



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
MAESTRÍA EN INGENIERÍA EN EXPLORACIÓN Y EXPLOTACIÓN DE RECURSOS
NATURALES

**REDES NEURONALES PARA LA ESTIMACIÓN DE PROPIEDADES PETROFÍSICAS A
ESCALA SÍSMICA USANDO ATRIBUTOS SÍSMICOS Y LA PRUEBA GAMMA.**

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
MONTES RIOS EDGAR ALAN

TUTOR PRINCIPAL
DRA. ÚRSULA X. ITURRARÁN VIVEROS
FACULTAD DE CIENCIAS, UNAM

CIUDAD UNIVERSITARIA, Cd. MX., ABRIL 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: Dr. Pérez Cruz Guillermo Alejandro

Secretario: Dr. Coconi Morales Enrique

Vocal: Dra. Iturrarán Viveros Úrsula X

1^{er.} Suplente: Dr. Aquino López Ambrosio

2^{do.} Suplente: Dr. Negrete Cadena Rocío

Ciudad Universitaria, Ciudad de México

TUTOR DE TESIS:

Dra. Úrsula Iturrarán Viveros.

FIRMA

AGRADECIMIENTOS

A CONACYT y a la UNAM por el apoyo brindado.

A mis padres;

Primeramente quiero agradecer a mis padres, Ricardo e Isabel, que sin duda alguna representan todo el infinito amor que tiene nuestro creador por nosotros, y que con sus valores, consejos y confianza brindados a lo largo de mi vida han hecho a un hombre generoso, recto, y humilde.

A mis hermanos;

También me gustaría agradecer a mis hermanos- Eder, Jasiel y Beto, de quienes aprendí a ser compartido, responsable y respetuoso. Además de ser las primeras personas con las que empecé a descubrir este apasionante y al mismo tiempo abrumador mundo. Así mismo, me gustaría hacer mención especial a mi hermano Eder que aún cuando ya no está con nosotros, me sigue inspirando por todo lo que es, era y seguirá siendo para todos aquellos que lo conocimos y que lo seguimos manteniendo en nuestros corazones y mentes.

A mi tía;

A mi tía Angeles, que cuidó de mi y mis hermanos cuando eramos chicos, y que ahora lo sigue haciendo desde el cielo.

A mis amigos;

Agradezco a la vida por haberme permitido, haberme cruzado en esta etapa de mi vida con estas personas que eran en un inicio unos desconocidos, pero que ahora puedo llamarlos amigos: Gerardo, Sergio, Ángel, Armando, Lalo, Luciano, José, Edmi, Lorena e Ieni. Y a mi antiquísima amiga Andrea, que ya vamos por una década de habernos conocido!

A mi Tutora;

No por ser de los últimos de mis agradecimientos, signifique que mi agradecimiento sea menor. Le agradezco mucho, Dra. Úrsula Iturrarán por haber confiado en mi cuando nadie lo hizo, y haber sido paciente.. muy paciente conmigo. Logramos el objetivo!!! Gracias.

A mis Sinodales;

Finalmente, a los sinodales; Dra. Rocío Negrete, Dr. Guillermo Pérez, Dr. Enrique Coconi y Dr. Ambrosio Aquino, que gracias a sus sus observaciones y comentarios este trabajo resulto mejor.

RESUMEN

Un objetivo principal en la industria del petróleo en México y en el mundo, es efectuar de manera concisa la caracterización estática de los yacimientos. Para llevar a cabo tal tarea en las últimas tres décadas, se ha empleado de manera muy generalizada los atributos sísmicos. Hoy en día, una decena o más de atributos son extraídos de los datos sísmicos para la determinación de las propiedades petrofísicas de los yacimientos. Así mismo, también se establecido una tendencia de adquirir más registros de pozo, que proporcionan aún más información para lograr dicho objetivo. No obstante, a medida que aumenta la cantidad de información disponible, aumenta el riesgo que los resultados obtenidos no sean tan satisfactorios, debido a que en ellos se crean artificios matemáticos que carecen de sentido geológico. Para evitar esto, se han empleado varias técnicas estadísticas y de inteligencia artificial. Mediante este trabajo se pretende dar a conocer en aquellos que trabajan en la industria, el potencial presente en el empleo de la herramienta de análisis estadístico, la **prueba Gamma**, en la construcción de modelos no lineales, concretamente usando redes neuronales.

Un enfoque de uso muy común usado en la industria petrolera para la caracterización estática de los yacimientos es la integración de datos de pozo y datos sísmicos. Estos trabajos están ampliamente documentados por mencionar algunos autores: **Gastaldi et al. (1997)**, **Russell et al. (1997)** y **Hampson et al. (2001)**. Uno de los enfoques que ha ganado mucha popularidad en las últimas décadas para llevar a cabo esta integración, es el empleo de redes neuronales, la razón de ello es porque éstas son aproximadores universales; estas herramientas pueden aproximar cualquier función continua con una precisión arbitraria. A pesar de ser una herramienta computacional que ya tiene un camino recorrido en el modelado no lineal, el diseño y entrenamiento de una red neuronal sigue siendo considerado más un arte que una ciencia, esto es debido a numerosas dificultades bien conocidas en el área de las redes neuronales. Entre estos problemas destacan; la determinación de la configuración óptima de la red (estructura), la distribución óptima de los pesos para una red en específico, y por supuesto la más transcendental de todas ellas, la garantía de una buena capacidad de generalización de las redes. En este trabajo, se propone una metodología, cuyo elemento principal es la prueba Gamma. La prueba Gamma es una herramienta de modelado no lineal que nos permite examinar la naturaleza hipotética de la relación existente entre los datos de entrada y salida (atributos sísmicos-registros de pozo) en un set de datos numéricos. La introducción de este elemento, en el proceso de integración de datos sísmicos y datos de pozo mediante redes neuronales, permitirá optimizar el diseño de las redes neuronales, en el sentido que permitirá determinar que combinación de datos de entrada (atributos sísmicos) es la mejor para modelar un set de datos de salida (propiedad petrofísica), el cual es un problema que se comparte entre los diferentes métodos que se usan para el modelado no lineal. Así mismo, este elemento estadístico, al poder determinar el ruido, asumiéndose que la variación de ruido en cada una de las salidas está limitada y es independiente de los valores de entrada, nos dará la posibilidad de establecer un criterio para detener el entrenamiento de la red, antes que empiece a modelar el ruido. Estas mejoras, se traducen, en redes neuronales

con mejor capacidad de generalización. Finalmente, y no por ello menos importante, la prueba gamma, al poder examinar la naturaleza hipotética de la relación existente entre los datos de entrada y salida, podrá ser usado en lugar de los análisis de factibilidad comunes, optimizando el proceso de construcción de modelos no lineales. Debido a la gran diversidad de tipos de redes, es importante mencionar que en este trabajo se usaron redes neuronales de flujo hacia delante supervisadas.

