problema concreto, el número de neuronas ocultas para alcanzar cierta cota de error sea tan elevado que su aplicación resulte inabordable en la práctica, por lo que algunos investigadores, como K. G. Mehrotra, C. K. Mohan y S. Ranka (1991) proponen mantener bajo el número de neuronas ocultas introduciendo para ello en el modelo otra capa oculta.

4.2.4 ASPECTOS PRÁCTICOS RELATIVOS AL PROCESO DE ENTRENAMIENTO DEL PERCEPTRÓN MULTINIVEL O MULTICAPA.

El objetivo del entrenamiento de una red de neuronas artificial es la construcción de un modelo estadístico que represente el proceso que genera los datos, obteniendo buenas predicciones para las nuevas entradas que se le proporcionen a la RNA y evitando que la red de neuronas artificial se especialice en los conjuntos de ejemplos proporcionados. A este problema se le conoce como pérdida de generalidad de la red de neuronas artificial, y para evitar que se produzca, la arquitectura de la RNA y el proceso de entrenamiento de ésta, deben ser diseñados de forma que la red de neuronas artificial aprenda de forma robusta y eficiente.

Con respecto a la arquitectura del MLP, el número de capas ocultas a emplear viene resuelto por cualquiera de los teoremas enunciados anteriormente y que determinan la propiedad de aproximador universal de un perceptrón multicapa con una única capa oculta, siempre que la capa oculta esté constituida por un número suficiente de neuronas.

En cuanto a ese número suficiente de neuronas ocultas, no existe ningún resultado definitivo que lo determine, sino tan sólo algunas reglas e indicaciones enumeradas por ciertos autores, como las proporcionadas, por ejemplo, por S. C. Huang e Y. F. Huang (1991), y por S. Y. Kung y J. N. Hwang (1988).

Por tanto, se hace preciso optimizar la complejidad del modelo, lo que algunos autores han resuelto mediante el problema conocido como sesgo-varianza.

Supuesto que se desea predecir una variable unidimensional T , sea

$$X^r = (x_1, \dots, x_m)^r, \quad r = 1, \dots, p,$$

un conjunto de ejemplos de datos de entrada e Y^r , $r = 1, \dots, p$, la salida que proporciona la red de neuronas artificial una vez que ésta ha sido entrenada a partir del conjunto de entrada X^r .

Definición 4.2.5.

Se define el error de generalización E_{gen} de una red de neuronas artificial como

$$E_{gen} = E_{X^r} [Y - T]^2 .$$

Definición 4.2.6.

Se define el error empírico E_{emp} de una red de neuronas artificial como

$$E_{emp} = \frac{1}{p} \sum_{r=1}^p (Y^r - T^r)^2 ,$$

siendo T^r , $r = 1, \dots, p$, el valor de la variable objetivo que se desea predecir e Y^r , $r = 1, \dots, p$, la salida que proporciona la red de neuronas artificial para cada uno de los elementos del conjunto de entrada X^r .

Desafortunadamente, el error empírico no constituye un buen estimador del error de generalización, ya que en su fórmula de cálculo intervienen únicamente los datos del conjunto de entrada utilizados en el proceso de entrenamiento de la red de neuronas artificial. Más adelante se analizará cómo proporcionar dicho estimador.

Proposición 4.2.1.

Sea una RNA con una arquitectura determinada y un vector de coeficientes sinápticos W . Entonces, el error de generalización de una RNA se puede descomponer en la suma del sesgo al cuadrado más la varianza, esto es,

$$E_{gen} = (E_X [Y] - T)^2 + E_X \{Y - E_X [Y]\}^2 .$$

Así, un modelo que sea demasiado simple o demasiado inflexible tendrá un sesgo grande, lo que indicaría una falta de representatividad de los datos de entrada, mientras que un modelo demasiado flexible presentará una gran varianza, pudiendo llegarse a una situación de sobreajuste y pérdida de generalidad, por lo que habrá que llegar a un balance óptimo entre sesgo y varianza para poder minimizar ambos.

Para controlar la complejidad del modelo, que en el caso de las redes de neuronas vendrá determinado por su número de parámetros, H. White (1990), demostró que la complejidad del modelo vendrá determinado por el número de datos de los que se disponga, pero no estableció ninguna relación directa entre ambos valores. Por ello, no es posible determinar mediante ninguna regla el número óptimo de parámetros de la red.

Respecto al número de neuronas de la capa de entrada y de la de salida, éstos van a depender de la naturaleza del problema que se pretende resolver y evidentemente de los datos de los que se disponga para ello. Hay que indicar aquí la conveniencia de transformar estos datos de entrada y salida de la red antes de comenzar con el proceso de entrenamiento, de forma que todas las variables tengan un rango de valores similar.

En cuanto al proceso de entrenamiento de la red, éste puede ser realizado mediante cualquiera de los algoritmos de entrenamiento existentes, y en el caso de utilizar un algoritmo de retropropagación de errores, el proceso de entrenamiento puede ser mejorado y agilizado introduciendo cualquiera de sus variaciones asegurando así una convergencia más rápida.

En este sentido, la selección de los pesos sinápticos iniciales también juega un papel determinante en la convergencia de la red, por lo que numerosos autores han propuesto diversas reglas para su selección, como por ejemplo, S. E. Fahlman (1988), A. P. Russo (1991) y D. Nguyen y B. Widrow (1990).

Una vez determinados todos estos aspectos, surge la necesidad de comprobar si la red está entrenada de forma apropiada o no. Realmente, en la práctica se recurre a buscar aquella red de neuronas artificiales que produzca predicciones buenas para los nuevos datos que se le introduzcan, esto es, hallar la red de neuronas que tenga la mejor precisión en el proceso de generalización, por lo que la función error cuadrático total tendrá que ser evaluada empleando un conjunto de datos distinto al que se utilice durante el proceso de entrenamiento, y por ello se emplea la validación cruzada. Esta idea, introducida por A. S. Weigend, B. A. Huberman y D. E. Rumelhart (1990), consiste en dividir el conjunto de datos en tres conjuntos disjuntos: uno de entrenamiento, uno de validación y uno de test. La red entonces es entrenada a partir de los datos del conjunto de entrenamiento y los pesos son modificados de forma iterativa con el objeto de minimizar el error cometido entre la salida obtenida por la red y los valores de las variables dependientes deseados, pero también se va obteniendo el error que comete la red ante los datos del conjunto de validación, calculándose los errores cuadráticos totales

sobre los conjuntos de entrenamiento, E_{ent} , y de validación, E_{val} . Tal y como puede observarse en la Figura 4.2.3, donde se representa el comportamiento de ambas medidas, éstas van decreciendo a medida que se avanza en el proceso de entrenamiento, y cuando se observa que E_{val} ha dejado de decrecer y comienza a incrementar su valor, eso indica que el proceso de entrenamiento debe terminar, porque la red está comenzando a especializarse en el conjunto de datos que se le ha proporcionado como entrenamiento, siendo menos precisa al incorporar nuevos datos.

Posteriormente, la red es evaluada sobre el conjunto test, cuyos datos, al no intervenir en el proceso de entrenamiento de la red, sirven para medir de una forma completamente objetiva el nivel de certeza de la respuesta de la red ante una nueva entrada, garantizándose así la capacidad de generalización de la red de neuronas. De esta forma, esta medida del error cuadrático total sobre el conjunto de datos del conjunto test, E_{test} , proporcionará una estimación insesgada del error de generalización.

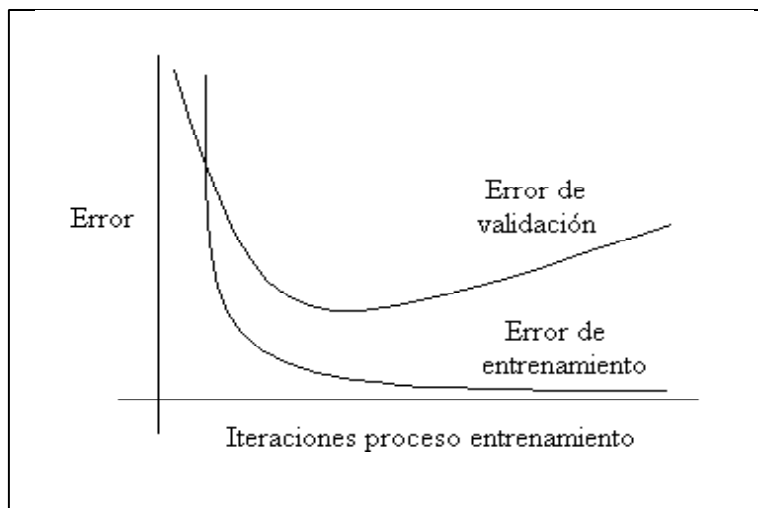


Figura 4.2.3. Comportamiento de los errores de entrenamiento y validación durante el proceso de entrenamiento de una red de neuronas.

5 CONSTRUCCIÓN DE REDES DE NEURONAS ARTIFICIALES.

5.1 ALGORITMOS DE ENTRENAMIENTO.

El algoritmo de aprendizaje BP es un método de gran generalidad. Su ventaja principal es que se puede aplicar a un gran número de problemas, proporcionando una solución aceptable en relativamente poco tiempo de desarrollo. Entre sus principales desventajas se encuentra una convergencia lenta, lo que surge como consecuencia de no precisar información apriori de la función a aproximar. Otro problema de este algoritmo de aprendizaje es el denominado sobreajuste, que puede provocar una especificación sobre el conjunto de entrenamiento en el proceso de aprendizaje, lo que implicaría una falta de generalidad de la red de neuronas. Una última consideración importante sobre este algoritmo, es que, al igual que los algoritmos de optimización no garantiza que se alcance el mínimo global de la función error, sino que es posible que se quede atrapado en un mínimo local, quedándose por tanto estancado el proceso de aprendizaje.

Observación:

En adelante se denotará por

$$\Delta W^{(\xi+1,\xi)} = W^{(\xi+1)} - W^{(\xi)}$$

al cambio producido en el vector de los pesos sinápticos W durante la iteración ξ , y por

$$\Delta (\nabla E(W^{(\xi+1,\xi)})) = \nabla E(W^{(\xi+1)}) - \nabla E(W^{(\xi)})$$

al cambio respectivo en los vectores gradientes, que denotando al vector gradiente por Γ , puede expresarse como

$$\Delta \Gamma^{(\xi+1,\xi)} = \Gamma^{(\xi+1)} - \Gamma^{(\xi)}.$$

Existen diversas variaciones del algoritmo BP propuestos para el perceptrón multinivel, pero básicamente pueden dividirse en dos grupos de métodos: aquellos basados en la aproximación de polinomios de Taylor (métodos

del descenso del gradiente y del gradiente conjugado), y aquellos que emplean aproximaciones cuadráticas (métodos quasi-Newton y de Levenberg-Marquardt).

Entre todas estas variantes del algoritmo de aprendizaje BP, en este trabajo se han seleccionado cinco métodos representativos de las distintas tendencias existentes: "gradient descent w/momentum & adaptative lr (GD)", "Fletcher-Reeves conjugate gradient (CGFR)", "Broyden-Fletcher-Goldfarb-Shanno quasi-Newton (BFGS)", "Levenberg-Marquardt (LM)" y "Resilient propagation (RP)", todos ellos disponibles en el módulo de redes de neuronas artificiales de MATLAB.

- **Gradient descent w/momentum & adaptative lr (GD).**

Para corregir el problema de la lenta convergencia del algoritmo BP, el mismo Rumelhart (1986) propuso la introducción de un término adicional, denominado término de momento, por lo que el incremento de los pesos sinápticos en la iteración ξ , se realiza de acuerdo a

$$\begin{aligned} W^{(\xi+1)} &= W^{(\xi)} - \eta^{(\xi)} \nabla E(W^{(\xi)}) + \mu (W^{(\xi)} - W^{(\xi-1)}) \\ &= W^{(\xi)} - \eta^{(\xi)} \Gamma^{(\xi)} + \mu \Delta W^{(\xi, \xi-1)}, \end{aligned}$$

siendo μ , $0 \leq \mu \leq 1$, el parámetro momento, y donde la tasa de aprendizaje $\eta^{(\xi)}$ se considerará diferente según el peso sináptico al que acompaña.

- **Fletcher-Reeves conjugate gradient (CGFR).**

Los métodos de gradiente conjugado fueron introducidos por M. R. Hestenes y E. L. Stiefel (1962), para el caso de sistemas lineales, y posteriormente fueron adaptadas por numerosos autores, entre ellos R. Fletcher y C. M. Reeves (1964), para resolver problemas de minimización no lineal.

Definición 5.1.1.

Dos vectores n -dimensionales $d^{(\xi+1)}$ y $d^{(\xi)}$ se dirán que son M -conjugados si se verifica

$$(d^{(\xi+1)})^t M d^{(\xi)} = 0,$$

siendo $M_{n \times n}$ una matriz simétrica y definida positiva.

Y el conjunto de direcciones $\{d^{(\xi)}\}$, donde cada par de direcciones no idénticas son M -conjugados, se dirá que es un conjunto M -conjugado.

La idea de este tipo de métodos consiste en actualizar el vector de pesos sinápticos actual, $W^{(\xi)}$, en la dirección H -conjugada $d^{(\xi)}$, para obtener un nuevo vector de pesos sinápticos, $W^{(\xi+1)}$, ya que la matriz Hessiana H es simétrica y definida positiva.

Según esta idea, se tomará

$$W^{(\xi+1)} = W^{(\xi)} + \beta^{(\xi)} d^{(\xi)},$$

y $\beta^{(\xi)}$ será elegido de forma que se minimice la función de error cuadrática E a lo largo de la dirección H -conjugada $d^{(\xi)}$, esto es,

$$\min_{\beta} E(W^{(\xi+1)}) = \min_{\beta} E(W^{(\xi)} + \beta^{(\xi)} d^{(\xi)}),$$

por lo que

$$\frac{\partial E(W^{(\xi)} + \beta^{(\xi)} d^{(\xi)})}{\partial \beta} = 0 \Rightarrow (\nabla E(W^{(\xi)} + \beta^{(\xi)} d^{(\xi)}))^t d^{(\xi)} = 0.$$

Mediante la aproximación de Taylor de primer orden para el gradiente

$$\begin{aligned} \Gamma^{(\xi+1)} &= \Gamma^{(\xi)} + H(W^{(\xi+1)} - W^{(\xi)}) \\ \Delta\Gamma^{(\xi+1, \xi)} &= H(W^{(\xi+1)} - W^{(\xi)}) \\ &= H((W^{(\xi)} + \beta^{(\xi)} d^{(\xi)}) - W^{(\xi)}) \\ &= \beta^{(\xi)} H d^{(\xi)}. \end{aligned}$$

Teniendo en cuenta que para una función cuadrática de error E , una dirección H -conjugada será generada mediante

$$d^{(\xi+1)} = -\Gamma^{(\xi+1)} + \beta^{(\xi)} d^{(\xi)}, \quad \xi = 0, \dots, n-1,$$

por lo que

$$[-\Gamma^{(\xi+1)} + \beta^{(\xi)} d^{(\xi)}]^t H d^{(\xi)} = 0,$$

entonces

$$\beta^{(\xi)} = \frac{[\Gamma^{(\xi+1)}]^t H d^{(\xi)}}{[d^{(\xi)}]^t H d^{(\xi)}},$$

que mediante la expresión obtenida a través de la aproximación de Taylor, es equivalente a

$$\beta^{(\xi)} = \frac{[\Gamma^{(\xi+1)}]^t \Delta\Gamma^{(\xi+1,\xi)}}{[d^{(\xi)}]^t \Delta\Gamma^{(\xi+1,\xi)}}.$$

Esta es la fórmula de Hestenes-Stiefel para el algoritmo de gradiente conjugado, que empleando la propiedad de ortogonalidad de los gradientes,

$$[\Gamma^{(\xi)}]^t \Gamma^{(\varphi)} = 0, \quad \forall \xi < \varphi \leq n,$$

se deriva la fórmula del algoritmo del gradiente conjugado de Fletcher-Reeves

$$\beta^{(\xi)} = \frac{[\Gamma^{(\xi+1)}]^t \Gamma^{(\xi+1)}}{[\Gamma^{(\xi)}]^t \Gamma^{(\xi)}}.$$

- **Broyden-Fletcher-Goldfarb-Shanno quasi-Newton (BFGS).**

Una alternativa a los métodos de gradiente conjugado en los que se precisa el cálculo de la matriz Hessiana, H , y de su inversa, H^{-1} , son los métodos conocidos como quasi-Newton, que en vez de calcular H directamente, y posteriormente evaluar su inversa, calculan una aproximación a la matriz inversa del Hessiano, $G = \hat{H}^{-1}$, de forma secuencial, empleando únicamente para ello las primeras derivadas de la función error.

De la fórmula de Newton se deduce

$$W^{(\xi+1)} - W^{(\xi)} = -H^{-1} (\Gamma^{(\xi+1)} - \Gamma^{(\xi)}),$$

que es conocida como la condición quasi-Newton, y cuyas soluciones proporcionan la aproximación a la inversa de la matriz Hessiana.

Existen numerosas soluciones a esta condición, y entre ellas la mejor aproximación fue propuesta por C. G. Broyden (1969), R. Fletcher (1970), D. Goldfarb (1970) y D. F. Shanno (1970), conocido como el método BFGS quasi-Newton, y cuya solución viene dada por la expresión

$$G^{(\xi+1)} = G^{(\xi)} + \frac{\Delta\Gamma^{(\xi+1,\xi)} (\Delta\Gamma^{(\xi+1,\xi)})^t}{(\Delta\Gamma^{(\xi+1,\xi)})^t \Delta W^{(\xi+1,\xi)}} - \frac{G^{(\xi)} \Delta W^{(\xi+1,\xi)} (\Delta W^{(\xi+1,\xi)})^t G^{(\xi)}}{(\Delta W^{(\xi+1,\xi)})^t G^{(\xi)} \Delta W^{(\xi+1,\xi)}}.$$

A partir de esta aproximación para la matriz inversa del Hessiano, el vector de pesos sinápticos en la iteración $(\xi + 1)$ se actualiza mediante la ecuación

$$W^{(\xi+1)} = W^{(\xi)} + \alpha^{(\xi)} G^{(\xi)} \Gamma^{(\xi)},$$

siendo $\alpha^{(\xi)}$ la dirección opuesta al gradiente del error cuadrático total E durante la iteración ξ .

- **Levenberg-Marquardt (LM).**

Al igual que los métodos quasi-Newton descritos anteriormente, este algoritmo emplea una aproximación de la matriz Hessiana y de su inversa.

De la aproximación de Taylor de primer orden para el vector de los errores, se deduce que

$$e^{(\xi+1)} = e^{(\xi)} + \nabla E(W^{(\xi)}) (W^{(\xi+1)} - W^{(\xi)}),$$

por lo que la función error cuadrático puede expresarse como

$$E(W^{(\xi+1)}) = \frac{1}{2} \|e^{(\xi+1)}\|^2 = \frac{1}{2} \|e^{(\xi)} + \nabla E(W^{(\xi)}) (W^{(\xi+1)} - W^{(\xi)})\|^2.$$

Minimizando este error respecto al vector de pesos sinápticos $W^{(\xi+1)}$, se obtiene que

$$W^{(\xi+1)} = W^{(\xi)} - \left[(\nabla E(W^{(\xi)}))^t (\nabla E(W^{(\xi)})) \right]^{-1} (\nabla E(W^{(\xi)}))^t e^{(\xi)},$$

por lo que el método de Levenberg-Marquardt (K. Levenberg (1944), D. W. Marquardt (1963)) aproxima la matriz Hessiana por el producto de las matrices que contienen las primeras derivadas de los errores respecto a los pesos sinápticos, esto es,

$$H \simeq (\nabla E(W^{(\xi)}))^t (\nabla E(W^{(\xi)})).$$

- **Resilient propagation (RP).**

Este método fue propuesto por M. Riedmiller y H. Braun (1993) para acelerar la convergencia del algoritmo de aprendizaje BP a través de valores de la tasa de aprendizaje adaptativos, esto es, que toman distintos valores según el peso sináptico al que acompaña.

Sea $w_{ij}^{(\xi)}$ un peso sináptico cualquiera de la red. Entonces, el algoritmo BP modifica el peso sináptico de acuerdo a la regla

$$w_{ij}^{(\xi+1)} = w_{ij}^{(\xi)} - \eta \left(\frac{\partial E(w_{ij}^{(\xi)})}{\partial w_{ij}} \Big|_{w_{ij}=w_{ij}^{(\xi)}} \right).$$

Denotando a $\left(\frac{\partial E(w_{ij}^{(\xi)})}{\partial w_{ij}} \Big|_{w_{ij}=w_{ij}^{(\xi)}} \right)$ por $\nabla E(w_{ij}^{(\xi)})$,

$$w_{ij}^{(\xi+1)} = w_{ij}^{(\xi)} - \eta \nabla E(w_{ij}^{(\xi)}) = w_{ij}^{(\xi)} - \Delta w_{ij}^{(\xi+1)}.$$

Se introducen los valores adaptadores de los pesos,

$$\delta_{ij}^{(\xi)}, \xi = 0, \dots, n-1,$$

que se inicializarán por $\delta_{ij}^{(0)} = 0.5$.

Durante el proceso de entrenamiento de la red los valores adaptadores de los pesos serán actualizados mediante las siguientes expresiones:

$$\begin{cases} \delta_{ij}^{(\xi+1)} = \eta^+ \delta_{ij}^{(\xi)}, & \text{si } \nabla E(w_{ij}^{(\xi+1)}) \nabla E(w_{ij}^{(\xi)}) > 0 \\ \delta_{ij}^{(\xi+1)} = \eta^- \delta_{ij}^{(\xi)}, & \text{si } \nabla E(w_{ij}^{(\xi+1)}) \nabla E(w_{ij}^{(\xi)}) < 0 \\ \delta_{ij}^{(\xi+1)} = \delta_{ij}^{(\xi)}, & \text{en otro caso} \end{cases},$$

donde $0 < \eta^- < 1 < \eta^+$.

Y los pesos sinápticos serán modificados según la siguiente relación

$$\begin{cases} \Delta w_{ij}^{(\xi+1)} = \text{signo} \left(\nabla E(w_{ij}^{(\xi+1)}) \right) \delta_{ij}^{(\xi+1)}, & \text{si } \nabla E(w_{ij}^{(\xi+1)}) \nabla E(w_{ij}^{(\xi)}) > 0 \\ \Delta w_{ij}^{(\xi+1)} = -\text{signo} \left(\nabla E(w_{ij}^{(\xi+1)}) \right) \delta_{ij}^{(\xi+1)}, & \text{si } \nabla E(w_{ij}^{(\xi+1)}) \nabla E(w_{ij}^{(\xi)}) < 0 \\ \Delta w_{ij}^{(\xi+1)} = 0, & \text{si } \nabla E(w_{ij}^{(\xi+1)}) \nabla E(w_{ij}^{(\xi)}) = 0 \end{cases}.$$

5.2 IMPLEMENTACIÓN Y DISEÑO EXPERIMENTAL.

En este trabajo se han construido dos perceptrones múltiples diferentes según la información de entrada empleada, que previamente ha sido preprocesada para normalizar sus valores. Ese procesamiento de los datos se ha realizado mediante un programa en MATLAB, cuyo código aparece en el Anexo 1.

Además de los conjuntos disjuntos de entrenamiento y test, se ha considerado un conjunto de validación para determinar cuándo se debe finalizar el proceso de entrenamiento, intentando evitar así la especialización de la red de neuronas artificiales en los ejemplos del conjunto de entrenamiento. La selección de los conjuntos disjuntos de entrenamiento, validación y test se ha llevado a cabo de forma aleatoria, tomando el 50%, 25% y 25% de los datos para cada uno de estos conjuntos respectivamente. Esta selección aleatoria de dichos conjuntos se ha repetido 20 veces para obtener una medida más robusta del rendimiento de las redes construidas.

Las redes de neuronas artificiales han sido entrenadas mediante los cinco métodos distintos de entrenamiento supervisado de retropropagación que se han indicado en el apartado anterior: BFGS quasi-Newton (BFGS), Fletcher-Reeves conjugate gradient (CGFR), Levenberg-Marquardt (LM), Gradient descent w/momentum & adaptive lr (GD) y Resilient propagation (RP). La selección de estos métodos de entrenamiento se ha hecho de forma representativa de cada una de las diversas tendencias existentes en la actualidad.

Con respecto al número de neuronas que constituyen la capa oculta, ante la inexistencia de una regla de selección válida, se han considerado varios tamaños de la capa oculta con el fin de determinar la estructura óptima de la RNA.

Como funciones de activación se han empleado la función logística para la capa oculta y la función identidad para la capa de salida, elecciones justificadas por la propiedad de aproximador universal.

El criterio escogido para determinar la selección de los parámetros óptimos consiste en minimizar el error cuadrático medio sobre el conjunto test.

Por último, los pesos de cada una de las capas de las RNAS han sido inicializados de forma aleatoria mediante la función de Nguyen-Widrow, y estas inicializaciones han sido repetidas 20 veces para evitar el efecto de dicha aleatoriedad en los resultados de la red.

A continuación se describe la regla de inicialización de pesos proporcionada por D. Nguyen y B. Widrow (1990), que es la empleada en este trabajo.

Definición 5.2.1.

Sea un MLP con m el número de neuronas de entrada y h el número de neuronas que constituye la capa oculta. Entonces, la regla de Nguyen-Widrow indica que los pesos sinápticos iniciales son escogidos según el siguiente algoritmo:

1. Sea $\beta = 0.7 \left(h^{\frac{1}{m}} \right)$.
2. Generar de forma independiente

$$\{w_{ik}, \quad i = 0, \dots, m, \quad k = 1, \dots, h\},$$

según una distribución $U(-1, 1)$.

3. Hallar los pesos iniciales, $\{w_{ik}^{(0)}, \quad i = 0, \dots, m, \quad k = 1, \dots, h\}$, según la expresión

$$w_{ik}^{(0)} = \left(\frac{\beta}{\|w_{01}, \dots, w_{0h}, w_{11}, \dots, w_{1h}, \dots, w_{m1}, \dots, w_{mh}\|} \right) w_{ik}.$$

5.3 MLP DE UNA CAPA CON DOS PREDICTORES.

La primera red que se ha construido, y que será referida como red_ algoritmo, consiste en un perceptrón múltiple con una única capa oculta.

La capa de entrada está formada por dos neuronas, correspondientes cada una de ellas a las variables "Temperatura del Agua" y "Profundidad", de forma similar a la información de la que se dispone para obtener las predicciones de la velocidad de sonido por los métodos de cálculo clásicos con datos procedentes de XBT.

El número de neuronas de la capa de salida es una, correspondiente a la variable "Velocidad de Sonido en el Agua del Mar", que constituye la variable que se pretende predecir, y que será referida como velocidad de sonido.

Con respecto al número de neuronas que constituyen la capa oculta, ante la inexistencia de una regla de selección válida, se han considerado cinco

tamaños de la capa oculta diferentes (2, 5, 10, 20 y 30), con el fin de determinar la estructura óptima de la red_ algoritmo.

Esta red_ algoritmo así descrita ha sido ejecutada en el paquete matemático MATLAB 5.3, empleando para ello el módulo denominado Neural Network Toolbox 3.0. El código del programa implementado aparece en el Anexo 1.

Se han realizado 20 iteraciones del programa, cada una de las cuales corresponde a las distintas selecciones de los conjuntos de entrenamiento, validación y test, obteniendo de esta manera $20 \times 20 \times 5 \times 5$ redes entrenadas, cada una de ellas con los diferentes métodos de entrenamiento y tamaños de capa oculta considerados, de entre las que se deducirá la red_ algoritmo óptima con respecto al criterio de optimalidad ya mencionado con anterioridad de minimización del error cuadrático medio.

La descripción y el análisis de los resultados obtenidos con este procedimiento ha sido realizado mediante representaciones gráficas y diversos contrastes de hipótesis ejecutados en el paquete estadístico SPSS 8.0.

5.3.1 RESUMEN DE LOS RESULTADOS.

Para determinar el método de entrenamiento y el tamaño de la capa oculta óptimos, se han analizado los estadísticos de las medias de los errores cuadráticos medios (ecm) de las 20 iteraciones realizadas sobre cada uno de los tres conjuntos de datos considerados, esto es, sobre los conjuntos de entrenamiento, validación y test, cuyos valores para cada una de las iteraciones aparecen en el Anexo 2.

El análisis ha sido realizado sobre cada uno de los conjuntos por separado, observándose que, independientemente del tamaño de la capa oculta, el mejor método de entrenamiento es el de Levenberg-Marquardt, a continuación el de BFGS quasi-Newton, el tercero es el de Fletcher-Reeves conjugate gradient, el cuarto el de Resilient propagation y por último el método GD. Estos resultados quedan reflejados en las representaciones gráficas de los estadísticos media, mínimo y máximo de las medias de los ecm para cada iteración, que aparecen en las Figuras 5.3.1, 5.3.2 y 5.3.3.

De estos gráficos cabe destacar además la estabilidad del método de Levenberg-Marquardt backpropagation con respecto a cada iteración, esto es, que la selección de los conjuntos de entrenamiento, validación y test no afecta de forma significativa a los resultados obtenidos al entrenar la red mediante este método. En cambio, puede observarse que el resto de métodos de

entrenamiento empleados sí que se ven afectados por dicha selección, especialmente el de Resilient propagation, lo que queda claramente reflejado en los resultados obtenidos con este método para la iteración número 7.

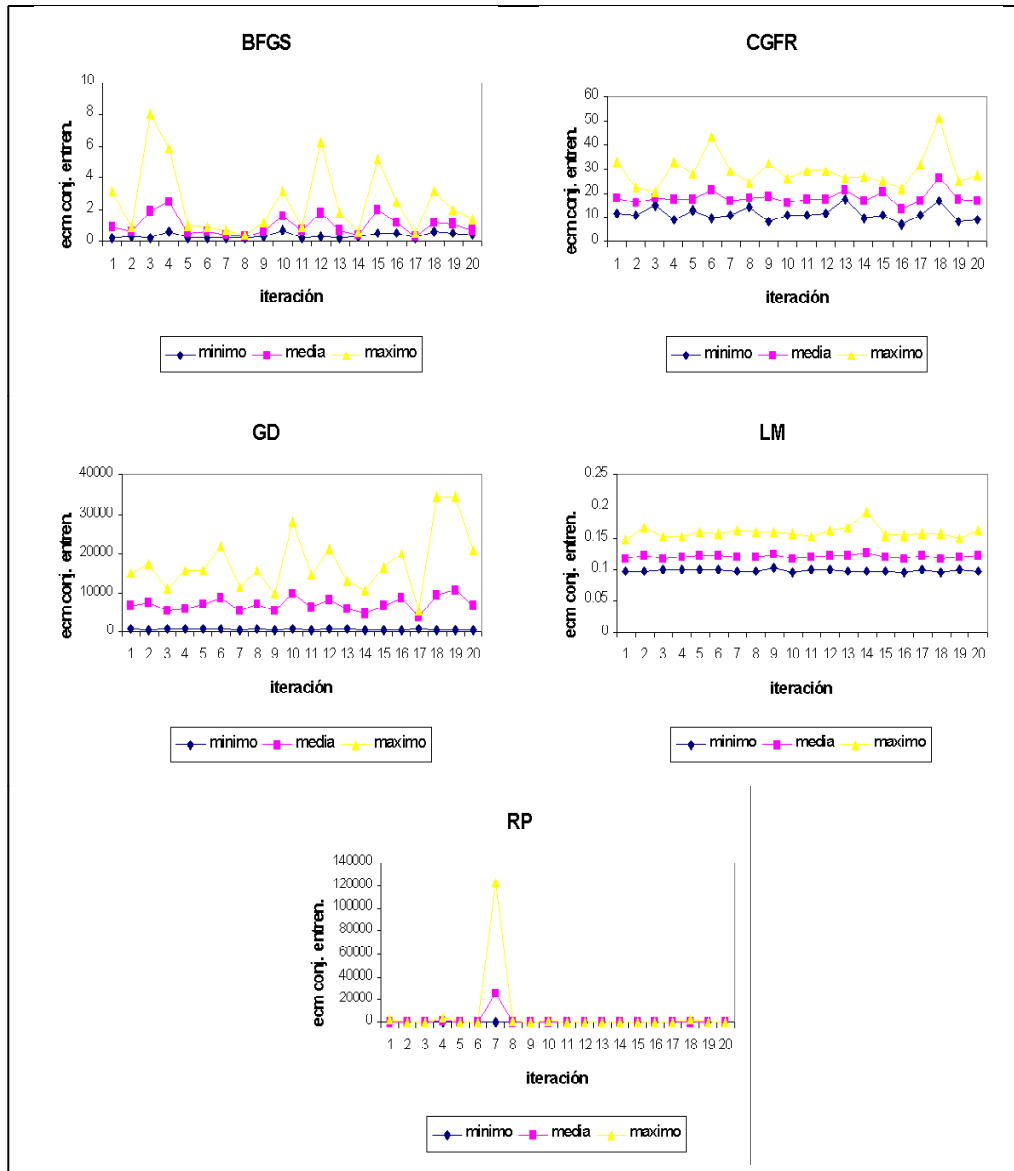


Figura 5.3.1. Representaciones gráficas de los estadísticos de los ecm para cada iteración sobre el conjunto de entrenamiento.

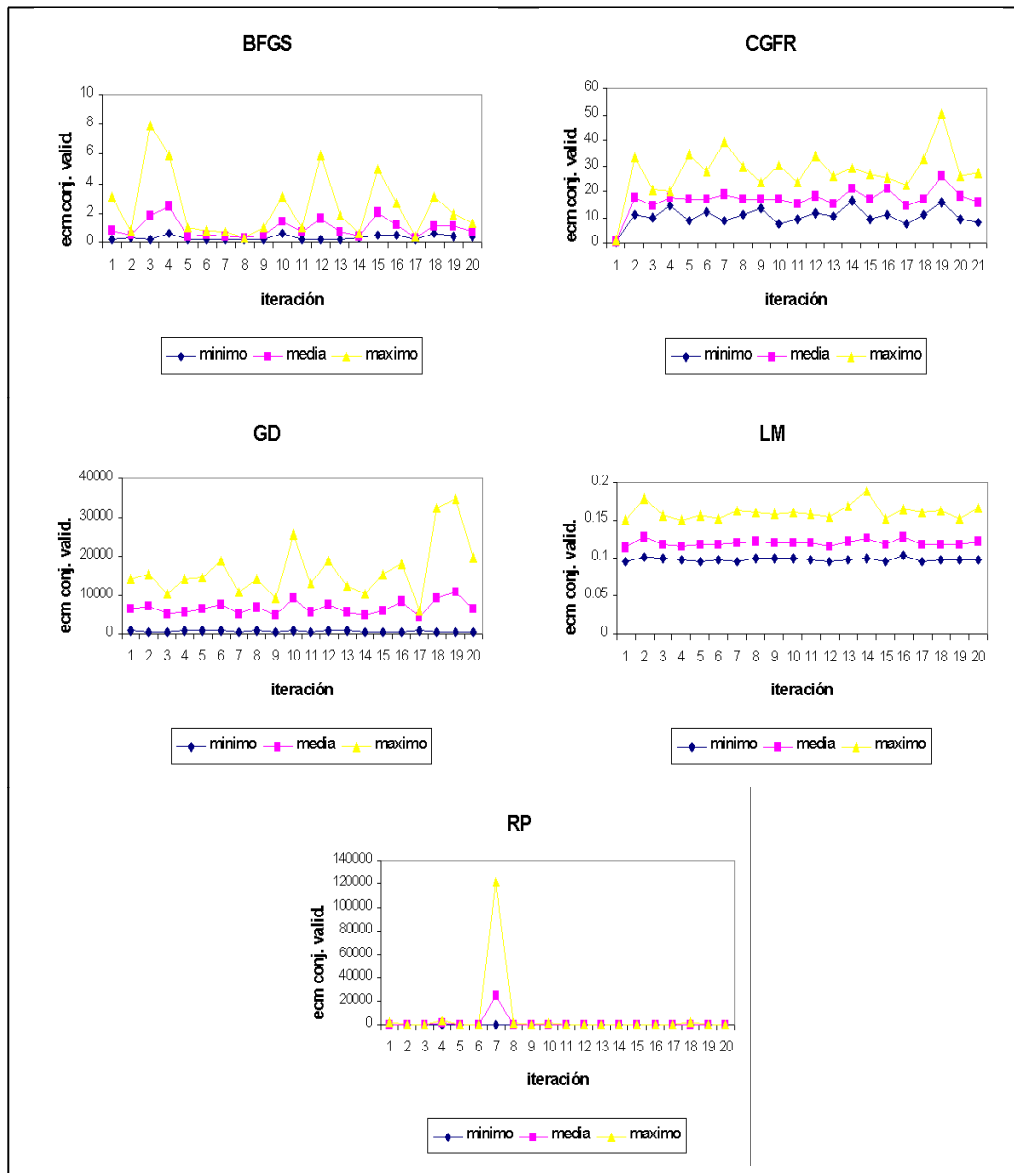


Figura 5.3.2. Representaciones gráficas de los estadísticos de los ecm para cada iteración sobre el conjunto de validación.

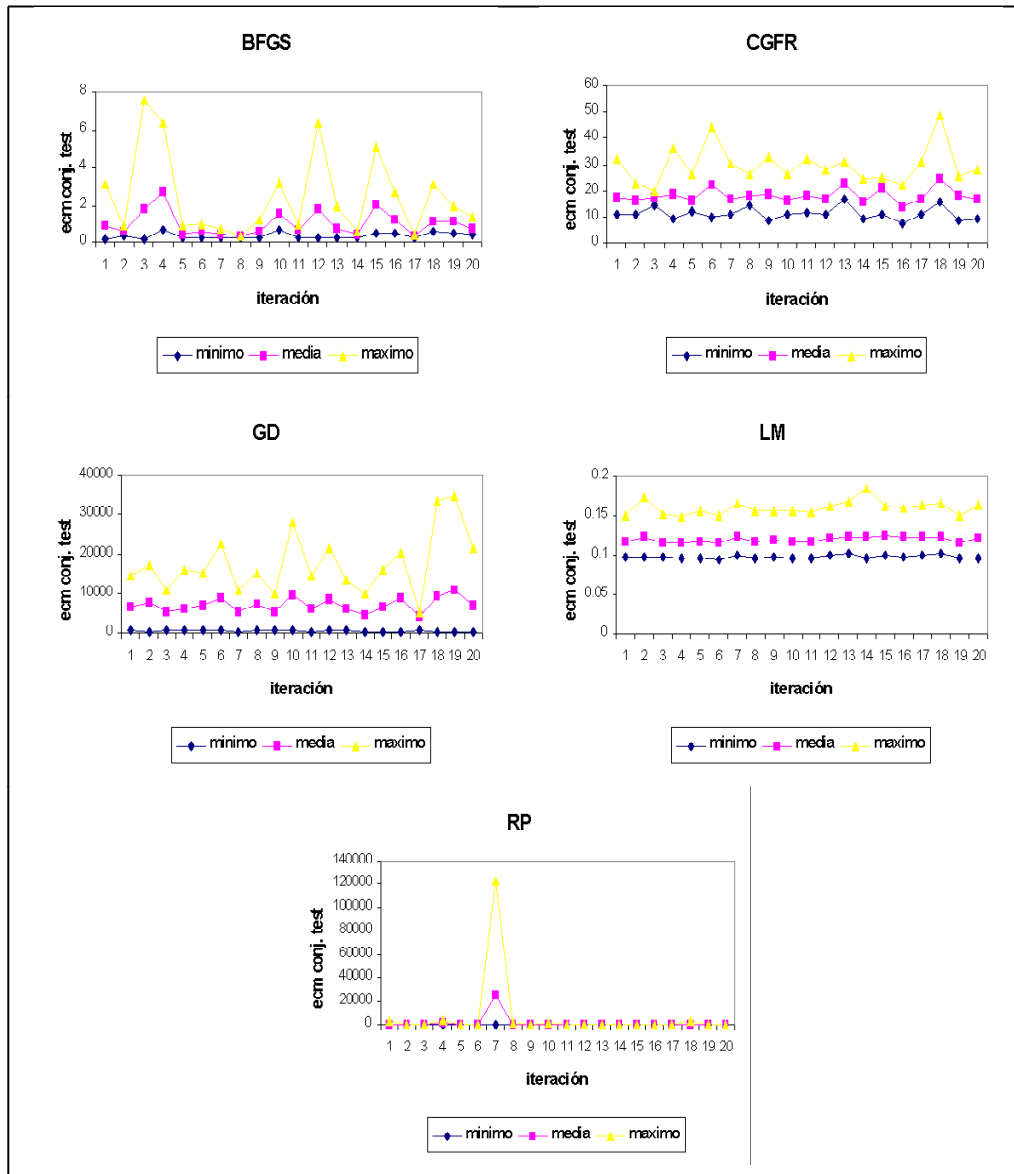


Figura 5.3.3. Representaciones gráficas de los estadísticos de los ecm para cada iteración sobre el conjunto test.