Esta metodología será aplicada en la estimación de las propiedades petrofísicas-volumen de arcilla (V_{cl}), densidad (Rho o Rho_b), porosidad efectiva ($phie$) y velocidad de onda P (V_p) a escala sísmica, a partir de la integración de datos sísmicos y datos de pozo. Este se efectuará sobre la **línea sísmica X-2D-3C**, la cual cuenta con una extensión de 9 km con 300 fuentes y 903 receptores, y pertenece a un cubo sísmico que cubre un área de aproximadamente 29 km². La geometría de adquisición es ortogonal. La línea X-2D-3C cruza el pozo X-2, que tiene la siguiente información de registros de pozos: potencial espontáneo (SP), resistividad e inducción (ILD), registro sónico de la onda P. Algunos de estos registros son más recientes como el rayo Gamma (GR), registros sísmicos dipolares, Check shots y un perfil sísmico vertical. El pozo X-2 tiene una profundidad de 2886 m (9468 ft).

El yacimiento del campo está conformado por sedimentos clásticos Terciarios, cuya caracterización se complica al ser una secuencia de lutitas y arenas; se encuentra aproximadamente a una profundidad de 2124m (6968 ft). El campo X produce en una trampa combinada estratigráfica-estructural en areniscas del Oligoceno. La trampa consiste en un anticlinal con fallas asimétricas, ubicado sobre el flanco occidental de una estructura anticlinal importante. El anticlinal también se ve afectado por algunas fallas inversas y retrocesos. El espesor del intervalo del yacimiento oscila entre 5.5 y 13.5 m, mientras que las profundidades oscilan entre 2690 y 2811 metros.

El yacimiento se encuentra dentro de la formación Macizo, cuyas características petrofísicas de este son: porosidad entre el [20-25%], una saturación de agua, S_w , entre [40-50%], una saturación de petróleo entre [30-40%] y una saturación de gas entre [10-30%] y la permeabilidad aproximada es 40.5 cp.

Los primeros capítulos del presente trabajo se refieren a aspectos generales tales como; área de estudio, datos disponibles y descripción de las propiedades petrofísicas a predecir. Posteriormente, se describe de manera detallada como operan las redes neuronales de flujo hacia delante. Seguido de este capítulo, se describe de manera breve los atributos sísmicos y su clasificación. Finalmente, los últimos dos capítulos se refieren a la descripción de la metodología, su aplicación y los resultados obtenidos. Es importante mencionar, que, dentro de los resultados, se hace una comparativa de los resultados obtenidos con redes neuronales y el método convencional estadístico, regresión multivariable convolucional, que hacen posible constatar la superioridad del primer enfoque sobre los métodos convencionales

ABSTRACT

A main objective in the oil industry in Mexico and the world, is to carry out in a concise way the static characterization of the reservoirs. To carry out this task in the last three decades, seismic attributes have been widely used. Today, a dozen or more attributes are extracted from the seismic data for the determination of the petrophysical properties of the reservoirs. Likewise, a tendency to acquire more well logs is also established, which provides even more information to achieve said objective. However, as the amount of information available increases, the risk that the results obtained are not as satisfactory increases, because mathematical artifices that lack geological meaning are created in them. To avoid this, several statistical and artificial intelligence techniques have been used. Through the present work, it is intended to make known to those working in the industry, the potential present in the use of the statistical analysis tool, the Gamma test, in the construction of non-linear models, specifically using neural networks.

A very common approach used in the oil industry, for the static characterization of reservoirs, is the integration of well data and seismic data. These works are widely documented to mention some authors, see for example: Gastaldi et al. (1997), Russell et al. (1997) and Hampson et al. (2001). One of the approaches that has gained much popularity in the last decades to carry out this integration, is the use of neural networks, the reason for this is because they are universal approximators, these tools can approximate any continuous function with arbitrary precision. Despite being a computational tool that already has a path traveled in non-linear modeling, the design and training of a neural network is still considered more an art than a science, this is due to numerous well-known difficulties in neural networks. Among these problems stand out; the determination of the optimal configuration of the network (structure), the optimal distribution of the weights for a specific network, and of course if not the most transcendental of all of them, the guarantee of a good generalization capacity of the networks. In this work, a methodology is proposed, whose main element is the Gamma test. The Gamma test is a non-linear modeling tool that allows us to examine the hypothetical nature of the relationship between input and output data (seismic attributes-well logs) in a set of numerical data. The insertion of this element, in the process of integrating seismic data and well data through neural networks, will allow optimizing the design of neural networks, in the sense that it will allow us to determine which combination of input data (seismic attributes) is the better to model an output data set (petrophysical property), which is a problem that is shared among the different methods used for nonlinear modeling. Also, this statistical element, to be able to determine the noise, assuming that the variation of noise in each of the outputs is limited and is independent of the input values, will give us the possibility of establishing a criterion to stop the training of the network, before starting modeling the noise. These improvements are translated, in neural networks with better generalization capacity. Finally, and not for that less important, the gamma test, to being able to examine the hypothetical nature of the relationship between input and output data, can be used instead of common feasibility analyzes, optimizing the model building process non-

linear Due to the great diversity of types of networks, it is important to mention that supervised forward-flow neural networks were used in this work.

This methodology will be applied in the estimation of the petrophysical properties-clay volume (V_{cl}), density (Rho or Rho_b), effective porosity ($phie$) and P wave velocity (V_p) -a seismic scale, from the integration of seismic data and well data. This will be done on the X-2D-3C seismic line, which has an extension of 9 km with 300 sources and 903 receivers and belongs to a seismic cube that covers an area of approximately 29 km². The acquisition geometry is orthogonal. Line X-2D-3C crosses well X-2, which has the following well-logging information: spontaneous potential (SP), resistivity and induction (ILD), sonic log of the P wave. Some of these logs are more recent such as the Gamma ray (GR), dipolar sonic logs, Check shots and a vertical seismic profile. Well X-2 has a depth of 2886 m (9468 ft).

The field is composed of Tertiary clastic sediments, whose characterization is complicated by being a sequence of shales and sands; It is approximately at a depth of 2124m (6968 ft). The X field produces in a combined stratigraphic-structural trap in Oligocene sandstones. The trap consists of an anticline with asymmetrical faults, located on the western flank of an important anticline structure. The anticline is also affected by some inverse faults and setbacks. The thickness of the reservoir interval oscillates between 5.5 and 13.5 m, while the depths range between 2690 and 2811 meters.

The reservoir is within the Macizo formation, whose petrophysical characteristics are: porosity between [20-25%], water saturation, S_w , between [40-50%], oil saturation between [30-40%] and a gas saturation between [10-30%] and the approximate permeability is 40.5 cp.

The first chapters of the present work refer to general aspects such as; area of study, available data and description of the petrophysical properties to be predicted. Subsequently, it is described in detail how work forward neural networks. Following this chapter, the seismic attributes and their classification are briefly described. Finally, the last two chapters refer to the description of the methodology, its application and the results obtained. It is important to mention that, within the results, a comparison is made of the results obtained with neural networks and the conventional statistical method, multivariable convolutional regression, which make it possible to confirm the superiority of the first approach over conventional methods

ÍNDICE

RESUMEN	II
ABSTRACT	IV
LISTA DE FIGURAS	VIII
LISTA DE TABLAS	XI
INTRODUCCIÓN	1
CAPITULO I.....	3
GENERALIDADES	3
1.1 ANTECEDENTES.	3
1.2 ÁREA DE ESTUDIO.	5
1.2.1 Marco Geológico.	5
1.2.1.1 Estratigrafía y evolución de la cuenca.	5
1.2.3 Objetivos de la investigación.....	8
1.2.4 Datos disponibles.	8
CAPITULO II.....	10
PARÁMETROS PETROFÍSICOS	10
2.1 POROSIDAD.	10
2.1.1 Porosidad de las rocas clásticas.	12
2.2 ARCILLAS.....	13
2.3 VELOCIDAD DE PROPAGACIÓN DE LAS ONDAS ELÁSTICAS.....	16
2.3.1 Influencia de la litología (mineralogía) y la porosidad sobre la velocidad.	21
2.4 DENSIDAD.	23
CAPITULO III.....	28
REDES NEURONALES ARTIFICIALES	28
3.1 FUNDAMENTOS.	28
3.2 REDES NEURONALES.....	32
3.2.1 Definición.	32
3.2.2 Modelo artificial de una neurona.....	33
3.2.3 Tipos de funciones de activación.....	36
3.2.4 Diferentes representaciones de las redes neuronales	38
3.2.5 Clasificación de redes neuronales.	42
3.2.5.1 Tipos de red de acuerdo con su arquitectura.....	42
3.2.5.2 Tipos de red de acuerdo con el aprendizaje.....	49
3.2.6 Beneficios de las redes neuronales.	51
3.2.7 Algoritmos de aprendizaje.	53
3.2.7 Tareas de aprendizaje.	57
CAPITULO IV	61
REDES NEURONALES MULTICAPA.	61
4.1 GENERALIDADES.	61
4.2 ALGORITMO DE RETROPROPAGACIÓN.....	64
4.2.1 Función de activación.....	72

4.2.2 Rapidez de aprendizaje.....	74
4.2.3 Criterios para detener el proceso de aprendizaje de la red (entrenamiento).....	76
4.2.4 Heurísticas para mejorar el desempeño del algoritmo de retropropagación.....	77
4.2.4 Capacidad de generalización.....	81
CAPITULO V	85
ATRIBUTOS SÍSMICOS	85
5.1 PERSPECTIVA HISTÓRICA.....	85
5.2 DEFINICIÓN.....	89
5.3 CLASIFICACIÓN.....	89
5.3.1 Criterios para la selección de atributos.....	91
CAPITULO VI	93
METODOLOGÍA.....	93
6.1 PROCESADO DE LOS DATOS.....	94
6.2 ANÁLISIS ESTADÍSTICO DE LOS DATOS A TRAVÉS DE LA PRUEBA GAMMA.....	95
6.3 ESTRUCTURA DE LA RED.....	100
6.4 ENTRENAMIENTO Y VALIDACIÓN DE LA RED.....	101
6.5 OPTIMIZACIÓN DEL MODELO.....	102
6.6 COMPARACIÓN DE LA RED NEURONAL CON MODELO DE REGRESIÓN ESTADÍSTICA CONVENCIONAL; REGRESIÓN LINEAL MULTIVARIABLE CONVOLUCIONAL.....	102
CAPITULO VII	103
RESULTADOS.....	103
7.1 RESULTADOS DEL ANÁLISIS ESTADÍSTICO.....	106
7.2 REDES NEURONALES: DISEÑO, ENTRENAMIENTO Y RESULTADOS.....	111
CONCLUSIONES Y RECOMENDACIONES	128
RECOMENDACIONES.....	129
FUENTES DE CONSULTA	130
APÉNDICE A	137
TÉCNICAS DE OPTIMIZACIÓN SIN RESTRICCIÓN	137
APÉNDICE B	143
ATRIBUTOS SÍSMICOS	143
APÉNDICE C.....	145
REGRESION LINEAL MULTIVARIABLE CONVOLUCIONAL.....	145
APÉNDICE D	147
PRUEBA GAMMA	147

LISTA DE FIGURAS

<i>Figura 1. Representación del concepto de porosidad.</i>	11
<i>Figura 2. Tendencia del decremento de la porosidad relacionado a diferentes series de rocas.</i>	12
<i>Figura 3. Minerales arcillosos, así como algunas de sus propiedades físicas.</i>	14
<i>Figura 4. Tipo de distribución que presentan las arcillas, y su influencia sobre la porosidad y permeabilidad de las rocas.</i>	15
<i>Figura 5. Módulo compresional y de corte para los elementos principales de las rocas.</i>	17
<i>Figura 6. Rango de valores que presentan las rocas, en la velocidad de propagación de las ondas compresionales (valores altos), y en las ondas de cizalla (valores bajos).</i>	21
<i>Figura 7. Velocidades de ondas compresivas y de corte versus porosidad (fracción) para 75 muestras de areniscas a una presión de confinamiento de 40 MPa y una presión de poro de 1.0 MPa.</i>	22
<i>Figura 8. Densidad del gas metano a diferentes presiones y temperaturas.</i>	25
<i>Figura 9. Rango de valores promedio de densidad para rocas sedimentarias más comunes.</i>	26
<i>Figura 10. Relación entre la densidad volumétrica y la porosidad para areniscas .</i>	27
<i>Figura 11. Representación en diagrama de bloques del sistema nervioso.</i>	29
<i>Figura 12. Modelo biológico de una neurona.</i>	30
<i>Figura 13. Niveles de organización estructural en el cerebro.</i>	31
<i>Figura 14. Modelo no lineal de una neurona real y artificial</i>	34
<i>Figura 15. Transformación llevada a cabo al incluir el sesgo en el diseño de la red neuronal.</i>	35
<i>Figura 16. Modelo no lineal de una neurona.</i>	36
<i>Figura 17. Representación gráfica de algunos tipos de función de activación.</i>	38
<i>Figura 18. Esquema ilustrativo de las reglas básicas para la construcción de gráficos de flujo de señal.</i>	40
<i>Figura 19. Gráfico de señal de flujo del modelo de una neurona.</i>	41
<i>Figura 20. Gráfico arquitectónico de una neurona.</i>	41
<i>Figura 21. Clasificación de las redes neuronales.</i>	42
<i>Figura 22. Diagrama que muestra los diferentes tipos de redes neuronales que existe.</i>	43
<i>Figura 23. Red de flujo hacia delante con una sola capa de neuronas.</i>	44
<i>Figura 24. Ilustración del hiperplano (en este ejemplo, la línea recta) como límite de decisión para un problema de clasificación de patrones en dos diferentes clases.</i>	45
<i>Figura 25. Representación del perceptrón.</i>	45
<i>Figura 26. Red de flujo hacia delante completamente conectada con una capa oculta y una capa de salida.</i>	47
<i>Figura 27. Red recurrente sin loops de retroalimentación, ni con neuronas ocultas.</i>	48
<i>Figura 28. Red recurrente con neuronas ocultas.</i>	48
<i>Figura 29. Diagrama de bloque del aprendizaje supervisado.</i>	49
<i>Figura 30. Diagrama de bloque del aprendizaje reforzado.</i>	50
<i>Figura 31. Diagrama de bloque del aprendizaje no supervisado.</i>	51
<i>Figura 32. Ilustración del aprendizaje por corrección de error.</i>	55
<i>Figura 33. Diagrama de bloque de un sistema de identificación.</i>	59
<i>Figura 34. Diagrama de bloque de un sistema de modelado.</i>	60
<i>Figura 35. Representación arquitectónica de una red neuronal multicapa, con dos capas ocultas.</i>	63
<i>Figura 36. Ilustración de las direcciones de dos flujos de señales básicas en un perceptrón multicapa.</i>	64
<i>Figura 37. Gráfico de señal de flujo que realiza los detalles de la salida para la neurona j.</i>	67
<i>Figura 38. Gráfico de flujo de señal que resalta los del detalle de la neurona k conectada la neurona oculta j.</i>	70
<i>Figura 39. Gráfico de señal de flujo de una parte del sistema adjunto perteneciente a la retropropagación de la señal error.</i>	72
<i>Figura 40. Ejemplos de funciones antisimétricas y no simétricas</i>	79