Varios estadísticos descriptivos han sido calculados para cada uno de los conjuntos de datos teniendo en cuenta además el tamaño de la capa oculta. Sus valores aparecen en las Tablas 5.3.1, 5.3.2 y 5.3.3, y en ellas puede observarse que el método LM es el mejor de todos los métodos de entrenamiento para todos los tamaños de la capa oculta considerados, presentando los resultados óptimos, esto es, los mínimos ecm para el tamaño de capa oculta 30.

En estas tablas también puede observarse que el segundo mejor método para todos los tamaños de capa oculta, es el BFGS, observándose unos resultados aparentemente peores para el tamaño de capa oculta 2, mientras que para el resto de tamaños los resultados parecen similares.

El tercer mejor método para todos los tamaños de capa oculta, según los estadísticos contenidos en las Tablas 5.3.1, 5.3.2 y 5.3.3, es el CGFR, donde la tendencia de los resultados con respecto al tamaño de la capa oculta se invierte, ya que con el tamaño de capa oculta 30 se obtienen los peores resultados, mientras que para el resto de tamaños los resultados parecen similares.

Según estas tablas, en el cuarto lugar se encuentra el método RP, cuyos resultados parecen que se pueden considerar similares para todos los tamaños de capa oculta considerados.

Y el peor de los métodos de entrenamiento empleados es el GD, cuyos resultados van empeorando progresivamente conforme aumenta el número de neuronas consideradas en la capa oculta.

Todas estas relaciones entre los distintos métodos y los tamaños de capa oculta considerados, son analizadas mediante gráficos del tipo de caja y bigotes, cuyas representaciones gráficas para cada uno de los conjuntos de datos aparecen en las Figuras 5.3.4, 5.3.5 y 5.3.6, y mediante contrastes de igualdad de medias.

A partir de modelos de análisis de la varianza de un factor, en este caso el tamaño de capa oculta, con medidas repetidas, se analizará la igualdad de las medias de los errores cuadráticos medios entre los tamaños de la capa oculta para cada uno de los métodos de entrenamiento empleados.

En estos gráficos de caja y bigotes se puede observar un comportamiento similar de los métodos de entrenamiento empleados con respecto a los conjuntos de entrenamiento, validación y test, sin embargo, los tests de igualdad de medias de los ecm entre los diversos tamaños de la capa oculta han sido realizados sobre los tres conjuntos.

Método	Tamaño	Media	Mínimo	Máximo	Desv. típica
BFGS	2	2.079885	0.206010	8.027643	2.466027
	5	0.583844	0.187260	2.723532	0.571546
	10	0.732797	0.263085	2.521691	0.633258
	20	0.565060	0.261563	1.089011	0.244092
	30	0.950576	0.363120	2.453343	0.591795
CGFR	2	15.956892	9.866445	25.097850	4.405632
	5	16.129047	8.567474	27.700045	4.994676
	10	15.451058	6.851852	26.616727	5.544003
	20	14.440501	8.499698	22.393097	4.072152
	30	27.858975	10.533231	51.264754	8.888234
GD	2	613.75465	407.62719	915.73127	152.67958
	5	1883.58687	1194.32611	3811.86449	603.43433
	10	4735.17886	2483.64828	10581.87900	1863.08773
	20	11166.99449	4784.19051	28388.10808	5135.40830
	30	5135.40830	5566.58043	34513.03553	7509.39307
LM	2	0.159812	0.148351	0.190656	0.009066
	5	0.125439	0.117636	0.132346	0.004201
	10	0.113583	0.107405	0.119373	0.003288
	20	0.102090	0.097199	0.104891	0.002152
	30	0.097775	0.094018	0.101469	0.001966
RP	2	130.689043	17.582145	445.770447	124.298046
	5	76.210558	16.413063	339.565637	90.199354
	10	6210.96353	10.391128	122596.5074	27394.7603
	20	314.819989	13.336497	2270.796751	709.131660
	30	417.978312	23.840628	4037.730118	1061.62069

Tabla 5.3.1. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de entrenamiento.

Método	Tamaño	Q ₁	Mediana (Q ₂)	Q ₃
BFGS	2	0.22183878	0.75169590	3.15990485
	5	0.22967042	0.46864002	0.66421833
	10	0.40937461	0.51725539	0.77428195
	20	0.34476685	0.50201483	0.71895124
	30	0.58058219	0.74571397	1.11752063
CGFR	2	12.3747009	14.4145133	19.9573711
	5	11.4652555	16.4491810	19.4237120
	10	10.9804314	14.5485330	18.9651415
	20	11.0203068	13.8964809	18.1449956
	30	21.7532652	27.5359874	32.3758795
GD	2	492.257136	586.821267	739.391437
	5	1456.72677	1747.12738	2184.53433
	10	3028.20375	5080.30526	5436.31387
	20	7962.06873	10623.9181	12670.6980
	30	10574.9389	15348.1802	20683.9089
LM	2	0.15395315	0.15818292	0.16304014
	5	0.12174238	0.12576299	0.12832705
	10	0.11026152	0.11414634	0.11580218
	20	0.10162370	0.10253612	0.10370067
	30	0.09647883	0.09772164	0.09931909
RP	2	29.3460841	81.4301735	209.123277
	5	24.3004011	41.6596217	61.5634116
	10	18.6959681	36.2688839	62.8622824
	20	21.5311537	31.7340917	65.6510481
	30	36.8433980	64.7482492	114.432384

Tabla 5.3.1-Cont. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de entrenamiento.

Método	Tamaño	Media	Mínimo	Máximo	Desv. típica
BFGS	2	2.053119	0.209907	7.876769	2.406347
	5	0.591619	0.188189	2.767408	0.588828
	10	0.739363	0.264124	2.748014	0.660812
	20	0.579694	0.244885	1.262765	0.277855
	30	0.945640	0.357285	2.914519	0.624137
CGFR	2	15.812265	9.203241	26.332065	4.549845
	5	15.908529	9.334646	27.207929	4.739755
	10	15.345722	7.651024	27.010559	5.651250
	20	14.478810	7.548322	25.451469	4.672294
	30	27.750346	11.032878	50.640510	8.567473
GD	2	603.2179	395.7002	913.5566	149.1553
	5	1846.0386	1150.8763	3851.8843	612.2881
	10	4481.2393	2257.0801	10432.445	1853.1095
	20	10327.188	4920.7216	25418.218	4509.6765
	30	15313.107	5709.6803	34328.237	7234.4594
LM	2	0.160279	0.149029	0.188445	0.009538
	5	0.125727	0.119777	0.137722	0.004683
	10	0.113889	0.107503	0.124678	0.003932
	20	0.102467	0.098140	0.108537	0.002743
	30	0.098175	0.095190	0.104384	0.002359
RP	2	129.9068	17.921064	441.9992	123.3616
	5	75.69551	15.596556	341.2734	90.2513
	10	6192.2417	10.821366	122230.79	27313.081
	20	315.7932	13.512031	2262.2210	708.5213
	30	418.8426	22.918739	4044.6113	1061.6427

Tabla 5.3.2. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de validación.

Método	Tamaño	Q ₁	Mediana (Q ₂)	Q ₃
BFGS	2	0.22227101	0.77429941	3.16613611
	5	0.23204429	0.46460957	0.67231304
	10	0.39803209	0.53369530	0.80595629
	20	0.34132755	0.50636180	0.71577718
	30	0.56422657	0.77522834	1.02540331
CGFR	2	12.5117626	14.1836331	19.6902689
	5	11.2962842	16.2783198	19.0143483
	10	10.9901439	14.1380911	17.8693637
	20	10.6453544	13.9435348	18.0192221
	30	21.6704628	27.3563811	33.1265032
GD	2	486.196441	579.519001	731.841981
	5	1467.76768	1651.35679	2129.55114
	10	2835.48545	4754.99236	5397.00877
	20	7513.56870	10087.9608	11506.8549
	30	10222.7348	14049.0609	18646.0622
LM	2	0.15305726	0.15889139	0.16357620
	5	0.12190616	0.12576158	0.12769342
	10	0.11222551	0.11345639	0.11559859
	20	0.10020390	0.10234366	0.10406676
	30	0.09670298	0.09773225	0.09937047
RP	2	28.8380204	82.2199494	205.414811
	5	23.9661888	41.7428841	60.0283020
	10	19.6014360	35.6756141	63.5903698
	20	20.7988383	30.5072103	81.7402333
	30	37.7061723	66.5443274	121.377337

Tabla 5.3.2-Cont. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de validación.

Método	Tamaño	Media	Mínimo	Máximo	Desv. típica
BFGS	2	2.085490	0.210301	7.583023	2.458155
	5	0.608913	0.182983	2.997187	0.636498
	10	0.752222	0.262278	2.691029	0.667669
	20	0.585127	0.258153	1.202936	0.272542
	30	0.979686	0.362573	2.710638	0.643501
CGFR	2	15.963693	10.121144	25.375573	4.242435
	5	16.227560	8.980934	28.144016	5.131644
	10	15.472519	7.436107	25.463950	5.336364
	20	14.726125	8.674837	23.946034	4.413507
	30	28.239125	11.397162	48.930873	8.561860
GD	2	616.07877	396.00179	933.1881	158.3418
	5	1901.7607	1162.8445	3901.0212	631.6482
	10	4762.7876	2537.0154	10656.373	1900.5144
	20	11191.6209	4858.1311	28165.4555	5122.3808
	30	16563.5907	5554.1377	34461.8542	7504.3165
LM	2	0.159545	0.147620	0.185237	0.009193
	5	0.125322	0.117304	0.132965	0.004398
	10	0.113638	0.109494	0.125171	0.003962
	20	0.102137	0.098584	0.108270	0.002537
	30	0.097747	0.094021	0.101590	0.002117
RP	2	131.0819	17.645659	446.0769	124.9554
	5	76.475127	16.072469	335.9897	89.5334
	10	6210.17333	10.591104	122569.524	27388.597
	20	314.1774	13.441916	2265.2883	706.6109
	30	419.8151	25.067681	4068.5149	1066.1324

Tabla 5.3.3. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto test.

Método	Tamaño	Q ₁	Mediana (Q ₂)	Q ₃
BFGS	2	0.21938666	0.78901341	3.15165047
	5	0.23006104	0.47427215	0.66498302
	10	0.41188739	0.52466311	0.81725340
	20	0.33565953	0.51635793	0.75941233
	30	0.55634554	0.77428921	1.14061141
CGFR	2	12.4905995	14.4902527	19.1650024
	5	11.2785155	15.7480955	19.5082900
	10	11.4752460	14.3745126	19.1129413
	20	11.2148508	13.6118768	18.3752909
	30	22.9174458	27.0376983	31.9303257
GD	2	487.045255	582.961068	743.881362
	5	1424.24295	1766.36770	2176.58943
	10	2951.50066	5003.37930	5583.52320
	20	7849.88878	10824.4245	13073.1862
	30	10323.7335	14971.2805	21191.8328
LM	2	0.15175236	0.15755192	0.16467963
	5	0.12142089	0.12573213	0.12907633
	10	0.11069225	0.11233395	0.11634586
	20	0.10058358	0.10203175	0.10325197
	30	0.09587838	0.09751468	0.09981897
RP	2	28.5088618	81.0687638	214.850213
	5	24.1957474	42.0801242	61.3343798
	10	19.4587877	35.9096671	61.2814194
	20	21.2021738	30.6209445	74.1017724
	30	36.9781763	65.9490354	115.628022

Tabla 5.3.3-Cont. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto test.

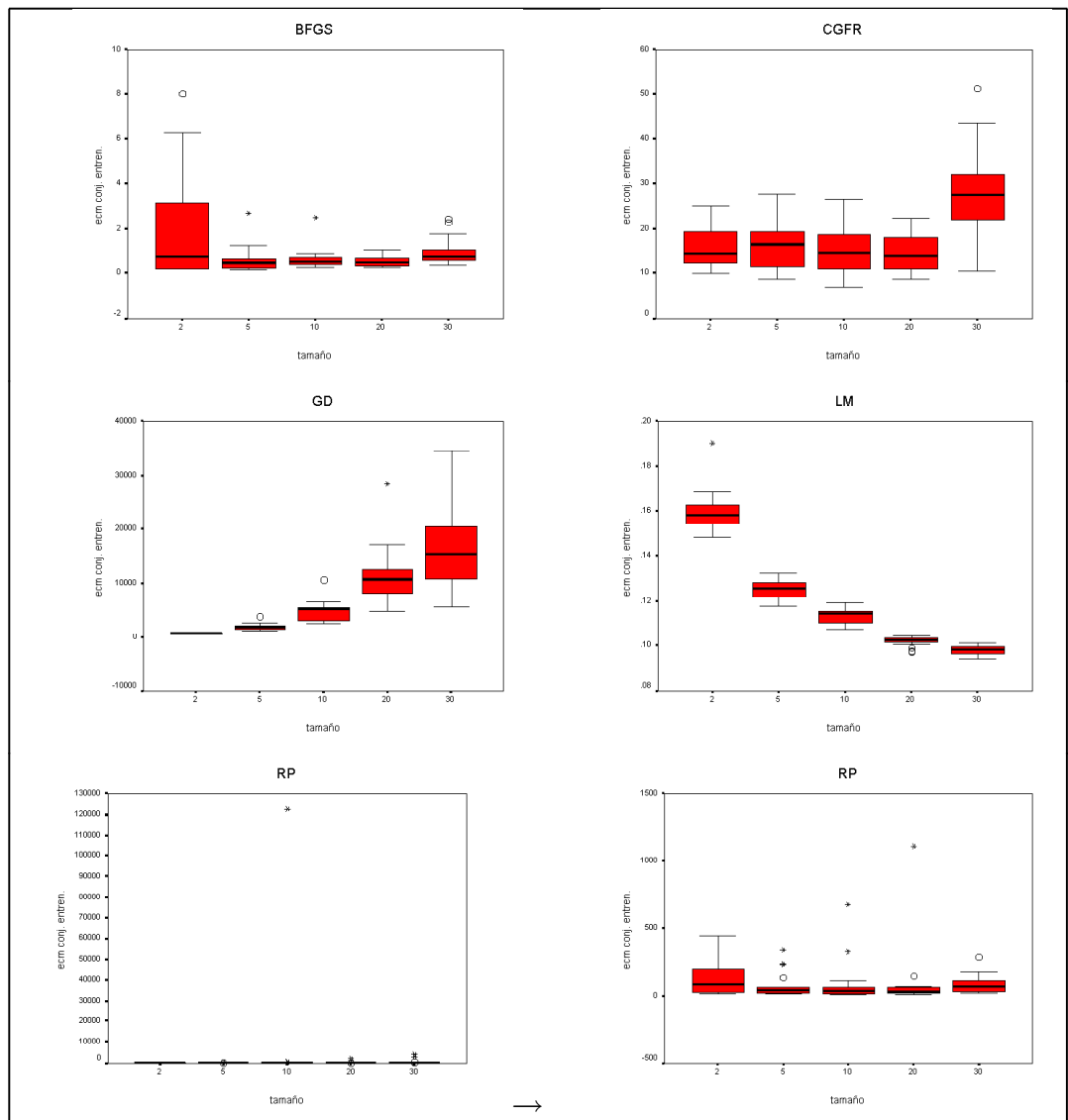


Figura 5.3.4. Gráficos de caja y bigotes de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de entrenamiento.

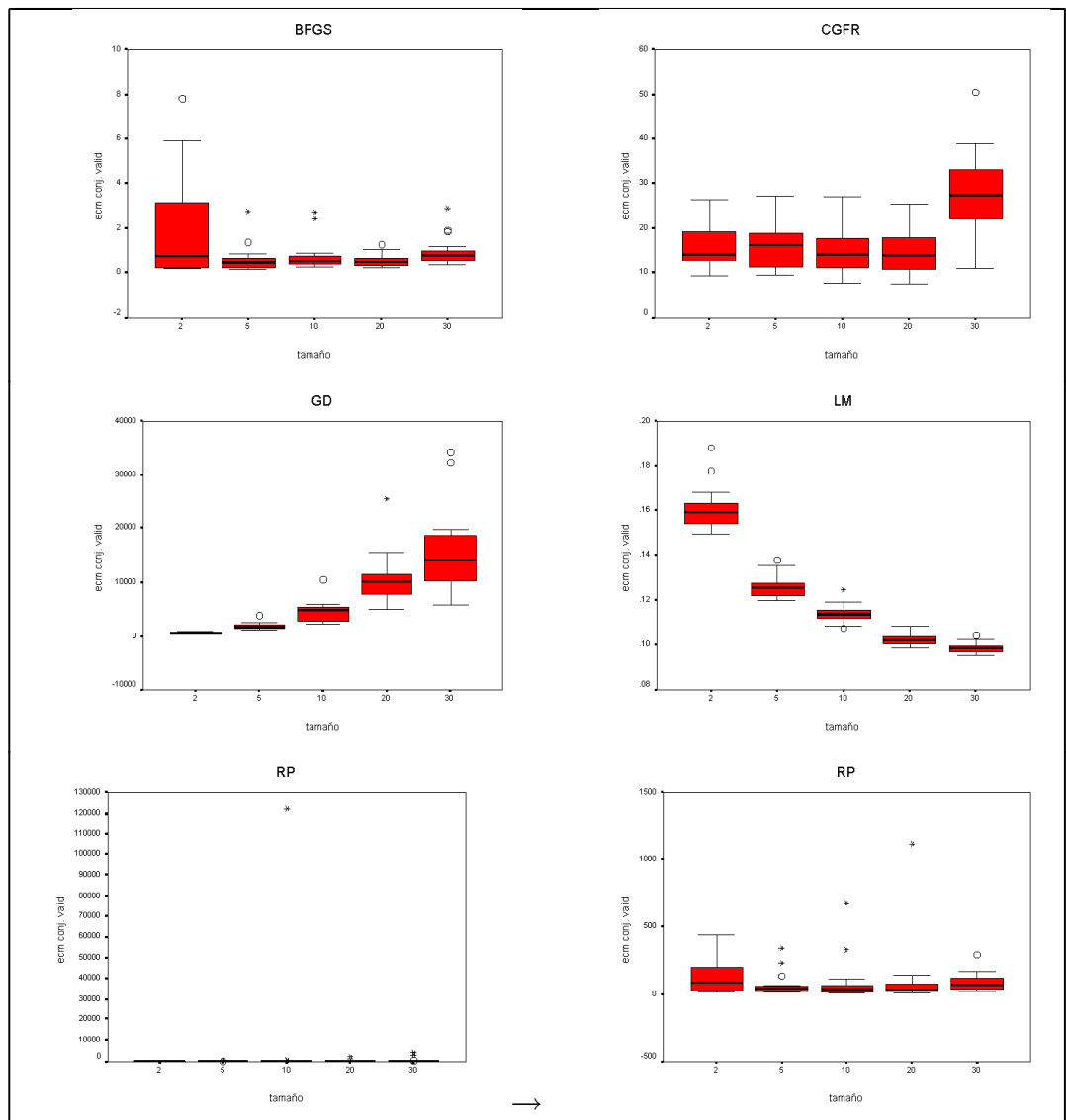


Figura 5.3.5. Gráficos de caja y bigotes de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de validación.

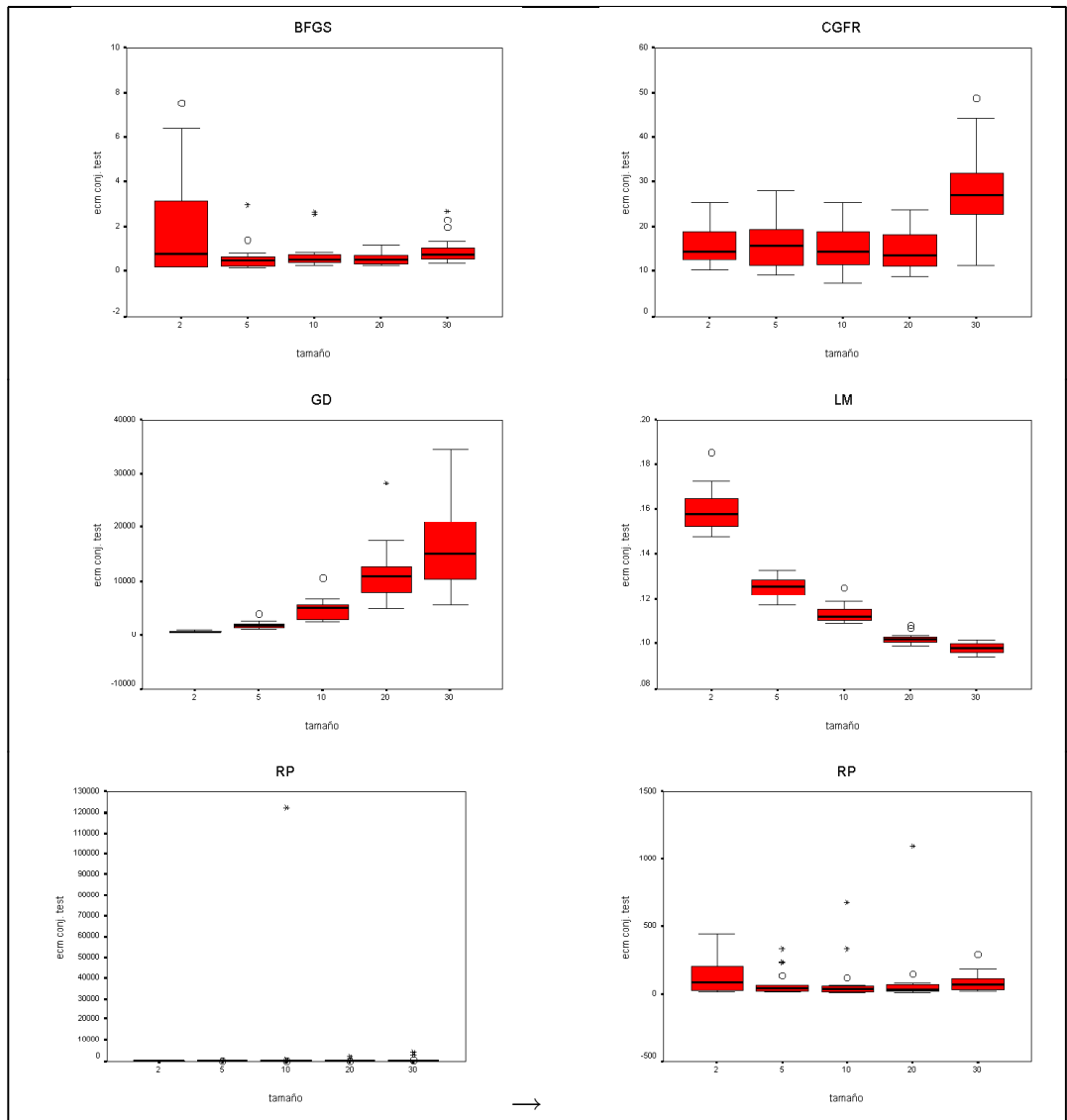


Figura 5.3.6. Gráficos de caja y bigotes de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto test.

5.3.2 ANÁLISIS DE LOS RESULTADOS.

En los modelos de medidas repetidas es necesario suponer que las diferencias entre cada dos niveles del factor son iguales, lo que es equivalente a afirmar que la matriz de varianzas-covarianzas es circular o esférica (véase R. E. Kirk (1982) o B. J. Winer, D. R. Brown y K. M. Michels (1991)). Esta hipótesis será contrastada mediante la prueba de esfericidad de Mauchly (J. W. Mauchly (1940)), y si dicha hipótesis es aceptada, entonces se recurre a emplear una aproximación mediante el estadístico F, pero si dicha hipótesis de esfericidad es rechazada, entonces se emplean una serie de estadísticos robustos ante dicha hipótesis, como son la Traza de Pillai, la Lambda de Wilks, la Traza de Hotelling y la Raíz mayor de Roy. Una descripción de dichos estadísticos puede consultarse en R. D. Bock (1975) o en B. G. Tabachnik y L. S. Fidel (1983).

Aunque el procedimiento de medidas repetidas no permite efectuar contrastes "post hoc" entre los niveles del factor, esto es, contrastes sobre la igualdad de las medias de los ecm para determinados tamaños de la capa oculta, para ver así cuál es la razón por la que se rechaza la hipótesis nula, lo que sí es posible realizar son comparaciones dos a dos entre ellos.

Por lo tanto, para cada uno de los métodos de entrenamiento se realizará primeramente el test de esfericidad de Mauchly, seguidamente se llevará a cabo el test de igualdad de medias de los ecm entre todos los tamaños de capa oculta, y finalmente los contrastes de igualdad de las medias de los ecm tomando los tamaños dos a dos y comparándolos.

Respecto al método de entrenamiento BFGS, sobre los tres conjuntos se rechaza la esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.3.4, por lo que se calculan los estadísticos robustos anteriormente citados. Los valores de dichos estadísticos aparecen en la Tabla 5.3.5, obteniéndose para un nivel de significación de 0.05 el mismo resultado de rechazo de la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos, circunstancia que quedaba claramente reflejada en sus correspondientes gráficos de caja y bigotes (Figuras 5.3.4, 5.3.5 y 5.3.6).

BFGS (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.007	87.547	9	0.000
ecm conj. valid.	0.009	82.473	9	0.000
ecm conj. test	0.010	80.769	9	0.000

Tabla 5.3.4. Prueba de esfericidad de Mauchly para el método de entrenamiento BFGS.

BFGS (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.474	4	16	0.028
	Lambda Wilks	0.526	4	16	0.028
	Traza Hotelling	0.903	4	16	0.028
	Raíz mayor Roy	0.903	4	16	0.028
ecm conj. valid.	Traza Pillai	0.462	4	16	0.033
	Lambda Wilks	0.538	4	16	0.033
	Traza Hotelling	0.860	4	16	0.033
	Raíz mayor Roy	0.860	4	16	0.033
ecm conj. test.	Traza Pillai	0.463	4	16	0.032
	Lambda Wilks	0.537	4	16	0.032
	Traza Hotelling	0.864	4	16	0.032
	Raíz mayor Roy	0.864	4	16	0.032

Tabla 5.3.5. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento BFGS.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento BFGS se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.3.6, 5.3.7 y 5.3.8, respectivamente.

Hay que tener en cuenta que el ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y que además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

Sobre el conjunto de entrenamiento se observa que, para un nivel de significación de 0.05, únicamente existen diferencias significativas entre los tamaños de capa oculta 20 y 30 (p-valor de 0.041), mientras que para los conjuntos de validación y test no existen diferencias significativas entre ninguno de los pares de tamaños de capa oculta, aunque los p-valores obtenidos para los tamaños de capa oculta 20 y 30, con valores 0.065 y 0.055 respectivamente, sólo superan ligeramente el nivel de significación establecido en 0.05.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	1.496	0.525	[-0.170, 3.162]	0.103
2 y 10	1.347	0.541	[-0.371, 3.065]	0.223
2 y 20	1.515	0.553	[-0.242, 3.271]	0.131
2 y 30	1.129	0.542	[-0.591, 2.850]	0.510
5 y 2	-1.496	0.525	[-3.162, 0.170]	0.103
5 y 10	-0.149	0.119	[-0.526, 0.228]	1.000
5 y 20	1.878E-02	0.116	[-0.350, 0.388]	1.000
5 y 30	-0.367	0.180	[-0.938, 0.205]	0.559
10 y 2	-1.347	0.541	[-3.065, 0.371]	0.223
10 y 5	0.149	0.119	[-0.228, 0.526]	1.000
10 y 20	0.168	0.141	[-0.279, 0.614]	1.000
10 y 30	-0.218	0.195	[-0.836, 0.401]	1.000
20 y 2	-1.515	0.553	[-3.271, 0.242]	0.131
20 y 5	-1.878E-02	0.116	[-0.388, 0.350]	1.000
20 y 10	-0.168	0.141	[-0.614, 0.279]	1.000
20 y 30	-0.386	0.118	[-0.760, -1.096E-02]	0.041
30 y 2	-1.129	0.542	[-2.850, 0.591]	0.510
30 y 5	0.367	0.180	[-0.205, 0.938]	0.559
30 y 10	0.218	0.195	[-0.401, 0.836]	1.000
30 y 20	0.386	0.118	[1.096E-02, 0.760]	0.041

Tabla 5.3.6. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento BFGS sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	1.462	0.509	[−0.154, 3.077]	0.098
2 y 10	1.314	0.531	[−0.373, 3.000]	0.230
2 y 20	1.473	0.539	[−0.238, 3.185]	0.132
2 y 30	1.107	0.530	[−0.574, 2.789]	0.503
5 y 2	− 1.462	0.509	[−3.077, 0.154]	0.098
5 y 10	− 0.148	0.131	[−0.563, 0.267]	1.000
5 y 20	1.193E-02	0.119	[−0.365, 0.389]	1.000
5 y 30	− 0.354	0.185	[−0.941, 0.233]	0.707
10 y 2	−1.314	0.531	[−3.000, 0.373]	0.230
10 y 5	0.148	0.131	[−0.267, 0.563]	1.000
10 y 20	0.160	0.144	[−0.296, 0.616]	1.000
10 y 30	−0.206	0.203	[−0.849, 0.437]	1.000
20 y 2	− 1.473	0.539	[−3.185, 0.238]	0.132
20 y 5	-1.193E-02	0.119	[−0.389, 0.365]	1.000
20 y 10	− 0.160	0.144	[−0.616, 0.296]	1.000
20 y 30	− 0.366	0.120	[−0.746, 1.390E-02]	0.065
30 y 2	−1.107	0.530	[−2.789, 0.574]	0.503
30 y 5	0.354	0.185	[−0.233, 0.941]	0.707
30 y 10	0.206	0.203	[−0.437, 0.849]	1.000
30 y 20	0.366	0.120	[−1.390E-02, 0.746]	0.065

Tabla 5.3.7. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento BFGS sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	1.477	0.515	[−0.158, 3.111]	0.099
2 y 10	1.333	0.535	[−0.366, 3.032]	0.222
2 y 20	1.500	0.550	[−0.244, 3.245]	0.133
2 y 30	1.106	0.540	[−0.609, 2.821]	0.548
5 y 2	− 1.477	0.515	[−3.111, 0.158]	0.099
5 y 10	− 0.143	0.131	[−0.559, 0.272]	1.000
5 y 20	2.378E-02	0.128	[−0.381, 0.429]	1.000
5 y 30	− 0.371	0.197	[−0.997, 0.256]	0.759
10 y 2	−1.333	0.535	[−3.032, 0.366]	0.222
10 y 5	0.143	0.131	[−0.272, 0.559]	1.000
10 y 20	0.167	0.146	[−0.296, 0.630]	1.000
10 y 30	−0.227	0.208	[−0.887, 0.432]	1.000
20 y 2	− 1.500	0.550	[−3.245, 0.244]	0.133
20 y 5	-2.378E-02	0.128	[−0.429, 0.381]	1.000
20 y 10	− 0.167	0.146	[−0.630, 0.296]	1.000
20 y 30	− 0.395	0.126	[−0.794, 5.120E-03]	0.055
30 y 2	−1.106	0.540	[−2.821, 0.609]	0.548
30 y 5	0.371	0.197	[−0.256, 0.997]	0.759
30 y 10	0.227	0.208	[−0.432, 0.887]	1.000
30 y 20	0.395	0.126	[−5.120E-03, 0.794]	0.055

Tabla 5.3.8. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento BFGS sobre el conjunto test.

En cuanto al método de entrenamiento CGFR, sobre los tres conjuntos se rechaza nuevamente la esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.3.9.

En los gráficos de caja y bigotes (Figuras 5.3.4, 5.3.5 y 5.3.6) vuelve a quedar claramente reflejada la diferencia entre las medias de los ecm para todos los tamaños de capa oculta, lo que también puede observarse en los tests realizados, cuyos valores aparecen en la Tabla 5.3.10, ya que se rechaza la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

CGFR (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.206	27.528	9	0.001
ecm conj. valid.	0.234	25.286	9	0.003
ecm conj. test	0.221	26.278	9	0.002

Tabla 5.3.9. Prueba de esfericidad de Mauchly para el método de entrenamiento CGFR.

CGFR (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.730	4	16	0.000
	Lambda Wilks	0.270	4	16	0.000
	Traza Hotelling	2.708	4	16	0.000
	Raíz mayor Roy	2.708	4	16	0.000
ecm conj. valid.	Traza Pillai	0.718	4	16	0.000
	Lambda Wilks	0.282	4	16	0.000
	Traza Hotelling	2.547	4	16	0.000
	Raíz mayor Roy	2.547	4	16	0.000
ecm conj. test.	Traza Pillai	0.726	4	16	0.000
	Lambda Wilks	0.274	4	16	0.000
	Traza Hotelling	2.651	4	16	0.000
	Raíz mayor Roy	2.651	4	16	0.000

Tabla 5.3.10. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento CGFR.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento CGFR se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.3.11, 5.3.12 y 5.3.13, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas

tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

Tal y como queda claramente reflejado en los gráficos de caja y bigotes, en estos tests de comparaciones dobles únicamente se observan diferencias significativas entre las medias de los ecm para el tamaño de capa oculta 30 y el resto de los tamaños, resultado que se obtiene sobre los tres conjuntos considerados.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-0.172	1.793	[-5.863, 5.518]	1.000
2 y 10	0.506	1.302	[-3.627, 4.639]	1.000
2 y 20	1.516	1.154	[-2.145, 5.178]	1.000
2 y 30	-11.902	2.599	[-20.150, -3.654]	0.002
5 y 2	0.172	1.793	[-5.518, 5.863]	1.000
5 y 10	0.678	1.946	[-5.499, 6.855]	1.000
5 y 20	1.689	1.571	[-3.297, 6.674]	1.000
5 y 30	-11.730	2.104	[-18.406, -5.054]	0.000
10 y 2	-0.506	1.302	[-4.639, 3.627]	1.000
10 y 5	-0.678	1.946	[-6.855, 5.499]	1.000
10 y 20	1.011	0.960	[-2.037, 4.059]	1.000
10 y 30	-12.408	2.195	[-19.375, -5.441]	0.000
20 y 2	-1.516	1.154	[-5.178, 2.145]	1.000
20 y 5	-1.689	1.571	[-6.674, 3.297]	1.000
20 y 10	-1.011	0.960	[-4.059, 2.037]	1.000
20 y 30	-13.418	2.150	[-20.243, -6.594]	0.000
30 y 2	11.902	2.599	[3.654, 20.150]	0.002
30 y 5	11.730	2.104	[5.054, 18.406]	0.000
30 y 10	12.408	2.195	[5.441, 19.375]	0.000
30 y 20	13.418	2.150	[6.594, 20.243]	0.000

Tabla 5.3.11. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento CGFR sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-9.626E-02	1.771	[-5.717, 5.525]	1.000
2 y 10	0.467	1.295	[-3.642, 4.576]	1.000
2 y 20	1.333	1.141	[-2.289, 4.956]	1.000
2 y 30	-11.938	2.508	[-19.899, -3.977]	0.001
5 y 2	9.626E-02	1.771	[-5.525, 5.717]	1.000
5 y 10	0.563	1.957	[-5.650, 6.775]	1.000
5 y 20	1.430	1.688	[-3.928, 6.788]	1.000
5 y 30	-11.842	2.128	[-18.594, -5.089]	0.000
10 y 2	-0.467	1.295	[-4.576, 3.642]	1.000
10 y 5	-0.563	1.957	[-6.775, 5.650]	1.000
10 y 20	0.867	0.942	[-2.123, 3.857]	1.000
10 y 30	-12.405	2.162	[-19.265, -5.544]	0.000
20 y 2	-1.333	1.141	[-4.956, 2.289]	1.000
20 y 5	-1.430	1.688	[-6.788, 3.928]	1.000
20 y 10	-0.867	0.942	[-3.857, 2.123]	1.000
20 y 30	-13.272	2.162	[-20.135, -6.409]	0.000
30 y 2	11.938	2.508	[3.977, 19.899]	0.001
30 y 5	11.842	2.128	[5.089, 18.594]	0.000
30 y 10	12.405	2.162	[5.544, 19.265]	0.000
30 y 20	13.272	2.162	[6.409, 20.135]	0.000

Tabla 5.3.12. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento CGFR sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-0.264	1.797	[-5.967, 5.439]	1.000
2 y 10	0.491	1.251	[-3.479, 4.462]	1.000
2 y 20	1.238	1.142	[-2.388, 4.863]	1.000
2 y 30	-12.275	2.492	[-20.184, -4.367]	0.001
5 y 2	0.264	1.797	[-5.439, 5.967]	1.000
5 y 10	0.755	1.953	[-5.443, 6.953]	1.000
5 y 20	1.501	1.696	[-3.881, 6.884]	1.000
5 y 30	-12.012	2.034	[-18.466, -5.557]	0.000
10 y 2	-0.491	1.251	[-4.462, 3.479]	1.000
10 y 5	-0.755	1.953	[-6.953, 5.443]	1.000
10 y 20	0.746	0.920	[-2.174, 3.667]	1.000
10 y 30	-12.767	2.184	[-19.697, -5.836]	0.000
20 y 2	-1.238	1.142	[-4.863, 2.388]	1.000
20 y 5	-1.501	1.696	[-6.884, 3.881]	1.000
20 y 10	-0.746	0.920	[-3.667, 2.174]	1.000
20 y 30	-13.513	2.192	[-20.471, -6.555]	0.000
30 y 2	12.275	2.492	[4.367, 20.184]	0.001
30 y 5	12.012	2.034	[5.557, 18.466]	0.000
30 y 10	12.767	2.184	[5.836, 19.697]	0.000
30 y 20	13.513	2.192	[6.555, 20.471]	0.000

Tabla 5.3.13. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento CGFR sobre el conjunto test.

En el caso del método de entrenamiento GD, sobre los tres conjuntos se rechaza la hipótesis de esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.3.14.

En los gráficos de caja y bigotes (Figuras 5.3.4, 5.3.5 y 5.3.6) se ve claramente la diferencia entre las medias de los ecm para todos los tamaños de capa oculta, lo que también puede observarse en los tests realizados, cuyos valores aparecen en la Tabla 5.3.15, ya que se rechaza la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

GD (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.003	102.970	9	0.000
ecm conj. valid.	0.003	99.417	9	0.000
ecm conj. test	0.003	100.492	9	0.000

Tabla 5.3.14. Prueba de esfericidad de Mauchly para el método de entrenamiento GD.

GD (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.965	4	16	0.000
	Lambda Wilks	0.035	4	16	0.000
	Traza Hotelling	27.664	4	16	0.000
	Raíz mayor Roy	27.664	4	16	0.000
ecm conj. valid.	Traza Pillai	0.961	4	16	0.000
	Lambda Wilks	0.039	4	16	0.000
	Traza Hotelling	24.953	4	16	0.000
	Raíz mayor Roy	24.953	4	16	0.000
ecm conj. test.	Traza Pillai	0.962	4	16	0.000
	Lambda Wilks	0.038	4	16	0.000
	Traza Hotelling	25.403	4	16	0.000
	Raíz mayor Roy	25.403	4	16	0.000

Tabla 5.3.15. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento GD.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento GD se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.3.16, 5.3.17 y 5.3.18, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

Tal y como queda claramente reflejado en los gráficos de caja y bigotes (Figuras 5.3.4, 5.3.5 y 5.3.6), en estos contrastes de comparaciones dobles

se observan diferencias significativas entre las medias de los ecm para cada par de tamaños de la capa oculta, salvo para los tamaños 20 y 30, para los que no puede rechazarse la hipótesis de igualdad entre las medias de los ecm, resultado que se obtiene sobre los tres conjuntos considerados.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-1269.832	129.361	[-1680.388, -859.276]	0.000
2 y 10	-4121.424	419.248	[-5452.001, -2790.847]	0.000
2 y 20	-10553.240	1136.743	[-14160.949, -6945.531]	0.000
2 y 30	-15883.725	1689.271	[-21245.005, -10522.445]	0.000
5 y 2	1269.832	129.361	[859.276, 1680.388]	0.000
5 y 10	-2851.592	438.263	[-4242.518, -1460.666]	0.000
5 y 20	-9283.408	1151.487	[-12937.911, -5628.904]	0.000
5 y 30	-14613.893	1719.022	[-20069.596, -9158.190]	0.000
10 y 2	4121.425	419.248	[2790.847, 5452.001]	0.000
10 y 5	2851.593	438.263	[1460.666, 4242.518]	0.000
10 y 20	-6431.816	1352.083	[-10722.953, -2140.678]	0.001
10 y 30	-11762.301	1684.195	[-17107.472, -6417.130]	0.000
20 y 2	10553.240	1136.743	[6945.531, 14160.949]	0.000
20 y 5	9283.408	1151.487	[5628.904, 12937.911]	0.000
20 y 10	6431.816	1352.083	[2140.678, 10722.953]	0.001
20 y 30	-5330.485	2108.627	[-12022.685, 1361.714]	0.205
30 y 2	15883.725	1689.271	[10522.445, 21245.005]	0.000
30 y 5	14613.893	1719.022	[9158.190, 20069.596]	0.000
30 y 10	11762.301	1684.195	[6417.130, 17107.472]	0.000
30 y 20	5330.485	2108.627	[-1361.714, 12022.685]	0.205

Tabla 5.3.16. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento GD sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-1242.821	129.987	[-1655.362, -830.279]	0.000
2 y 10	-3878.021	416.453	[-5199.730, -2556.313]	0.000
2 y 20	-9723.970	997.145	[-12888.634, -6559.306]	0.000
2 y 30	-14709.890	1627.260	[-19874.366, -9545.414]	0.000
5 y 2	1242.822	129.987	[830.279, 1655.362]	0.000
5 y 10	-2635.201	428.105	[-3993.888, -1276.514]	0.000
5 y 20	-8481.149	1015.124	[-11702.874, -5259.425]	0.000
5 y 30	-13467.069	1649.628	[-18702.535, -8231.603]	0.000
10 y 2	3878.022	416.453	[2556.313, 5199.730]	0.000
10 y 5	2635.201	428.105	[1276.514, 3993.888]	0.000
10 y 20	-5845.949	1209.198	[-9683.609, -2008.289]	0.001
10 y 30	-10831.868	1622.902	[-15982.513, -5681.223]	0.000
20 y 2	9723.971	997.145	[6559.306, 12888.634]	0.000
20 y 5	8481.149	1015.124	[5259.425, 11702.874]	0.000
20 y 10	5845.950	1209.198	[2008.289, 9683.609]	0.001
20 y 30	-4985.920	1977.247	[-11261.157, 1289.317]	0.208
30 y 2	14709.890	1627.260	[9545.415, 19874.366]	0.000
30 y 5	13467.069	1649.628	[8231.604, 18702.535]	0.000
30 y 10	10831.868	1622.902	[5681.224, 15982.513]	0.000
30 y 20	4985.920	1977.247	[-1289.317, 11261.157]	0.208

Tabla 5.3.17. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento GD sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-1285.682	133.522	[-1709.444, -861.920]	0.000
2 y 10	-4146.709	425.985	[-5498.667, -2794.750]	0.000
2 y 20	-10575.542	1132.615	[-14170.151, -6980.934]	0.000
2 y 30	-15947.512	1687.083	[-21301.848, -10593.176]	0.000
5 y 2	1285.682	133.522	[861.920, 1709.444]	0.000
5 y 10	-2861.027	441.878	[-4263.427, -1458.627]	0.000
5 y 20	-9289.860	1148.035	[-12933.406, -5646.314]	0.000
5 y 30	-14661.830	1713.705	[-20100.657, -9223.002]	0.000
10 y 2	4146.709	425.985	[2794.751, 5498.667]	0.000
10 y 5	2861.027	441.878	[1458.628, 4263.427]	0.000
10 y 20	-6428.833	1346.019	[-10700.727, -2156.940]	0.001
10 y 30	-11800.803	1674.202	[-17114.259, -6487.347]	0.000
20 y 2	10575.542	1132.615	[6980.934, 14170.151]	0.000
20 y 5	9289.860	1148.035	[5646.314, 12933.406]	0.000
20 y 10	6428.833	1346.019	[2156.940, 10700.727]	0.001
20 y 30	-5371.970	2090.245	[-12005.833, 1261.894]	0.187
30 y 2	15947.512	1687.083	[10593.176, 21301.848]	0.000
30 y 5	14661.830	1713.705	[9223.002, 20100.657]	0.000
30 y 10	11800.803	1674.202	[6487.347, 17114.259]	0.000
30 y 20	5371.970	2090.245	[-1261.893, 12005.833]	0.187

Tabla 5.3.18. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento GD sobre el conjunto test.

Para el método de entrenamiento LM, que ya ha sido determinado como el mejor de todos los métodos de entrenamiento empleados, sobre los tres conjuntos se rechaza la hipótesis de esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.3.19.

En los gráficos de caja y bigotes (Figuras 5.3.4, 5.3.5 y 5.3.6) queda claramente reflejada la diferencia entre las medias de los ecm para todos los tamaños de capa oculta, lo que también puede observarse en los tests realizados, cuyos valores aparecen en la Tabla 5.3.20, ya que se rechaza la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

LM (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.023	65.762	9	0.000
ecm conj. valid.	0.034	58.681	9	0.000
ecm conj. test	0.041	55.605	9	0.000

Tabla 5.3.19. Prueba de esfericidad de Mauchly para el método de entrenamiento LM.

LM (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.991	4	16	0.000
	Lambda Wilks	0.009	4	16	0.000
	Traza Hotelling	111.336	4	16	0.000
	Raíz mayor Roy	111.336	4	16	0.000
ecm conj. valid.	Traza Pillai	0.991	4	16	0.000
	Lambda Wilks	0.009	4	16	0.000
	Traza Hotelling	116.508	4	16	0.000
	Raíz mayor Roy	116.508	4	16	0.000
ecm conj. test.	Traza Pillai	0.990	4	16	0.000
	Lambda Wilks	0.010	4	16	0.000
	Traza Hotelling	98.288	4	16	0.000
	Raíz mayor Roy	98.288	4	16	0.000

Tabla 5.3.20. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento LM.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento LM se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.3.21, 5.3.22 y 5.3.23, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

Tal y como queda claramente reflejado en los gráficos de caja y bigotes (Figuras 5.3.4, 5.3.5 y 5.3.6), en estos tests de comparaciones dobles se observan diferencias significativas entre las medias de los ecm para todos los pares de tamaños de la capa oculta, resultado que se obtiene sobre los tres conjuntos considerados.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	3.437E-02	0.002	[2.749E-02, 4.126E-02]	0.000
2 y 10	4.623E-02	0.002	[3.926E-02, 5.320E-02]	0.000
2 y 20	5.772E-02	0.002	[5.107E-02, 6.438E-02]	0.000
2 y 30	6.204E-02	0.002	[5.533E-02, 6.875E-02]	0.000
5 y 2	-3.437E-02	0.002	[-4.126E-02, -2.749E-02]	0.000
5 y 10	1.186E-02	0.001	[8.837E-03, 1.487E-02]	0.000
5 y 20	2.335E-02	0.001	[2.040E-02, 2.630E-02]	0.000
5 y 30	2.766E-02	0.001	[2.506E-02, 3.026E-02]	0.000
10 y 2	-4.623E-02	0.002	[-5.320E-02, -3.926E-02]	0.000
10 y 5	-1.186E-02	0.001	[-1.487E-02, -8.837E-03]	0.000
10 y 20	1.149E-02	0.001	[9.055E-03, 1.393E-02]	0.000
10 y 30	1.581E-02	0.001	[1.357E-02, 1.805E-02]	0.000
20 y 2	-5.772E-02	0.002	[-6.438E-02, -5.107E-02]	0.000
20 y 5	-2.335E-02	0.001	[-2.630E-02, -2.040E-02]	0.000
20 y 10	-1.149E-02	0.001	[-1.393E-02, -9.055E-03]	0.000
20 y 30	4.314E-03	0.000	[3.182E-03, 5.446E-03]	0.000
30 y 2	-6.204E-02	0.002	[-6.875E-02, -5.533E-02]	0.000
30 y 5	-2.766E-02	0.001	[-3.026E-02, -2.506E-02]	0.000
30 y 10	-1.581E-02	0.001	[-1.805E-02, -1.357E-02]	0.000
30 y 20	-4.314E-03	0.000	[-5.446E-03, -3.182E-03]	0.000

Tabla 5.3.21. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento LM sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	3.455E-02	0.002	[2.810E-02, 4.100E-02]	0.000
2 y 10	4.639E-02	0.002	[3.979E-02, 5.299E-02]	0.000
2 y 20	5.781E-02	0.002	[5.159E-02, 6.404E-02]	0.000
2 y 30	6.210E-02	0.002	[5.591E-02, 6.830E-02]	0.000
5 y 2	-3.455E-02	0.002	[-4.100E-02, -2.810E-02]	0.000
5 y 10	1.184E-02	0.001	[8.938E-03, 1.474E-02]	0.000
5 y 20	2.326E-02	0.001	[2.034E-02, 2.618E-02]	0.000
5 y 30	2.755E-02	0.001	[2.494E-02, 3.016E-02]	0.000
10 y 2	-4.639E-02	0.002	[-5.299E-02, -3.979E-02]	0.000
10 y 5	-1.184E-02	0.001	[-1.474E-02, -8.938E-03]	0.000
10 y 20	1.142E-02	0.001	[8.906E-03, 1.394E-02]	0.000
10 y 30	1.572E-02	0.001	[1.344E-02, 1.799E-02]	0.000
20 y 2	-5.781E-02	0.002	[-6.404E-02, -5.159E-02]	0.000
20 y 5	-2.326E-02	0.001	[-2.618E-02, -2.034E-02]	0.000
20 y 10	-1.142E-02	0.001	[-1.394E-02, -8.906E-03]	0.000
20 y 30	4.294E-03	0.000	[3.107E-03, 5.481E-03]	0.000
30 y 2	-6.210E-02	0.002	[-6.830E-02, -5.591E-02]	0.000
30 y 5	-2.755E-02	0.001	[-3.016E-02, -2.494E-02]	0.000
30 y 10	-1.572E-02	0.001	[-1.799E-02, -1.344E-02]	0.000
30 y 20	-4.294E-03	0.000	[-5.481E-03, -3.107E-03]	0.000

Tabla 5.3.22. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento LM sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	3.422E-02	0.002	[2.792E-02, 4.052E-02]	0.000
2 y 10	4.591E-02	0.002	[3.939E-02, 5.242E-02]	0.000
2 y 20	5.741E-02	0.002	[5.120E-02, 6.362E-02]	0.000
2 y 30	6.180E-02	0.002	[5.557E-02, 6.803E-02]	0.000
5 y 2	-3.422E-02	0.002	[-4.052E-02, -2.792E-02]	0.000
5 y 10	1.168E-02	0.001	[8.674E-03, 1.470E-02]	0.000
5 y 20	2.319E-02	0.001	[2.024E-02, 2.613E-02]	0.000
5 y 30	2.757E-02	0.001	[2.489E-02, 3.026E-02]	0.000
10 y 2	-4.591E-02	0.002	[-5.242E-02, -3.939E-02]	0.000
10 y 5	-1.168E-02	0.001	[-1.470E-02, -8.674E-03]	0.000
10 y 20	1.150E-02	0.001	[8.969E-03, 1.403E-02]	0.000
10 y 30	1.589E-02	0.001	[1.340E-02, 1.838E-02]	0.000
20 y 2	-5.741E-02	0.002	[-6.362E-02, -5.120E-02]	0.000
20 y 5	-2.319E-02	0.001	[-2.613E-02, -2.024E-02]	0.000
20 y 10	-1.150E-02	0.001	[-1.403E-02, -8.969E-03]	0.000
20 y 30	4.389E-03	0.000	[3.199E-03, 5.580E-03]	0.000
30 y 2	-6.180E-02	0.002	[-6.803E-02, -5.557E-02]	0.000
30 y 5	-2.757E-02	0.001	[-3.026E-02, -2.489E-02]	0.000
30 y 10	-1.589E-02	0.001	[-1.838E-02, -1.340E-02]	0.000
30 y 20	-4.389E-03	0.000	[-5.580E-03, -3.199E-03]	0.000

Tabla 5.3.23. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento LM sobre el conjunto test.

Y por último, para el método de entrenamiento RP, sobre los tres conjuntos se rechaza la hipótesis de esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.3.24.

En los gráficos de caja y bigotes correspondientes (Figuras 5.3.4, 5.3.5 y 5.3.6), parece que las medias de los ecm para todos los tamaños de capa oculta pueden considerarse iguales, lo que también puede observarse en los contrastes realizados, cuyos valores aparecen en la Tabla 5.3.25, ya que no es posible rechazar la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

RP (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.000	355.293	9	0.000
ecm conj. valid.	0.000	355.595	9	0.000
ecm conj. test	0.000	355.308	9	0.000

Tabla 5.3.24. Prueba de esfericidad de Mauchly para el método de entrenamiento RP.

RP (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.327	4	16	0.152
	Lambda Wilks	0.673	4	16	0.152
	Traza Hotelling	0.487	4	16	0.152
	Raíz mayor Roy	0.487	4	16	0.152
ecm conj. valid.	Traza Pillai	0.333	4	16	0.144
	Lambda Wilks	0.667	4	16	0.144
	Traza Hotelling	0.499	4	16	0.144
	Raíz mayor Roy	0.499	4	16	0.144
ecm conj. test.	Traza Pillai	0.328	4	16	0.150
	Lambda Wilks	0.672	4	16	0.150
	Traza Hotelling	0.489	4	16	0.150
	Raíz mayor Roy	0.489	4	16	0.150

Tabla 5.3.25. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento RP.

Por tanto, de todos estos gráficos, tablas y contrastes, se puede establecer la siguiente relación de optimalidad de resultados teniendo en cuenta los distintos métodos de entrenamiento y los diversos tamaños de capa oculta utilizados.

El mejor de los métodos de entrenamiento es el LM, incluso para cualquiera de los tamaños de capa oculta considerados, observándose que los resultados van mejorando progresivamente conforme aumenta el número de neuronas consideradas en la capa oculta, no pudiendo ser considerados similares para ninguno de los pares de tamaños de la capa oculta considerados, y los resultados óptimos se obtienen para un tamaño de capa oculta de 30 neuronas.

El método BFGS es el segundo mejor método de entrenamiento, pero sus resultados no pueden ser considerados similares para cualquier tamaño de capa oculta debido a las diferencias existentes entre los resultados obtenidos para los tamaños de capa oculta 20 y 30, mientras que el resto de los tamaños de capa oculta comparados por pares sí que pueden ser considerados similares.

El tercer mejor método de entrenamiento es el CGFR, y sus resultados no pueden ser considerados similares para todos los tamaños de capa oculta porque con el tamaño 30 sus resultados empeoran de forma abrumadora, siendo rechazada en todas las comparaciones por pares la igualdad de medias de los ecm entre este tamaño de capa oculta y el resto.

En el cuarto lugar se encuentra el método RP, cuyos resultados se pueden considerar similares para todos los tamaños de capa oculta considerados.

Y el peor de los métodos de entrenamiento empleados es el GD, cuyos resultados van empeorando progresivamente conforme aumenta el número de neuronas consideradas en la capa oculta, pudiendo ser considerados similares únicamente para los tamaños de capa oculta 20 y 30, mientras que para el resto de los pares de tamaños de la capa oculta considerados, sus resultados difieren.

Una vez determinada la red_ algoritmo óptima como la obtenida mediante el método de entrenamiento de Levenberg-Marquardt y una estructura de perceptrón con una única capa oculta de 30 neuronas, una capa de entrada con dos neuronas y una capa de salida formada por una única neurona, esto es, una estructura óptima (2:30:1), a continuación se procede a la búsqueda, entre todas las realizaciones llevadas a cabo, de aquella que proporciona el menor de los ecm. Para ello, se hallan los mínimos ecm sobre el conjunto test para el método de entrenamiento LM y tamaño de capa oculta 30, observándose que este mínimo se alcanza durante la iteración número 6, en su ejecución número 11, presentando un ecm sobre el conjunto test de **0.08794365**. Los valores de estos ecm mínimos se recogen en la Tabla 5.3.26.

Iteración	Ejecución	ecm conj. entren.	ecm conj. valid.	ecm conj. test
1	18	0.0921184	0.09155368	0.09341449
2	8	0.09046805	0.09696565	0.09293312
3	9	0.09347201	0.09394534	0.09201448
4	17	0.09227147	0.09206939	0.09002821
5	9	0.09470303	0.08984377	0.09024146
6	11	0.09477306	0.09222674	0.08794365
7	13	0.09200926	0.09171065	0.09457193
8	4	0.09243153	0.09376322	0.09128568
9	15	0.09379996	0.0935848	0.08964269
10	4	0.09109517	0.09593542	0.09273399
11	5	0.09357236	0.09301624	0.08998471
12	3	0.0931868	0.09011547	0.0938063
13	20	0.09124661	0.0906599	0.09478287
14	12	0.09142171	0.09489248	0.09079017
15	20	0.09202989	0.09151142	0.09424694
16	7	0.08882512	0.09823228	0.09354697
17	2	0.0930138	0.0910018	0.09495343
18	1	0.09000966	0.09271339	0.09547272
19	19	0.09362927	0.09130152	0.09047117
20	2	0.09284057	0.09251657	0.09171999

Tabla 5.3.26. Mínimos ecm sobre el conjunto test para el método de entrenamiento LM y tamaño de capa oculta 30.

5.4 MLP DE UNA CAPA CON ONCE PREDICTORES.

La segunda red que se ha construido, y que será referida como red_información, utiliza toda la información adicional de la que se dispone a bordo del barco, y consiste también en un perceptrón múltiple con una única capa oculta.

La capa de entrada está formada por once neuronas, correspondientes cada una de ellas a las variables "Latitud", "Longitud", "Profundidad", "Fondo existente en la localización", "Hora en que se realiza la medición", "Temperatura en la superficie del agua", "Presión atmosférica", "Velocidad del viento", "Dirección del viento", "Temperatura del aire" y "Temperatura del agua".

El número de neuronas de la capa de salida es una, correspondiente a la variable "Velocidad de Sonido en el Agua del Mar", que constituye la variable que se pretende predecir, y que será referida como velocidad de sonido.

Con respecto al número de neuronas que constituyen la capa oculta, ante la inexistencia de una regla de selección válida, se han considerado seis tamaños de la capa oculta diferentes (2, 5, 10, 20, 30 y 40), con el fin de determinar la estructura óptima de la red_información.

Esta red_información descrita ha sido ejecutada en el paquete matemático MATLAB 5.3, empleando para ello el módulo denominado Neural Network Toolbox 3.0. El código del programa implementado aparece en el Anexo 1.

Se han realizado 20 iteraciones del programa, cada una de las cuales corresponde a las distintas selecciones de los conjuntos de entrenamiento, validación y test, obteniendo de esta manera $20 \times 20 \times 5 \times 6$ redes entrenadas, cada una de ellas con los diferentes métodos de entrenamiento y tamaños de capa oculta considerados, de entre las que se deducirá la red_información óptima con respecto al criterio de optimalidad ya mencionado con anterioridad de minimización del error cuadrático medio.

La descripción y el análisis de los resultados obtenidos con este procedimiento ha sido realizado mediante representaciones gráficas y diversos contrastes de hipótesis ejecutados en el paquete estadístico SPSS 8.0.

5.4.1 RESUMEN DE LOS RESULTADOS.

Para determinar el método de entrenamiento y el tamaño de la capa oculta óptimos, se han analizado los estadísticos de las medias de los errores cuadráticos medios (ecm) sobre cada uno de los tres conjuntos de datos considerados, esto es, sobre los conjuntos de entrenamiento, validación y test, cuyos valores para cada una de las iteraciones aparecen en el Anexo 3.

El análisis ha sido realizado sobre cada uno de los conjuntos por separado, observándose que, independientemente del tamaño de la capa oculta, el mejor método de entrenamiento es el de Levenberg-Marquardt. En cambio, las relaciones entre el resto de métodos de entrenamiento empleados no se pueden anticipar, por lo que más adelante éstas serán analizadas en profundidad. Estos resultados quedan reflejados en las representaciones gráficas de los estadísticos media, mínimo y máximo de las medias de los ecm para cada iteración, que aparecen en las Figuras 5.4.1, 5.4.2 y 5.4.3.

De estos gráficos hay que destacar además la estabilidad del método de Levenberg-Marquardt con respecto a cada iteración, esto es, que la selección de los conjuntos de entrenamiento, validación y test no afecta de forma significativa a los resultados obtenidos al entrenar la red mediante este método. En cambio, puede observarse que el resto de métodos de entrenamiento empleados sí que se ven afectados por dicha selección.



Figura 5.4.1. Representaciones gráficas de los estadísticos de los ecm para cada iteración sobre el conjunto de entrenamiento.

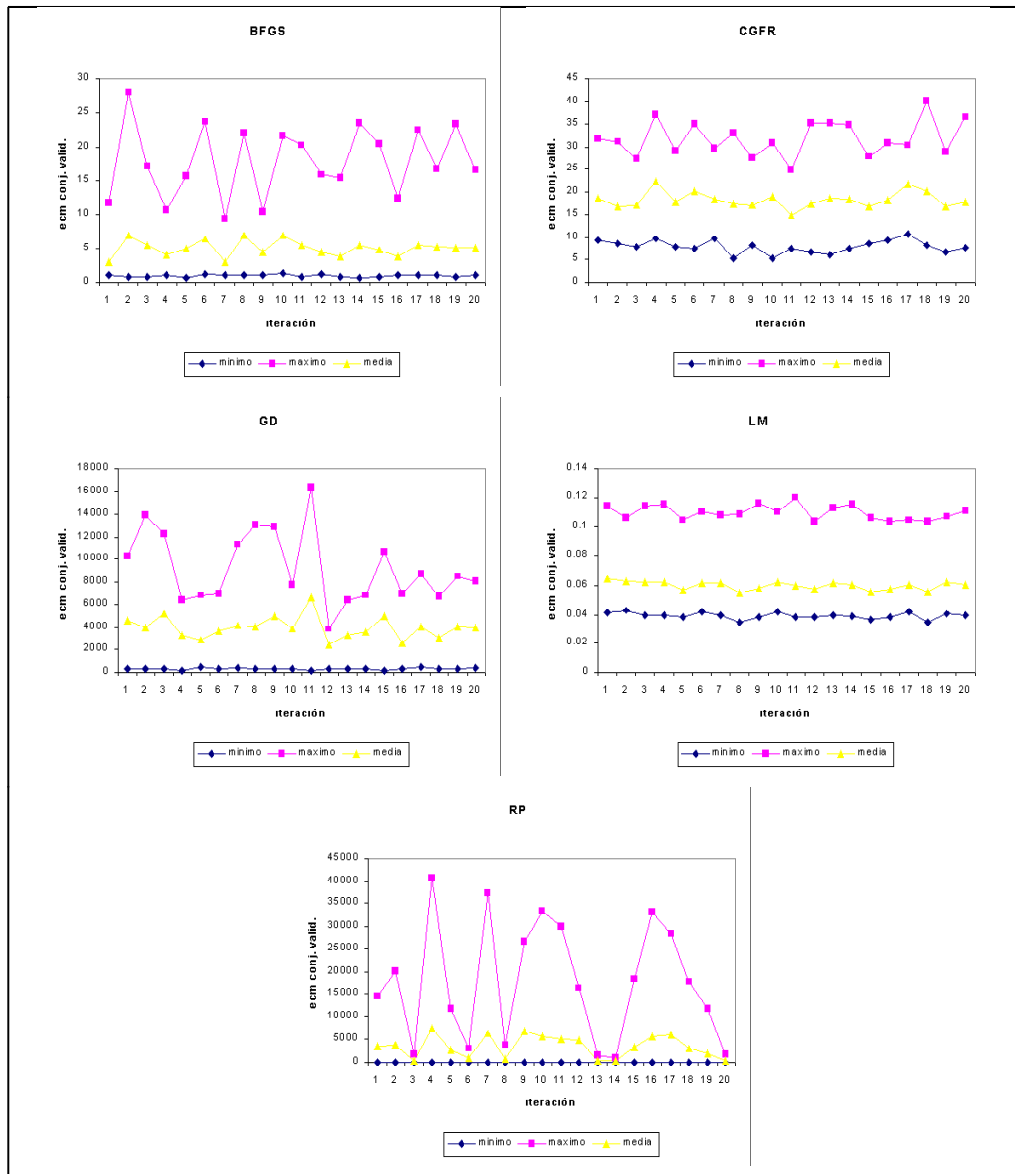


Figura 5.4.2. Representaciones gráficas de los estadísticos de los ecm para cada iteración sobre el conjunto de validación.



Figura 5.4.3. Representaciones gráficas de los estadísticos de los ecm para cada iteración sobre el conjunto test.

Varios estadísticos descriptivos han sido calculados para cada uno de los conjuntos de datos teniendo en cuenta además el tamaño de la capa oculta. Sus valores aparecen en las Tablas 5.4.1, 5.4.2 y 5.4.3, y en ellas puede observarse que el método LM es el mejor de todos los métodos de entrenamiento para todos los tamaños de la capa oculta considerados, presentando para el tamaño 40 la menor de las medias de los ecm.

El segundo mejor método para todos los tamaños de capa oculta, es el BFGS, observándose unos resultados mucho peores para los tamaños de capa oculta 2 y 5, mientras que para el resto de tamaños los resultados son similares.

El tercer mejor método para todos los tamaños de capa oculta, es el CGFR, donde la tendencia de los resultados con respecto al incremento en el tamaño de la capa oculta decrece hasta el tamaño 10 y luego se invierte, llegando nuevamente en el tamaño 40, que es el mayor tamaño de capa oculta considerado, casi a equiparar los resultados obtenidos con el tamaño 2.

Y nuevamente surgen complicaciones a la hora de determinar en función de los tamaños de la capa oculta cuál de los métodos GD y RP es mejor, por lo que la relación entre estos métodos será analizada con detalle más adelante.

Todas estas relaciones entre los distintos métodos y los tamaños de capa oculta considerados, son analizadas mediante gráficos del tipo de caja y bigotes, cuyas representaciones gráficas para cada uno de los conjuntos de datos aparecen en las Figuras 5.4.4, 5.4.5 y 5.4.6, y mediante contrastes de igualdad de medias.

A partir de modelos de análisis de la varianza de un factor, en este caso el tamaño de capa oculta, con medidas repetidas, se analizará la igualdad de las medias de los errores cuadráticos medios entre los tamaños de la capa oculta para cada uno de los métodos de entrenamiento empleados.

En estos gráficos de caja y bigotes se puede observar un comportamiento similar de los métodos de entrenamiento empleados con respecto a los conjuntos de entrenamiento, validación y test. Sin embargo, los tests de igualdad de medias de los ecm entre los diversos tamaños de la capa oculta han sido realizados sobre los tres conjuntos.

Método	Tamaño	Media	Mínimo	Máximo	Desv. típica
BFGS	2	17.732173	9.740709	30.368943	5.810610
	5	7.210186	1.608860	15.759990	4.024845
	10	1.733950	0.748396	5.768698	1.423867
	20	1.134283	0.756265	3.404822	0.558681
	30	1.249081	1.044258	1.657564	0.142341
	40	1.526524	1.196514	1.788482	0.169819
CGFR	2	30.053406	20.603645	40.120338	5.186495
	5	12.211312	9.123164	17.305452	2.797760
	10	8.325294	5.462999	12.625683	1.984843
	20	10.704941	7.283119	15.222667	1.798516
	30	21.542958	14.906327	35.307476	5.821723
	40	27.547826	21.606127	37.437879	4.836644
GD	2	339.150275	217.030787	529.109356	81.644467
	5	857.726877	487.943012	1572.65257	313.053937
	10	2448.29304	1543.59690	4238.49936	684.218313
	20	5991.01869	2438.96195	10938.1721	2325.68122
	30	6188.44271	2785.52826	13054.5789	2894.97618
	40	9233.27884	4031.11171	17417.5330	3763.79929
LM	2	0.109007	0.099339	0.119209	0.005179
	5	0.070540	0.062223	0.078749	0.005160
	10	0.052393	0.047305	0.055265	0.002173
	20	0.042025	0.040002	0.044848	0.001400
	30	0.039431	0.037660	0.041652	0.001189
	40	0.037969	0.035560	0.040546	0.001350
RP	2	369.679829	31.246094	3679.86262	851.146158
	5	143.848229	42.545761	1032.88204	217.441889
	10	714.077918	50.790109	4485.34763	1245.98809
	20	2993.94171	165.185651	18753.5440	5116.83860
	30	5790.05189	224.566519	40502.5660	10554.6988
	40	11326.3619	355.759145	37233.0257	14278.7101

Tabla 5.4.1. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de entrenamiento.

Método	Tamaño	Q ₁	Mediana (Q ₂)	Q ₃
BFGS	2	11.9829042	17.3251188	22.6162608
	5	4.32312360	6.17021926	9.94886716
	10	0.94291924	1.10501889	1.50947133
	20	0.90637188	1.00922175	1.16890928
	30	1.13337430	1.24303344	1.31768786
	40	1.42335578	1.56999483	1.64783590
CGFR	2	25.5452883	29.4590829	34.0107722
	5	9.54503556	11.8670304	14.6461049
	10	6.86145411	8.21047284	9.75805087
	20	9.80461140	10.1801988	11.9543077
	30	17.4828639	19.3890913	22.7268141
	40	23.4578564	26.4782209	31.0371154
GD	2	282.897758	332.857547	367.121922
	5	613.558635	780.832494	1087.05995
	10	1882.47675	2423.03720	2839.97965
	20	3708.51033	6383.42754	7605.35681
	30	3969.75686	5753.15057	7725.67841
	40	6699.33749	7190.06781	12695.3842
LM	2	0.10450525	0.10958003	0.11302849
	5	0.06607476	0.06964315	0.07555072
	10	0.05091267	0.05258470	0.05448604
	20	0.04084478	0.04173260	0.04313514
	30	0.03858028	0.03908867	0.04059440
	40	0.03700203	0.03781739	0.03891330
RP	2	43.1658947	124.055273	207.232730
	5	66.3110727	83.8971782	113.804499
	10	85.8276319	111.732142	946.181810
	20	217.180718	277.634684	3703.05060
	30	305.479905	349.839597	12268.1163
	40	484.749941	1342.97307	28135.3026

Tabla 5.4.1-Cont. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de entrenamiento.

Método	Tamaño	Media	Mínimo	Máximo	Desv. típica
BFGS	2	17.574750	9.395308	27.979490	5.546875
	5	7.119242	1.673833	15.664817	3.923352
	10	1.726886	0.740970	5.664046	1.400136
	20	1.137876	0.744205	3.523973	0.585455
	30	1.254726	1.029988	1.617304	0.142323
	40	1.537029	1.192330	1.808199	0.176523
CGFR	2	29.824874	20.076614	40.137315	5.113290
	5	12.143948	8.670819	17.151674	2.835955
	10	8.214823	5.411307	12.266962	1.910962
	20	10.681113	7.256899	15.401517	1.808318
	30	21.455753	14.936765	34.907564	5.712522
	40	27.379743	21.808163	37.027730	4.721370
GD	2	330.882554	216.910487	521.665800	79.839014
	5	825.289924	486.008445	1454.84735	285.153743
	10	2374.33513	1455.24509	4100.15199	680.394554
	20	5588.24406	2362.73731	10275.6185	2111.74106
	30	5953.01071	2606.83835	12956.8947	2822.72198
	40	8755.24179	3931.32297	16408.5196	3480.01873
LM	2	0.110053	0.103890	0.120451	0.004867
	5	0.071769	0.058723	0.082675	0.006252
	10	0.053663	0.047937	0.058926	0.003647
	20	0.043227	0.039500	0.047635	0.002349
	30	0.040825	0.036301	0.044745	0.002482
	40	0.039411	0.035153	0.043116	0.002243
RP	2	374.398856	31.432196	3749.85648	866.871725
	5	143.350534	43.021950	1016.99840	214.217522
	10	707.918208	50.830649	4319.44073	1220.42083
	20	2964.30252	166.999134	18523.5557	5077.39586
	30	5756.87603	219.446755	40722.7409	10564.3402
	40	11291.4973	349.654511	37426.9975	14262.3798

Tabla 5.4.2. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de validación.

Método	Tamaño	Q ₁	Mediana (Q ₂)	Q ₃
BFGS	2	11.8871688	16.8867616	22.3657101
	5	4.27387355	6.09284496	9.92470478
	10	0.93098602	1.09533819	1.56056718
	20	0.89123852	1.00335167	1.17787570
	30	1.13460383	1.24863655	1.33218816
	40	1.44384833	1.58006953	1.63556418
CGFR	2	25.7058130	29.8879406	33.5645898
	5	9.47724704	11.8840885	14.6124392
	10	6.75189011	7.97650816	9.60167488
	20	9.69455648	10.2886954	11.7656710
	30	17.4944292	19.3578708	22.2567665
	40	23.4836162	26.2800199	30.7146866
GD	2	277.240766	326.815426	346.627358
	5	590.301228	765.456236	1004.40312
	10	1809.59196	2301.23513	2826.01759
	20	3478.82223	6029.40143	7194.87632
	30	3759.05312	5537.46898	7182.91211
	40	6504.88762	6934.99254	12066.8684
LM	2	0.10533501	0.10975884	0.11455538
	5	0.06712880	0.07213658	0.07567024
	10	0.04999786	0.05465175	0.05698485
	20	0.04128957	0.04332560	0.04513106
	30	0.03903324	0.04009562	0.04317753
	40	0.03806518	0.03942973	0.04135029
RP	2	41.6836104	124.232028	208.285824
	5	66.0847621	84.4984482	115.212983
	10	83.4213090	112.142795	944.934085
	20	216.121148	269.497749	3604.80190
	30	309.445565	345.638181	11838.0441
	40	481.812569	1316.75696	27887.2276

Tabla 5.4.2-Cont. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de validación.

Método	Tamaño	Media	Mínimo	Máximo	Desv. típica
BFGS	2	17.712270	9.664552	29.181202	5.671702
	5	7.215190	1.576349	16.611203	4.063616
	10	1.731410	0.761205	5.699627	1.410702
	20	1.142272	0.763678	3.415837	0.560477
	30	1.266726	1.042849	1.675376	0.144428
	40	1.539058	1.199906	1.778211	0.160871
CGFR	2	30.028793	20.754904	40.156821	5.127500
	5	12.177778	8.855515	16.958084	2.740110
	10	8.322701	5.533805	12.713748	2.000396
	20	10.736583	7.162346	14.970814	1.751551
	30	21.625879	15.044868	35.707717	5.785923
	40	27.646630	21.844015	37.236405	5.006397
GD	2	338.598904	217.384121	539.029214	79.885933
	5	863.172019	492.174286	1608.55242	317.151284
	10	2450.84190	1549.39852	4161.98510	667.536847
	20	5984.21642	2412.24298	10918.5479	2323.90690
	30	6220.22390	2785.24206	13285.8876	2934.92515
	40	9212.74752	3995.35761	17259.4503	3790.32529
LM	2	0.111691	0.099386	0.124122	0.006187
	5	0.073069	0.062039	0.083528	0.005037
	10	0.054965	0.049567	0.059772	0.002570
	20	0.044516	0.041172	0.048560	0.001861
	30	0.042005	0.038898	0.046889	0.002177
	40	0.040618	0.037693	0.043746	0.001761
RP	2	373.892496	32.643722	3739.04040	864.339815
	5	143.903896	43.246926	1032.73147	217.302500
	10	720.713021	49.698467	4518.76117	1257.099129
	20	2998.39108	165.960018	18615.4029	5102.05821
	30	5796.33650	220.841161	40194.4113	10528.1485
	40	11380.2710	363.351118	38244.4011	14385.8799

Tabla 5.4.3. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto test.

Método	Tamaño	Q ₁	Mediana (Q ₂)	Q ₃
BFGS	2	11.9442489	17.0004434	22.3469549
	5	4.29299882	6.21335289	9.90004474
	10	0.93913571	1.08569063	1.49772528
	20	0.90366608	1.02568260	1.18285634
	30	1.14628194	1.26180093	1.33778820
	40	1.46218836	1.59273191	1.64276336
CGFR	2	26.3249867	29.9780500	34.5484562
	5	9.60246677	11.9579334	15.0512429
	10	6.76609059	8.24463163	9.67197535
	20	9.74824931	10.3916155	11.8951925
	30	17.4909041	19.4889331	22.7430835
	40	23.1855372	26.8340886	31.1068224
GD	2	282.948179	335.505685	354.220242
	5	610.253578	778.581713	1104.63510
	10	1914.35035	2440.04439	2879.85695
	20	3713.45794	6344.79180	7663.03529
	30	3994.33148	5840.52816	7846.63197
	40	6547.35369	7320.79656	12806.1675
LM	2	0.10735029	0.11225795	0.11478905
	5	0.07003571	0.07384270	0.07620027
	10	0.05284573	0.05463573	0.05718514
	20	0.04315785	0.04454005	0.04578369
	30	0.04062882	0.04198288	0.04316780
	40	0.03941821	0.04067231	0.04183909
RP	2	42.9166062	124.525148	206.433235
	5	67.9717296	84.5611145	112.199504
	10	86.4661994	112.827604	947.886159
	20	217.605598	266.111362	3713.59489
	30	308.320726	349.805543	12442.0304
	40	485.726900	1335.86864	28119.0441

Tabla 5.4.3-Cont. Estadísticos de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto test.