Figura 41. Ilustración de las operaciones de normalizado: remoción de la media, decorrelación y eculización de la covarianza para un espacio de entrada 2D.	80
Figura 42. Capacidad de generalización de una red neuronal artificial	83
Figura 43. Línea de tiempo del desarrollo de los atributos sísmicos, y su relación con los avances claves en la tecnología de exploración sísmica.	85
Figura 44. Traza sísmica; composición.	88
Figura 45. Clasificación de los atributos sísmicos.	91
Figura 46. Diagrama de la metodología propuesta para la obtención de modelos no lineales usando redes neurales y la prueba Gamma.	94
Figura 47. Representación gráfica del estadístico Gamma, y su gradiente.	97
Figura 48. Ejemplo del empleo de la prueba Gamma para la selección de los datos de entrada para la construcción de un modelo no lineal.	98
Figura 49. Flujo de trabajo para llevar a cabo el análisis de datos con la prueba Gamma.	99
Figura 50. Gráfico cruzado de algunos de los atributos que aparecen en la Tabla 5.	104
Figura 51. Resultados preliminares de la prueba Gamma usando un número de vecinos igual a 10 para cada una de las propiedades petrofísicas.	107
Figura 52. Prueba de incremento del número de vecinos más cercanos para cada una de las redes usadas para la estimación de las propiedades de interés.	108
Figura 53. Histograma de los valores de Gamma para la búsqueda de los mejores atributos para cada uno de los parámetros petrofísicos.	109
Figura 54. Prueba M para los 4 modelos. La prueba M se usa para mostrar cómo varía la estadística Gamma a medida que se utilizan más datos para calcularla. Eventualmente, si se utilizan suficientes datos, la estadística Gamma debe sugerir una variación de ruido real en la salida para la cual se ha calculado. Esta prueba también nos puede decir cuántos datos es probable que se necesiten para obtener un modelo de una calidad determinada, en el sentido de predecir con un MSE alrededor del nivel de ruido. Se tomó la máscara que produce el MSE más bajo para cada uno de los parámetros petrofísicos para ejecutar la prueba M.	111
Figura 55. Gráfico donde se compara el valor objetivo y el valor estimado de V_{cl} por la red durante el entrenamiento, con su respectivo histograma de errores.	113
Figura 56. Gráfico donde se compara el valor objetivo y el valor estimado de ϕ_{ie} por la red durante el entrenamiento, con su respectivo histograma de errores.	114
Figura 57. Gráfico donde se compara el valor objetivo y el valor estimado de R_{ho} por la red durante el entrenamiento, con su respectivo histograma de errores.	115
Figura 58. Gráfico donde se compara el valor objetivo y el valor estimado de V_p por la red durante el entrenamiento, con su respectivo histograma de errores.	116
Figura 59. Secciones de V_{cl} que abarcan la ventana de tiempo 1126-1863 ms.	
Figura 60. Gráficos cruzados de R_{hob} vs V_{cl} y ϕ_{ie} vs V_{cl}	120
Figura 61. Evaluación petrofísica correspondiente al área que cubre cada una de las secciones de propiedades obtenidas (1129 a 1862 ms), 1560-2855 m.	121
Figura 62. Secciones de las propiedades petrofísicas R_{ho} , ϕ_{ie} y V_p , obtenidas con las redes neuronales con mejor desempeño	122
Figura 63. Sobreposición del valor de registro de cada una de las propiedades petrofísicas, sobre paleta de colores que representa el valor de salida de su respectiva red, durante el entrenamiento.	123
Figura 64. Secciones de R_{hob} , ϕ_{ie} y V_p , respectivamente, donde se sobreponen los datos sísmicos.	124
Figura 65. A) Histograma donde se compara la salida de la red para V_p (color rojo), y los valores del registro V_p (color azul). B) Histograma de errores, de la red para predecir V_p , durante el segundo entrenamiento.	125
Figura 66. Secciones de R_{hob} y V_{cl} , respectivamente, obtenidas por el método de regresión lineal convolucional, con un número de operadores igual a 7. ás, no se conservan los echados de los estratos.	126
Figura 67. Trayectoria del método de descenso más pronunciado en un espacio 2D para dos diferentes valores de η .	129

<i>Figura 68. Ejemplo de combinación lineal para el caso de 3 atributos en el mismo tiempo</i>	<u>145</u>
<i>Figura 69. Ejemplo de regresión lineal convolucional, donde se emplea un operador convolucional de 5 puntos para relacionar los atributos sísmicos al pozo objetivo.</i>	<u>146</u>