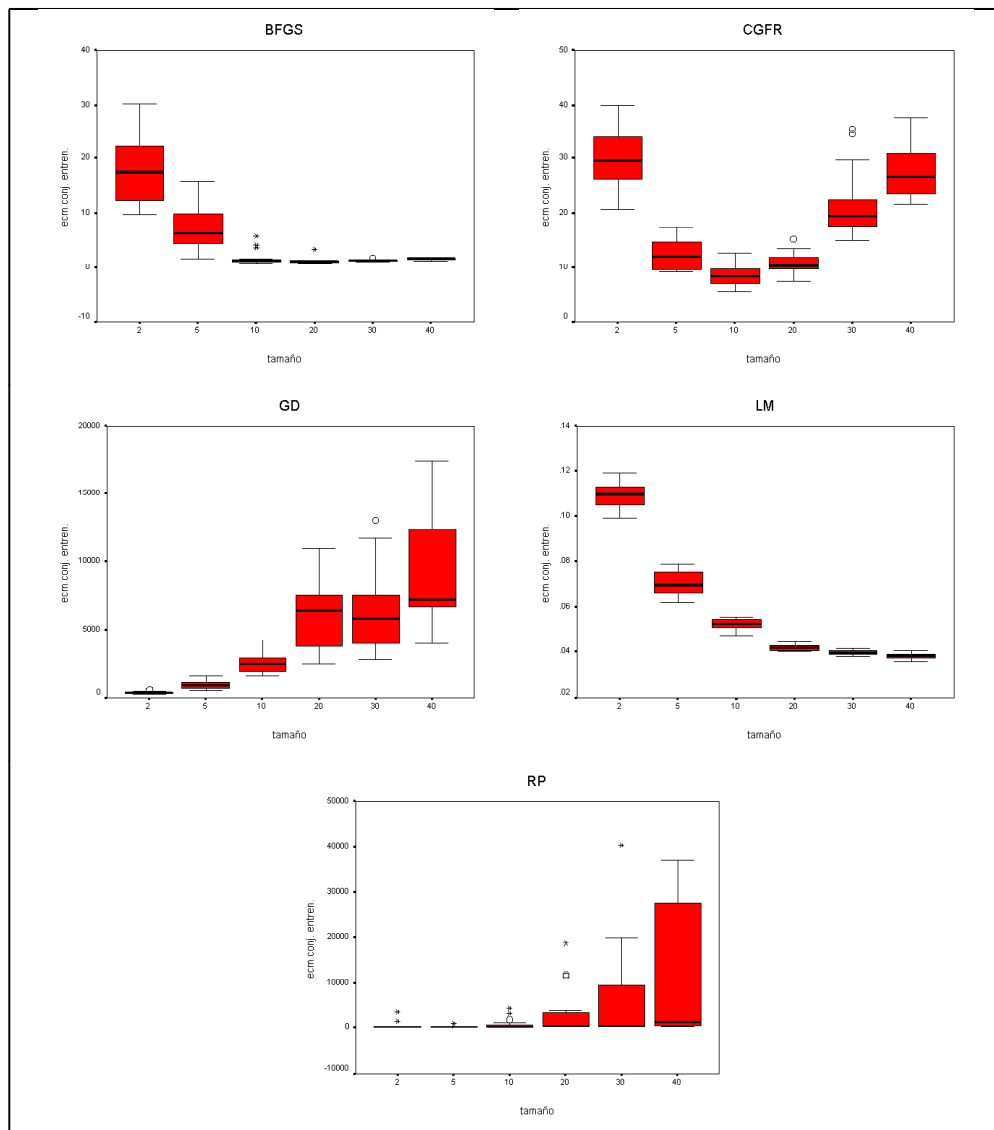


Figura 5.4.4. Gráficos de caja y bigotes de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de entrenamiento.

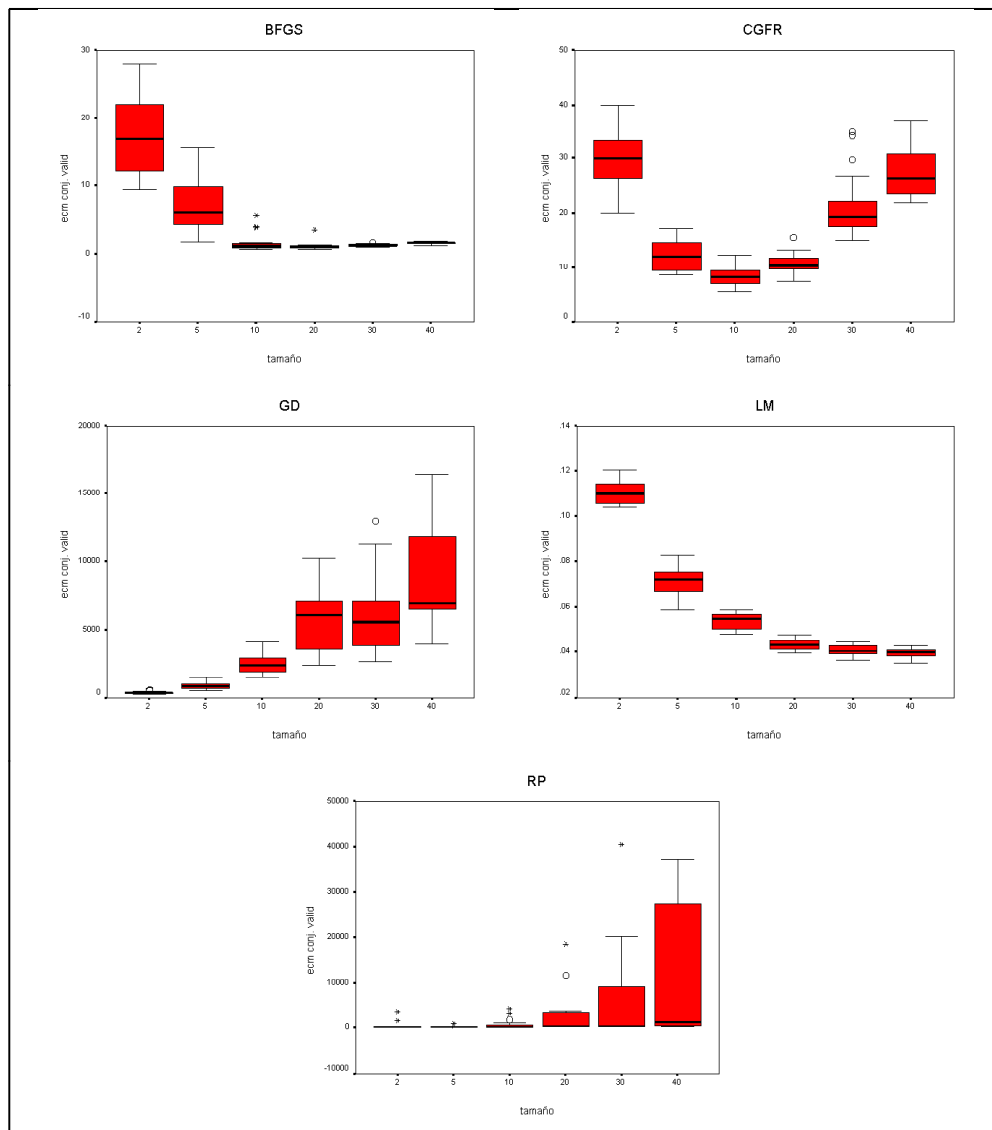


Figura 5.4.5. Gráficos de caja y bigotes de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto de validación.

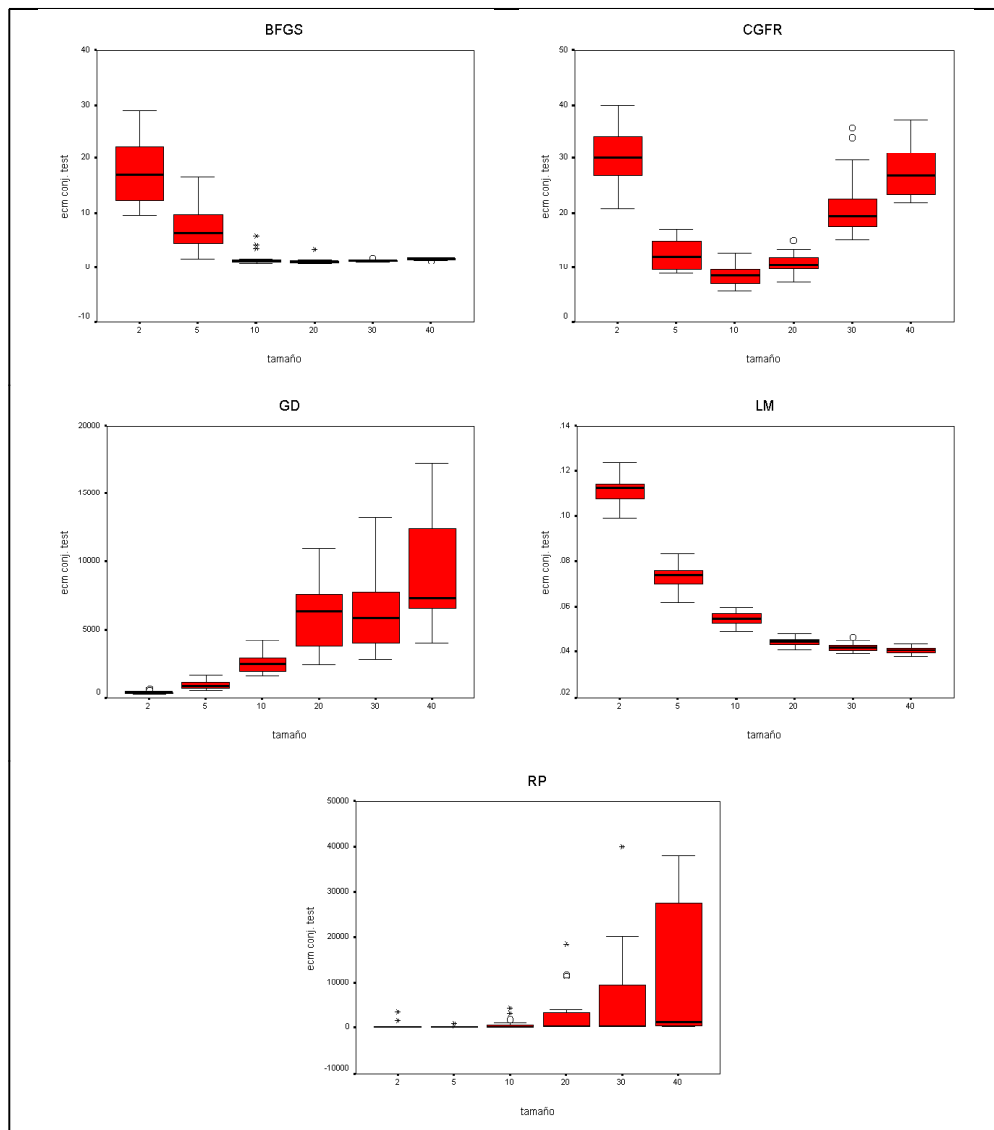


Figura 5.4.6. Gráficos de caja y bigotes de las medias de los ecm según el método de entrenamiento empleado y el tamaño de la capa oculta para el conjunto test.

5.4.2 ANÁLISIS DE LOS RESULTADOS.

De forma similar al caso de los resultados de la red_ algoritmo, para la red_ información también se analizarán sus resultados mediante un modelo de análisis de la varianza de un factor con medidas repetidas. Así, para cada uno de los métodos de entrenamiento se realizará primeramente el test de esfericidad de Mauchly, seguidamente se llevará a cabo el test de igualdad de medias de los ecm entre todos los tamaños de capa oculta, y finalmente los tests de igualdad de las medias de los ecm tomando los tamaños dos a dos y comparándolos.

Respecto al método de entrenamiento BFGS, sobre los tres conjuntos se rechaza la esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.4.4.

Los valores de los estadísticos para contrastar la igualdad de las medias de los ecm para todos los tamaños de capa oculta aparecen en la Tabla 5.4.5, rechazándose esta hipótesis sobre los tres conjuntos para un nivel de significación de 0.05, circunstancia que quedaba claramente reflejada en sus correspondientes gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6).

BFGS (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.007	179.199	14	0.000
ecm conj. valid.	0.009	175.561	14	0.000
ecm conj. test	0.010	179.037	14	0.000

Tabla 5.4.4. Prueba de esfericidad de Mauchly para el método de entrenamiento BFGS.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento BFGS se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.4.6, 5.4.7 y 5.4.8, respectivamente.

Hay que tener en cuenta que el ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y que además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

BFGS (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.943	5	15	0.000
	Lambda Wilks	0.057	5	15	0.000
	Traza Hotelling	16.681	5	15	0.000
	Raíz mayor Roy	16.681	5	15	0.000
ecm conj. valid.	Traza Pillai	0.949	5	15	0.000
	Lambda Wilks	0.051	5	15	0.000
	Traza Hotelling	18.466	5	15	0.000
	Raíz mayor Roy	18.466	5	15	0.000
ecm conj. test.	Traza Pillai	0.948	5	15	0.000
	Lambda Wilks	0.052	5	15	0.000
	Traza Hotelling	18.301	5	15	0.000
	Raíz mayor Roy	18.301	5	15	0.000

Tabla 5.4.5. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento BFGS.

Tal y como puede observarse en los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6), en estos tests de comparaciones dobles se observan diferencias significativas entre las medias de los ecm para el tamaño de capa oculta 2 y el resto de los tamaños, al igual que para el tamaño de capa oculta 5 y el resto de los tamaños. Respecto a las medias de los ecm de los tamaños de capa oculta 10 y 20, éstas pueden considerarse similares por pares a las obtenidas con los tamaños 30 y 40, pero existen diferencias significativas entre las medias de los ecm para los tamaños de capa oculta 30 y 40. Estos resultados se obtienen sobre los tres conjuntos considerados.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	10.522	1.565	[5.274, 15.770]	0.000
2 y 10	15.998	1.322	[11.563, 20.434]	0.000
2 y 20	16.598	1.286	[12.285, 20.911]	0.000
2 y 30	16.483	1.305	[12.107, 20.859]	0.000
2 y 40	16.206	1.299	[11.848, 20.563]	0.000
5 y 2	-10.522	1.565	[-15.770, -5.274]	0.000
5 y 10	5.476	0.900	[2.459, 8.493]	0.000
5 y 20	6.076	0.926	[2.971, 9.181]	0.000
5 y 30	5.961	0.884	[2.995, 8.927]	0.000
5 y 40	5.684	0.918	[2.606, 8.762]	0.000
10 y 2	-15.998	1.322	[-20.434, -11.563]	0.000
10 y 5	-5.476	0.900	[-8.493, -2.459]	0.000
10 y 20	0.600	0.297	[-0.397, 1.596]	0.869
10 y 30	0.485	0.321	[-0.592, 1.562]	1.000
10 y 40	0.207	0.324	[-0.880, 1.295]	1.000
20 y 2	-16.598	1.286	[-20.911, -12.285]	0.000
20 y 5	-6.076	0.926	[-9.181, -2.971]	0.000
20 y 10	-0.600	0.297	[-1.596, 0.397]	0.869
20 y 30	-0.115	0.130	[-0.552, 0.322]	1.000
20 y 40	-0.392	0.133	[-0.838, 5.374E-02]	0.123
30 y 2	-16.483	1.305	[-20.859, -12.107]	0.000
30 y 5	-5.961	0.884	[-8.927, -2.995]	0.000
30 y 10	-0.485	0.321	[-1.562, 0.592]	1.000
30 y 20	0.115	0.130	[-0.322, 0.552]	1.000
30 y 40	-0.277	0.051	[-0.449, -0.106]	0.000
40 y 2	-16.206	1.299	[-20.563, -11.848]	0.000
40 y 5	-5.684	0.918	[-8.762, -2.606]	0.000
40 y 10	-0.207	0.324	[-1.295, 0.880]	1.000
40 y 20	0.392	0.133	[-5.374E-02, 0.838]	0.123
40 y 30	0.277	0.051	[0.106, 0.449]	0.000

Tabla 5.4.6. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento BFGS sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	10.456	1.550	[5.257, 15.654]	0.000
2 y 10	15.848	1.264	[11.608, 20.087]	0.000
2 y 20	16.437	1.220	[12.346, 20.528]	0.000
2 y 30	16.320	1.246	[12.142, 20.498]	0.000
2 y 40	16.038	1.239	[11.881, 20.194]	0.000
5 y 2	-10.456	1.550	[-15.654, -5.257]	0.000
5 y 10	5.392	0.886	[2.420, 8.365]	0.000
5 y 20	5.981	0.903	[2.954, 9.008]	0.000
5 y 30	5.865	0.863	[2.969, 8.760]	0.000
5 y 40	5.582	0.897	[2.574, 8.590]	0.000
10 y 2	-15.848	1.264	[-20.087, -11.608]	0.000
10 y 5	-5.392	0.886	[-8.365, -2.420]	0.000
10 y 20	0.589	0.288	[-0.376, 1.554]	0.822
10 y 30	0.472	0.314	[-0.582, 1.527]	1.000
10 y 40	0.190	0.316	[-0.871, 1.251]	1.000
20 y 2	-16.437	1.220	[-20.528, -12.346]	0.000
20 y 5	-5.981	0.903	[-9.008, -2.954]	0.000
20 y 10	-0.589	0.288	[-1.554, 0.376]	0.822
20 y 30	-0.117	0.134	[-0.567, 0.333]	1.000
20 y 40	-0.399	0.137	[-0.859, 6.026E-02]	0.133
30 y 2	-16.320	1.246	[-20.498, -12.142]	0.000
30 y 5	-5.865	0.863	[-8.760, -2.969]	0.000
30 y 10	-0.472	0.314	[-1.527, 0.582]	1.000
30 y 20	0.117	0.134	[-0.333, 0.567]	1.000
30 y 40	-0.282	0.051	[-0.455, -0.110]	0.000
40 y 2	-16.038	1.239	[-20.194, -11.881]	0.000
40 y 5	-5.582	0.897	[-8.590, -2.574]	0.000
40 y 10	-0.190	0.316	[-1.251, 0.871]	1.000
40 y 20	0.399	0.137	[-6.026E-02, 0.859]	0.133
40 y 30	0.282	0.051	[0.110, 0.455]	0.000

Tabla 5.4.7. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento BFGS sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	10.497	1.580	[5.196, 15.798]	0.000
2 y 10	15.981	1.299	[11.626, 20.336]	0.000
2 y 20	16.570	1.258	[12.352, 20.788]	0.000
2 y 30	16.446	1.273	[12.177, 20.715]	0.000
2 y 40	16.173	1.268	[11.919, 20.427]	0.000
5 y 2	-10.497	1.580	[-15.798, -5.196]	0.000
5 y 10	5.484	0.911	[2.427, 8.540]	0.000
5 y 20	6.073	0.935	[2.938, 9.208]	0.000
5 y 30	5.948	0.892	[2.956, 8.941]	0.000
5 y 40	5.676	0.926	[2.571, 8.781]	0.000
10 y 2	-15.981	1.299	[-20.336, -11.626]	0.000
10 y 5	-5.484	0.911	[-8.540, -2.427]	0.000
10 y 20	0.589	0.293	[-0.395, 1.573]	0.885
10 y 30	0.465	0.317	[-0.599, 1.528]	1.000
10 y 40	0.192	0.318	[-0.873, 1.257]	1.000
20 y 2	-16.570	1.258	[-20.788, -12.352]	0.000
20 y 5	-6.073	0.935	[-9.208, -2.938]	0.000
20 y 10	-0.589	0.293	[-1.573, 0.395]	0.885
20 y 30	-0.124	0.131	[-0.562, 0.313]	1.000
20 y 40	-0.397	0.132	[-0.840, 4.648E-02]	0.110
30 y 2	-16.446	1.273	[-20.715, -12.177]	0.000
30 y 5	-5.948	0.892	[-8.941, -2.956]	0.000
30 y 10	-0.465	0.317	[-1.528, 0.599]	1.000
30 y 20	0.124	0.131	[-0.313, 0.562]	1.000
30 y 40	-0.272	0.051	[-0.443, -0.102]	0.001
40 y 2	-16.173	1.268	[-20.427, -11.919]	0.000
40 y 5	-5.676	0.926	[-8.781, -2.571]	0.000
40 y 10	-0.192	0.318	[-1.257, 0.873]	1.000
40 y 20	0.397	0.132	[-4.648E-02, 0.840]	0.110
40 y 30	0.272	0.051	[0.102, 0.443]	0.001

Tabla 5.4.8. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento BFGS sobre el conjunto test.

En cuanto al método de entrenamiento CGFR, sobre los tres conjuntos se rechaza nuevamente la esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.4.9.

En los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6) queda claramente reflejada la diferencia entre las medias de los ecm para todos los tamaños de capa oculta, lo que también puede observarse en los tests realizados, cuyos valores aparecen en la Tabla 5.4.10, ya que se rechaza la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

CGFR (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.110	37.680	14	0.001
ecm conj. valid.	0.147	32.831	14	0.003
ecm conj. test	0.119	36.352	14	0.001

Tabla 5.4.9. Prueba de esfericidad de Mauchly para el método de entrenamiento CGFR.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento CGFR se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.4.11, 5.4.12 y 5.4.13, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

En estos tests de comparaciones dobles no se observan diferencias significativas entre las medias de los ecm para los pares de tamaños de capa oculta 2 y 40, y 5 y 20, mientras que para el resto de los pares de tamaños de capa oculta las medias de sus ecm no pueden considerarse similares, resultado que se obtiene sobre los tres conjuntos considerados.

CGFR (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.968	5	15	0.000
	Lambda Wilks	0.032	5	15	0.000
	Traza Hotelling	30.039	5	15	0.000
	Raíz mayor Roy	30.039	5	15	0.000
ecm conj. valid.	Traza Pillai	0.967	5	15	0.000
	Lambda Wilks	0.033	5	15	0.000
	Traza Hotelling	29.374	5	15	0.000
	Raíz mayor Roy	29.374	5	15	0.000
ecm conj. test.	Traza Pillai	0.966	5	15	0.000
	Lambda Wilks	0.034	5	15	0.000
	Traza Hotelling	28.806	5	15	0.000
	Raíz mayor Roy	28.806	5	15	0.000

Tabla 5.4.10. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento CGFR.

En el caso del método de entrenamiento GD, sobre los tres conjuntos se rechaza la hipótesis de esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.4.14.

En los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6) se ve claramente la diferencia entre las medias de los ecm para todos los tamaños de capa oculta, lo que también puede observarse en los tests realizados, cuyos valores aparecen en la Tabla 5.4.15, ya que se rechaza la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	17.842	1.425	[13.063, 22.621]	0.000
2 y 10	21.728	1.363	[17.156, 26.300]	0.000
2 y 20	19.348	1.345	[14.837, 23.860]	0.000
2 y 30	8.510	2.032	[1.695, 15.326]	0.007
2 y 40	2.506	1.553	[-2.702, 7.713]	1.000
5 y 2	-17.842	1.425	[-22.621, -13.063]	0.000
5 y 10	3.886	0.729	[1.441, 6.331]	0.001
5 y 20	1.506	0.764	[-1.058, 4.070]	0.953
5 y 30	-9.332	1.442	[-14.168, -4.495]	0.000
5 y 40	-15.337	1.312	[-19.737, -10.936]	0.000
10 y 2	-21.728	1.363	[-26.300, -17.156]	0.000
10 y 5	-3.886	0.729	[-6.331, -1.441]	0.001
10 y 20	-2.380	0.499	[-4.052, -0.707]	0.002
10 y 30	-13.218	1.223	[-17.321, -9.114]	0.000
10 y 40	-19.223	1.098	[-22.906, -15.539]	0.000
20 y 2	-19.348	1.345	[-23.860, -14.837]	0.000
20 y 5	-1.506	0.764	[-4.070, 1.058]	0.953
20 y 10	2.380	0.499	[0.707, 4.052]	0.002
20 y 30	-10.838	1.197	[-14.854, -6.822]	0.000
20 y 40	-16.843	1.162	[-20.739, -12.946]	0.000
30 y 2	-8.510	2.032	[-15.326, -1.695]	0.007
30 y 5	9.332	1.442	[4.495, 14.168]	0.000
30 y 10	13.218	1.223	[9.114, 17.321]	0.000
30 y 20	10.838	1.197	[6.822, 14.854]	0.000
30 y 40	-6.005	1.307	[-10.388, -1.621]	0.003
40 y 2	-2.506	1.553	[-7.713, 2.702]	1.000
40 y 5	15.337	1.312	[10.936, 19.737]	0.000
40 y 10	19.223	1.098	[15.539, 22.906]	0.000
40 y 20	16.843	1.162	[12.946, 20.739]	0.000
40 y 30	6.005	1.307	[1.621, 10.388]	0.003

Tabla 5.4.11. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento CGFR sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	17.681	1.380	[13.051, 22.311]	0.000
2 y 10	21.610	1.333	[17.141, 26.079]	0.000
2 y 20	19.144	1.315	[14.735, 23.553]	0.000
2 y 30	8.369	1.920	[1.930, 14.808]	0.005
2 y 40	2.445	1.508	[-2.614, 7.504]	1.000
5 y 2	-17.681	1.380	[-22.311, -13.051]	0.000
5 y 10	3.929	0.740	[1.449, 6.410]	0.001
5 y 20	1.463	0.769	[-1.117, 4.042]	1.000
5 y 30	-9.312	1.394	[-13.987, -4.637]	0.000
5 y 40	-15.236	1.290	[-19.563, -10.909]	0.000
10 y 2	-21.610	1.333	[-26.079, -17.141]	0.000
10 y 5	-3.929	0.740	[-6.410, -1.449]	0.001
10 y 20	-2.466	0.506	[-4.162, -0.771]	0.002
10 y 30	-13.241	1.191	[-17.235, -9.247]	0.000
10 y 40	-19.165	1.083	[-22.797, -15.533]	0.000
20 y 2	-19.144	1.315	[-23.533, -14.735]	0.000
20 y 5	-1.463	0.769	[-4.042, 1.117]	1.000
20 y 10	2.466	0.506	[0.771, 4.162]	0.002
20 y 30	-10.775	1.175	[-14.716, -6.833]	0.000
20 y 40	-16.699	1.132	[-20.497, -12.900]	0.000
30 y 2	-8.369	1.920	[-14.808, -1.930]	0.005
30 y 5	9.312	1.394	[4.637, 13.987]	0.000
30 y 10	13.241	1.191	[9.247, 17.235]	0.000
30 y 20	10.775	1.175	[6.833, 14.716]	0.000
30 y 40	-5.924	1.305	[-10.302, -1.546]	0.003
40 y 2	-2.445	1.508	[-7.504, 2.614]	1.000
40 y 5	15.236	1.290	[10.909, 19.563]	0.000
40 y 10	19.165	1.083	[15.533, 22.797]	0.000
40 y 20	16.699	1.132	[12.900, 20.497]	0.000
40 y 30	5.924	1.305	[1.546, 10.302]	0.003

Tabla 5.4.12. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento CGFR sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	17.851	1.414	[13.109, 22.593]	0.000
2 y 10	21.706	1.354	[17.165, 26.247]	0.000
2 y 20	19.292	1.355	[14.747, 23.838]	0.000
2 y 30	8.403	1.992	[1.722, 15.084]	0.007
2 y 40	2.382	1.544	[-2.798, 7.562]	1.000
5 y 2	-17.851	1.414	[-22.593, -13.109]	0.000
5 y 10	3.855	0.727	[1.415, 6.295]	0.001
5 y 20	1.441	0.751	[-1.078, 3.960]	1.000
5 y 30	-9.448	1.427	[-14.233, -4.663]	0.000
5 y 40	-15.469	1.317	[-19.888, -11.050]	0.000
10 y 2	-21.706	1.354	[-26.247, -17.165]	0.000
10 y 5	-3.855	0.727	[-6.295, -1.415]	0.001
10 y 20	-2.414	0.493	[-4.067, -0.761]	0.002
10 y 30	-13.303	1.218	[-17.390, -9.217]	0.000
10 y 40	-19.324	1.130	[-23.114, -15.534]	0.000
20 y 2	-19.292	1.355	[-23.838, -14.747]	0.000
20 y 5	-1.441	0.751	[-3.960, 1.078]	1.000
20 y 10	2.414	0.493	[0.761, 4.067]	0.002
20 y 30	-10.889	1.185	[-14.864, -6.915]	0.000
20 y 40	-16.910	1.184	[-20.881, -12.939]	0.000
30 y 2	-8.403	1.992	[-15.084, -1.722]	0.007
30 y 5	9.448	1.427	[4.663, 14.233]	0.000
30 y 10	13.303	1.218	[9.217, 17.390]	0.000
30 y 20	10.889	1.185	[6.915, 14.864]	0.000
30 y 40	-6.021	1.343	[-10.526, -1.516]	0.004
40 y 2	-2.382	1.544	[-7.562, 2.798]	1.000
40 y 5	15.469	1.317	[11.050, 19.888]	0.000
40 y 10	19.324	1.130	[15.534, 23.114]	0.000
40 y 20	16.910	1.184	[12.939, 20.881]	0.000
40 y 30	6.021	1.343	[1.516, 10.526]	0.004

Tabla 5.4.13. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento CGFR sobre el conjunto test.

GD (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.002	110.475	14	0.000
ecm conj. valid.	0.002	109.386	14	0.000
ecm conj. test	0.002	110.924	14	0.000

Tabla 5.4.14. Prueba de esfericidad de Mauchly para el método de entrenamiento GD.

GD (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.974	5	15	0.000
	Lambda Wilks	0.026	5	15	0.000
	Traza Hotelling	37.851	5	15	0.000
	Raíz mayor Roy	37.851	5	15	0.000
ecm conj. valid.	Traza Pillai	0.974	5	15	0.000
	Lambda Wilks	0.026	5	15	0.000
	Traza Hotelling	38.098	5	15	0.000
	Raíz mayor Roy	38.098	5	15	0.000
ecm conj. test.	Traza Pillai	0.975	5	15	0.000
	Lambda Wilks	0.025	5	15	0.000
	Traza Hotelling	38.805	5	15	0.000
	Raíz mayor Roy	38.805	5	15	0.000

Tabla 5.4.15. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento GD.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento GD se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.4.16, 5.4.17 y 5.4.18, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-518.577	76.588	[-775.457, -261.696]	0.000
2 y 10	-2109.143	155.788	[-2631.662, -1586.624]	0.000
2 y 20	-5651.868	519.786	[-7395.248, -3908.488]	0.000
2 y 30	-5849.292	653.703	[-8041.834, -3656.750]	0.000
2 y 40	-8894.129	847.085	[-11735.282, -6052.976]	0.000
5 y 2	518.577	76.588	[261.696, 775.457]	0.000
5 y 10	-1590.566	140.445	[-2061.623, -1119.509]	0.000
5 y 20	-5133.292	542.571	[-6953.094, -3313.489]	0.000
5 y 30	-5330.716	649.958	[-7510.700, -3150.732]	0.000
5 y 40	-8375.552	869.222	[-11290.952, -5460.152]	0.000
10 y 2	2109.143	155.788	[1586.624, 2631.662]	0.000
10 y 5	1590.566	140.445	[1119.509, 2061.623]	0.000
10 y 20	-3542.726	599.885	[-5554.762, -1530.689]	0.000
10 y 30	-3740.150	680.643	[-6023.052, -1457.248]	0.000
10 y 40	-6784.986	904.124	[-9817.451, -3752.520]	0.000
20 y 2	5651.868	519.786	[3908.488, 7395.248]	0.000
20 y 5	5133.292	542.571	[3313.489, 6953.094]	0.000
20 y 10	3542.726	599.885	[1530.689, 5554.762]	0.000
20 y 30	-197.424	651.460	[-2382.445, 1987.596]	1.000
20 y 40	-3242.260	903.837	[-6273.764, -210.757]	0.029
30 y 2	5849.292	653.703	[3656.750, 8041.834]	0.000
30 y 5	5330.716	649.958	[3150.732, 7510.700]	0.000
30 y 10	3740.150	680.643	[1457.248, 6023.052]	0.000
30 y 20	197.424	651.460	[-1987.596, 2382.445]	1.000
30 y 40	-3044.836	837.452	[-5853.682, -235.990]	0.026
40 y 2	8894.129	847.085	[6052.976, 11735.282]	0.000
40 y 5	8375.552	869.222	[5460.152, 11290.952]	0.000
40 y 10	6784.986	904.124	[3752.520, 9817.451]	0.000
40 y 20	3242.260	903.837	[210.757, 6273.764]	0.029
40 y 30	3044.836	837.452	[235.990, 5853.682]	0.026

Tabla 5.4.16. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento GD sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-494.407	70.882	[-732.150, -256.665]	0.000
2 y 10	-2043.453	154.601	[-2561.988, -1524.917]	0.000
2 y 20	-5257.362	472.072	[-6840.708, -3674.015]	0.000
2 y 30	-5622.128	637.728	[-7761.092, -3483.165]	0.000
2 y 40	-8424.359	782.855	[-11050.083, -5798.635]	0.000
5 y 2	494.407	70.882	[256.665, 732.150]	0.000
5 y 10	-1549.045	138.363	[-2013.121, -1084.970]	0.000
5 y 20	-4762.954	492.764	[-6415.702, -3110.207]	0.000
5 y 30	-5127.721	631.952	[-7247.312, -3008.130]	0.000
5 y 40	-7929.952	801.652	[-10618.720, -5241.184]	0.000
10 y 2	2043.453	154.601	[1524.917, 2561.988]	0.000
10 y 5	1549.045	138.363	[1084.970, 2013.121]	0.000
10 y 20	-3213.909	555.565	[-5077.293, -1350.525]	0.000
10 y 30	-3578.676	663.519	[-5804.141, -1353.210]	0.000
10 y 40	-6380.907	837.602	[-9190.256, -3571.558]	0.000
20 y 2	5257.362	472.072	[3674.015, 6840.708]	0.000
20 y 5	4762.954	492.764	[3110.207, 6415.702]	0.000
20 y 10	3213.909	555.565	[1350.525, 5077.293]	0.000
20 y 30	-364.767	626.961	[-2467.617, 1738.084]	1.000
20 y 40	-3166.998	832.438	[-5959.023, -374.972]	0.018
30 y 2	5622.128	637.728	[3483.165, 7761.092]	0.000
30 y 5	5127.721	631.952	[3008.130, 7247.312]	0.000
30 y 10	3578.676	663.519	[1353.210, 5804.141]	0.000
30 y 20	364.767	626.961	[-1738.084, 2467.617]	1.000
30 y 40	-2802.231	795.514	[-5470.413, -134.049]	0.034
40 y 2	8424.359	782.855	[5798.635, 11050.083]	0.000
40 y 5	7929.952	801.652	[5241.184, 10618.720]	0.000
40 y 10	6380.907	837.602	[3571.558, 9190.256]	0.000
40 y 20	3166.998	832.438	[374.972, 5959.023]	0.018
40 y 30	2802.231	795.514	[134.049, 5470.413]	0.034

Tabla 5.4.17. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento GD sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	-524.573	78.084	[-786.469, -262.678]	0.000
2 y 10	-2112.243	152.492	[-2623.706, -1600.780]	0.000
2 y 20	-5645.618	519.903	[-7389.390, -3901.845]	0.000
2 y 30	-5881.625	662.776	[-8104.600, -3658.650]	0.000
2 y 40	-8874.149	852.731	[-11734.241, -6014.057]	0.000
5 y 2	524.573	78.084	[262.678, 786.469]	0.000
5 y 10	-1587.670	138.782	[-2053.150, -1122.189]	0.000
5 y 20	-5121.044	542.894	[-6941.930, -3300.159]	0.000
5 y 30	-5357.052	659.512	[-7569.079, -3145.025]	0.000
5 y 40	-8349.575	876.022	[-11287.786, -5411.365]	0.000
10 y 2	2112.243	152.492	[1600.780, 2623.706]	0.000
10 y 5	1587.670	138.782	[1122.189, 2053.150]	0.000
10 y 20	-3533.375	598.706	[-5541.457, -1525.292]	0.000
10 y 30	-3769.382	689.163	[-6080.860, -1457.904]	0.000
10 y 40	-6761.906	907.425	[-9805.443, -3718.369]	0.000
20 y 2	5645.618	519.903	[3901.845, 7389.390]	0.000
20 y 5	5121.044	542.894	[3300.159, 6941.930]	0.000
20 y 10	3533.375	598.706	[1525.292, 5541.457]	0.000
20 y 30	-236.007	654.561	[-2431.430, 1959.415]	1.000
20 y 40	-3228.531	911.123	[-6284.471, -172.591]	0.033
30 y 2	5881.625	662.776	[3658.650, 8104.600]	0.000
30 y 5	5357.052	659.512	[3145.025, 7569.079]	0.000
30 y 10	3769.382	689.163	[1457.904, 6080.860]	0.000
30 y 20	236.007	654.561	[-1959.415, 2431.430]	1.000
30 y 40	-2992.524	852.378	[-5851.429, -133.618]	0.035
40 y 2	8874.149	852.731	[6014.057, 11734.241]	0.000
40 y 5	8349.575	876.022	[5411.365, 11287.786]	0.000
40 y 10	6761.906	907.425	[3718.369, 9805.443]	0.000
40 y 20	3228.531	911.123	[172.591, 6284.471]	0.033
40 y 30	2992.524	852.378	[133.618, 5851.429]	0.035

Tabla 5.4.18. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento GD sobre el conjunto test.

Tal y como queda claramente reflejado en los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6), en estos tests de comparaciones dobles se observan diferencias significativas entre las medias de los ecm para cada par de tamaños de la capa oculta, salvo para los tamaños 20 y 30, para los que no puede rechazarse la hipótesis de igualdad entre las medias de los ecm, resultado que se obtiene sobre los tres conjuntos considerados.

Para el método de entrenamiento LM, que ya ha sido determinado como el mejor de todos los métodos de entrenamiento empleados, sobre los tres conjuntos se rechaza la hipótesis de esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.4.19.

En los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6) queda claramente reflejada la diferencia entre las medias de los ecm para todos los tamaños de capa oculta, lo que también puede observarse en los tests realizados, cuyos valores aparecen en la Tabla 5.4.20, ya que se rechaza la igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

LM (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.004	94.053	14	0.000
ecm conj. valid.	0.003	98.991	14	0.000
ecm conj. test	0.004	93.853	14	0.000

Tabla 5.4.19. Prueba de esfericidad de Mauchly para el método de entrenamiento LM.

LM (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.998	5	15	0.000
	Lambda Wilks	0.002	5	15	0.000
	Traza Hotelling	503.334	5	15	0.000
	Raíz mayor Roy	503.334	5	15	0.000
ecm conj. valid.	Traza Pillai	0.998	5	15	0.000
	Lambda Wilks	0.002	5	15	0.000
	Traza Hotelling	458.716	5	15	0.000
	Raíz mayor Roy	458.716	5	15	0.000
ecm conj. test.	Traza Pillai	0.998	5	15	0.000
	Lambda Wilks	0.002	5	15	0.000
	Traza Hotelling	406.531	5	15	0.000
	Raíz mayor Roy	406.531	5	15	0.000

Tabla 5.4.20. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento LM.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento LM se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.4.21, 5.4.22 y 5.4.23, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

Tal y como queda claramente reflejado en los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6), en estos tests de comparaciones dobles se observan diferencias significativas entre las medias de los ecm para todos los pares de tamaños de la capa oculta, resultado que se obtiene sobre los tres conjuntos considerados.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	3.847E-02	0.002	[3.319E-02, 4.375E-02]	0.000
2 y 10	5.661E-02	0.001	[5.278E-02, 6.045E-02]	0.000
2 y 20	6.698E-02	0.001	[6.305E-02, 7.091E-02]	0.000
2 y 30	6.958E-02	0.001	[6.574E-02, 7.341E-02]	0.000
2 y 40	7.104E-02	0.001	[6.747E-02, 7.461E-02]	0.000
5 y 2	-3.847E-02	0.002	[-4.375E-02, -3.319E-02]	0.000
5 y 10	1.815E-02	0.001	[1.460E-02, 2.170E-02]	0.000
5 y 20	2.851E-02	0.001	[2.471E-02, 3.232E-02]	0.000
5 y 30	3.111E-02	0.001	[2.719E-02, 3.502E-02]	0.000
5 y 40	3.257E-02	0.001	[2.882E-02, 3.632E-02]	0.000
10 y 2	-5.661E-02	0.001	[-6.045E-02, -5.278E-02]	0.000
10 y 5	-1.815E-02	0.001	[-2.170E-02, -1.460E-02]	0.000
10 y 20	1.037E-02	0.000	[8.727E-03, 1.201E-02]	0.000
10 y 30	1.296E-02	0.000	[1.136E-02, 1.456E-02]	0.000
10 y 40	1.442E-02	0.000	[1.289E-02, 1.595E-02]	0.000
20 y 2	-6.698E-02	0.001	[-7.091E-02, -6.305E-02]	0.000
20 y 5	-2.851E-02	0.001	[-3.232E-02, -2.471E-02]	0.000
20 y 10	-1.037E-02	0.000	[-1.201E-02, -8.727E-03]	0.000
20 y 30	2.594E-03	0.000	[1.799E-03, 3.389E-03]	0.000
20 y 40	4.055E-03	0.000	[3.348E-03, 4.762E-03]	0.000
30 y 2	-6.958E-02	0.001	[-7.341E-02, -6.574E-02]	0.000
30 y 5	-3.111E-02	0.001	[-3.502E-02, -2.719E-02]	0.000
30 y 10	-1.296E-02	0.000	[-1.456E-02, -1.136E-02]	0.000
30 y 20	-2.594E-03	0.000	[-3.389E-03, -1.799E-03]	0.000
30 y 40	1.461E-03	0.000	[4.383E-04, 2.484E-03]	0.002
40 y 2	-7.104E-02	0.001	[-7.461E-02, -6.747E-02]	0.000
40 y 5	-3.257E-02	0.001	[-3.632E-02, -2.882E-02]	0.000
40 y 10	-1.442E-02	0.000	[-1.595E-02, -1.289E-02]	0.000
40 y 20	-4.055E-03	0.000	[-4.762E-03, -3.348E-03]	0.000
40 y 30	-1.461E-03	0.000	[-2.484E-03, -4.383E-04]	0.002

Tabla 5.4.21. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento LM sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	3.828E-02	0.002	[3.284E-02, 4.373E-02]	0.000
2 y 10	5.639E-02	0.001	[5.240E-02, 6.038E-02]	0.000
2 y 20	6.682E-02	0.001	[6.259E-02, 7.106E-02]	0.000
2 y 30	6.923E-02	0.001	[6.522E-02, 7.323E-02]	0.000
2 y 40	7.064E-02	0.001	[6.672E-02, 7.456E-02]	0.000
5 y 2	-3.828E-02	0.002	[-4.373E-02, -3.284E-02]	0.000
5 y 10	1.811E-02	0.001	[1.447E-02, 2.175E-02]	0.000
5 y 20	2.854E-02	0.001	[2.463E-02, 3.246E-02]	0.000
5 y 30	3.094E-02	0.001	[2.694E-02, 3.495E-02]	0.000
5 y 40	3.236E-02	0.001	[2.848E-02, 3.624E-02]	0.000
10 y 2	-5.639E-02	0.001	[-6.038E-02, -5.240E-02]	0.000
10 y 5	-1.811E-02	0.001	[-2.175E-02, -1.447E-02]	0.000
10 y 20	1.043E-02	0.001	[8.730E-03, 1.214E-02]	0.000
10 y 30	1.284E-02	0.000	[1.123E-02, 1.444E-02]	0.000
10 y 40	1.425E-02	0.000	[1.259E-02, 1.591E-02]	0.000
20 y 2	-6.682E-02	0.001	[-7.106E-02, -6.259E-02]	0.000
20 y 5	-2.854E-02	0.001	[-3.246E-02, -2.463E-02]	0.000
20 y 10	-1.043E-02	0.001	[-1.214E-02, -8.730E-03]	0.000
20 y 30	2.403E-03	0.000	[1.594E-03, 3.212E-03]	0.000
20 y 40	3.816E-03	0.000	[3.146E-03, 4.487E-03]	0.000
30 y 2	-6.923E-02	0.001	[-7.323E-02, -6.522E-02]	0.000
30 y 5	-3.094E-02	0.001	[-3.495E-02, -2.694E-02]	0.000
30 y 10	-1.284E-02	0.000	[-1.444E-02, -1.123E-02]	0.000
30 y 20	-2.403E-03	0.000	[-3.212E-03, -1.594E-03]	0.000
30 y 40	1.413E-03	0.000	[3.877E-04, 2.439E-03]	0.003
40 y 2	-7.064E-02	0.001	[-7.456E-02, -6.672E-02]	0.000
40 y 5	-3.236E-02	0.001	[-3.624E-02, -2.848E-02]	0.000
40 y 10	-1.425E-02	0.000	[-1.591E-02, -1.259E-02]	0.000
40 y 20	-3.816E-03	0.000	[-4.487E-03, -3.146E-03]	0.000
40 y 30	-1.413E-03	0.000	[-2.439E-03, -3.877E-04]	0.003

Tabla 5.4.22. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento LM sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	3.862E-02	0.002	[3.319E-02, 4.405E-02]	0.000
2 y 10	5.673E-02	0.001	[5.270E-02, 6.075E-02]	0.000
2 y 20	6.718E-02	0.001	[6.300E-02, 7.135E-02]	0.000
2 y 30	6.969E-02	0.001	[6.571E-02, 7.366E-02]	0.000
2 y 40	7.107E-02	0.001	[6.711E-02, 7.503E-02]	0.000
5 y 2	-3.862E-02	0.002	[-4.405E-02, -3.319E-02]	0.000
5 y 10	1.810E-02	0.001	[1.463E-02, 2.158E-02]	0.000
5 y 20	2.855E-02	0.001	[2.492E-02, 3.219E-02]	0.000
5 y 30	3.106E-02	0.001	[2.743E-02, 3.469E-02]	0.000
5 y 40	3.245E-02	0.001	[2.886E-02, 3.604E-02]	0.000
10 y 2	-5.673E-02	0.001	[-6.075E-02, -5.270E-02]	0.000
10 y 5	-1.810E-02	0.001	[-2.158E-02, -1.463E-02]	0.000
10 y 20	1.045E-02	0.001	[8.721E-03, 1.218E-02]	0.000
10 y 30	1.296E-02	0.000	[1.140E-02, 1.452E-02]	0.000
10 y 40	1.435E-02	0.000	[1.267E-02, 1.602E-02]	0.000
20 y 2	-6.718E-02	0.001	[-7.135E-02, -6.300E-02]	0.000
20 y 5	-2.855E-02	0.001	[-3.219E-02, -2.492E-02]	0.000
20 y 10	-1.045E-02	0.001	[-1.218E-02, -8.721E-03]	0.000
20 y 30	2.510E-03	0.000	[1.754E-03, 3.266E-03]	0.000
20 y 40	3.898E-03	0.000	[3.192E-03, 4.603E-03]	0.000
30 y 2	-6.969E-02	0.001	[-7.366E-02, -6.571E-02]	0.000
30 y 5	-3.106E-02	0.001	[-3.469E-02, -2.743E-02]	0.000
30 y 10	-1.296E-02	0.000	[-1.452E-02, -1.140E-02]	0.000
30 y 20	-2.510E-03	0.000	[-3.266E-03, -1.754E-03]	0.000
30 y 40	1.388E-03	0.000	[3.782E-04, 2.397E-03]	0.003
40 y 2	-7.107E-02	0.001	[-7.503E-02, -6.711E-02]	0.000
40 y 5	-3.245E-02	0.001	[-3.604E-02, -2.886E-02]	0.000
40 y 10	-1.435E-02	0.000	[-1.602E-02, -1.267E-02]	0.000
40 y 20	-3.898E-03	0.000	[-4.603E-03, -3.192E-03]	0.000
40 y 30	-1.388E-03	0.000	[-2.397E-03, -3.782E-04]	0.003

Tabla 5.4.23. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento LM sobre el conjunto test.

Y por último, para el método de entrenamiento RP, sobre los tres conjuntos se rechaza la hipótesis de esfericidad de la matriz de varianzas-covarianzas para un nivel de significación de 0.05, tal y como puede apreciarse en la Tabla 5.4.24.

En los gráficos de caja y bigotes (Figuras 5.4.4, 5.4.5 y 5.4.6) no puede apreciarse si las medias de los ecm para todos los tamaños de capa oculta pueden considerarse iguales, por lo que se realizan los contrastes correspondientes, cuyos valores aparecen en la Tabla 5.4.25, y en los que se observa que se rechaza la hipótesis de igualdad entre las medias de los ecm para todos los tamaños de capa oculta sobre los tres conjuntos y con un nivel de significación de 0.05.

RP (Factor tamaño)	W de Mauchly	Aprox. χ^2	g.l.	p-valor
ecm conj. entren.	0.000	157.169	14	0.000
ecm conj. valid.	0.000	157.545	14	0.000
ecm conj. test	0.000	156.975	14	0.000

Tabla 5.4.24. Prueba de esfericidad de Mauchly para el método de entrenamiento RP.

Con la finalidad de descubrir la razón por la que se rechaza la hipótesis nula, para el método de entrenamiento RP se realizan las comparaciones por pares de tamaños para las medias de los ecm sobre los conjuntos de entrenamiento, validación y test, cuyos resultados aparecen en las Tablas 5.4.26, 5.4.27 y 5.4.28, respectivamente.

El ajuste para realizar las comparaciones dos a dos se ha realizado mediante la corrección de Bonferroni, y además se proporcionan en dichas tablas los intervalos de confianza al 95% para la diferencia entre las medias de los ecm para cada par de tamaños de capa oculta.

RP (Factor tamaño)	Estadístico	Valor	g.l. H_0	g.l. error	p-valor
ecm conj. entren.	Traza Pillai	0.731	5	15	0.001
	Lambda Wilks	0.269	5	15	0.001
	Traza Hotelling	2.717	5	15	0.001
	Raíz mayor Roy	2.717	5	15	0.001
ecm conj. valid.	Traza Pillai	0.734	5	15	0.001
	Lambda Wilks	0.266	5	15	0.001
	Traza Hotelling	2.759	5	15	0.001
	Raíz mayor Roy	2.759	5	15	0.001
ecm conj. test.	Traza Pillai	0.732	5	15	0.001
	Lambda Wilks	0.268	5	15	0.001
	Traza Hotelling	2.734	5	15	0.001
	Raíz mayor Roy	2.734	5	15	0.001

Tabla 5.4.25. Contrastes para la hipótesis de igualdad de medias de los ecm entre los tamaños de capa oculta para el método de entrenamiento RP.

En estos tests de comparaciones dobles se observan diferencias significativas entre las medias de los ecm para los pares de tamaños de la capa oculta 2 y 40, y 5 y 40, mientras que para el resto de pares de tamaños no puede rechazarse la hipótesis de igualdad entre las medias de sus ecm, resultado que se obtiene sobre los tres conjuntos considerados.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	225.832	200.988	[-448.290, 899.953]	1.000
2 y 10	-344.398	360.080	[-1552.120, 863.324]	1.000
2 y 20	-2624.262	1191.421	[-6620.331, 1371.808]	0.603
2 y 30	-5420.372	2303.171	[-13145.292, 2304.548]	0.443
2 y 40	-10956.682	3216.235	[-21744.050, -169.314]	0.044
5 y 2	-225.832	200.988	[-899.953, 448.290]	1.000
5 y 10	-570.230	289.023	[-1539.622, 399.163]	0.948
5 y 20	-2850.093	1153.643	[-6719.453, 1019.267]	0.347
5 y 30	-5646.204	2370.724	[-13597.699, 2305.293]	0.418
5 y 40	-11182.514	3196.310	[-21903.052, -461.976]	0.036
10 y 2	344.398	360.080	[-863.325, 1552.121]	1.000
10 y 5	570.230	289.023	[-399.164, 1539.623]	0.948
10 y 20	-2279.864	1188.256	[-6265.313, 1705.586]	1.000
10 y 30	-5075.974	2393.747	[-13104.685, 2952.738]	0.710
10 y 40	-10612.284	3212.064	[-21385.658, 161.090]	0.056
20 y 2	2624.263	1191.421	[-1371.808, 6620.331]	0.603
20 y 5	2850.094	1153.643	[-1019.266, 6719.453]	0.347
20 y 10	2279.865	1188.255	[-1705.585, 6265.313]	1.000
20 y 30	-2796.110	2842.320	[-12329.355, 6737.135]	1.000
20 y 40	-8332.420	3754.368	[-20924.705, 4259.865]	0.582
30 y 2	5420.372	2303.171	[-2304.548, 13145.292]	0.443
30 y 5	5646.204	2370.724	[-2305.292, 13597.699]	0.418
30 y 10	5075.974	2393.746	[-2952.737, 13104.685]	0.710
30 y 20	2796.110	2842.320	[-6737.135, 12329.355]	1.000
30 y 40	-5536.310	4310.444	[-19993.694, 8921.074]	1.000
40 y 2	10956.682	3216.235	[169.314, 21744.050]	0.044
40 y 5	11182.514	3196.310	[461.975, 21903.052]	0.036
40 y 10	10612.284	3212.063	[-161.090, 21385.658]	0.056
40 y 20	8332.420	3754.368	[-4259.865, 20924.705]	0.582
40 y 30	5536.310	4310.444	[-8921.074, 19993.694]	1.000

Tabla 5.4.26. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento RP sobre el conjunto de entrenamiento.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	231.048	204.173	[−453.755, 915.852]	1.000
2 y 10	−333.519	357.753	[−1533.437, 866.398]	1.000
2 y 20	−2589.904	1183.772	[−6560.318, 1380.511]	0.621
2 y 30	−5382.477	2303.987	[−13110.134, 2345.179]	0.459
2 y 40	−10917.098	3213.146	[−21694.107, −140.090]	0.045
5 y 2	−231.048	204.173	[−915.852, 453.755]	1.000
5 y 10	−564.568	283.218	[−1514.492, 385.357]	0.912
5 y 20	−2820.952	1144.636	[−6660.102, 1018.198]	0.351
5 y 30	−5613.526	2372.774	[−13571.895, 2344.844]	0.432
5 y 40	−11148.147	3192.448	[−21855.734, −440.560]	0.037
10 y 2	333.519	357.753	[−866.399, 1533.437]	1.000
10 y 5	564.568	283.218	[−385.358, 1514.492]	0.912
10 y 20	−2256.384	1180.613	[−6216.200, 1703.431]	1.000
10 y 30	−5048.958	2397.319	[−13089.651, 2991.735]	0.731
10 y 40	−10583.579	3211.367	[−21354.615, 187.457]	0.057
20 y 2	2589.904	1183.772	[−1380.511, 6560.318]	0.621
20 y 5	2820.952	1144.636	[−1018.198, 6660.102]	0.351
20 y 10	2256.384	1180.612	[−1703.431, 6216.200]	1.000
20 y 30	−2792.574	2839.712	[−12317.069, 6731.922]	1.000
20 y 40	−8327.195	3748.151	[−20898.628, 4244.238]	0.580
30 y 2	5382.477	2303.987	[−2345.179, 13110.134]	0.459
30 y 5	5613.526	2372.774	[−2344.844, 13571.895]	0.432
30 y 10	5048.958	2397.318	[−2991.735, 13089.651]	0.731
30 y 20	2792.574	2839.712	[−6731.922, 12317.069]	1.000
30 y 40	−5534.621	4306.811	[−19979.821, 8910.578]	1.000
40 y 2	10917.098	3213.146	[140.090, 21694.107]	0.045
40 y 5	11148.147	3192.448	[440.560, 21855.734]	0.037
40 y 10	10583.579	3211.366	[−187.457, 21354.615]	0.057
40 y 20	8327.195	3748.151	[−4244.238, 20898.628]	0.580
40 y 30	5534.621	4306.811	[−8910.578, 19979.821]	1.000

Tabla 5.4.27. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento RP sobre el conjunto de validación.

Tamaños	Dif. medias	Error típico	IC dif. medias al 95% [Lím. Inf., Lím. Sup.]	p-valor
2 y 5	229.989	203.823	[-453.642, 913.619]	1.000
2 y 10	-346.821	364.218	[-1568.422, 874.781]	1.000
2 y 20	-2624.499	1189.104	[-6612.797, 1363.800]	0.597
2 y 30	-5422.444	2296.114	[-13123.693, 2278.805]	0.435
2 y 40	-11006.379	3240.981	[-21876.746, -136.011]	0.045
5 y 2	-229.989	203.823	[-913.619, 453.642]	1.000
5 y 10	-576.809	291.505	[-1554.526, 400.908]	0.938
5 y 20	-2854.487	1150.186	[-6712.251, 1003.277]	0.339
5 y 30	-5652.433	2364.793	[-13584.034, 2279.168]	0.410
5 y 40	-11236.367	3220.310	[-22037.402, -435.332]	0.037
10 y 2	346.821	364.218	[-874.781, 1568.422]	1.000
10 y 5	576.809	291.505	[-400.908, 1554.526]	0.938
10 y 20	-2277.678	1184.996	[-6252.198, 1696.842]	1.000
10 y 30	-5075.623	2388.841	[-13087.883, 2936.636]	0.704
10 y 40	-10659.558	3237.109	[-21516.938, 197.822]	0.057
20 y 2	2624.499	1189.104	[-1363.800, 6612.797]	0.597
20 y 5	2854.487	1150.186	[-1003.277, 6712.251]	0.339
20 y 10	2277.678	1184.996	[-1696.842, 6252.198]	1.000
20 y 30	-2797.945	2836.237	[-12310.786, 6714.895]	1.000
20 y 40	-8381.880	3774.841	[-21042.833, 4279.073]	0.581
30 y 2	5422.444	2296.114	[-2278.805, 13123.693]	0.435
30 y 5	5652.433	2364.793	[-2279.168, 13584.034]	0.410
30 y 10	5075.623	2388.841	[-2936.636, 13087.883]	0.704
30 y 20	2797.945	2836.237	[-6714.895, 12310.786]	1.000
30 y 40	-5583.935	4328.212	[-20100.913, 8933.044]	1.000
40 y 2	11006.379	3240.981	[136.011, 21876.746]	0.045
40 y 5	11236.367	3220.310	[435.332, 22037.402]	0.037
40 y 10	10659.558	3237.109	[-197.822, 21516.938]	0.057
40 y 20	8381.880	3774.841	[-4279.073, 21042.833]	0.581
40 y 30	5583.935	4328.212	[-8933.044, 20100.913]	1.000

Tabla 5.4.28. Comparaciones por pares de tamaños de capa oculta para el método de entrenamiento RP sobre el conjunto test.

Por tanto, de todos estos gráficos, tablas y contrastes, se puede establecer la siguiente relación de optimalidad de resultados teniendo en cuenta los distintos métodos de entrenamiento y los diversos tamaños de capa oculta utilizados.

El mejor de los métodos de entrenamiento es el LM, incluso para cualquiera de los tamaños de capa oculta considerados, observándose que los resultados van mejorando progresivamente conforme aumenta el número de neuronas consideradas en la capa oculta, no pudiendo ser considerados similares para ninguno de los pares de tamaños de la capa oculta considerados, y los resultados óptimos se obtienen para un tamaño de capa oculta de 40 neuronas.

El método BFGS es el segundo mejor método de entrenamiento, pero sus resultados no pueden ser considerados similares para cualquier tamaño de capa oculta debido a las diferencias existentes entre los resultados obtenidos para los tamaños de capa oculta 2 y 5 con el resto de los tamaños de capa oculta, y a las diferencias existentes entre los tamaños de capa oculta 30 y 40, mientras que el resto de los tamaños de capa oculta comparados por pares sí que pueden ser considerados similares. Obsérvese en sus gráficos de caja y bigotes la gran diferencia existente entre las medias de los ecm para los tamaños 2 y 5 con respecto al resto de tamaños.

El tercer mejor método de entrenamiento es el CGFR, y sus resultados no pueden ser considerados similares para todos los tamaños de capa oculta. Únicamente pueden ser considerados similares los resultados para los tamaños de capa oculta 2 y 40, y para los tamaños 5 y 20, observándose una tendencia decreciente en la media de sus ecm hasta alcanzar el tamaño de capa oculta 10, para el que se obtienen los resultados óptimos de este método, y a partir de este tamaño, la media de los ecm comienza a crecer nuevamente.

En el cuarto lugar se encuentra el método RP, cuyos resultados no pueden ser considerados similares para todos los tamaños de capa oculta debido a las diferencias existentes entre los resultados obtenidos para los tamaños de capa oculta 2 y 5 con el tamaño de capa oculta 40, mientras que el resto de los tamaños de capa oculta comparados por pares sí que pueden ser considerados similares.

Y el peor de los métodos de entrenamiento empleados es el GD, cuyos resultados van empeorando progresivamente conforme aumenta el número de neuronas consideradas en la capa oculta, pudiendo ser considerados similares únicamente para los tamaños de capa oculta 20 y 30, mientras que para el resto de los pares de tamaños de la capa oculta considerados, sus resultados difieren.

Una vez determinada la red_información óptima como la obtenida mediante el método de entrenamiento de Levenberg-Marquardt y una estructura de perceptrón con una única capa oculta de 40 neuronas, una capa de entrada con once neuronas y una capa de salida formada por una única neurona, esto es, una estructura óptima (11:40:1), a continuación se procede a la búsqueda, entre todas las realizaciones llevadas a cabo, de aquella que proporciona el menor de los ecm. Para ello, se hallan los mínimos ecm sobre el conjunto test para el método de entrenamiento LM y tamaño de capa oculta 40, observándose que este mínimo se alcanza durante la iteración número 10, en su ejecución número 13, presentando un ecm sobre el conjunto test de **0.031667295**. Los valores de estos ecm mínimos se recogen en la Tabla 5.4.29.

Iteración	Ejecución	ecm conj. entren.	ecm conj. valid.	ecm conj. test
1	4	0.03057263	0.035270922	0.035194309
2	19	0.029032497	0.03595441	0.032526701
3	5	0.030120462	0.03237563	0.034487029
4	12	0.033128114	0.036631487	0.037503822
5	16	0.031010612	0.031763321	0.035806689
6	9	0.031347318	0.034990606	0.032396327
7	16	0.028070144	0.032240614	0.036599793
8	7	0.034780504	0.031760225	0.03775525
9	17	0.032408654	0.032641093	0.035114995
10	13	0.030455064	0.035545825	0.031667295
11	14	0.03317714	0.033688826	0.038939312
12	5	0.031849758	0.031533656	0.03235687
13	16	0.034033366	0.035197846	0.03550329
14	18	0.034332221	0.033116738	0.034314855
15	2	0.031358505	0.030150741	0.036683371
16	17	0.032063168	0.033579988	0.03853176
17	5	0.030436475	0.036680003	0.032389186
18	19	0.033038884	0.027505761	0.033184716
19	14	0.03208186	0.03463654	0.032206014
20	20	0.031645077	0.035323882	0.033496328

Tabla 5.4.29. Mínimos ecm sobre el conjunto test para el método de entrenamiento LM y tamaño de capa oculta 40.

5.5 COMPARACIÓN CON OTROS MODELOS DE PREDICCIÓN.

En este apartado se van a analizar las predicciones de la velocidad de sonido en el agua del mar obtenidas mediante las dos redes de neuronas artificiales construidas, comparándose con las que se obtienen mediante los métodos clásicos de cálculo de la velocidad de sonido en el agua del mar.

Además de las predicciones de la velocidad de sonido en el agua del mar obtenidas a partir de las redes de neuronas artificiales construidas y denominadas *red_algoritmo* y *red_información*, se ha procedido a su cálculo mediante las ecuaciones de Chen y Millero (1977) y de Del Grosso (1974), ambas actualizadas a la escala de temperatura internacional de 1990 por Wong y Zhu (1995), cuyos códigos de implementación aparecen en el Anexo 1, y que para su aplicación precisan de las variables temperatura, presión y salinidad del agua.

Las mediciones de temperatura y presión del agua se han realizado directamente mediante el perfilador de sonido "Sound Velocity, Temperature and Depth Profiler Model SVPlus" disponible a bordo del B/O Vizconde de Eza durante la realización de las campañas de investigación oceanográfica en las que se recogieron todos los datos necesarios para la realización de este estudio.

En cambio, las mediciones de la salinidad del agua necesarias para aplicar estos algoritmos se han obtenido de las bases de datos históricas recogidas en *Mediterranean Hydrological Atlas 1997*, (*MEDATLAS*). Tal y como ya se indicó en el apartado 1.5.1 sobre el cálculo de la velocidad de sonido, aunque la precisión de la ecuación de Chen y Millero es de 0.05 m/s, y la del algoritmo de Del Grosso es de 0.19 m/s, al no emplear los valores reales de la salinidad del agua, los errores de predicción aumentan de forma considerable, lo que queda claramente reflejado en los resultados que se analizan en este apartado.

En la Tabla 5.5.1 se recogen los estadísticos descriptivos de los valores reales de la velocidad de sonido y los de las distintas predicciones obtenidas con cada una de las técnicas empleadas.

Comparando los valores de los estadísticos descriptivos de la velocidad de sonido en el agua del mar con los de las predicciones obtenidas mediante las técnicas de redes de neuronas artificiales se observan similitudes entre ellos, aunque sus valores máximos superan ligeramente en ambos casos a los valores máximos reales, pero puede afirmarse que los estadísticos descriptivos

de dichas variables pueden considerarse similares.

En cambio, si se analizan los valores de los estadísticos descriptivos de la velocidad de sonido con los de las predicciones obtenidas mediante los procedimientos clásicos de cálculo de la velocidad de sonido, se observan grandes diferencias en todos sus valores, sobre todo para los valores de la media, el mínimo y la desviación típica que superan de forma alarmante a los estadísticos correspondientes a la velocidad de sonido en el agua del mar, por lo que los valores de los estadísticos descriptivos de las predicciones obtenidas mediante los procedimientos clásicos de cálculo de la velocidad de sonido no sirven para caracterizar con precisión y fiabilidad a la velocidad de sonido en el agua del mar.

Variable	Media	Mínimo	Máximo	Desv. típica
Velocidad de sonido	1515.807	1507.23	1542.58	7.650
Pred. red_algoritmo	1515.807	1507.53	1542.79	7.642
Pred. red_información	1515.809	1507.27	1542.99	7.653
Pred. Chen y Millero	1533.704	1511.25	1668.58	24.112
Pred. Del Grosso	1533.631	1511.21	1668.87	24.144

Tabla 5.5.1. Estadísticos descriptivos de los valores reales de la velocidad de sonido y de las distintas predicciones obtenidas con cada una de las técnicas empleadas.

Las representaciones gráficas de los valores reales de la velocidad de sonido en el agua del mar frente a las predicciones obtenidas con cada una de las técnicas empleadas se muestran en la Figura 5.5.1, observándose una tendencia lineal en los gráficos correspondientes a las estimaciones obtenidas mediante las redes de neuronas artificiales construidas, mientras que esa tendencia lineal deseada no se obtiene con los procedimientos clásicos de cálculo de la velocidad de sonido, lo que nuevamente indica la misma conclusión ya enunciada.

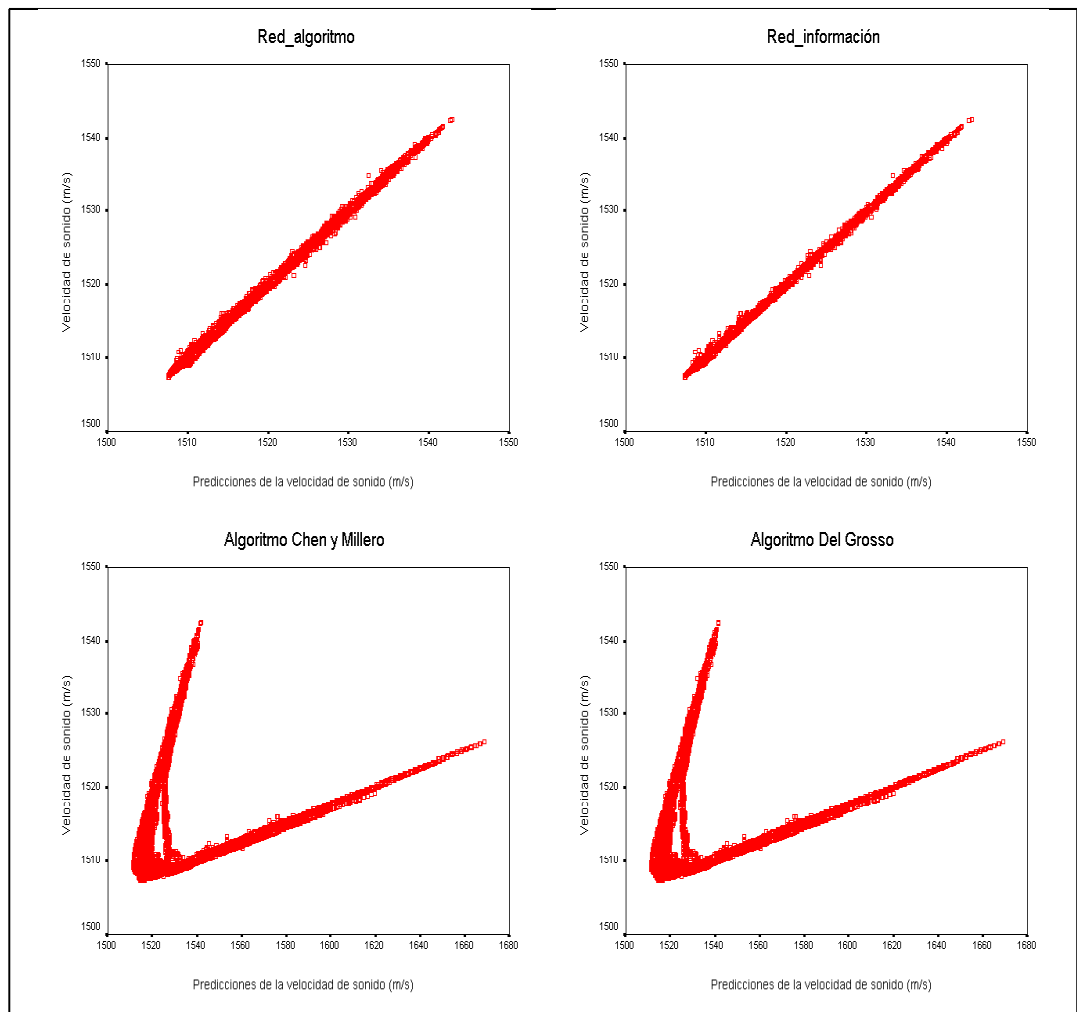


Figura 5.5.1. Representación gráfica de los valores reales de la velocidad de sonido frente a las predicciones obtenidas con cada una de las técnicas empleadas.

También se han calculado los coeficientes de correlación de Pearson entre los valores reales de la velocidad de sonido y las distintas predicciones obtenidas con cada una de las técnicas empleadas, cuyos valores aparecen en la Tabla 5.5.2. En dicha tabla pueden observarse unos coeficientes de correlación próximos a la unidad entre la velocidad de sonido en el agua del mar y las predicciones resultantes de las técnicas de redes de neuronas artificiales, lo que indica una relación de tipo lineal y directa entre ambas

variables, mientras que entre la velocidad de sonido en el agua del mar y las predicciones resultantes de las técnicas clásicas de cálculo de la velocidad de sonido el resultado es completamente distinto, obteniéndose valores próximos a cero, por lo que no existe relación de tipo lineal entre ambas variables. Estos coeficientes de correlación vienen a corroborar el resultado ya enunciado de que las predicciones obtenidas mediante las técnicas clásicas de cálculo de la velocidad de sonido no son representativas de la velocidad de sonido real, lo que sí ocurre con las predicciones resultantes de las técnicas de redes de neuronas artificiales.

Variabes	Coef. correl. Pearson
Velocidad de sonido y Pred. red_ algoritmo	0.999
Velocidad de sonido y Pred. red_ información	1.000
Velocidad de sonido y Pred. Chen y Millero	0.143
Velocidad de sonido y Pred. Del Grosso	0.144

Tabla 5.5.2. Coeficientes de correlación de Pearson entre los valores reales de la velocidad de sonido y las distintas predicciones.

A continuación se va a proceder a realizar un análisis de los residuos de las distintas predicciones obtenidas con cada una de las técnicas empleadas.

En la Tabla 5.5.3 se recogen los estadísticos descriptivos de los residuos de las distintas predicciones obtenidas con cada una de las técnicas empleadas, y en la Figura 5.5.2 sus gráficos de caja y bigotes. De ambas puede deducirse que la media de los residuos de las predicciones obtenidas con las técnicas de redes de neuronas artificiales podrían ser consideradas nulas, mientras que para los residuos de las predicciones resultantes de las técnicas clásicas de cálculo de la velocidad de sonido no puede afirmarse esta conclusión.

Variable	Media	Mínimo	Máximo	Desv. típica
Resid. red_ algoritmo	5.755E-04	-1.97	2.27	0.3040
Resid. red_ información	-2.095E-03	-1.66	2.14	0.1790
Resid. Chen y Millero	-18.3102	-142.46	2.81	24.2341
Resid. Del Grosso	-18.2372	-142.75	2.84	24.2562

Tabla 5.5.3. Estadísticos descriptivos de los residuos de las distintas predicciones obtenidas con cada una de las técnicas empleadas.

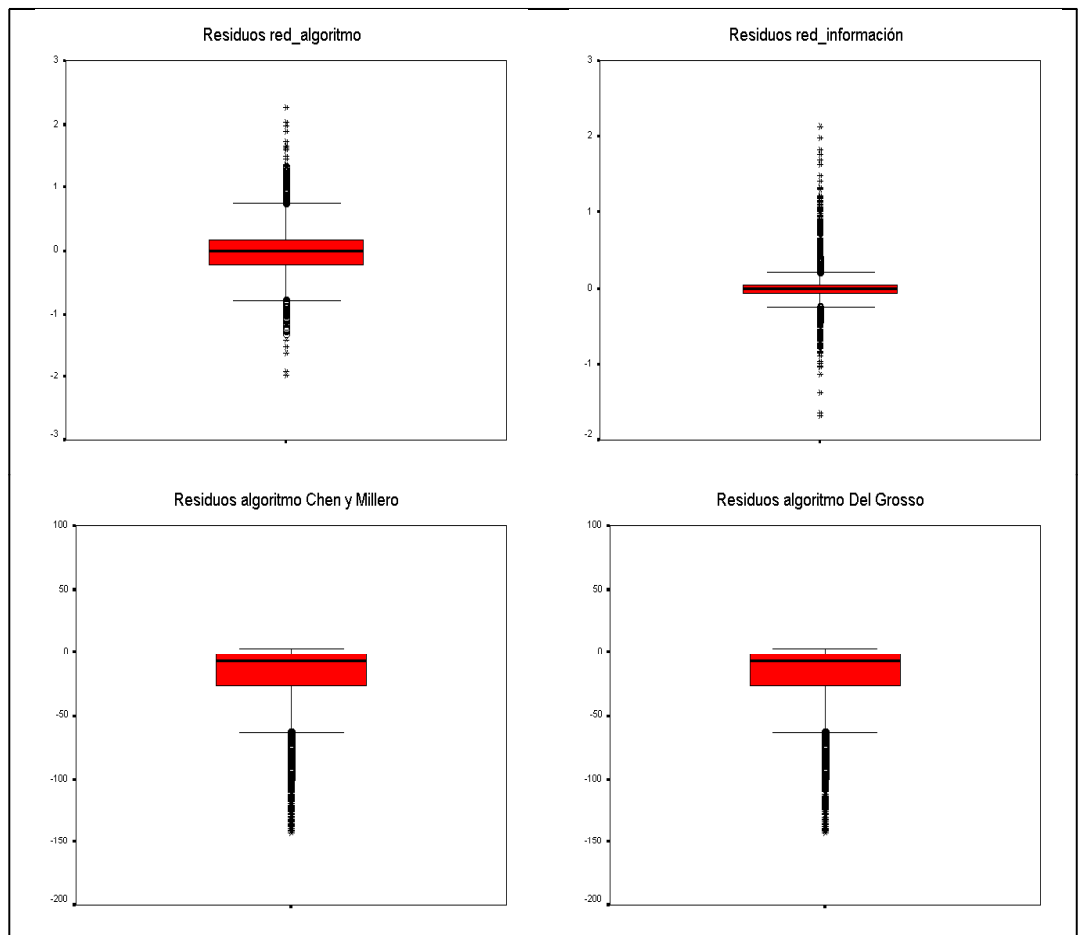


Figura 5.5.2. Gráficos de caja y bigotes de los residuos de las distintas predicciones obtenidas con cada una de las técnicas empleadas.

La normalidad de los residuos de las distintas predicciones obtenidas con cada una de las técnicas empleadas se analiza gráficamente mediante los gráficos cuantil-cuantil de probabilidad, que aparecen en la Figura 5.5.3. En los correspondientes a los residuos de las predicciones obtenidas con las técnicas de redes de neuronas artificiales puede observarse una tendencia aproximadamente lineal en la zona central, pero en los extremos se desvía un poco. Si se observa el gráfico correspondiente a la red_algoritmo, éste es el que más se aproxima a la tendencia normal, pudiendo concluirse la normalidad de sus residuos, mientras que a partir del gráfico correspondiente

a la red_información no puede llegarse a esa misma conclusión, ya que los resultados son dudosos. En cambio, en los correspondientes a los residuos de las predicciones resultantes de las técnicas clásicas de cálculo de la velocidad de sonido no se obtiene esa tendencia lineal deseada, por lo que se puede afirmar rotundamente que en estos casos los residuos no se distribuyen normalmente.

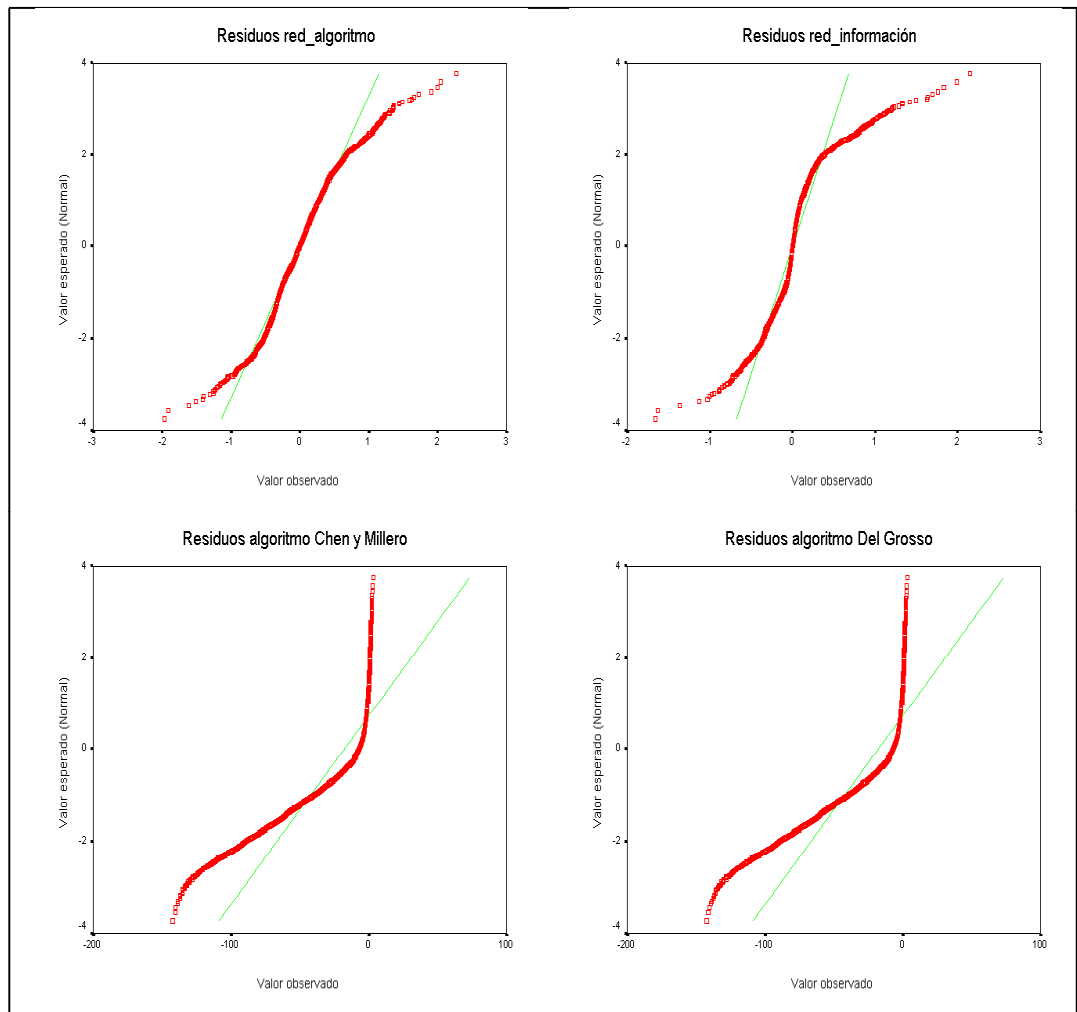


Figura 5.5.3. Gráficos cuantil-cuantil de probabilidad para los residuos de las distintas predicciones obtenidas con cada una de las técnicas empleadas.