LISTA DE TABLAS

<i>Tabla 1. Rango de valores que presentan los fluidos de los yacimientos, en su módulo de compresibilidad, k, y su densidad</i>	18
<i>Tabla 2. Densidad, módulos elásticos y velocidad de onda P de los minerales más comunes que conforman a las rocas.</i>	19
<i>Tabla 3. Velocidad de onda compresional para algunos fluidos.</i>	20
<i>Tabla 4. Conjunto de atributos sísmicos considerados para construir los modelos.</i>	103
<i>Tabla 5. Estadísticos descriptivos de los atributos a la traza más cercana al pozo, los cuales fueron los usados durante el entrenamiento de las redes, seguido de los estadísticos de la totalidad de los atributos en el resto de las trazas.</i>	105
<i>Tabla 6. Número de vecinos óptimo para cada uno de los parámetros petrofísicos.</i>	108
<i>Tabla 7. Mejor combinación de atributos para la predicción de V_{cl}. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario</i>	110
<i>Tabla 8. Mejor combinación de atributos para la predicción de porosidad efectiva. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario.</i>	110
<i>Tabla 9. Mejor combinación de atributos para la predicción de V_p. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario.</i>	110
<i>Tabla 10. Mejor combinación de atributos para la predicción de R_{hob}. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario.</i>	110
<i>Tabla 11. Arquitectura, algoritmo de entrenamiento y MSE alcanzado para cada una de las redes, según la propiedad petrofísica a estimar. La arquitectura corresponde al número de neuronas en cada capa oculta.</i>	112

INTRODUCCIÓN

Al menos dos ingredientes esenciales se requieren para el desarrollo de “**Tecnología**”: El **concepto** y la **implementación de ésta**. Primero, se debe contar con un concepto, una forma innovadora de observar un tema en específico; es decir, un punto de vista que vislumbre una solución jamás antes planteada a algún problema. Esto puede ser una simple idea, o bien algo más específico, tal como la formulación de una ecuación o de un modelo matemático. Y lo segundo, se refiere a contar con la capacidad tecnológica, económica y demás, para poder llevar eso intangible al mundo físico, implicando su aplicación a la solución de problemas reales. Es así precisamente, como surge lo que es la **Inteligencia Artificial**, que ha asumido en los últimos años el nuevo significado del concepto de “**Tecnologías Avanzadas**”, que dan solución a una infinidad de problemas, que van desde problemas superfluos, hasta aquellos problemas complejos, como la construcción de modelos no lineales entre dos variables o más, tal como es el caso de la predicción de propiedades petrofísicas a partir de atributos sísmicos.

La Inteligencia Artificial se divide generalmente en dos categorías básicas (**Mohaghegh et al., 1995**); **sistemas basados en reglas** (expertos), los cuales en la actualidad su capacidad de solución está siendo sobrepasada, y los **sistemas adaptativos**. En este trabajo se concentrará en el uso de redes neuronales para dar solución a uno de los problemas que comúnmente se presentan en la industria petrolera, dado el mayor grado de anisotropía en las formaciones; la **caracterización estática de los yacimientos**. Las redes neuronales, constituyen un enfoque computacional el cual está inspirado en sistemas biológicos, específicamente en el cerebro humano. Las redes neuronales, son sistemas análogos, adaptativos y distributivos capaces de resolver una gran cantidad de problemas, y que han sido empleados ya en numerosas disciplinas con éxito, en gran medida por su capacidad de encontrar relaciones no lineales (**Ali J.K, 1994**), así como de poseer la habilidad de generalizar, y ser robustas al ruido. No obstante, se debe aceptar el hecho que las redes neuronales no pueden generalizar perfectamente todos los casos de prueba que se le presenten, y que son susceptibles a errores, este hecho abre las puertas de oportunidad de mejora en su funcionamiento. En este trabajo se pretende dar a conocer las mejoras que se pueden obtener en la construcción de modelos no lineales usando redes neuronales en la predicción de propiedades petrofísicas, mediante el empleo de una herramienta estadística no lineal, llamada **Prueba Gamma**. Estas mejoras, son esencialmente efectuadas sobre la capacidad de generalización de las redes, cuyo factor limitante principal de ésta, en la mayoría de los casos, es el ruido presente en los datos, o una cantidad insuficiente de éstos para poder establecer un modelo aceptable. Los aspectos en los cuales la prueba Gamma ayudará en la construcción de los modelos son; en la selección de los datos de entrada para la construcción de los modelos, y a evitar el sobre entrenamiento de las redes, al permitir ésta tener conocimiento del ruido presente en los datos.

Históricamente, la construcción de **modelos no lineales** a partir de datos muestreados ha sido en gran medida un proceso subjetivo, esto en parte se debe a la enorme diversidad de posibles técnicas para el modelado y la dificultad de evaluar la calidad de los datos (**Iturrarán, 2011**). Son pocos los trabajos llevados a cabo para la estimación de la precisión de ajuste para modelos no lineales sin conocer el modelo (la función subyacente entre los datos), tal como se hace en la regresión lineal, donde la precisión de ajuste se hace, sabiendo de antemano que la función representa una línea recta. Sin embargo, existe una prueba llamada **Prueba Gamma**, desconocida por la mayoría y por ende poco usada en la industria petrolera, que permite llevar a cabo la precisión de ajuste del modelo no lineal construido, entre otras cosas que se mencionarán más adelante. La Prueba Gamma es una herramienta de modelado y análisis no lineal que nos permite examinar la naturaleza de una relación hipotética existente entre las entradas/salidas en un conjunto de datos numéricos. Ésta se basa en el cálculo de la **estadística Gamma**, Γ , que mide el ruido en un conjunto de datos relativo a todas las funciones suaves, que pueda existir (**Haythorn, 2011**). Así, si la estadística Gamma es pequeña, existirá una fuerte relación predictiva entre las variables de entrada-salida. Al contrario, si la estadística Gamma presenta un valor alto, no existirá una relación predictiva entre las variables, es decir, las entradas son irrelevantes para la determinación de la salida (**Parra et al., 2015**). Entonces, con base a lo anterior la prueba Gamma constituye una regresión no lineal, ya que proporciona una estimación de la precisión de ajuste contra la clase de modelos no lineales.

Como se sabe la integración de datos de registro de pozo con datos sísmicos, en la predicción de propiedades petrofísicas, está bien documentado en la literatura **Gastaldi et al. (1997)**, **Russell et al. (1997)**, **Hampson et al. (2001)**. El objetivo de este trabajo es dar fe de la eficacia de la prueba Gamma, de mejorar la capacidad de generalización de las redes neuronales, para ello, se propone una metodología, cuyo elemento principal es la prueba Gamma. Dicha metodología, será empleada en la caracterización estática de un yacimiento, cuya heterogeneidad hace que la eficacia de métodos tradicionales sea pobre. Para ello, se utilizarán datos sísmicos como datos de entrada, específicamente atributos sísmicos, tomando la ventaja la no linealidad de éstos, así como también, la reducción de la dimensionalidad de los datos al emplearlos, y datos de registro de pozo, como datos de salida.