Una vez determinada, desde un punto de vista matemático, la fiabilidad y precisión de las predicciones de la velocidad de sonido en el agua del mar obtenidas mediante las técnicas de redes de neuronas artificiales, se procederá a analizar estas características desde un punto de vista oceanográfico. Para ello, se debe comprobar si la distribución de dichas estimaciones mantiene también las propiedades físicas de la variable a la que pretende estimar, esto es, si mantiene las características y propiedades de la velocidad de sonido en el agua del mar, lo que determina a su vez, a través de los denominados perfiles de sonido, las características y propiedades del océano y de las diferentes capas de agua que lo constituyen.

Tal y como se indicó en el apartado 2.3, se han detectado en la zona estudiada en el mar de Alborán diversos tipos de perfiles acústicos, que se han denotado como TC1, TC2 día, TC2 noche, TC3, TC4 día, TC4 noche, TC5 norte, TC5 sur, TC6, LAOR, LAOC norte y LAOC sur, así que se deberá analizar el mantenimiento para cada uno de ellos de las propiedades de la velocidad de sonido en el agua del mar.

Concretamente se ha tomado al azar una localización de cada uno de los distintos perfiles acústicos detectados en cada una de las unidades fisiográficas y se han representado conjuntamente los valores de la velocidad de sonido en el agua del mar para dicha localización junto con sus predicciones derivadas de las redes de neuronas artificiales construidas, para comprobar así si los perfiles de sonido obtenidos a partir de las estimaciones difieren o no de los originales.

Nótese que en este proceso únicamente se han considerado las estimaciones provenientes de las redes de neuronas artificiales construidas, debido a que el análisis de las estimaciones procedentes de las técnicas clásicas de cálculo de velocidad de sonido ha puesto de manifiesto su invalidez e imprecisión en el caso de no disponer de los valores reales de la salinidad del agua del mar. Esto ocurre con mucha frecuencia aún en la actualidad, ya que a veces la única instrumentación disponible para la medición de las propiedades físicas del agua del mar son los batitermógrafos extensibles (XBT).

Estas representaciones gráficas se muestran en la Figura 5.5.4, y en ellas puede observarse que, en todos los casos estudiados, los perfiles de sonido reales y sus estimaciones obtenidas a través de las técnicas de redes de neuronas artificiales pueden ser considerados similares, por lo que siguen verificando las propiedades físicas del agua del mar, confirmando la validez de las técnicas de predicción mediante redes de neuronas artificiales para estimar la velocidad de sonido en el agua del mar.

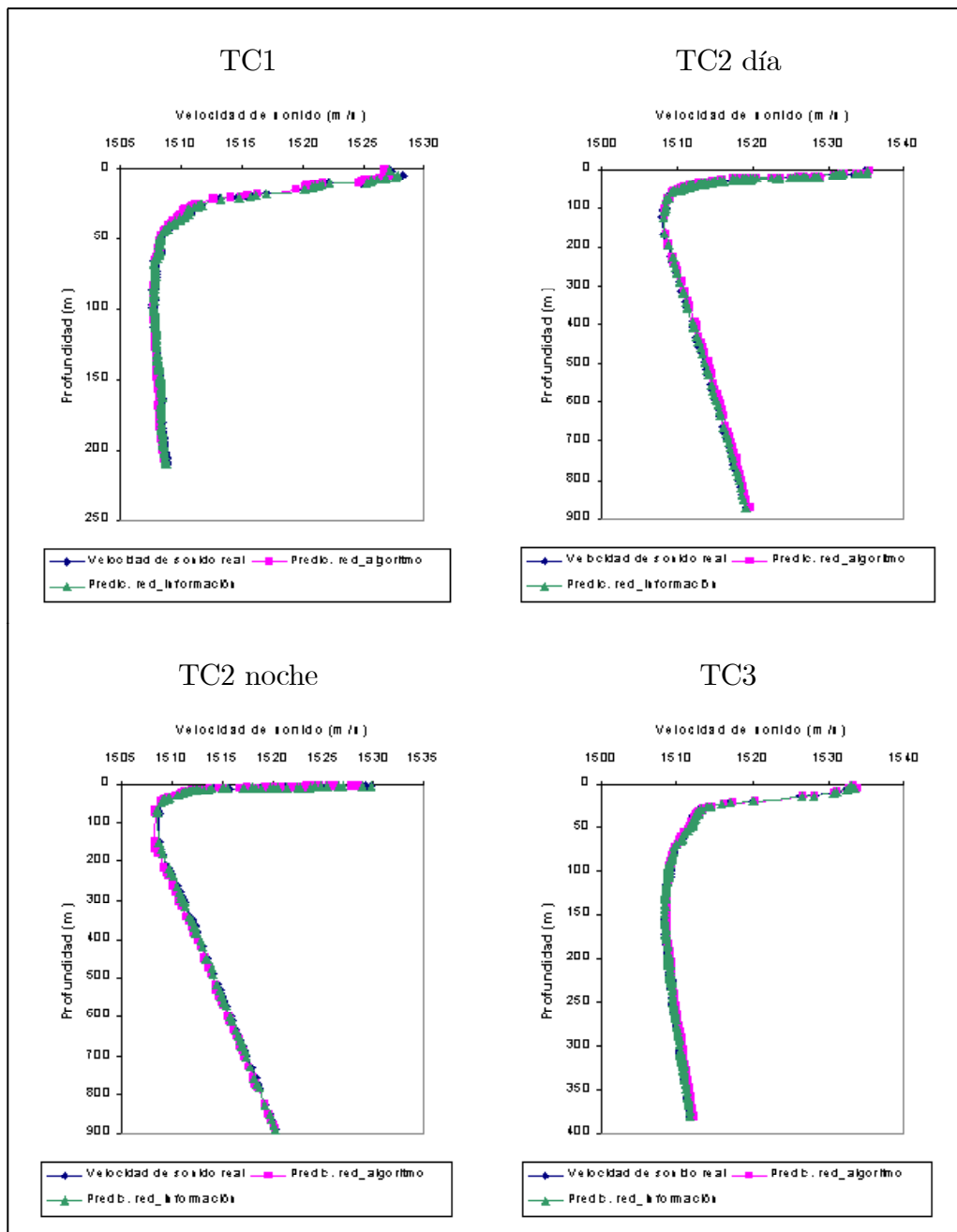


Figura 5.5.4. Representación gráfica de los perfiles de sonido junto con sus estimaciones mediante las redes de neuronas artificiales construidas.

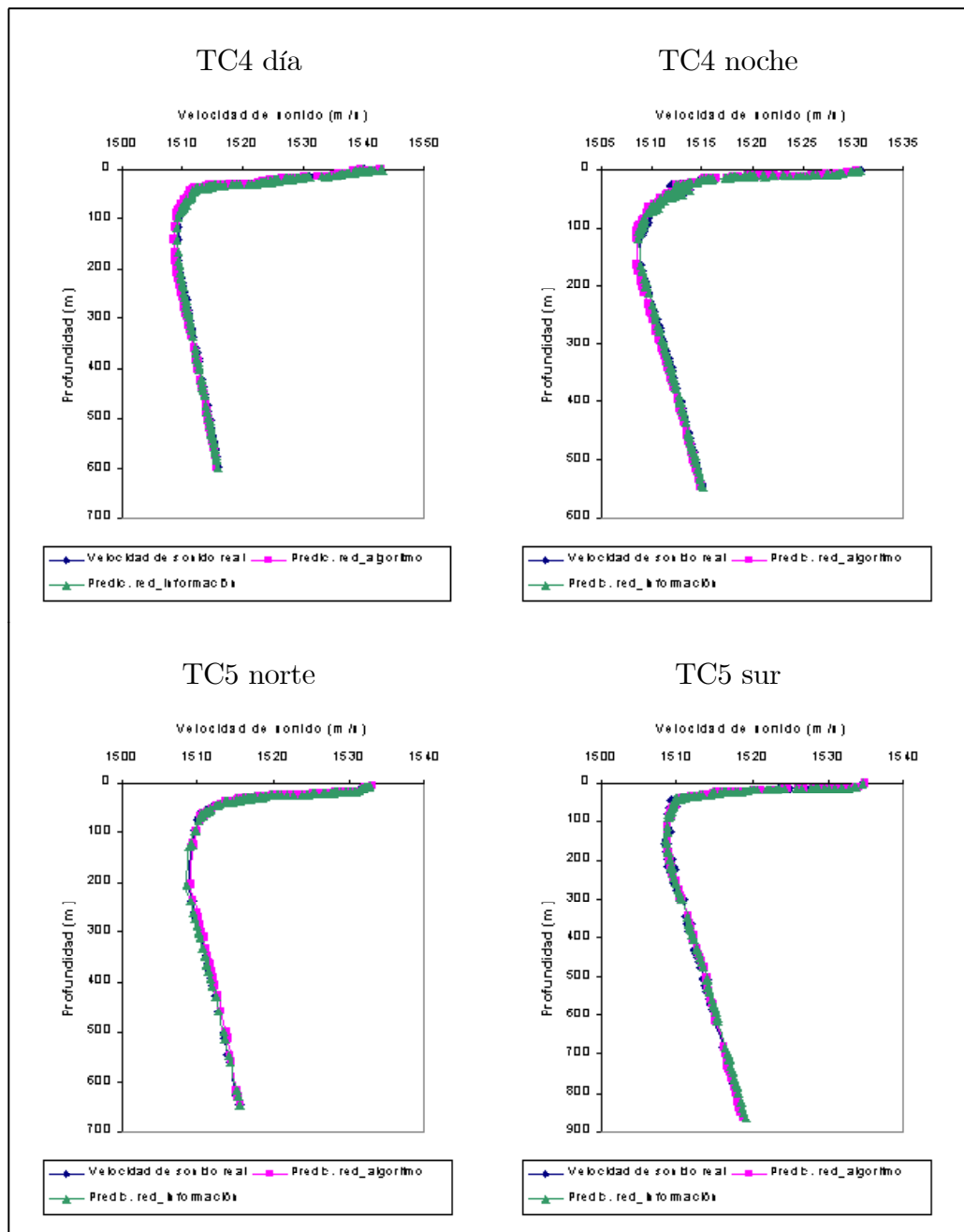


Figura 5.5.4-Cont. Representación gráfica de los perfiles de sonido junto con sus estimaciones mediante las redes de neuronas artificiales construidas.

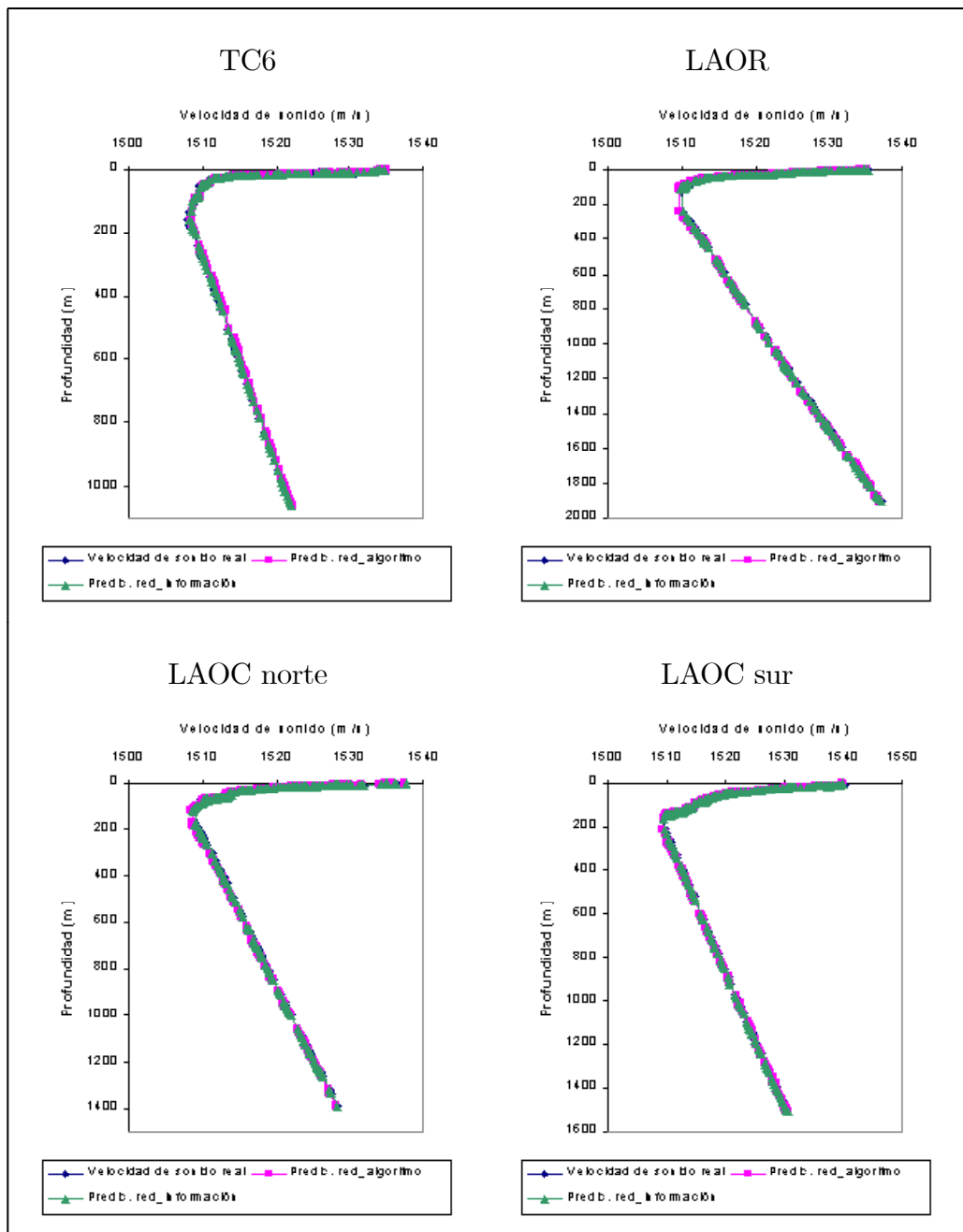


Figura 5.5.4-Cont. Representación gráfica de los perfiles de sonido junto con sus estimaciones mediante las redes de neuronas artificiales construidas.

6 CONCLUSIONES.

En este trabajo se ha abordado el problema de la predicción de una variable oceanográfica desde una perspectiva totalmente diferente a la que se ha venido empleando a lo largo de la historia, lo que lo convierte en un problema de gran interés y completamente novedoso. En concreto, se ha modelizado la velocidad de propagación del sonido en el agua del mar a través de las técnicas de redes de neuronas artificiales.

Se han construido dos redes de neuronas artificiales diferentes, cada una de ellas tomando distintas variables de entrada, aunque ambas son perceptrones multinivel o multicapa con una única capa oculta. Estas dos redes de neuronas artificiales han sido entrenadas con diversos algoritmos de entrenamiento y para varios tamaños de capa oculta, analizando los resultados obtenidos en cada una de estas simulaciones.

En concreto, se han considerado cinco algoritmos de entrenamiento, que, a partir del ecm medido sobre el conjunto test, admiten la siguiente ordenación de menor a mayor rendimiento: "gradient descent w/momentum & adaptative lr" (GD), "Resilient propagation" (RP), "Fletcher-Reeves conjugate gradient" (CGFR), "Broyden-Fletcher-Goldfarb-Shanno quasi-Newton" (BFGS) y "Levenberg-Marquardt" (LM).

La comparación entre estos métodos de entrenamiento ha sido efectuada no sólo de forma descriptiva, sino que se han realizado los correspondientes contrastes de hipótesis con un nivel de significación del 0.05.

Aunque en ambas redes coincide el orden de optimalidad de los métodos de entrenamiento utilizados, el comportamiento de estos métodos con respecto al número de nodos de la capa oculta considerados difiere según el número de predictores empleados, lo que pone de manifiesto la importancia de la realización de este estudio exhaustivo.

Para los métodos GD, RP, CGFR y BFGS se observa en general que para los tamaños de capa oculta más elevados el perceptrón multinivel comienza a exhibir cierto sobreajuste, a pesar de que durante el proceso de entrenamiento se ha utilizado un conjunto de validación para intentar evitar dicho problema.

Para el método CGFR, en el caso de la red `_información`, se observa que la tendencia de los resultados con respecto al tamaño de capa oculta va decreciendo hasta el tamaño 10 y luego cambia su tendencia creciendo hasta equiparar los resultados entre los tamaños de capa oculta 2 y 40, correspondientes al menor y al mayor tamaño de capa oculta considerados respectivamente.

El método LM ha resultado para ambas redes el mejor de todos los métodos de entrenamiento considerados. Además, se ha observado que esta regla de aprendizaje es menos sensible al problema del sobreajuste, puesto que el ecm medido sobre el conjunto test disminuye al aumentar el tamaño de la red. De hecho, se han observado diferencias significativas entre las medias de los ecm para todos los pares de tamaños de capa oculta considerados.

De todo esto se puede afirmar que el hecho de considerar una estructura más compleja no implica necesariamente una mejora en la precisión de los resultados obtenidos, sino que cada uno de los métodos de entrenamiento empleados presenta un comportamiento diferente.

A partir de todas las redes de neuronas artificiales entrenadas se han seleccionado aquellas que producen los mejores resultados, que como ya se ha comentado, se obtienen para el método de entrenamiento de Levenberg-Marquardt y para el tamaño de capa oculta máximo considerado, que en el caso de la red_ algoritmo está formada por 30 neuronas y en el caso de la red_ información consta de 40 neuronas.

El hecho de que con este método de entrenamiento los resultados vayan mejorando conforme aumenta el tamaño de capa oculta hace que se pueda plantear la cuestión de realizar otro estudio con el que se intente determinar si este comportamiento se produce de forma indefinida o si, por el contrario, existe un tamaño de capa oculta a partir del que esta tendencia se invierte o las mejoras se hacen significativamente despreciables. Por otro lado, también se podría analizar la conveniencia o no de aumentar el tamaño de capa oculta, ya que esto supondría un mayor tiempo de ejecución, por lo que se plantearía si esa mejora en los resultados compensa un mayor tiempo empleado.

Las precisiones de estas redes de neuronas óptimas vienen determinadas por sus errores cuadráticos medios calculados para el conjunto test, que presentan los valores 0.0879 (m/s)^2 para la red_ algoritmo y 0.0317 (m/s)^2 para la red_ información. Estas precisiones pueden compararse con las de los algoritmos clásicos empleados para la obtención de la velocidad de propagación del sonido en el agua del mar, cuyas varianzas son 0.0025 (m/s)^2 para el algoritmo de Del Grosso y 0.0361 (m/s)^2 para el algoritmo de Chen y Millero.

A simple vista únicamente la red_ información óptima mejora los resultados del algoritmo de Chen y Millero, pero como ya se ha comentado, las precisiones de estos procedimientos empeoran de forma considerable si no se dispone de toda la información necesaria para su utilización, lo que sucede en las ocasiones en que los datos proceden de batitermógrafos extensibles (XBT), en cuyo caso únicamente se dispone de la temperatura respecto a

la profundidad, pero no de la salinidad. En estos casos, ambas redes de neuronas artificiales proporcionan predicciones más precisas y fiables de la velocidad de propagación del sonido en el agua del mar.

Tras este estudio surgen varias cuestiones que se podrían abordar en un futuro:

- La realización de este estudio exhaustivo ha reflejado de forma evidente la influencia de la información de entrada empleada. La introducción de un mayor número de variables de entrada provoca una mejora de los resultados, pero esto conlleva un mayor tiempo de ejecución, por lo que, en este sentido, se podría plantear la posibilidad de realizar un estudio respecto a las variables de entrada, para ver cómo la introducción de una o más variables en la red de neuronas artificiales pueden influir en los resultados y si dicha introducción compensa con el tiempo adicional que esto supondría.
- Estas predicciones para la velocidad de propagación del sonido en el agua del mar obtenidas mediante las redes de neuronas artificiales son válidas únicamente para el Mar de Alborán durante los meses de verano (los datos proceden de campañas de investigación oceanográfica realizadas durante los meses de julio y agosto), por lo que se podrían realizar modelos análogos para el Mar de Alborán en el resto de las estaciones del año, analizando las posibles similitudes y diferencias entre dichos modelos.
- El Mar de Alborán es un mar cerrado, por lo que la distribución de la velocidad de propagación del sonido en el agua del mar en los primeros metros estará muy influenciada por unos valores elevados de salinidad. Por lo tanto, se podría realizar un modelo de predicción mediante redes de neuronas artificiales en otras zonas del océano que no se vean tan influenciadas por estos valores de salinidad tan elevados, existiendo la posibilidad de caracterizar globalmente a esta variable.
- Por último, nótese que en este trabajo se han construido únicamente dos redes de neuronas artificiales del mismo tipo, esto es, se han construido dos perceptrones multinivel o multicapa con una única capa oculta, pero existe la posibilidad de emplear otros modelos de redes de neuronas artificiales para obtener predicciones de la velocidad de propagación del sonido en el agua del mar, e incluso, de otras variables oceanográficas.

ANEXO 1: CÓDIGOS DE LOS PROGRAMAS IMPLEMENTADOS EN MATLAB.

Veloc-Ch-M

Este programa computa la velocidad de sonido a partir del algoritmo de Chen y Millero (1977), cuyos coeficientes han sido recalculados por Wong y Zhu (1995) para adaptar el algoritmo a la Escala Internacional de Temperatura de 1990.

En este algoritmo se precisan las variables Presión (en Bares), Temperatura (°C) y Salinidad(PSU), cuyos rangos de validez son de 0 a 1000 bares de presión, de 0 a 40°C de temperatura y de 0 a 40 de salinidad.

```
load ('C:\Pres_Ch_M.txt');  
load ('C:\Temp.txt');  
load ('C:\Sal.txt');
```

```
C00=1402.388;  
C01=5.03830;  
C02=-5.81090*10^-2;  
C03=3.3432*10^-4;  
C04=-1.47797*10^-6;  
C05=3.1419*10^-9;  
C10=0.153563;  
C11=6.8999*10^-4;  
C12=-8.1829*10^-6;  
C13=1.3632*10^-7;  
C14=-6.1260*10^-10;  
C20=3.1260*10^-5;  
C21=-1.7111*10^-6;  
C22=2.5986*10^-8;  
C23=-2.5353*10^-10;  
C24=1.0415*10^-12;  
C30=-9.7729*10^-9;  
C31=3.8513*10^-10;  
C32=-2.3654*10^-12;  
A00=1.389;  
A01=-1.262*10^-2;  
A02=7.166*10^-5;  
A03=2.008*10^-6;  
A04=-3.21*10^-8;  
A10=9.4742*10^-5;  
A11=-1.2583*10^-5;  
A12=-6.4928*10^-8;  
A13=1.0515*10^-8;  
A14=-2.0142*10^-10;  
A20=-3.9064*10^-7;  
A21=9.1061*10^-9;  
A22=-1.6009*10^-10;  
A23=7.994*10^-12;
```

```

A30=1.100*10^-10;
A31=6.651*10^-12;
A32=-3.391*10^-13;
B00=-1.922*10^-2;
B01=-4.42*10^-5;
B10=7.3637*10^-5;
B11=1.7950*10^-7;
D00=1.727*10^-3;
D10=-7.9836*10^-6;

```

```

dim=size(Temp);
aux=ones(dim(:,1),1);

```

```

Cw=(C00*aux+Temp.*C01+Temp.^2*C02+Temp.^3*C03+Temp.^4*C04+
Temp.^5*C05)+Pres_Ch_M.*(C10*aux+Temp.*C11+Temp.^2*C12+
Temp.^3*C13+Temp.^4*C14)+(Pres_Ch_M.^2).*(C20*aux+
Temp.*C21+Temp.^2*C22+Temp.^3*C23+Temp.^4*C24)+
(Pres_Ch_M.^3).*(C30*aux+Temp.*C31+Temp.^2*C32);

```

```

A=(A00*aux+Temp.*A01+Temp.^2*A02+Temp.^3*A03+Temp.^4*A04)+
Pres_Ch_M.*(A10*aux+Temp.*A11+Temp.^2*A12+Temp.^3*A13+
Temp.^4*A14)+(Pres_Ch_M.^2).*(A20*aux+Temp.*A21+
Temp.^2*A22+Temp.^3*A23)+(Pres_Ch_M.^3).*(A30*aux+Temp.*A31+
Temp.^2*A32);

```

```

B=B00*aux+Temp.*B01+Pres_Ch_M.*(B10*aux+Temp.*B11);

```

```

D=D00*aux+Pres_Ch_M.*D10;

```

```

Vel_Ch_M=Cw+Sal.*A+(Sal.^(3/2)).*B+(Sal.^2).*D;

```

```

save 'C:\Vel_Ch_M.txt' Vel_Ch_M -ASCII;

```

Veloc-G

Este programa calcula la velocidad de sonido a partir de las ecuaciones proporcionadas por Del Grosso (1974), cuyos coeficientes han sido recalculados por Wong y Zhu (1995) para adaptar el algoritmo a la Escala Internacional de Temperatura de 1990.

Para este algoritmo se precisan las variables Presión (en Kg/cm²), Temperatura (°C) y Salinidad(PSU), cuyos rangos de validez son de 0 a 1000 Kg/cm² de presión, de 0 a 30°C de temperatura y de 30 a 40 de salinidad.

```
load ('C:\Pres_G.txt');
load ('C:\Temp.txt');
load ('C:\Sal.txt');

C000=1402.392;
CT1=0.5012285*10^1;
CT2=-0.551184*10^-1;
CT3=0.221649*10^-3;
CS1=0.1329530*10^1;
CS2=0.1288598*10^-3;
CP1=0.1560592;
CP2=0.2449993*10^-4;
CP3=-0.8833959*10^-8;
CS1T1=-0.1275936*10^-1;
CS1T2=0.9688441*10^-4;
CT1P1=0.6353509*10^-2;
CT2P2=0.2656174*10^-7;
CT1P2=-0.1593895*10^-5;
CT1P3=0.5222483*10^-9;
CT3P1=-0.4383615*10^-6;
CS2P2=-0.1616745*10^-8;
CS2T1P1=0.4857614*10^-5;
CS1T1P1=-0.3406824*10^-3;

dim=size(Temp);
aux=ones(dim(:,1),1);

CT=Temp.*CT1+Temp.^2*CT2+Temp.^3*CT3;
CS=Sal.*CS1+Sal.^2*CS2;
CP=Pres_G.*CP1+Pres_G.^2*CP2+Pres_G.^3*CP3;

CSPT=CT1P1*Temp.*Pres_G+CT3P1*(Temp.^3).*Pres_G+CT1P2*
      Temp.*(Pres_G.^2)+CT2P2*(Temp.^2).*(Pres_G.^2)+CT1P3*
      Temp.*(Pres_G.^3)+CS1T1*Sal.*Temp+CS1T2*Sal.*(Temp.^2)+
      CS1T1P1*(Sal.*Temp).*Pres_G+
      CS2T1P1*((Sal.^2).*Temp).*Pres_G+CS2P2*(Sal.^2).*(Pres_G.^2);

Vel_G=C000*aux+CT+CS+CP+CSPT;

save 'C:\Vel_G.txt' Vel_G -ASCII;
```

Prof-Pres

Este programa convierte profundidad en presión (megaPascal=dbar) según el algoritmo de Leroy y Parthiot (1998). Para ello, se precisan las variables Profundidad (en metros) y Latitud.

Existen unos términos correctores para cada área de aplicación, y en este caso se empleará el correspondiente al Mar Mediterráneo.

```
load ('C:\Prof.txt');  
load ('C:\Latitud.txt');
```

```
Correc=-8.5*10^-6*Prof+1.4*10^-9*(Prof.^2);
```

```
Pres=((1.00818*10^2*Prof+2.465*10^-8*(Prof.^2)-  
1.25*10^13*(Prof.^3)+2.8*10^-19*(Prof.^4)).*(0.7803*  
(1+5.3*10^-3*sin(Latitud).^2)-2*10^-5*Prof)./  
(9.80612-2*10^-5*Prof))-Correc;
```

```
save 'C:\Pres.txt' Pres -ASCII;
```

Pres-Prof

Este programa convierte presión en profundidad según el algoritmo de Leroy y Parthiot (1998). Para ello se precisan las matrices de Presión (en MPa) y Latitud.

Existen unos términos correctores para cada área de aplicación, y en este caso se empleará el correspondiente al Mar Mediterráneo.

```
load ('C:\Pres_L_P.txt');  
load ('C:\Latitud.txt');
```

```
Correc=-7*10^-2*Pres_L_P+2*10^-3*(Pres_L_P.^2);
```

```
Prof=Correc+(9.7266*10^2*Pres_L_P-2.512*10^-1*(Pres_L_P.^2) +  
2.28*10^-4*(Pres_L_P.^3)-1.8*10^-7*(Pres_L_P.^4))./  
(9.7803*(1+5.3*10^-3*sin(Latitud).^2)+1.1*10^-4*Pres_L_P);
```

```
save 'C:\Prof.txt' Prof -ASCII;
```

Transf-datos

Este programa transforma los datos de las variables que se van a considerar al computar las redes de neuronas artificiales para que todos sus valores se encuentren dentro de un mismo rango, que en este caso se ha fijado en [0,10].

```
datos=load('C:\sonido_alb.txt');
```

```
hora=(1/1000)*datos(:,1);  
fondo=(1/1000)*datos(:,2);  
latitud=(1/10)*datos(:,3);  
longitud=(-1)*datos(:,4);  
temp_sup=(1/10)*datos(:,5);  
v_viento=(1/10)*datos(:,6);  
d_viento=(1/100)*datos(:,7);  
temp_aire=(1/10)*datos(:,8);  
pres_atm=(1/1000)*datos(:,9);  
prof=(1/1000)*datos(:,10);  
v_sonido=(1/1000)*datos(:,11);  
temp_agua=(1/10)*datos(:,12);
```

```
datos_transf=[hora fondo latitud longitud temp_sup v_viento d_viento temp_aire  
pres_atm prof v_sonido temp_agua];
```

```
save C:\sonido_alb_transf datos_transf -ASCII;
```


Div-datos

Este programa realiza aleatoriamente la división de los datos en los diversos conjuntos necesarios para proceder al proceso de entrenamiento de la red de neuronas artificial: un 50% de los datos para el conjunto de entrenamiento de la red, un 25% de los datos para el conjunto de validación y el 25% restante para el conjunto test.

```
resto_datos=load('C:\sonido_alb_transf.txt');
n=length(resto_datos);
conj_entren=[];
conj_valid=[];
rand('state',sum(100*clock));

for i=1:(1/2*n)
    aleat=fix(length(resto_datos)*rand)+1;
    conj_entren=[conj_entren ; resto_datos(aleat,:)];
    resto_datos(aleat,:)=[];
end

for j=1:(1/4*n)
    aleat=fix(length(resto_datos)*rand)+1;
    conj_valid=[conj_valid ; resto_datos(aleat,:)];
    resto_datos(aleat,:)=[];
end

conj_test=resto_datos;

save C:\conj_entren conj_entren -ASCII;
save C:\conj_valid conj_valid -ASCII;
save C:\conj_test conj_test -ASCII;
```

Red-algoritmo

Esta red de neuronas artificial consiste en un perceptrón multinivel o multicapa con una única capa oculta que toma como nodos de entrada la temperatura del agua y la profundidad, y como variable de salida, la velocidad del sonido.

```
div_datos;
neur_ocult=[2 5 10 20 30];

pred_entren_bfg=[];
pred_valid_bfg=[];
pred_test_bfg=[];
pred_entren_cgf=[];
pred_valid_cgf=[];
pred_test_cgf=[];
pred_entren_gdx=[];
pred_valid_gdx=[];
pred_test_gdx=[];
pred_entren_lm=[];
pred_valid_lm=[];
pred_test_lm=[];
pred_entren_rp=[];
pred_valid_rp=[];
pred_test_rp=[];
ECM=[];
param=[];

inputs_entren=[conj_entren(:,10),conj_entren(:,12)];
outputs_entren=[conj_entren(:,11)];
inputs_valid=[conj_valid(:,10),conj_valid(:,12)];
outputs_valid=[conj_valid(:,11)];
inputs_test=[conj_test(:,10),conj_test(:,12)];
outputs_test=[conj_test(:,11)];

clear conj_entren conj_valid conj_test;

valid=struct('P',inputs_valid,'T',outputs_valid,'Pi',[],'Ai',[]);
test=struct('P',inputs_test,'T',outputs_test,'Pi',[],'Ai',[]);

for i=1:20

    for k=1:length(neur_ocult)

        red=newff([0 10;0 10],[neur_ocult(k) 1], {'logsig' 'purelin'},
            'trainbfg');
        [entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
            valid,test);
        ECM_entren=getfield(tr,'perf');
        ECM_valid=getfield(tr,'vperf');
        ECM_test=getfield(tr,'tperf');
```

```

ECM=[ECM;i 1 neur_ocult(k) ECM_entren(1,end)
      ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 1 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_bfg=[pred_entren_bfg;
                 i*ones(length(inputs_entren),1)
                 neur_ocult(k)*ones(length(inputs_entren),1)
                 sim_entren];
pred_valid_bfg=[pred_valid_bfg; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_bfg=[pred_test_bfg; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10],[neur_ocult(k) 1], {'logsig' 'purelin'},
          'traincgf');
[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
                    valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 2 neur_ocult(k) ECM_entren(1,end)
      ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 2 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_cgf=[pred_entren_cgf; i*ones(length(inputs_entren),1)
                 neur_ocult(k)*ones(length(inputs_entren),1)
                 sim_entren];
pred_valid_cgf=[pred_valid_cgf; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_cgf=[pred_test_cgf; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10],[neur_ocult(k) 1], {'logsig' 'purelin'},
          'traingdx');
[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
                    valid,test);
ECM_entren=getfield(tr,'perf');

```

```

ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 3 neur_ocult(k) ECM_entren(1,end)
      ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 3 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_gdx=[pred_entren_gdx;
                 i*ones(length(inputs_entren),1)
                 neur_ocult(k)*ones(length(inputs_entren),1)
                 sim_entren];
pred_valid_gdx=[pred_valid_gdx; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_gdx=[pred_test_gdx; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10],[neur_ocult(k) 1], {'logsig' 'purelin'},
          'trainlm');
[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
                    valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 4 neur_ocult(k) ECM_entren(1,end)
      ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 4 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_lm=[pred_entren_lm;
                i*ones(length(inputs_entren),1)
                neur_ocult(k)*ones(length(inputs_entren),1)
                sim_entren];
pred_valid_lm=[pred_valid_lm; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_lm=[pred_test_lm; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10],[neur_ocult(k) 1], {'logsig' 'purelin'},
          'trainrp');

```

```

[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
    valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 5 neur_ocult(k) ECM_entren(1,end)
    ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 5 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_rp=[pred_entren_rp; i*ones(length(inputs_entren),1)
    neur_ocult(k)*ones(length(inputs_entren),1)
    sim_entren];
pred_valid_rp=[pred_valid_rp; i*ones(length(inputs_valid),1)
    neur_ocult(k)*ones(length(inputs_valid),1)
    sim_valid];
pred_test_rp=[pred_test_rp; i*ones(length(inputs_test),1)
    neur_ocult(k)*ones(length(inputs_test),1)
    sim_test];
end
end

save C:\ECM ECM -ASCII;
save C:\param param -ASCII;
save C:\pred_entren_bfg pred_entren_bfg -ASCII;
save C:\pred_valid_bfg pred_valid_bfg -ASCII;
save C:\pred_test_bfg pred_test_bfg -ASCII;
save C:\pred_entren_cgf pred_entren_cgf -ASCII;
save C:\pred_valid_cgf pred_valid_cgf -ASCII;
save C:\pred_test_cgf pred_test_cgf -ASCII;
save C:\pred_entren_gdx pred_entren_gdx -ASCII;
save C:\pred_valid_gdx pred_valid_gdx -ASCII;
save C:\pred_test_gdx pred_test_gdx -ASCII;
save C:\pred_entren_lm pred_entren_lm -ASCII;
save C:\pred_valid_lm pred_valid_lm -ASCII;
save C:\pred_test_lm pred_test_lm -ASCII;
save C:\pred_entren_rp pred_entren_rp -ASCII;
save C:\pred_valid_rp pred_valid_rp -ASCII;
save C:\pred_test_rp pred_test_rp -ASCII;

```

Red-información

Esta red de neuronas artificial consiste en un perceptrón multinivel o multicapa con una única capa oculta. Como variables de entrada incluye la latitud, la longitud, la profundidad, la hora, el fondo existente en esa localización, la temperatura en la superficie del agua, la presión atmosférica, la velocidad del viento, la dirección del viento y la temperatura del aire. Como variable de salida tiene la velocidad del sonido.

```
div_datos;
neur_ocult=[2 5 10 20 30 40];

pred_entren_bfg=[];
pred_valid_bfg=[];
pred_test_bfg=[];
pred_entren_cgf=[];
pred_valid_cgf=[];
pred_test_cgf=[];
pred_entren_gdx=[];
pred_valid_gdx=[];
pred_test_gdx=[];
pred_entren_lm=[];
pred_valid_lm=[];
pred_test_lm=[];
pred_entren_rp=[];
pred_valid_rp=[];
pred_test_rp=[];
ECM=[];
param=[];

inputs_entren=[conj_entren(:,1:10),conj_entren(:,12)'];
outputs_entren=[conj_entren(:,11)'];
inputs_valid=[conj_valid(:,1:10),conj_valid(:,12)'];
outputs_valid=[conj_valid(:,11)'];
inputs_test=[conj_test(:,1:10),conj_test(:,12)'];
outputs_test=[conj_test(:,11)'];

clear conj_entren conj_valid conj_test;

valid=struct('P',inputs_valid,'T',outputs_valid,'Pi',[],'Ai',[]);
test=struct('P',inputs_test,'T',outputs_test,'Pi',[],'Ai',[]);

for i=1:20
    for k=1:length(neur_ocult)

        red=newff([0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10],
            [neur_ocult(k) 1], {'logsig' 'purelin'}, 'trainbfg');
        [entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
            valid,test);
        ECM_entren=getfield(tr,'perf');
```

```

ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 1 neur_ocult(k) ECM_entren(1,end)
      ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 1 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_bfg=[pred_entren_bfg;
                i*ones(length(inputs_entren),1)
                neur_ocult(k)*ones(length(inputs_entren),1)
                sim_entren];
pred_valid_bfg=[pred_valid_bfg; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_bfg=[pred_test_bfg; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10],
          [neur_ocult(k) 1], {'logsig' 'purelin'}, 'traincgf');
[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
                    valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 2 neur_ocult(k) ECM_entren(1,end)
      ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 2 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_cgf=[pred_entren_cgf; i*ones(length(inputs_entren),1)
                neur_ocult(k)*ones(length(inputs_entren),1)
                sim_entren];
pred_valid_cgf=[pred_valid_cgf; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_cgf=[pred_test_cgf; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10],
          [neur_ocult(k) 1], {'logsig' 'purelin'}, 'traingdx');

```

```

[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
                    valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 3 neur_ocult(k) ECM_entren(1,end)
     ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 3 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_gdx=[pred_entren_gdx;
                i*ones(length(inputs_entren),1)
                neur_ocult(k)*ones(length(inputs_entren),1)
                sim_entren];
pred_valid_gdx=[pred_valid_gdx; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_gdx=[pred_test_gdx; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

red=newff([0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10],
          [neur_ocult(k) 1], {'logsig' 'purelin'}, 'trainlm');
[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
                    valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 4 neur_ocult(k) ECM_entren(1,end)
     ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
    param=[param;i 4 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_lm=[pred_entren_lm;
                i*ones(length(inputs_entren),1)
                neur_ocult(k)*ones(length(inputs_entren),1)
                sim_entren];
pred_valid_lm=[pred_valid_lm; i*ones(length(inputs_valid),1)
                neur_ocult(k)*ones(length(inputs_valid),1)
                sim_valid];
pred_test_lm=[pred_test_lm; i*ones(length(inputs_test),1)
                neur_ocult(k)*ones(length(inputs_test),1)
                sim_test];

```



```

red=newff([0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10;0 10],
[neur_ocult(k) 1], {'logsig' 'purelin'}, 'trainrp');
[entren_red,tr]=train(red,inputs_entren,outputs_entren,[],[],
valid,test);
ECM_entren=getfield(tr,'perf');
ECM_valid=getfield(tr,'vperf');
ECM_test=getfield(tr,'tperf');
ECM=[ECM;i 5 neur_ocult(k) ECM_entren(1,end)
ECM_valid(1,end) ECM_test(1,end)];
estruct=getx(entren_red);
for l=1:length(estruct)
param=[param;i 5 neur_ocult(k) l estruct(l,1)];
end
sim_entren=[sim(entren_red,inputs_entren)];
sim_valid=[sim(entren_red,inputs_valid)];
sim_test=[sim(entren_red,inputs_test)];
pred_entren_rp=[pred_entren_rp; i*ones(length(inputs_entren),1)
neur_ocult(k)*ones(length(inputs_entren),1)
sim_entren];
pred_valid_rp=[pred_valid_rp; i*ones(length(inputs_valid),1)
neur_ocult(k)*ones(length(inputs_valid),1)
sim_valid];
pred_test_rp=[pred_test_rp; i*ones(length(inputs_test),1)
neur_ocult(k)*ones(length(inputs_test),1)
sim_test];

end
end

save C:\ECM ECM -ASCII;
save C:\param param -ASCII;
save C:\pred_entren_bfg pred_entren_bfg -ASCII;
save C:\pred_valid_bfg pred_valid_bfg -ASCII;
save C:\pred_test_bfg pred_test_bfg -ASCII;
save C:\pred_entren_cgf pred_entren_cgf -ASCII;
save C:\pred_valid_cgf pred_valid_cgf -ASCII;
save C:\pred_test_cgf pred_test_cgf -ASCII;
save C:\pred_entren_gdx pred_entren_gdx -ASCII;
save C:\pred_valid_gdx pred_valid_gdx -ASCII;
save C:\pred_test_gdx pred_test_gdx -ASCII;
save C:\pred_entren_lm pred_entren_lm -ASCII;
save C:\pred_valid_lm pred_valid_lm -ASCII;
save C:\pred_test_lm pred_test_lm -ASCII;
save C:\pred_entren_rp pred_entren_rp -ASCII;
save C:\pred_valid_rp pred_valid_rp -ASCII;
save C:\pred_test_rp pred_test_rp -ASCII;

```


ANEXO 2. MEDIAS DE LOS ECM DE LA RED-ALGORITMO.

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
1	BFGS	2	3.16383	3.14228	3.13051
1	BFGS	5	0.19785	0.19787	0.19687
1	BFGS	10	0.26308	0.26412	0.26227
1	BFGS	20	0.32511	0.32179	0.32915
1	BFGS	30	0.48427	0.49255	0.48928
1	CGFR	2	13.3020	13.1309	13.0195
1	CGFR	5	11.5554	11.3451	11.1542
1	CGFR	10	12.9368	12.6184	12.4401
1	CGFR	20	17.7903	17.7113	17.7848
1	CGFR	30	33.2838	33.2689	31.7637
1	GD	2	768.255	744.610	741.131
1	GD	5	1513.21	1458.16	1440.62
1	GD	10	3129.38	2891.26	3010.30
1	GD	20	12270.6	11548.7	11953.3
1	GD	30	15076.4	13965.7	14630.5
1	LM	2	0.14835	0.14916	0.15021
1	LM	5	0.12171	0.12068	0.12223
1	LM	10	0.10927	0.10846	0.11048
1	LM	20	0.10208	0.10092	0.10369
1	LM	30	0.09649	0.09518	0.09813
1	RP	2	33.4925	33.2599	33.1477
1	RP	5	16.4130	15.5965	16.0724
1	RP	10	54.4280	53.7804	53.6293
1	RP	20	28.8139	28.8962	29.9706
1	RP	30	2901.39	2893.59	2894.27

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
2	BFGS	2	0.29261	0.29756	0.30175
2	BFGS	5	0.45233	0.44487	0.45983
2	BFGS	10	0.87731	0.85160	0.88923
2	BFGS	20	0.53282	0.51125	0.52501
2	BFGS	30	0.65414	0.61087	0.63633
2	CGFR	2	13.4342	12.8811	13.7051
2	CGFR	5	18.9345	18.0560	19.6053
2	CGFR	10	14.4375	13.6940	14.8488
2	CGFR	20	10.6468	9.90869	10.9304
2	CGFR	30	22.6541	21.0515	22.9136
2	GD	2	457.016	441.854	475.003
2	GD	5	1682.86	1584.13	1723.16
2	GD	10	2483.64	2257.08	2537.01
2	GD	20	17183.2	15483.0	17459.7
2	GD	30	16020.7	14605.5	16528.8
2	LM	2	0.16842	0.17797	0.17257
2	LM	5	0.12878	0.13543	0.12968
2	LM	10	0.10967	0.11601	0.11023
2	LM	20	0.10244	0.10853	0.10308
2	LM	30	0.09647	0.10264	0.09772
2	RP	2	311.617	302.605	314.331
2	RP	5	52.1302	49.5355	51.9657
2	RP	10	39.0086	36.5521	38.7534
2	RP	20	32.0514	30.5534	30.7726
2	RP	30	31.4908	31.0009	31.8040

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
3	BFGS	2	8.02764	7.87676	7.58302
3	BFGS	5	0.18726	0.18818	0.18298
3	BFGS	10	0.42295	0.41562	0.41716
3	BFGS	20	0.31353	0.31250	0.31359
3	BFGS	30	0.41379	0.40897	0.43313
3	CGFR	2	20.4800	20.1973	19.5186
3	CGFR	5	16.5873	16.4283	15.9778
3	CGFR	10	14.6594	14.5820	14.4704
3	CGFR	20	18.5186	18.3841	18.9894
3	CGFR	30	19.2564	18.6823	19.0711
3	GD	2	596.308	580.868	578.733
3	GD	5	1422.18	1397.33	1418.78
3	GD	10	4975.40	4728.88	4853.17
3	GD	20	11057.8	10138.2	10878.1
3	GD	30	9655.63	9005.98	9372.62
3	LM	2	0.15377	0.15595	0.15117
3	LM	5	0.12146	0.12278	0.11949
3	LM	10	0.11088	0.11155	0.10949
3	LM	20	0.10262	0.10346	0.10132
3	LM	30	0.09880	0.09959	0.09730
3	RP	2	28.8571	28.0742	27.7123
3	RP	5	34.8283	33.9029	33.2491
3	RP	10	12.9678	12.5995	12.9120
3	RP	20	20.4960	19.9746	20.8119
3	RP	30	65.6047	65.0841	68.1193

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
4	BFGS	2	5.84490	5.91294	6.39020
4	BFGS	5	2.72353	2.76740	2.99718
4	BFGS	10	2.52169	2.43723	2.56736
4	BFGS	20	0.58731	0.62403	0.66073
4	BFGS	30	0.68309	0.72047	0.72453
4	CGFR	2	11.0790	11.1705	12.0664
4	CGFR	5	21.5965	21.5394	23.1242
4	CGFR	10	10.6710	10.7334	11.4283
4	CGFR	20	8.64655	9.05223	9.26279
4	CGFR	30	32.9535	34.3558	35.9061
4	GD	2	874.625	824.634	896.070
4	GD	5	1843.21	1793.47	1950.50
4	GD	10	5397.23	4940.52	5685.18
4	GD	20	5689.29	5164.03	5980.67
4	GD	30	15689.9	14132.3	16089.5
4	LM	2	0.15270	0.14902	0.14762
4	LM	5	0.12650	0.12274	0.12276
4	LM	10	0.11511	0.11229	0.11168
4	LM	20	0.10295	0.10238	0.10120
4	LM	30	0.09848	0.09782	0.09639
4	RP	2	213.732	209.276	221.671
4	RP	5	48.1556	49.6968	56.1175
4	RP	10	43.7848	45.1227	48.8634
4	RP	20	13.4385	14.4781	15.3680
4	RP	30	4037.73	4044.61	4068.51

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
5	BFGS	2	0.22109	0.21920	0.21663
5	BFGS	5	0.21511	0.20887	0.20652
5	BFGS	10	0.56223	0.55370	0.53205
5	BFGS	20	0.33701	0.33553	0.32313
5	BFGS	30	1.00451	1.01693	0.91344
5	CGFR	2	12.5119	12.3886	12.4641
5	CGFR	5	14.8336	14.9426	14.6129
5	CGFR	10	14.7222	14.7473	14.2785
5	CGFR	20	15.5435	15.6391	14.3909
5	CGFR	30	28.0803	28.1369	26.1240
5	GD	2	677.138	685.754	665.440
5	GD	5	1300.81	1273.73	1294.45
5	GD	10	5712.95	5534.21	5606.61
5	GD	20	11279.4	10321.3	11110.8
5	GD	30	15619.8	14376.6	15312.0
5	LM	2	0.16004	0.15669	0.15599
5	LM	5	0.13033	0.12768	0.12651
5	LM	10	0.11436	0.11220	0.11104
5	LM	20	0.10207	0.09836	0.09861
5	LM	30	0.09955	0.09548	0.09571
5	RP	2	35.6981	35.8011	35.0522
5	RP	5	48.7695	49.4181	48.2742
5	RP	10	23.0375	23.6746	22.8993
5	RP	20	22.4425	22.5698	22.4083
5	RP	30	106.819	111.474	104.745

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
6	BFGS	2	0.23606	0.23146	0.22762
6	BFGS	5	0.75189	0.67949	0.75207
6	BFGS	10	0.41524	0.38629	0.41256
6	BFGS	20	0.68207	0.63690	0.66274
6	BFGS	30	0.92419	0.78832	0.94356
6	CGFR	2	9.86644	9.20324	10.1211
6	CGFR	5	18.6187	16.9050	19.2169
6	CGFR	10	16.7813	15.2323	17.1929
6	CGFR	20	18.2632	14.7446	18.9682
6	CGFR	30	43.3797	38.9402	44.2303
6	GD	2	641.497	595.414	660.238
6	GD	5	2621.07	2476.09	2734.69
6	GD	10	3822.22	3431.95	3977.79
6	GD	20	13405.1	11381.1	13867.0
6	GD	30	21794.8	19000.8	22867.1
6	LM	2	0.15691	0.15268	0.14929
6	LM	5	0.13234	0.12779	0.12583
6	LM	10	0.11830	0.11406	0.11238
6	LM	20	0.10428	0.10130	0.09858
6	LM	30	0.10019	0.09754	0.09402
6	RP	2	19.7193	18.1588	20.1002
6	RP	5	50.8195	47.4394	51.2139
6	RP	10	13.0651	12.4798	12.8111
6	RP	20	31.8234	29.8567	30.9523
6	RP	30	179.966	167.947	183.042

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
7	BFGS	2	0.21077	0.21284	0.21568
7	BFGS	5	0.21031	0.21347	0.21606
7	BFGS	10	0.45616	0.47070	0.47803
7	BFGS	20	0.42465	0.43287	0.44451
7	BFGS	30	0.68057	0.72271	0.71969
7	CGFR	2	11.8458	11.6879	11.5894
7	CGFR	5	19.1042	19.1917	19.1369
7	CGFR	10	13.1113	13.1998	13.1097
7	CGFR	20	10.9444	11.0848	11.1950
7	CGFR	30	29.4633	29.9516	30.2123
7	GD	2	423.003	405.709	414.318
7	GD	5	2187.57	2039.71	2139.60
7	GD	10	2726.47	2495.64	2648.60
7	GD	20	11180.6	10507.3	10815.1
7	GD	30	11038.2	10417.4	10946.2
7	LM	2	0.16185	0.16188	0.16565
7	LM	5	0.12742	0.12769	0.13074
7	LM	10	0.10740	0.10750	0.11051
7	LM	20	0.10412	0.10413	0.10720
7	LM	30	0.09745	0.09667	0.10026
7	RP	2	30.8129	31.1293	30.8983
7	RP	5	18.9275	19.6989	19.6650
7	RP	10	122596.5	122230.7	122569.5
7	RP	20	2256.71	2262.22	2250.16
7	RP	30	290.243	299.095	296.478

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
8	BFGS	2	0.21865	0.21701	0.21368
8	BFGS	5	0.24135	0.23657	0.23594
8	BFGS	10	0.33388	0.32702	0.33136
8	BFGS	20	0.36803	0.35870	0.35518
8	BFGS	30	0.36311	0.35728	0.36257
8	CGFR	2	14.5555	14.3968	14.4012
8	CGFR	5	17.4890	17.2466	17.5147
8	CGFR	10	18.4009	17.9811	18.3598
8	CGFR	20	14.0570	13.6952	14.8974
8	CGFR	30	24.3134	23.5271	26.2138
8	GD	2	705.113	693.535	708.968
8	GD	5	2175.42	2138.71	2188.91
8	GD	10	3484.13	3297.67	3446.07
8	GD	20	15552.8	14292.4	15419.4
8	GD	30	13765.2	12745.7	13863.9
8	LM	2	0.15995	0.16048	0.15491
8	LM	5	0.12501	0.12692	0.12100
8	LM	10	0.11702	0.11912	0.11348
8	LM	20	0.10154	0.10396	0.09960
8	LM	30	0.09713	0.09929	0.09566
8	RP	2	91.5940	91.1516	91.0434
8	RP	5	62.9669	60.4869	62.3816
8	RP	10	678.088	674.178	678.001
8	RP	20	31.6447	30.5677	30.4692
8	RP	30	100.164	98.3595	97.7662

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
9	BFGS	2	0.25397	0.24489	0.24659
9	BFGS	5	0.33812	0.31791	0.34142
9	BFGS	10	0.60592	0.57664	0.61132
9	BFGS	20	0.46417	0.45663	0.47142
9	BFGS	30	1.15518	1.02822	1.20629
9	CGFR	2	18.0811	17.1323	18.1041
9	CGFR	5	22.4200	21.1192	22.4756
9	CGFR	10	10.9247	10.2336	10.9714
9	CGFR	20	8.49969	7.54832	8.67483
9	CGFR	30	32.6075	30.1074	32.8540
9	GD	2	577.334	550.791	587.188
9	GD	5	1765.99	1633.28	1790.27
9	GD	10	5429.89	5026.15	5479.28
9	GD	20	9350.59	8420.38	9529.55
9	GD	30	9889.45	8832.94	10069.2
9	LM	2	0.16114	0.15805	0.15543
9	LM	5	0.13174	0.12891	0.12747
9	LM	10	0.11493	0.11282	0.11156
9	LM	20	0.10389	0.10229	0.10045
9	LM	30	0.10146	0.10025	0.09811
9	RP	2	114.367	110.835	113.873
9	RP	5	22.4153	21.5932	22.1496
9	RP	10	15.2026	14.7695	15.1133
9	RP	20	19.8712	19.7919	20.2903
9	RP	30	49.8793	49.3531	51.7483

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
10	BFGS	2	3.12669	3.17408	3.15869
10	BFGS	5	0.67024	0.66904	0.66507
10	BFGS	10	0.89273	0.87552	0.88125
10	BFGS	20	0.83055	0.74206	0.83677
10	BFGS	30	2.27261	1.85309	2.29587
10	CGFR	2	15.7458	15.8318	15.9914
10	CGFR	5	13.4649	13.2848	13.6846
10	CGFR	10	13.6953	13.2215	13.8863
10	CGFR	20	10.8173	9.54006	10.8956
10	CGFR	30	25.9742	23.9044	26.3482
10	GD	2	915.731	913.556	933.188
10	GD	5	2330.74	2298.80	2353.20
10	GD	10	2918.21	2738.80	2931.90
10	GD	20	28388.1	25418.2	28165.4
10	GD	30	14333.8	13125.6	14556.4
10	LM	2	0.15795	0.16063	0.15573
10	LM	5	0.11944	0.12335	0.12115
10	LM	10	0.10867	0.11251	0.11057
10	LM	20	0.10041	0.10410	0.10185
10	LM	30	0.09515	0.09939	0.09689
10	RP	2	253.861	253.981	254.207
10	RP	5	339.565	341.273	335.989
10	RP	10	37.3203	38.7413	36.7550
10	RP	20	1116.49	1122.22	1106.28
10	RP	30	38.9786	41.3375	37.3005

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
11	BFGS	2	0.22406	0.23345	0.22802
11	BFGS	5	0.64612	0.67340	0.66472
11	BFGS	10	0.80494	0.89700	0.85770
11	BFGS	20	0.90807	0.99614	0.94644
11	BFGS	30	0.70582	0.76212	0.72757
11	CGFR	2	20.9697	20.5290	20.9572
11	CGFR	5	11.9953	12.0925	12.1761
11	CGFR	10	11.1476	11.7601	11.6160
11	CGFR	20	11.6365	12.7941	12.3570
11	CGFR	30	29.5515	33.8865	31.9858
11	GD	2	492.830	483.939	492.406
11	GD	5	2407.95	2370.03	2404.20
11	GD	10	5418.14	5017.55	5407.06
11	GD	20	7766.29	7214.26	7723.76
11	GD	30	14395.1	13072.9	14429.0
11	LM	2	0.15207	0.15784	0.15349
11	LM	5	0.12474	0.12504	0.12276
11	LM	10	0.11603	0.11466	0.11280
11	LM	20	0.10233	0.10057	0.09870
11	LM	30	0.09909	0.09763	0.09541
11	RP	2	171.271	171.370	171.597
11	RP	5	35.1635	36.0463	35.8860
11	RP	10	10.3911	10.8213	10.5911
11	RP	20	145.218	142.421	144.173
11	RP	30	116.969	124.678	119.255

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
12	BFGS	2	6.24750	5.92517	6.37103
12	BFGS	5	0.31504	0.28994	0.31029
12	BFGS	10	0.68143	0.63353	0.69294
12	BFGS	20	0.26156	0.24488	0.25815
12	BFGS	30	1.37374	1.21694	1.34015
12	CGFR	2	14.2734	13.5481	14.5792
12	CGFR	5	11.4351	10.7701	11.6514
12	CGFR	10	19.1532	17.5339	19.3639
12	CGFR	20	11.4962	10.4988	11.2743
12	CGFR	30	29.1895	25.9965	27.7271
12	GD	2	729.629	691.973	744.797
12	GD	5	1660.51	1566.26	1698.55
12	GD	10	6617.05	5943.58	6779.07
12	GD	20	10769.6	9654.84	10919.5
12	GD	30	21004.0	18784.9	21412.3
12	LM	2	0.16343	0.15417	0.16118
12	LM	5	0.12821	0.12137	0.12562
12	LM	10	0.11401	0.10838	0.11208
12	LM	20	0.10361	0.09960	0.10271
12	LM	30	0.10056	0.09679	0.09993
12	RP	2	445.770	441.999	446.076
12	RP	5	27.6292	25.6302	28.8300
12	RP	10	115.298	110.667	118.655
12	RP	20	21.2567	20.4506	22.3729
12	RP	30	23.8406	22.9187	25.0676

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
13	BFGS	2	0.21323	0.20990	0.21030
13	BFGS	5	0.29693	0.30410	0.31586
13	BFGS	10	0.55542	0.56957	0.58623
13	BFGS	20	0.47120	0.50146	0.50770
13	BFGS	30	1.78702	1.91611	1.99149
13	CGFR	2	17.1329	16.3731	16.9566
13	CGFR	5	19.5302	18.4820	19.1345
13	CGFR	10	20.4385	20.3956	21.5669
13	CGFR	20	22.3930	23.0162	23.9460
13	CGFR	30	26.1097	29.2508	30.6625
13	GD	2	770.751	756.007	778.679
13	GD	5	1728.26	1669.42	1742.45
13	GD	10	4603.32	4277.81	4667.67
13	GD	20	12804.0	12032.0	13446.4
13	GD	30	9350.92	8641.58	9483.04
13	LM	2	0.16856	0.16790	0.16819
13	LM	5	0.12496	0.12682	0.12856
13	LM	10	0.11292	0.11586	0.11747
13	LM	20	0.09862	0.10008	0.10313
13	LM	30	0.09658	0.09760	0.10102
13	RP	2	23.1238	22.4102	22.8451
13	RP	5	29.3981	28.4928	29.1200
13	RP	10	22.1246	21.9285	22.4194
13	RP	20	47.1556	47.9641	47.1985
13	RP	30	74.7850	78.0385	75.6212

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
14	BFGS	2	0.25964	0.26268	0.25609
14	BFGS	5	0.58471	0.59703	0.58438
14	BFGS	10	0.40741	0.39564	0.39518
14	BFGS	20	0.32409	0.31564	0.31413
14	BFGS	30	0.56949	0.54867	0.53600
14	CGFR	2	12.3289	13.0982	12.5699
14	CGFR	5	20.4041	21.2281	20.2613
14	CGFR	10	9.47499	9.66512	9.12105
14	CGFR	20	14.2696	14.1918	12.9730
14	CGFR	30	26.9915	26.5758	24.7331
14	GD	2	504.430	524.684	485.258
14	GD	5	1437.89	1496.58	1388.58
14	GD	10	2855.28	2778.62	2896.88
14	GD	20	8549.37	8411.49	8228.26
14	GD	30	10420.5	10157.8	10116.2
14	LM	2	0.19065	0.18844	0.18523
14	LM	5	0.12182	0.12162	0.12020
14	LM	10	0.11416	0.11435	0.11250
14	LM	20	0.10267	0.10427	0.10095
14	LM	30	0.09672	0.09911	0.09543
14	RP	2	43.0984	44.8885	43.9190
14	RP	5	21.1255	21.1761	21.2816
14	RP	10	35.2174	34.7990	34.9691
14	RP	20	22.3542	21.8434	19.4215
14	RP	30	62.1146	62.0143	62.3553

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
15	BFGS	2	5.20935	4.95903	5.08005
15	BFGS	5	0.80460	0.85883	0.83954
15	BFGS	10	0.47908	0.51368	0.51727
15	BFGS	20	1.08901	1.26276	1.20293
15	BFGS	30	2.45334	2.91451	2.71063
15	CGFR	2	22.7060	21.9551	22.2419
15	CGFR	5	11.1199	11.1964	11.0956
15	CGFR	10	25.0169	25.3710	25.2043
15	CGFR	20	21.4254	25.4514	23.7544
15	CGFR	30	21.4529	23.8501	22.9287
15	GD	2	407.627	395.700	396.001
15	GD	5	1194.32	1150.87	1162.84
15	GD	10	5244.42	4781.10	5153.57
15	GD	20	9621.39	8726.60	9438.27
15	GD	30	16478.2	15175.8	16119.6
15	LM	2	0.15449	0.15196	0.16158
15	LM	5	0.12354	0.12310	0.12924
15	LM	10	0.11937	0.11873	0.12517
15	LM	20	0.10372	0.10330	0.10827
15	LM	30	0.09608	0.09565	0.09963
15	RP	2	179.693	180.829	178.573
15	RP	5	239.355	236.302	239.217
15	RP	10	17.5530	18.8257	18.4719
15	RP	20	67.4533	87.4792	78.0258
15	RP	30	26.2343	27.4569	26.6530

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
16	BFGS	2	1.21077	1.25103	1.27626
16	BFGS	5	0.48494	0.52278	0.51063
16	BFGS	10	2.49916	2.74801	2.69102
16	BFGS	20	0.73124	0.83574	0.79163
16	BFGS	30	0.78560	0.86997	0.82100
16	CGFR	2	21.9327	22.8149	21.9433
16	CGFR	5	10.0475	10.3911	10.1394
16	CGFR	10	6.85185	7.65102	7.43610
16	CGFR	20	11.2478	12.0802	11.8178
16	CGFR	30	18.5288	20.9613	19.9300
16	GD	2	492.065	492.967	499.873
16	GD	5	1893.75	1890.72	1924.13
16	GD	10	10581.8	10432.4	10656.3
16	GD	20	10478.1	10390.2	10833.6
16	GD	30	20135.1	18229.3	20530.2
16	LM	2	0.15486	0.16413	0.15911
16	LM	5	0.12830	0.13772	0.13296
16	LM	10	0.11402	0.12467	0.11921
16	LM	20	0.09719	0.10776	0.10221
16	LM	30	0.09401	0.10438	0.09878
16	RP	2	71.2663	73.2882	71.0940
16	RP	5	133.111	135.071	134.539
16	RP	10	44.8547	45.3905	44.6066
16	RP	20	60.2439	64.5233	62.3294
16	RP	30	63.8917	68.0045	63.7787

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
17	BFGS	2	0.20600	0.21125	0.21437
17	BFGS	5	0.22577	0.23053	0.22809
17	BFGS	10	0.33716	0.34001	0.33644
17	BFGS	20	0.44832	0.45632	0.43917
17	BFGS	30	0.43937	0.44056	0.42614
17	CGFR	2	11.2346	11.4341	11.5884
17	CGFR	5	10.8654	11.2799	10.9456
17	CGFR	10	16.0835	16.6902	16.1503
17	CGFR	20	13.7359	14.3530	13.5683
17	CGFR	30	31.6810	32.6991	30.8458
17	GD	2	742.645	766.775	753.048
17	GD	5	3811.86	3851.88	3901.02
17	GD	10	5438.45	5520.62	5514.25
17	GD	20	4784.19	4920.72	4858.13
17	GD	30	5566.58	5709.68	5554.13
17	LM	2	0.15798	0.15972	0.16441
17	LM	5	0.12833	0.12648	0.13191
17	LM	10	0.11642	0.11300	0.11875
17	LM	20	0.10184	0.09813	0.10328
17	LM	30	0.09854	0.09524	0.09988
17	RP	2	17.5821	17.9210	17.6456
17	RP	5	24.2357	24.6771	24.2046
17	RP	10	65.6736	66.8603	63.8321
17	RP	20	44.6991	45.1037	43.3512
17	RP	30	82.5803	83.8662	80.3113

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
18	BFGS	2	3.14812	3.10246	3.06913
18	BFGS	5	0.58289	0.57416	0.55708
18	BFGS	10	0.68230	0.66900	0.69588
18	BFGS	20	0.63479	0.63586	0.64955
18	BFGS	30	0.61383	0.61145	0.61735
18	CGFR	2	18.3891	18.1690	17.7081
18	CGFR	5	16.3110	16.1282	15.5183
18	CGFR	10	26.6167	27.0105	23.5052
18	CGFR	20	18.5449	18.8593	16.6138
18	CGFR	30	51.2647	50.6405	48.9308
18	GD	2	434.267	426.022	415.286
18	GD	5	1558.83	1535.45	1504.24
18	GD	10	2994.47	2816.89	2889.01
18	GD	20	7505.46	7074.62	7165.44
18	GD	30	34513.0	32348.7	33508.0
18	LM	2	0.15838	0.16119	0.16476
18	LM	5	0.11763	0.12096	0.12450
18	LM	10	0.11005	0.11282	0.11719
18	LM	20	0.09745	0.09992	0.10389
18	LM	30	0.09528	0.09788	0.10159
18	RP	2	24.6920	24.4168	24.3966
18	RP	5	24.4942	23.7292	24.1927
18	RP	10	34.0423	34.2031	35.0642
18	RP	20	2270.79	2260.96	2265.28
18	RP	30	31.4635	31.0883	32.6718

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
19	BFGS	2	1.94545	2.02018	1.95289
19	BFGS	5	1.26281	1.37350	1.42492
19	BFGS	10	0.45082	0.45710	0.47742
19	BFGS	20	1.02023	1.06411	1.10668
19	BFGS	30	0.79427	0.80409	0.83224
19	CGFR	2	25.0978	26.3320	25.3755
19	CGFR	5	8.56747	9.33464	8.98093
19	CGFR	10	24.8633	26.0024	25.4639
19	CGFR	20	16.8169	18.1218	18.5721
19	CGFR	30	10.5332	11.0328	11.3971
19	GD	2	538.313	578.169	558.323
19	GD	5	1888.23	2102.05	1997.21
19	GD	10	5685.76	5778.65	5829.17
19	GD	20	10005.2	10037.6	10227.0
19	GD	30	34334.8	34328.2	34461.8
19	LM	2	0.15101	0.15236	0.15053
19	LM	5	0.11938	0.11977	0.11730
19	LM	10	0.11412	0.11390	0.11228
19	LM	20	0.10489	0.10392	0.10255
19	LM	30	0.09939	0.09794	0.09663
19	RP	2	308.233	312.909	309.067
19	RP	5	237.352	235.490	236.958
19	RP	10	28.6576	28.3879	28.6159
19	RP	20	13.3364	13.5120	13.4419
19	RP	30	36.1316	36.9463	36.8707

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
20	BFGS	2	1.33726	1.35811	1.36718
20	BFGS	5	0.48499	0.48434	0.48870
20	BFGS	10	0.40694	0.40519	0.41166
20	BFGS	20	0.54734	0.54860	0.56384
20	BFGS	30	0.85348	0.82885	0.86637
20	CGFR	2	14.1700	13.9704	14.3718
20	CGFR	5	27.7000	27.2079	28.1440
20	CGFR	10	9.03355	8.59031	9.03577
20	CGFR	20	13.5155	12.9004	13.6554
20	CGFR	30	19.9096	18.1865	20.0037
20	GD	2	526.505	511.387	537.617
20	GD	5	1246.97	1194.01	1277.71
20	GD	10	5185.20	4935.28	5286.71
20	GD	20	5698.22	5406.08	5812.17
20	GD	30	20866.8	19604.0	21420.3
20	LM	2	0.16364	0.16525	0.16377
20	LM	5	0.12703	0.12759	0.12643
20	LM	10	0.11485	0.11480	0.11378
20	LM	20	0.10297	0.10227	0.10137
20	LM	30	0.09799	0.09730	0.09636
20	RP	2	195.295	193.828	194.385
20	RP	5	57.3528	58.6523	58.1925
20	RP	10	332.046	330.263	336.978
20	RP	20	30.0994	30.4609	30.4438
20	RP	30	39.2836	39.9856	39.9178

ANEXO 3. MEDIAS DE LOS ECM DE LA RED-INFORMACIÓN.

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
1	BFGS	2	11.7402	11.7090	11.6448
1	BFGS	5	1.98952	1.99824	2.01583
1	BFGS	10	1.04763	1.05842	1.05319
1	BFGS	20	1.34467	1.30837	1.35119
1	BFGS	30	1.23010	1.21977	1.23664
1	BFGS	40	1.19651	1.19232	1.19990
1	CGFR	2	32.3217	32.0682	32.0762
1	CGFR	5	9.56610	9.50688	9.55498
1	CGFR	10	10.1364	9.99407	10.2656
1	CGFR	20	10.6088	10.4938	10.6683
1	CGFR	30	26.7706	26.7457	27.2417
1	CGFR	40	23.2623	22.9164	22.7350
1	GD	2	336.146	328.834	336.697
1	GD	5	487.943	486.008	492.174
1	GD	10	1564.23	1543.68	1560.02
1	GD	20	10938.1	10275.6	10918.5
1	GD	30	6683.84	6317.41	6735.61
1	GD	40	8744.66	8484.27	8764.72
1	LM	2	0.11137	0.11418	0.11845
1	LM	5	0.07874	0.08267	0.08352
1	LM	10	0.05318	0.05773	0.05812
1	LM	20	0.04082	0.04518	0.04505
1	LM	30	0.03852	0.04296	0.04284
1	LM	40	0.03693	0.04157	0.04124
1	RP	2	219.109	220.202	217.972
1	RP	5	89.0593	88.9487	87.9949
1	RP	10	3331.77	3308.94	3357.82
1	RP	20	2166.83	2125.09	2221.66
1	RP	30	15301.5	14485.5	15271.1
1	RP	40	521.686	513.836	504.937

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
2	BFGS	2	30.3689	27.9794	29.1812
2	BFGS	5	9.58893	8.89083	9.22069
2	BFGS	10	0.93367	0.91791	0.93313
2	BFGS	20	1.06764	1.06401	1.06572
2	BFGS	30	1.33712	1.32368	1.34815
2	BFGS	40	1.50608	1.51100	1.47775
2	CGFR	2	33.5786	31.2177	32.2695
2	CGFR	5	9.12316	8.67081	8.85551
2	CGFR	10	9.63678	9.40545	9.49607
2	CGFR	20	11.7215	11.3839	11.5951
2	CGFR	30	17.4855	17.3966	17.3888
2	CGFR	40	23.9736	23.4394	23.8285
2	GD	2	295.258	292.254	300.773
2	GD	5	612.952	607.138	617.117
2	GD	10	2848.44	2831.25	2902.03
2	GD	20	3625.60	3406.48	3646.66
2	GD	30	2785.52	2606.83	2785.24
2	GD	40	14879.7	13943.0	15116.7
2	LM	2	0.10293	0.10681	0.10533
2	LM	5	0.07319	0.07842	0.07608
2	LM	10	0.05137	0.05805	0.05556
2	LM	20	0.04084	0.04763	0.04447
2	LM	30	0.03765	0.04463	0.04101
2	LM	40	0.03608	0.04311	0.03941
2	RP	2	1605.95	1624.68	1621.45
2	RP	5	71.3779	70.7530	70.9493
2	RP	10	70.9218	69.4920	70.0651
2	RP	20	280.469	277.166	274.072
2	RP	30	19965.8	20181.7	20245.0
2	RP	40	568.995	551.438	553.308

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
3	BFGS	2	17.9011	17.0126	17.2752
3	BFGS	5	11.4367	10.8358	10.9329
3	BFGS	10	1.08364	1.06005	1.07397
3	BFGS	20	0.90495	0.90039	0.90348
3	BFGS	30	1.35684	1.33466	1.35665
3	BFGS	40	1.65805	1.63680	1.64385
3	CGFR	2	28.9948	27.3535	27.7896
3	CGFR	5	13.5977	12.9952	13.0911
3	CGFR	10	8.40328	8.00051	8.36566
3	CGFR	20	12.0319	11.8929	11.9952
3	CGFR	30	17.3756	17.0563	17.3078
3	CGFR	40	25.4168	25.3686	25.5632
3	GD	2	331.438	329.265	337.063
3	GD	5	674.967	648.644	681.734
3	GD	10	2277.35	2245.80	2308.78
3	GD	20	4698.85	4550.54	4788.30
3	GD	30	11706.9	11240.8	11758.5
3	GD	40	13005.6	12299.4	13163.0
3	LM	2	0.11083	0.11467	0.11232
3	LM	5	0.07845	0.08174	0.08128
3	LM	10	0.05433	0.05602	0.05707
3	LM	20	0.04084	0.04205	0.04395
3	LM	30	0.03896	0.04022	0.04234
3	LM	40	0.03825	0.03942	0.04178
3	RP	2	117.190	117.497	117.552
3	RP	5	104.541	104.219	104.451
3	RP	10	1198.71	1197.10	1201.37
3	RP	20	195.806	193.264	196.983
3	RP	30	258.331	257.702	260.592
3	RP	40	1994.18	1915.61	1981.55

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
4	BFGS	2	10.3768	10.6545	10.8077
4	BFGS	5	9.16994	9.46419	9.55479
4	BFGS	10	1.05638	1.08712	1.08102
4	BFGS	20	0.98918	0.99080	0.99503
4	BFGS	30	1.23877	1.24560	1.23729
4	BFGS	40	1.41953	1.44953	1.45882
4	CGFR	2	29.1785	30.0838	30.2747
4	CGFR	5	11.9965	12.2733	12.2113
4	CGFR	10	9.79847	9.66708	9.77073
4	CGFR	20	10.7665	10.8436	10.8507
4	CGFR	30	34.5878	34.1953	33.8592
4	CGFR	40	37.4378	37.0277	37.2364
4	GD	2	257.476	250.063	255.093
4	GD	5	1125.97	1128.18	1141.88
4	GD	10	4238.49	4100.15	4161.98
4	GD	20	2892.13	2742.53	2857.76
4	GD	30	6643.69	6433.94	6537.99
4	GD	40	5523.71	5248.07	5482.64
4	LM	2	0.11305	0.11534	0.11743
4	LM	5	0.07249	0.07326	0.07633
4	LM	10	0.05459	0.05615	0.05787
4	LM	20	0.04097	0.04347	0.04415
4	LM	30	0.03950	0.04229	0.04291
4	LM	40	0.03658	0.03960	0.04029
4	RP	2	52.0846	53.2460	53.9786
4	RP	5	42.5457	43.0219	43.2469
4	RP	10	188.590	188.409	187.428
4	RP	20	249.768	247.669	245.143
4	RP	30	40502.5	40722.7	40194.4
4	RP	40	3654.28	3677.26	3682.68

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
5	BFGS	2	9.97835	9.91367	10.5067
5	BFGS	5	15.7599	15.6648	16.6112
5	BFGS	10	0.76574	0.75044	0.77761
5	BFGS	20	0.83965	0.82343	0.84371
5	BFGS	30	1.65756	1.61730	1.67537
5	BFGS	40	1.33352	1.28558	1.33552
5	CGFR	2	24.8447	24.4743	25.8367
5	CGFR	5	14.6523	14.3757	15.3201
5	CGFR	10	8.19845	7.95250	8.28047
5	CGFR	20	9.80118	9.63483	9.99656
5	CGFR	30	21.7079	21.4471	21.7950
5	CGFR	40	29.7065	29.0440	29.9331
5	GD	2	529.109	521.665	539.029
5	GD	5	625.293	590.724	629.640
5	GD	10	1940.20	1866.45	1968.42
5	GD	20	3957.21	3695.82	3913.84
5	GD	30	3928.33	3694.68	3968.70
5	GD	40	7143.70	6868.86	7346.67
5	LM	2	0.10615	0.10482	0.11434
5	LM	5	0.06508	0.06529	0.07084
5	LM	10	0.05008	0.05087	0.05545
5	LM	20	0.04084	0.04177	0.04576
5	LM	30	0.03787	0.03918	0.04266
5	LM	40	0.03726	0.03857	0.04196
5	RP	2	232.496	229.993	235.351
5	RP	5	64.8498	64.5286	66.9791
5	RP	10	114.698	115.290	116.876
5	RP	20	11886.2	11787.6	11964.9
5	RP	30	4033.22	3895.48	4011.65
5	RP	40	472.437	471.138	479.323