CAPITULO I

GENERALIDADES

1.1 Antecedentes.

La historia de las redes neuronales ha progresado a través de innovaciones conceptuales y desarrollos tecnológicos. Estos avances, sin embargo, parecen haber ocurrido de manera intermitente, más que en una evolución constante.

Algunos de los primeros trabajos en el campo de las redes neuronales ocurrieron a fines del siglo **XIX** y principios del **XX**. Estos consistieron principalmente en trabajos interdisciplinarios en física, psicología y neurofisiología llevados a cabo por científicos como Hermann Von Helmholtz, Ernst Mach e Iván Pavlov. Estos trabajos iniciales enfatizaron las teorías generales de aprendizaje, visión, acondicionamiento, etc., y no incluyeron modelos matemáticos específicos de operación de neuronas.

La visión moderna de las redes neuronales comenzó en la década de 1940 con el trabajo de **Warren McCulloch y Walter Pitts (1943)**, quienes demostraron que las redes de neuronas artificiales podían, en principio, calcular cualquier función aritmética o lógica. Su trabajo a menudo se reconoce como el origen del campo de la red neuronal.

McCulloch y Pitts fueron seguidos por **Donald Hebb (1949)**, quien propuso que el condicionamiento clásico (tal como lo descubrió Pavlov) está presente debido a las propiedades de las neuronas individuales. Él propuso un mecanismo de aprendizaje para las neuronas biológicas.

La primera aplicación práctica de redes neuronales artificiales se produjo a fines de la década de 1950, con la invención de la red de perceptrones y la regla de aprendizaje asociada por **Frank Rosenblatt (1958)**. Rosenblatt y sus colegas construyeron una red de perceptrón y demostraron su capacidad para realizar reconocimiento de patrones. Este éxito inicial generó un gran interés en la investigación de redes neuronales. Desafortunadamente, más tarde se demostró que la red básica del perceptrón podía resolver solo una clase limitada de problemas, como se verá más adelante en este trabajo.

Aproximadamente al mismo tiempo, **Bernard Widrow y Ted Hoff (1960)** introdujeron un nuevo algoritmo de aprendizaje y lo utilizaron para entrenar redes neuronales lineales adaptativas, que eran similares en estructura y capacidad al perceptrón de Rosenblatt. La regla de aprendizaje de Widrow-Hoff todavía está en uso hoy en día.

Desafortunadamente, las redes de Rosenblatt y Widrow sufrieron las mismas limitaciones inherentes, que fueron ampliamente publicitadas en un libro de **Marvin Minsky y Seymour Papert (1969)**. Rosenblatt y Widrow eran conscientes de estas limitaciones y propusieron

nuevas redes que los superarían. Sin embargo, no pudieron modificar con éxito sus algoritmos de aprendizaje para capacitar a las redes más complejas.

Muchas personas, influenciadas por Minsky y Papert, creían que la investigación adicional sobre redes neuronales era un callejón sin salida. Esto, combinado con el hecho de que no había computadoras digitales potentes para experimentar, causó que muchos investigadores abandonaran el campo. Durante una década, la investigación de redes neuronales se suspendió en gran medida.

Sin embargo, un trabajo importante continuó durante la década de 1970. En 1972 **Teuvo Kohonen y James Anderson (1972)** de forma independiente y por separado desarrollaron nuevas redes neuronales que podrían actuar como recuerdos. **Stephen Grossberg (1976)** también estuvo muy activo durante este período en la investigación de redes autoorganizadas.

El interés en las redes neuronales había fallado a fines de la década de 1960 debido a la falta de nuevas ideas y poderosas computadoras con las que experimentar. Durante la década de 1980, ambos impedimentos se superaron y la investigación en redes neuronales aumentó drásticamente. Nuevas computadoras personales y estaciones de trabajo, que crecieron rápidamente en capacidad, se volvieron ampliamente disponibles. Además, se introdujeron nuevos conceptos importantes.

Dos nuevos conceptos fueron los principales responsables del renacimiento de las redes neuronales. El primero fue el uso de mecanismos estadísticos para explicar el funcionamiento de una cierta clase de red recurrente, que podría usarse como memoria asociativa. Esto fue descrito en un artículo seminal por el físico **John Hopfield (1982)**.

El segundo desarrollo clave de la década de 1980 fue el algoritmo de retropropagación para el entrenamiento de redes de perceptrón multicapa, que fue descubierto de forma independiente por varios investigadores diferentes. La publicación más influyente del algoritmo de retropropagación fue por **David Rumelhart y James McClelland (1986)**. Este algoritmo fue la respuesta a las críticas que Minsky y Papert habían hecho en la década de 1960. Estos nuevos desarrollos revitalizaron el campo de las redes neuronales. Desde la década de 1980, se han escrito miles de artículos, las redes neuronales han encontrado innumerables aplicaciones, y el campo ha estado zumbando con nuevos trabajos teóricos y prácticos.

En un inicio, las principales aplicaciones de las redes neuronales estaban en el campo financiero, sin embargo, en la actualidad están presente en numerosas áreas, desde la apantallante industria espacial, hasta en la gran industria del entretenimiento, solo por mencionar un gran ejemplo, Spotify.

El uso de las redes neuronales en la industria petrolera, que es donde estamos interesados, es sin embargo, relativamente reciente. Principalmente se han usado en la identificación de facies, algunos trabajos de ello son; **Rogers et al. (1992)** aplicaron una red neuronal de flujo hacia delante de retropropagación para clasificar la litología de los registros de pozos. En un estudio comparativo de **Wong et al. (1995)** se demostró que un enfoque de red neuronal

similar es mejor que el análisis discriminante en la clasificación de registros de facies. **Saggaf y Nebrija (2000)** aplicaron con éxito redes neuronales autoorganizadas (**Kohonen, 1989**) para la predicción de registros de facies.

Así mismo, otras aplicaciones importantes que se han llevado a cabo son: En el reconocimiento de formas de onda y determinación de los primeros arribos (**Murat y Rudman, 1992; McCormack et al., 1993**); para propósitos de inversión sísmica (**Roth y Tarantola, 1994; Langer et al., 1996; Calderón–Macías et al., 1998**); para análisis de registro de pozo (**Huang et al., 1996**), etc.

1.2 Área de estudio.