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
6	BFGS	2	22.9649	23.6776	22.5458
6	BFGS	5	4.38083	4.54776	4.35365
6	BFGS	10	3.61206	3.77125	3.55674
6	BFGS	20	3.40482	3.52397	3.41583
6	BFGS	30	1.22495	1.26154	1.24717
6	BFGS	40	1.57292	1.61008	1.59666
6	CGFR	2	24.3632	25.1565	24.0100
6	CGFR	5	9.14179	9.32478	9.03873
6	CGFR	10	7.29779	7.53069	7.31474
6	CGFR	20	15.2226	15.4015	14.9708
6	CGFR	30	35.3074	34.9075	35.7077
6	CGFR	40	29.2787	29.4065	29.5210
6	GD	2	343.269	336.242	351.465
6	GD	5	907.540	883.788	925.636
6	GD	10	2507.36	2438.77	2603.11
6	GD	20	6392.29	6079.80	6464.51
6	GD	30	7268.39	7003.43	7683.91
6	GD	40	5646.75	5365.86	5893.81
6	LM	2	0.10719	0.11029	0.10694
6	LM	5	0.06320	0.06724	0.06203
6	LM	10	0.05453	0.05892	0.05425
6	LM	20	0.04269	0.04694	0.04215
6	LM	30	0.03896	0.04324	0.03889
6	LM	40	0.03798	0.04224	0.03814
6	RP	2	39.3916	40.0311	40.0398
6	RP	5	144.779	146.241	145.764
6	RP	10	1887.28	1886.67	1881.42
6	RP	20	3127.81	3189.12	3113.71
6	RP	30	413.692	427.881	409.982
6	RP	40	691.758	717.902	690.186

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
7	BFGS	2	9.74070	9.39530	9.66455
7	BFGS	5	4.30388	4.18257	4.27277
7	BFGS	10	0.97064	0.97021	0.95714
7	BFGS	20	1.15658	1.15418	1.16324
7	BFGS	30	1.27240	1.28995	1.28615
7	BFGS	40	1.78822	1.80819	1.77821
7	CGFR	2	30.3026	29.6920	30.1293
7	CGFR	5	17.3054	17.1516	16.9580
7	CGFR	10	9.86672	9.66797	9.73060
7	CGFR	20	12.4166	12.3176	12.4205
7	CGFR	30	19.0206	19.3588	19.0952
7	CGFR	40	22.2309	22.2424	22.0263
7	GD	2	427.665	427.356	423.934
7	GD	5	795.782	809.089	795.842
7	GD	10	2814.57	2810.30	2813.32
7	GD	20	6374.56	6031.62	6358.20
7	GD	30	4094.01	3952.14	4071.22
7	GD	40	11764.7	11369.0	11735.5
7	LM	2	0.09933	0.10810	0.10988
7	LM	5	0.06736	0.07342	0.07623
7	LM	10	0.05263	0.05747	0.05977
7	LM	20	0.04087	0.04475	0.04703
7	LM	30	0.04065	0.04474	0.04688
7	LM	40	0.03555	0.03943	0.04180
7	RP	2	34.7218	34.0091	34.4362
7	RP	5	123.904	120.526	122.923
7	RP	10	50.7901	50.8306	49.6984
7	RP	20	256.025	245.939	251.597
7	RP	30	224.566	219.446	220.841
7	RP	40	37233.0	37426.9	38244.4

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
8	BFGS	2	23.3595	21.9080	23.1974
8	BFGS	5	13.0570	12.2147	13.0332
8	BFGS	10	4.10763	3.88522	4.10111
8	BFGS	20	1.01993	1.00610	1.04197
8	BFGS	30	1.24729	1.23733	1.27642
8	BFGS	40	1.28106	1.28817	1.32439
8	CGFR	2	35.0096	33.0470	35.0666
8	CGFR	5	9.16985	8.85066	9.22266
8	CGFR	10	5.56930	5.41130	5.66657
8	CGFR	20	10.1052	10.0245	10.1319
8	CGFR	30	19.3379	19.1355	19.6968
8	CGFR	40	27.6314	27.1913	28.1049
8	GD	2	334.276	324.796	334.313
8	GD	5	811.277	771.731	817.156
8	GD	10	2027.90	1931.12	2045.08
8	GD	20	5342.78	5028.82	5331.71
8	GD	30	3329.95	3229.03	3337.43
8	GD	40	13946.1	13064.4	13976.0
8	LM	2	0.11066	0.10922	0.11469
8	LM	5	0.06222	0.05872	0.06564
8	LM	10	0.05125	0.04793	0.05450
8	LM	20	0.04322	0.03950	0.04578
8	LM	30	0.04165	0.03827	0.04415
8	LM	40	0.03852	0.03515	0.04105
8	RP	2	42.8885	41.2481	43.0903
8	RP	5	62.8704	62.4425	63.1097
8	RP	10	85.0057	82.5068	85.7347
8	RP	20	3894.79	3743.36	3913.55
8	RP	30	361.195	353.963	367.454
8	RP	40	355.759	349.654	363.351

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
9	BFGS	2	10.8493	10.4539	10.6901
9	BFGS	5	7.30556	7.02089	7.21967
9	BFGS	10	5.76869	5.66404	5.69962
9	BFGS	20	1.07334	1.07082	1.10420
9	BFGS	30	1.16077	1.15548	1.17987
9	BFGS	40	1.57963	1.60030	1.63947
9	CGFR	2	22.0625	21.4259	21.6074
9	CGFR	5	14.6272	14.1981	14.2445
9	CGFR	10	8.37951	8.09359	8.31869
9	CGFR	20	10.2440	10.3668	10.6122
9	CGFR	30	21.7902	21.6696	22.1543
9	CGFR	40	27.4666	27.7371	28.3542
9	GD	2	281.008	273.029	281.574
9	GD	5	1145.66	1125.54	1154.58
9	GD	10	1543.59	1455.24	1549.39
9	GD	20	7633.43	7247.87	7716.92
9	GD	30	7878.10	7242.73	7900.87
9	GD	40	13643.1	12859.7	13678.2
9	LM	2	0.11920	0.11584	0.12164
9	LM	5	0.06811	0.06552	0.07089
9	LM	10	0.05078	0.04904	0.05365
9	LM	20	0.04223	0.04057	0.04490
9	LM	30	0.04054	0.03917	0.04325
9	LM	40	0.03897	0.03787	0.04184
9	RP	2	40.1790	38.5541	39.9561
9	RP	5	53.8118	51.9937	53.6728
9	RP	10	88.2931	85.0292	88.6604
9	RP	20	186.316	181.684	188.187
9	RP	30	15013.0	14825.8	15252.1
9	RP	40	26747.9	26566.4	26588.9

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
10	BFGS	2	21.3799	21.6434	20.9203
10	BFGS	5	12.2534	12.4463	11.9572
10	BFGS	10	4.04600	4.02675	4.08064
10	BFGS	20	1.27545	1.29771	1.30068
10	BFGS	30	1.43746	1.49553	1.48187
10	BFGS	40	1.43481	1.49451	1.48660
10	CGFR	2	28.8663	28.9689	28.3394
10	CGFR	5	16.0698	16.1294	15.8699
10	CGFR	10	5.46299	5.46459	5.53380
10	CGFR	20	10.1163	10.1545	10.1710
10	CGFR	30	21.9159	21.7623	22.0950
10	CGFR	40	30.3805	30.9686	30.5629
10	GD	2	341.374	337.340	342.174
10	GD	5	757.565	750.047	761.320
10	GD	10	2441.27	2241.01	2435.24
10	GD	20	8401.95	7774.30	8482.63
10	GD	30	5887.66	5749.22	5945.14
10	GD	40	7177.42	6573.84	7173.52
10	LM	2	0.10761	0.11082	0.10721
10	LM	5	0.07020	0.07512	0.07069
10	LM	10	0.04889	0.05387	0.04956
10	LM	20	0.04066	0.04514	0.04117
10	LM	30	0.03873	0.04363	0.03929
10	LM	40	0.03693	0.04176	0.03769
10	RP	2	53.1525	53.8058	52.7941
10	RP	5	114.898	116.644	111.942
10	RP	10	76.5131	76.6503	74.8969
10	RP	20	229.540	226.998	222.355
10	RP	30	318.658	319.981	314.243
10	RP	40	33230.3	33214.1	33424.4

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
11	BFGS	2	19.6014	20.3205	19.6575
11	BFGS	5	8.08566	8.35699	8.10885
11	BFGS	10	0.88658	0.90617	0.90820
11	BFGS	20	1.20828	1.23326	1.22539
11	BFGS	30	1.09188	1.12764	1.11592
11	BFGS	40	1.26049	1.27637	1.27386
11	CGFR	2	23.9509	24.7850	24.0907
11	CGFR	5	9.53801	9.77275	9.90275
11	CGFR	10	7.37683	7.51454	7.54046
11	CGFR	20	10.0004	10.2105	10.0836
11	CGFR	30	14.9063	14.9367	15.0448
11	CGFR	40	21.6061	21.8081	21.8440
11	GD	2	217.030	217.283	217.384
11	GD	5	579.736	585.318	591.384
11	GD	10	2628.63	2552.39	2648.53
11	GD	20	7521.12	7035.87	7501.37
11	GD	30	13054.5	12956.8	13285.8
11	GD	40	17417.5	16408.5	17259.4
11	LM	2	0.11646	0.12045	0.12412
11	LM	5	0.06564	0.06896	0.07386
11	LM	10	0.05079	0.05240	0.05763
11	LM	20	0.04000	0.04112	0.04681
11	LM	30	0.03825	0.03920	0.04487
11	LM	40	0.03720	0.03805	0.04374
11	RP	2	166.228	168.483	167.256
11	RP	5	322.756	326.974	321.626
11	RP	10	126.039	127.694	126.945
11	RP	20	197.937	199.556	201.176
11	RP	30	356.232	349.526	353.691
11	RP	40	29844.2	29903.4	29932.4

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
12	BFGS	2	15.4155	15.8842	15.9849
12	BFGS	5	5.00539	5.14459	5.20703
12	BFGS	10	1.12639	1.16913	1.17301
12	BFGS	20	1.17301	1.18577	1.18939
12	BFGS	30	1.30223	1.32990	1.33114
12	BFGS	40	1.59560	1.62941	1.61605
12	CGFR	2	34.0552	35.2102	35.3692
12	CGFR	5	9.61030	9.96050	9.93450
12	CGFR	10	6.44008	6.53452	6.53224
12	CGFR	20	9.81487	9.87372	9.82093
12	CGFR	30	18.6330	18.8019	18.7928
12	CGFR	40	23.9020	24.1788	23.8721
12	GD	2	322.728	306.882	321.156
12	GD	5	1143.46	1043.88	1144.14
12	GD	10	3423.68	3330.98	3394.60
12	GD	20	3421.56	3243.93	3407.76
12	GD	30	2914.92	2935.13	2948.26
12	GD	40	4031.11	3931.32	3995.35
12	LM	2	0.10712	0.10398	0.10926
12	LM	5	0.07283	0.07101	0.07247
12	LM	10	0.05280	0.05077	0.05236
12	LM	20	0.04453	0.04267	0.04460
12	LM	30	0.04012	0.03881	0.04060
12	LM	40	0.03925	0.03808	0.03959
12	RP	2	3679.86	3749.85	3739.04
12	RP	5	70.6946	71.3447	71.2754
12	RP	10	108.765	108.995	108.778
12	RP	20	295.010	295.204	292.478
12	RP	30	16454.0	16473.0	16402.8
12	RP	40	8910.16	8840.08	8841.34

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
13	BFGS	2	14.7677	15.3977	14.4164
13	BFGS	5	1.60885	1.67383	1.57634
13	BFGS	10	1.53927	1.60193	1.52985
13	BFGS	20	0.79338	0.80678	0.80208
13	BFGS	30	1.29946	1.34206	1.32162
13	BFGS	40	1.68771	1.72848	1.70802
13	CGFR	2	33.8772	35.3027	32.9938
13	CGFR	5	14.1073	14.6913	13.8184
13	CGFR	10	6.03312	6.02254	5.78188
13	CGFR	20	12.5176	12.8711	12.5772
13	CGFR	30	18.4218	18.7426	18.6709
13	CGFR	40	23.7001	23.8061	23.7141
13	GD	2	347.487	339.917	352.345
13	GD	5	511.571	505.227	520.579
13	GD	10	2404.80	2356.66	2444.84
13	GD	20	5673.01	5517.27	5694.91
13	GD	30	4599.67	4328.64	4685.61
13	GD	40	6695.75	6481.90	6671.50
13	LM	2	0.11293	0.11315	0.11323
13	LM	5	0.07510	0.07585	0.07572
13	LM	10	0.05526	0.05577	0.05673
13	LM	20	0.04287	0.04375	0.04491
13	LM	30	0.03916	0.03996	0.04125
13	LM	40	0.03960	0.04068	0.04159
13	RP	2	136.861	138.361	137.187
13	RP	5	99.5049	101.402	98.2945
13	RP	10	80.8981	82.8853	81.5277
13	RP	20	1618.36	1650.98	1663.24
13	RP	30	287.843	300.885	290.283
13	RP	40	438.079	454.535	447.059

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
14	BFGS	2	24.7097	23.5751	25.0564
14	BFGS	5	4.78802	4.62200	4.88030
14	BFGS	10	0.74839	0.74096	0.76120
14	BFGS	20	0.75626	0.74420	0.77330
14	BFGS	30	1.24792	1.25166	1.29509
14	BFGS	40	1.56706	1.56717	1.59850
14	CGFR	2	35.1876	33.7370	35.7156
14	CGFR	5	10.4199	10.1246	10.6364
14	CGFR	10	7.35793	7.34305	7.57360
14	CGFR	20	8.90825	8.80689	9.04658
14	CGFR	30	15.7047	15.3594	15.9520
14	CGFR	40	35.7164	34.8932	36.7589
14	GD	2	274.292	263.354	280.917
14	GD	5	970.319	885.961	992.894
14	GD	10	1872.20	1800.37	1927.10
14	GD	20	7098.67	6515.78	7380.50
14	GD	30	5618.63	5325.71	5735.90
14	GD	40	7193.32	6856.80	7294.92
14	LM	2	0.11354	0.11528	0.11218
14	LM	5	0.07569	0.07433	0.07381
14	LM	10	0.05472	0.05276	0.05260
14	LM	20	0.04373	0.04214	0.04200
14	LM	30	0.04061	0.03898	0.03901
14	LM	40	0.04049	0.03872	0.03888
14	RP	2	125.735	125.652	126.232
14	RP	5	1032.88	1016.99	1032.73
14	RP	10	103.342	105.062	105.966
14	RP	20	223.971	222.427	228.144
14	RP	30	327.355	327.073	339.052
14	RP	40	535.672	536.233	548.710

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
15	BFGS	2	20.1885	20.4046	20.9625
15	BFGS	5	3.25030	3.21310	3.36827
15	BFGS	10	0.92022	0.89394	0.90971
15	BFGS	20	0.91063	0.88818	0.90419
15	BFGS	30	1.32283	1.33294	1.34000
15	BFGS	40	1.78848	1.80094	1.76150
15	CGFR	2	27.6470	27.8685	28.5339
15	CGFR	5	15.2962	15.2666	15.4125
15	CGFR	10	8.81027	8.64262	8.80320
15	CGFR	20	8.72987	8.62586	8.84897
15	CGFR	30	19.4402	19.3569	19.2810
15	CGFR	40	22.4045	22.2068	22.3967
15	GD	2	226.486	216.910	228.330
15	GD	5	925.291	884.985	921.987
15	GD	10	1801.84	1668.05	1792.22
15	GD	20	8394.31	7377.45	8266.86
15	GD	30	9247.28	9055.25	9210.48
15	GD	40	11752.2	10615.9	11698.9
15	LM	2	0.11377	0.10646	0.11482
15	LM	5	0.06788	0.06339	0.06981
15	LM	10	0.05165	0.04859	0.05455
15	LM	20	0.04252	0.04044	0.04599
15	LM	30	0.03901	0.03745	0.04280
15	LM	40	0.03804	0.03680	0.04223
15	RP	2	31.2460	31.4321	32.6437
15	RP	5	71.4263	72.3964	73.8655
15	RP	10	139.497	141.811	141.840
15	RP	20	18753.5	18523.5	18615.4
15	RP	30	303.147	308.150	306.346
15	RP	40	450.864	458.909	453.987

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
16	BFGS	2	12.7108	12.4214	12.8425
16	BFGS	5	4.76977	4.68408	4.80208
16	BFGS	10	1.28906	1.26961	1.31244
16	BFGS	20	0.95088	0.94486	0.95799
16	BFGS	30	1.12423	1.11103	1.13218
16	BFGS	40	1.61716	1.59296	1.60286
16	CGFR	2	20.6036	20.0766	20.7549
16	CGFR	5	9.80025	9.46736	9.74491
16	CGFR	10	12.6256	12.2669	12.7137
16	CGFR	20	13.4765	13.2041	13.4307
16	CGFR	30	22.9971	22.4215	22.9393
16	CGFR	40	31.2559	30.8509	31.2881
16	GD	2	317.292	307.013	315.946
16	GD	5	615.377	590.160	607.965
16	GD	10	1913.29	1837.24	1910.09
16	GD	20	2879.38	2718.01	2837.25
16	GD	30	2866.00	2725.38	2824.91
16	GD	40	7186.80	6983.62	7157.36
16	LM	2	0.10275	0.10388	0.11014
16	LM	5	0.06885	0.07042	0.07638
16	LM	10	0.04730	0.04870	0.05355
16	LM	20	0.04236	0.04353	0.04856
16	LM	30	0.03876	0.03980	0.04494
16	LM	40	0.03728	0.03849	0.04353
16	RP	2	352.778	355.330	355.137
16	RP	5	110.522	110.919	112.285
16	RP	10	127.491	126.638	127.565
16	RP	20	214.916	214.018	216.022
16	RP	30	343.446	341.749	345.919
16	RP	40	33120.9	33060.0	33026.1

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
17	BFGS	2	21.5703	22.5182	21.7504
17	BFGS	5	5.03487	5.16479	5.02763
17	BFGS	10	1.23941	1.24197	1.24099
17	BFGS	20	0.99850	0.97685	1.00939
17	BFGS	30	1.05268	1.02998	1.08387
17	BFGS	40	1.48389	1.46908	1.51946
17	CGFR	2	29.4183	30.4880	29.6731
17	CGFR	5	15.9654	16.3482	16.0444
17	CGFR	10	12.3515	12.2231	12.4386
17	CGFR	20	11.0122	10.7204	11.1632
17	CGFR	30	29.6724	29.6543	29.7335
17	CGFR	40	31.4220	30.3058	31.9547
17	GD	2	483.802	464.230	468.221
17	GD	5	590.230	566.722	578.276
17	GD	10	1803.12	1733.76	1765.84
17	GD	20	6484.28	6027.18	6331.37
17	GD	30	7196.50	6758.71	7141.97
17	GD	40	9558.23	8733.25	9220.37
17	LM	2	0.10204	0.10495	0.09938
17	LM	5	0.06908	0.07326	0.06877
17	LM	10	0.05191	0.05569	0.05228
17	LM	20	0.04122	0.04509	0.04229
17	LM	30	0.03794	0.04210	0.03924
17	LM	40	0.03872	0.04273	0.04011
17	RP	2	125.538	127.250	125.960
17	RP	5	75.5896	76.4134	76.1875
17	RP	10	4485.34	4319.44	4518.76
17	RP	20	4076.51	3843.71	4155.84
17	RP	30	327.558	332.220	335.864
17	RP	40	28597.7	28327.4	28629.0

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
18	BFGS	2	16.7491	16.7608	16.7256
18	BFGS	5	10.0283	10.0324	10.0151
18	BFGS	10	0.99092	0.98240	0.98684
18	BFGS	20	0.98936	1.00059	0.99198
18	BFGS	30	1.10101	1.09502	1.13508
18	BFGS	40	1.43832	1.44195	1.47228
18	CGFR	2	40.1203	40.1373	40.1568
18	CGFR	5	9.41980	9.39116	9.40482
18	CGFR	10	8.22248	8.28975	8.20879
18	CGFR	20	9.34026	9.56677	9.46141
18	CGFR	30	17.4819	17.7878	17.7970
18	CGFR	40	35.2965	35.5668	35.2223
18	GD	2	288.565	289.875	287.069
18	GD	5	1572.65	1454.84	1608.55
18	GD	10	2757.96	2755.34	2710.14
18	GD	20	2438.96	2362.73	2412.24
18	GD	30	4476.38	4508.96	4399.39
18	GD	40	6710.08	6778.50	6505.97
18	LM	2	0.11055	0.10401	0.10774
18	LM	5	0.07598	0.06999	0.07478
18	LM	10	0.05516	0.04973	0.05471
18	LM	20	0.04484	0.03966	0.04415
18	LM	30	0.04125	0.03630	0.04070
18	LM	40	0.04054	0.03536	0.04021
18	RP	2	171.601	172.534	171.816
18	RP	5	84.5224	85.5286	84.9846
18	RP	10	102.317	103.605	102.848
18	RP	20	165.185	166.999	165.960
18	RP	30	395.360	403.507	400.647
18	RP	40	18126.8	17817.5	18190.7

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
19	BFGS	2	23.5437	23.3173	24.0589
19	BFGS	5	2.67605	2.62531	2.60951
19	BFGS	10	1.42007	1.43646	1.40134
19	BFGS	20	0.76810	0.77147	0.76367
19	BFGS	30	1.04425	1.05738	1.04284
19	BFGS	40	1.71015	1.72580	1.69856
19	CGFR	2	29.4997	28.8536	29.8267
19	CGFR	5	11.7375	11.4948	11.7045
19	CGFR	10	6.71600	6.55483	6.58320
19	CGFR	20	7.28311	7.25689	7.16234
19	CGFR	30	21.8724	21.7082	21.7624
19	CGFR	40	25.4897	25.0193	25.0062
19	GD	2	373.666	348.863	354.845
19	GD	5	1535.05	1428.60	1524.99
19	GD	10	3084.46	2952.11	3029.11
19	GD	20	6658.60	6043.47	6567.36
19	GD	30	8937.49	8493.95	8894.10
19	GD	40	5642.17	5351.92	5346.29
19	LM	2	0.10395	0.10703	0.10154
19	LM	5	0.07653	0.07959	0.07452
19	LM	10	0.05253	0.05542	0.05176
19	LM	20	0.04326	0.04588	0.04306
19	LM	30	0.04099	0.04369	0.04076
19	LM	40	0.03765	0.04062	0.03775
19	RP	2	122.572	122.811	123.089
19	RP	5	83.2719	83.4682	84.1376
19	RP	10	98.8317	98.7536	99.9912
19	RP	20	11584.9	11689.7	11579.2
19	RP	30	308.990	313.331	315.018
19	RP	40	620.119	625.264	632.389

<u>Iteración</u>	<u>Método</u>	<u>Tamaño</u>	<u>ecm conj. entren.</u>	<u>ecm conj. valid.</u>	<u>ecm conj. test</u>
20	BFGS	2	16.7265	16.5470	16.3557
20	BFGS	5	9.71044	9.60141	9.53654
20	BFGS	10	1.12652	1.10354	1.09035
20	BFGS	20	1.06095	1.06568	1.04292
20	BFGS	30	1.23177	1.23595	1.21110
20	BFGS	40	1.61119	1.63184	1.58879
20	CGFR	2	37.1849	36.5498	36.0611
20	CGFR	5	13.0810	12.8846	12.5850
20	CGFR	10	7.82205	7.71665	7.53517
20	CGFR	20	9.98032	9.97146	9.72401
20	CGFR	30	16.4290	16.6703	16.2016
20	CGFR	40	23.3770	23.6160	23.0093
20	GD	2	454.628	442.471	443.640
20	GD	5	765.881	759.180	755.566
20	GD	10	3072.38	3035.95	3046.88
20	GD	20	8993.43	8089.71	8805.53
20	GD	30	4650.82	4501.26	4553.21
20	GD	40	7002.62	6886.35	6773.69
20	LM	2	0.10860	0.11167	0.11308
20	LM	5	0.06407	0.06709	0.06764
20	LM	10	0.05400	0.05726	0.05722
20	LM	20	0.04109	0.04318	0.04343
20	LM	30	0.03940	0.04178	0.04162
20	LM	40	0.03748	0.03988	0.03943
20	RP	2	43.9977	42.9900	42.8586
20	RP	5	53.1544	52.2447	51.6548
20	RP	10	1816.43	1882.54	1886.05
20	RP	20	274.800	261.828	258.150
20	RP	30	304.309	297.725	289.589
20	RP	40	412.119	401.981	390.246

BIBLIOGRAFÍA.

Amari, S. (1995). The EM algorithm and information geometry in neural network learning. *Neural Computation*, 7(1), 13-18.

Anderson, D. Z. (ed) (1988). *Neural Information Processing Systems*. American Institute of Physics, New York.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford Clarendon Press.

Bock, R. D. (1975). *Multivariate Statistical Methods In Behavioral Research*. Mc Graw-Hill, New York.

Broomhead, D. S. y Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321-355.

Broyden, C. G. (1969). A new double-rank minimization algorithm. *Notices of the American Mathematical Society*, 16, 670.

Bryden, H. L. y Stommel, H. M. (1982). Origin of Mediterranean outflow. *Journal Marine Research*, (40), 55-71.

Carpenter, G. A. y Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing*, 37, 54-115.

Carpenter, G. A. y Grossberg, S. (1987b). ART 2: stable self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26, 4919-4930.

Carpenter, G. A. y Grossberg, S. (1990). ART 3: hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, 3, 129-152.

Carpenter, G. A. y Grossberg, S. (1994). Self-organizing neural networks for supervised and unsupervised learning and prediction. En Cherkassky et al. (1994), p. 319-348.

Carter, D. J. T. (1980). *Echo-sounding correction tables*, 3rd edition. Hydrographic Department, Ministry of Defense, United Kingdom.

Carter, T. G., Flanagan, J. P., Jones, C. R., Marchant, F. L., Murchison, R. R., Rebman, J. A., Sylvester, J. C. y Whitney, J. C. (1972). *A New Bathymetric Chart and Physiography of the Mediterranean Sea*". En Stanley, D. J. (1972), p. 1-23.

Chen, C. T. y Millero, F. J. (1977). Speed of sound in seawater at high pressures. *Journal of the Acoustical Society of America*, 62(5), 1129-1135.

- Cheney, R. E. y Doblár, R. A. (1982). Structure and Variability of the Alboran Sea Frontal System. *Journal Geophysical Research*, 87, 585-594.
- Cheng, B. y Titterton, D.M. (1994). Neural Networks: a review from a statistical perspective. *Statistical Science*, 9(1), 2-54.
- Cherkassky, V., Friedman, J. H. y Wechsler, H. (eds) (1994). *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, Berlin-Springer.
- Cubiles de la Vega, M. D. (1999). *Redes de neuronas artificiales en el análisis de series temporales*. Tesis Doctoral Universidad de Sevilla.
- Cybenco, G. (1989). Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signals & Systems*, 2, 4, 303-314.
- Del Grosso, V. A. y Mader, C. W. (1972). Speed of sound in pure water. *Journal of the Acoustical Society of America*, 52, 1442-1446.
- Del Grosso, V. A. (1974). New equation for the speed of sound in natural waters (with comparisons to other equations). *Journal of the Acoustical Society of America*, 56(4), 1084-1091.
- Demuth, H. y Beale, M. (1994). *Neural Network TOOLBOX for Use with Matlab. User's Guide*. The math Works Inc.
- Denker, J. S., Schwartz, D., Wittner, B., Solla, S., Howard, R., Jackel, L. y Hopfield, J. (1987). *Complex Systems* 1, 877.
- Dushaw, B. D., Worcester, P. F., Cornuelle, B. D. y Howe, B. M. (1993). On equations for the speed of sound in sea water. *Journal of the Acoustical Society of America*, 93(1), 255-275.
- Fahlman, S. E. (1988). Faster-learning variations on back-propagation: an empirical study. En Touretzky et al. (1988), p. 38-51.
- Fine, T. L. (1999). *Feedforward Neural Network Methodology*. Springer Verlag.
- Fletcher, R. (1970). A new approach to variable metric methods. *Computer Journal*, 13, 317-322.
- Fletcher, R. y Reeves, C. M. (1964). Function minimization by conjugate gradients. *Computer Journal*, 7, 149-154.
- Fofonoff, N. P. y Millard R. C. Jr. (1983). Algorithms for computation of fundamental properties of seawater. UNESCO technical papers in marine science, nº 44, Division of Marine Sciences.
- Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2, 183-192.

- Gascard, J. C. y Richez, C. (1985). Water masses and circulation in the Western Alboran Sea and in the Strait of Gibraltar. *Progress in Oceanography*, 15, 157-215.
- Geman, S., Bienenstock, E. y Doursat, R. (1992). Neural Networks and the bias/variance dilemma. *Neural Computation*, 4, 1-58.
- Gill, P. E., Murry, W. y Wright, M. H. (1981). *Practical Optimization*. New York Academic Press.
- Goldfarb, D. (1970). A family of variable metric methods derived by variational means. *Mathematics of Computation*, 24, 23-26.
- Hastie, T., Tibshirani, R. y Friedman, J. (2001). *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Springer.
- Hatabian, G. (1994). Are Artificial Neural Networks a Dangerous Adversary for Statistical Data Analysis?. *Proceedings in Computational Statistics, XI Symposium, COMPSTAT'94*. Springer-Verlag.
- Hebb, D. O. (1949). *The Organization of Behaviour*. New York Wiley.
- Hecht-Nielsen, R. (1987). Kolmogorov's Mapping Neural Network Existence Theorem. *IEEE First Annual International Conference on Neural Networks*, artículo III-11.
- Hecht-Nielsen, R. (1990). *Neurocomputing*. Addison-Wesley, Reading, MA.
- Hertz, J., Krogh, A. y Palmer, R. (1991). *Introduction to the Theory of Neural Computation*. Addison Wesley.
- Hestenes, M. R. y Stiefel, E. L. (1962). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49 (6), 409-436.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational facilities. *Proceedings of the National Academy of Sciences of the USA*, 79, 2554-2558.
- Hornik, K., Stinchcombe, M. y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.
- Huang, S. C. y Huang, Y. F. (1991). Bounds on the number of hidden neurons in multilayer perceptrons. *IEEE Transactions on Neural Networks*, 2 (1), 47-55.
- Huang, T. C. y Stanley, D. J. (1972). Western Alboran Sea: Sediment Dispersal, Ponding and Reversal of Currents. En Stanley, D. J. (1972), p. 521-559.
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences*. Monterey: Brooks/Cole.

- Kohonen, T. (1982a). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Kohonen, T. (1982b). Analysis of a simple self-organizing process. *Biological Cybernetics*, 43, 135-140.
- Kohonen, T. (1984). *Self-organization and Associative Memory*. Berlin Springer, Heidelberg.
- Kung, S. Y. y Hwang, J. N. (1988). An algebraic projection analysis for optimal hidden units size and learning rate in backpropagation learning. *IEEE International Conference on Neural Networks*, San Diego, 1, 363-370.
- Lacombe, H. y Tchernia, P. (1972). Caracteres hydrologiques et circulation des eaux en Méditerranée. En Stanley, D. J. (1972), p. 26-36.
- Lapedes, A. y Farber, R. (1988). How neural nets work. En Anderson (1988) p. 442-456.
- Leroy, C. C. y Parthiot, F. (1998). Depth-pressure relationship in the oceans and seas. *Journal of the Acoustical Society of America*, 103(3), 1346-1352.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics II* (2), 164-168.
- Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, abril, 4-22.
- Looney, C. (1997). *Pattern recognition using neural networks*. Oxford University Press.
- Lurton, X. (2002). *An Introduction to Underwater Acoustic: Principles and Applications*. Chichester, United Kingdom, Springer/Praxis.
- Marquardt, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. *Journal of the Society of Industrial and Applied Mathematics* 11 (2), 431-441.
- Martín del Brío, B. y Sanz-Molina, A. (1998). *Redes Neuronales y Sistemas Borrosos*, 2ª edición. RA-MA.
- Matthews, D. J. (1939). *Tables of the velocity of sound in pure water and sea water for use in Echo-sounding and Sound-ranging*, 2nd edition. Hydrographic Department, Ministry of Defense, United Kingdom.
- Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics*, 11, 204-209.
- McCulloch, W. S. y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.

- McLellan, H. J. (1965). *Elements of Physical Oceanography*. Pergamon Press.
- Mediterranean Hydrological Atlas (MEDATLAS) (1997). Proyecto en el que colaboran los centros HNODC (Grecia), IEO (España), SHOM (Francia), IFREMER/SISMER (Francia) e ICES (Dinamarca).
- Mehrotra, K. G., Mohan, C. K. y Ranka, S. (1991). Bounds on the number of samples needed for neural learning. *IEES Transactions on Neural Networks*, 2 (6), 548-558.
- Meinen, C. S. y Watts, D. R. (1997). Further evidence that the sound-speed algorithm of Del Grosso is more accurate than that of Chen and Millero. *Journal of the Acoustical Society of America*, 102(4), 2058-2062.
- Miller, A. R. (1972). Speculations Concerning Bottom Circulation in the Mediterranean Sea. En Stanley, D. J. (1972), p. 37-42.
- Millero, F. J. y Xu Li (1994). Comments on "On equations for the speed of sound in seawater". *Journal of the Acoustical Society of America*, 95(5), 2757-2759.
- Milliman, J. D., Weiler, Y. y Stanley, D. J. (1972). Morphology and Carbonate Sedimentation on Shallow Banks in the Alboran Sea. En Stanley, D. J. (1972), p. 241-259.
- Minsky, M. L. y Papert, S. A. (1969). *Perceptrons: an Introduction to Computational Geometry*. MIT Press, Cambridge. Edición extendida en 1988.
- Müller, B., Reinhardt, J. y Strickland, M. (1995). *Neural Networks, An Introduction*, 2nd edition. Springer Verlag, Heidelberg.
- Munk, W., Worcester, P. y Wunsch, C. (1995). *Ocean Acoustic Tomography*. Cambridge University Press.
- Nguyen, D. y Widrow, B. (1990). Improving the learning speed of two-layer neural networks by choosing initial values of the adaptive weights. *IEEE International Conference on Neural Networks*, San Diego, 3, 21-26.
- Nordbotten, S. (1995). Editing Statistical Records by Neural Networks. *Journal of Official Statistics*, 11, 391-411.
- Nordbotten, S. (1996). Neural Network Imputation Applied to the Norwegian 1990 Population Census Data. *Journal of Official Statistics*, 12(4), 385-401.
- Norusis, M. J. (1998). *SPSS 8.0 Guide to Data Analysis*, 1st ed. Prentice Hall.
- Ocean Circulation* (1989). The Open University, Pergamon Press.

- Parrilla, G. y Kinder, T. H. (1987). Oceanografía física del mar de Alborán. Boletín del Instituto Español de Oceanografía, vol. 4, nº 1.
- Pickard, G. L. (1975). Descriptive Physical Oceanography, 2nd edition. Pergamon Press.
- Pike, J. M. y Beiboer, F. L. (1993). A comparison between algorithms for the speed of sound in seawater. The Hydrographic Society, Special Publication nº 34.
- Urick, R. J. (1975). Principles of Underwater Sound, 2nd edition. McGraw Hill.
- Riedmiller, M. y Braun, H. (1993). A direct adaptive method for faster backpropagation learning: the RPROP algorithm. IEEE International Conference on Neural Networks, San Francisco, 1, 586-591.
- Ripley, B. D. (1993). Statistical aspects of neural networks. Capítulo segundo del libro "Networks and Chaos. Statistical and Probability Aspects", editado por Barndorff-Nielsen et al. Chapman and Hall.
- Ripley, B. D. (1994). Neural Networks and Related Methods for Classification. Journal of the Royal Statistical Society, Series B, 56(3), 409-456.
- Ripley, B. D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.
- Rosenblatt, F. (1957). The perceptron: a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review 65, 386-408.
- Rosenblatt, F. (1962). Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington D. C. Spartan.
- Rumelhart, D. E., McClelland, J. L., Hinton, G. E. y Williams, R. J. (1986). Learning representations by backpropagating errors. Nature, 323, 533-536. También en Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1, 318-362. Cambridge, MA, MIT Press.
- Rumelhart, D. E., Widrow, B. y Lehr, M. A. (1994). The Basic Ideas in Neural Networks. Communications of the ACM, 37(3), 87-92.
- Russo, A. P. (1991). Neural networks for sonar signal processing. IEEE Conference on Neural Networks for Ocean Engineering, tutorial 8. Washington D. C.

Schumacher, M., Roßner, R. y Vach, W. (1996). Neural Networks and logistic regression: Part I. *Computational Statistics and Data Analysis* (21), 6661-682.

Seawater: Its composition, properties and behaviour (1989). The Open University, Pergamon Press.

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24, 145-160.

Smith, M. (1993). *Neural Networks for Statistical Modeling*. Van Nostrand Reinhold, New York.

Speisberger, J. L. y Metzger, K. (1991a). New estimates of sound speed in water. *Journal of the Acoustical Society of America*, 89(4), 1697-1700.

Speisberger, J. L. y Metzger, K. (1991b). A new algorithm for sound speed in seawater. *Journal of the Acoustical Society of America*, 89(6), 2677-2687.

Speisberger, J. L. (1993). Is Del Grosso's sound-speed algorithm correct?. *Journal of the Acoustical Society of America*, 93(4), 2235-2237.

Stanley, D. J. (ed) (1972). *The Mediterranean Sea: A Natural Sedimentation Laboratory*. Symposium on the Mediterranean Sea, Heidelberg, 1971. Dowden, Hutchinson & Ross, Inc.

Summerhayes, C. P. y Thorpe, S. A. (eds) (1996). *Oceanography. An Illustrated Guide*. Manson Publishing Ltd.

Tabachnik, B. G. y Fidel L. S. (1983). *Using Multivariate Statistics*. New York, Harper and Row.

Thisted, R. A. (1988). *Elements of Statistical Computing. Numerical Computation*. Chapman and Hall.

Tibshirani, R. (1995). A comparison of some error estimates for neural networks models. Department of Preventive Medicine and Department of Statistics, Universidad de Toronto. Technical Report.

Touretzky, D., Hinton, G. y Sejnowski, T. (eds) (1988). *Proceedings of the 1988 Connectionist Models Summer School*, Pittsburg. San Mateo, CA: Morgan Kaufmann.

Towell, G. y Shavlik, J. (1994). Knowledge-based artificial neural networks. *Artificial Intelligence*, 70, 119-165.

Weigend, A. S., Huberman, B. A. y Rumelhart, D. E. (1990). *Predicting the Future: A Connectionist Approach*. Stanford PDP Research Group Report 90-91.

Weigend, A. S. y LeBaron, A. (1994). Evaluating neural network predictors by bootstrapping. Proceedings of the International Conference on Neural Information Processing (ICONIP'94), Seúl.

White, H. (1990). Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings. Neural Networks, 3, 535-550.

Widrow, B. y Hoff, M. E. (1960). Adaptive switching circuits. IRE WESCON Convention Record, 4, 94-104.

Wilson, W. (1960). Equation for the speed of sound in sea water. Journal of the Acoustical Society of America, 32(10), 1357.

Winer, B. J., Brown, D. R. y Michels, K. M. (1991). Statistical principles in experimental design. New York, McGraw-Hill, Inc.

Wong, G. S. K. y Zhu, S. (1995). Speed of sound in seawater as a function of salinity, temperature and pressure. Journal of the Acoustical Society of America, 97(3), 1732-1736.