El campo X fue descubierto en 1971 después de los resultados positivos de la perforación del pozo X-1. La estrategia de evaluación fue seguida por la perforación de los pozos X-2 y X-3. El desarrollo del campo se detuvo más tarde debido a resultados fallidos en X-3. Los pozos X-1 y X-2 tenían una producción de alrededor de 100 STBO / día de 22.8 API de petróleo crudo; Sin embargo, están casi totalmente agotados actualmente. El campo X produce en una trampa combinada estratigráfica-estructural en areniscas del Oligoceno. La trampa consiste en un anticlinal con fallas asimétricas, ubicado sobre el flanco occidental de una estructura anticlinal importante. El anticlinal también se ve afectado por algunas fallas inversas y retrocesos. El espesor del intervalo del yacimiento oscila entre 5.5 y 13.5 m, mientras que las profundidades oscilan entre 2795 y 3040 metros

1.2.1 Marco Geológico.

La cuenca bajo estudio es una cuenca Inter montaña con un echado hacia el este con una historia geológica compleja. Una serie de eventos tectónicos, hacen a ésta, una cuenca poli-histórica, usando el término de la definición establecida por **Kingston et al. (1983)**. Esta cuenca se superpone al basamento continental compuesto de rocas metamórficas de granulita, sobre las cuales yacen rocas meta-sedimentarias de baja presión, como lo establece **Restrepo (1995)**, que también constituye el basamento cristalino de la Cordillera Central (CC) y la Cordillera Oriental (CO).

1.2.1.1 Estratigrafía y evolución de la cuenca.

Antes de dar inicio a esta sección, es importante decir que los nombres de las formaciones reales fueron modificados, por lo que pueden o no existir, esto para salvar la confidencialidad de los datos.

Durante el Mesozoico temprano, el área fue controlada por un entorno extensional evidenciado por un cambio intracontinental relacionado con la ruptura de Pangea (**Pindell y Dewey, 1982; Cediel et al., 2003**), dicho periodo extensional produce rellenos Syn-rift representado por secuencias fluviales-aluviales y rocas sedimentarias lacustres durante las

Formaciones Triásico-Jurásico (formaciones Jordan, Girón y Santos). Estos depósitos continentales consisten principalmente en lechos rojos y depósitos volcánicos y piroclásticos.

La sedimentación en la cuenca continuó en el Cretácico en un entorno de arco inverso al este de la zona de subducción andina, dominado por la sedimentación marina poco profunda (**Cooper et al., 1995**). Para el periodo Aptiense (125-113 Ma), la extensión pura finalizó y una fase de sedimentación posterior a la extensión (post-rift) comenzó a controlarse por subsidencia térmica (**Villamil, 1993; Etayo-Serna, 1994; Sarmiento-Rojas, 2001**). Por lo tanto, los depósitos Syn-rift están cubiertos por calizas marinas y margas depositadas durante el Cretácico Inferior (Formaciones Rojo-Negro y Leña, respectivamente). Durante esta fase también se depositaron las siguientes Formaciones: **Trampolín**, que consiste en calizas interconectadas con arcillas lutitas negras; **Sumatra**, principalmente constituida de arcillas depositadas en una plataforma interna a media; **Libertad**, que consiste en la intercalación de calizas y lutitas negras ricas en materia orgánica; y **Unión**, que marca una transición a un ambiente más clástico y está compuesto de esquistos grises y areniscas con intervalos de carbón. A gran escala, la sedimentación marina de las rocas del Cretácico representa un gran ciclo transgresivo (regresivo con una superficie de inundación máxima cercana al límite Cenomaniano-Turaniano, que corresponde al máximo nivel eustático mesozoico (**Villamil, 1993**), produciendo una excelente roca generadora regional (**Formación Libertad**).

La evolución compleja durante el Cenozoico se ha atribuido a las condiciones más tempranas de cuenca antepaís relacionadas con el levantamiento de la Cordillera Central (CC) durante el Cretácico Tardío-Cenozoico Temprano (**Dengo y Covey, 1993; Cooper et al., 1995; Gómez et al., 2005**), y a la transpresión causada por la acreción oblicua de la Cordillera Occidental (CO) hasta el Eoceno Temprano (**McCourt et al., 1984; Cooper et al., 1995**). Por lo tanto, la sedimentación marina se desplazó lentamente hacia un ambiente continental. La Formación Lizama, sobrepone a la Formación Unión marina poco profunda del Maastrichtiense con un contacto transicional y registra sedimentación regresiva en llanuras deltaicas y aluviales (**Gómez, 2001**), y, por lo tanto, la primera unidad con sedimentos continentales después de los depósitos marinos predominantes del Cretácico. La fase de deformación principal del CC que afecta a la cuenca marca una discordancia regional llamada Discordancia del Eoceno Medio (**DEM**). Es el límite más importante dentro del registro Cenozoico y se puede observar en toda la cuenca.

El subsiguiente establecimiento de un sistema fluvial meándrico se registra en los estratos de la **Formación California** durante el Eoceno Medio. De acuerdo con **Dengo y Covey (1993)** y **Cooper et al. (1995)** la sedimentación en la cuenca y CO del Eoceno al Mioceno ocurrió en una cuenca clásica de antepaís. Por lo tanto, los depósitos fluviales de las Formaciones California y Rubí registran el levantamiento inicial de la CO. Ambas unidades muestran un adelgazamiento sistemático hacia el oeste que representa un solapamiento progresivo hacia la antigua CC. Los cambios en los movimientos tectónicos de las placas documentados en el Mioceno superior al Oligoceno tardío produjeron la reactivación de las estructuras del Eoceno medio y crearon una discordancia en el Oligoceno superior (**Schamel, 1991**). De acuerdo con esta configuración tectónica, se desarrollaron sistemas fluviales meándricos y se depositaron las Formaciones Macizo y Dorado.

Además, una colisión del arco de isla Panamá-Chocó con el noroeste del margen de América del Sur también se produjo durante el Mioceno medio. Los sedimentos fluviales-aluviales de la Formación denominado **grupo La auténtica**, representa el último registro sedimentario asociado con la inversión de la CO (Velásquez, 2012).

Dado que este trabajo se centra en la caracterización geofísica de los yacimientos del Cenozoico, las Formaciones pertenecientes del Cenozoico se describirán con cierto detalle (Cambios en el registro estratigráfico del Cenozoico dentro de la cuenca). Varios autores, trabajando con diferentes grupos fósiles, llegaron a interpretaciones contradictorias de las eras del Eoceno Medio al Mioceno medio: **California, Rubí, Macizo y Formación el Dorado**.

Gómez et al., 2005 señalaron que la mayoría de las areniscas cenozoicas de esta cuenca son litoarenitas y sublitoarenitas, con la excepción de Macizo y Dorado, que tienen predominantemente areniscas feldespáticas. Los componentes de las areniscas cenozoicas son granos de cuarzo policristalino y mono cristalino, fragmentos metamórficos (esquisto, gneis micáceo y filita), clastos sedimentarios, granos de feldespato y escasos fragmentos ígneos (Mesa, 1995).

A continuación, se describirán de manera más detallada cada una de las Formaciones pertenecientes al Cenozoico:

- ✓ **Formación California (Eoceno medio):** Esta unidad se superpone a la Disc. del Eoceno Medio y consiste en intercalaciones de areniscas de grano grueso con conglomerados y arcillas de color gris. Esta unidad fue depositada en abanicos aluviales en la parte occidental de la cuenca (Medina et al., 1992 en Suarez, 1997). El mayor desarrollo de areniscas ocurrió en el flanco oriental del Sinclinal llamado Nuevo Mundo donde las litoarenitas amalgamadas pertenecientes a la Formación California marca la posición de los canales fluviales principales (Gómez et al., 2005). El cuarzo policristalino y los granos sedimentarios por lo general disminuyen hacia arriba en la parte superior de ésta, y continúan disminuyendo hacia arriba en las areniscas de la Formación Rubí.
- ✓ **Formación Rubí (Eoceno tardío):** Se compone principalmente de capas gruesas de mudstones y siltstone, depositadas en planicies de un sistema meándrico (Rubiano, 1995 en Suárez, 1997). Eventualmente, esta unidad también incluye capas de areniscas de estratificación cruzada. La mayoría de las areniscas son litoarenitas y sublitoarenitas. (Mesa, 1995).
- ✓ **Formación Macizo (Oligoceno-Mioceno inferior):** Se superpone discordantemente a la Formación Rubí. Formalmente definido como el intervalo de rocas entre los horizontes fósiles \ Los Corros \ "y \ Mugrosa \", y consiste en capas intercaladas de areniscas y mudstone depositadas en un sistema fluvial meándrico (Rubiano, 1995 en Suarez, 1997). Mesa (1995) revela una composición homogénea de las areniscas feldespáticas pertenecientes a las Formaciones Macizo y Dorado. Sin embargo, Gómez et al. (2005) describieron una asociación de facies distintiva de areniscas. Con base en la relación de espesor neto-bruto, los geólogos de yacimientos han

dividido informalmente la Formación macizo en dos grandes bloques de arenisca: la Zona D en la parte superior y la Zona E en la base.

- ✓ **Formación el Dorado (Mioceno Inferior a Mioceno Medio Inferior):** Se superpone discordantemente a la Formación Macizo. Se compone predominantemente de lutitas masivas, de color gris claro a rojo púrpura, intercaladas con areniscas fluviales de grano fino a grueso y bien clasificadas (**Morales, 1958 en Suarez, 1997**), que son difíciles de diferenciar de las areniscas de la Formación Macizo (**Mesa, 1995**). Un aumento en los granos sedimentarios y una disminución en el porcentaje de feldespato permite diferenciar el Grupo La auténtica de las areniscas pertenecientes a la Formación el Dorado (**Gómez et al., 2005**). **Gómez et al. (2005)** también señalaron que la asociación de facies de capas ascendentes de areniscas con estratificación cruzada y mudstones variegadas indica una disminución de la sinuosidad de los canales y un aumento de la energía baja en comparación con la Formación Macizo.
- ✓ **Grupo La auténtica (Mioceno superior):** Se superpone discordantemente a la Formación el Dorado. El grupo La auténtica es una molasa compuesta principalmente de conglomerados, areniscas conglomeráticas y arcillas grises, depositadas en ambientes fluviales-aluviales (**Rubiano, 1995 en Suárez, 1997**). Las capas de depósitos volcánicos retrabajados aparecen en la mitad superior este grupo. **Gómez et al. (2005)** interpretaron estas facies como un sistema fluvial trenzado con avulsiones de canales y anexión a la llanura de inundación.

1.2.3 Objetivos de la investigación.

El objetivo principal de este trabajo es la predicción de las propiedades petrofísicas: volumen de arcilla (Vcl), porosidad efectiva (ϕ_{ie}), velocidad de onda P (V_p) y densidad de la roca (Rho o Rho_b) a escala sísmica. Para lograr este objetivo, se combinó la información de los registros de pozos del pozo X-2 y la línea sísmica X-2D-3C, mediante el empleo de redes neuronales de flujo hacia delante, optimizando su capacidad de generalización mediante el uso de la herramienta de análisis estadístico, Prueba Gamma.

1.2.4 Datos disponibles.

Se cuenta con una línea sísmica 2D de 3 componentes, línea X-2D-3C, de onda P reflejada apilada, la cual tiene una extensión de 9 km.

La línea X-2D-3C cruza el pozo X-2, que tiene la siguiente información de registros de pozos: potencial espontáneo (SP), resistividad e inducción (ILD), registro sónico de la onda P. Algunos de estos registros son más recientes como el rayo Gamma (GR), registros sísmicos dipolares, Check shots y un perfil sísmico vertical. El pozo X-2 tiene una profundidad de 2886 m (9468 ft), la información disponible a partir del análisis básico muestra que X-2 cruza algunas unidades de Formación como el Dorado (profundidad 1350 m (4429 ft)), la parte superior de la zona de Macizo (profundidad 1957m (6421 ft)), la unidad Macizo D (profundidad 2125 m (6972 ft)) y finalmente la Formación Rubí (zona F) (profundidad 2886

m (9468 ft)). La unidad de la Formación Macizo E se divide en: las arenas Macizo E y las lutitas Macizo E. La primera unidad tiene más arenas petrolíferas, este hecho define a las arenas Macizo E como el principal yacimiento de la cuenca. Es una arenisca llena de petróleo, tiene una porosidad entre el [20-25%] y una saturación de agua, S_w , entre [40-50%], una saturación de petróleo entre [30-40%] y saturación de gas entre [10-30%] y la permeabilidad aproximada es 40.5 cp. Con respecto al pozo X-2 la zona de arenas Macizo E se encuentra localizado entre las profundidades (2690-2811 m (8826-9222 ft)), en la parte superior de esta Formación hay capas delgadas de arcilla con profundidades no mayores a 10 pies, con contenido mayor a 80%, 10% de arcilla arenosa, pirita y otros minerales. Desde esta parte superior a la inferior tenemos grandes cuerpos de arena de más de 100 pies de profundidad con altas porosidades, mayores al 15% y pequeñas intercalaciones de arcilla. Por otro lado, las lutitas Macizo E definen el fondo de las arenas Macizo, tiene una profundidad de 74 m (244 ft), está constituido de un 80% de arcilla tipo pizarra, con una porosidad menor al 15%; saturación de hidrocarburos baja o nula y muy baja permeabilidad. La ventana de tiempo de las secciones sísmicas corresponde a [1000-1864 ms], de las cuales se obtendrán las propiedades petrofísicas. Estas secciones consideran las unidades de Formación desde la parte superior de la Formación Macizo D hasta la base de las lutitas de la Formación Macizo E.

CAPITULO II

PARÁMETROS PETROFÍSICOS

Antes de abordar la construcción de secciones de propiedades petrofísicas a través de la construcción de modelos no lineales entre datos sísmicos y registros de pozo mediante el empleo de redes neuronales, es importante tener bien en claro la definición de los parámetros petrofísicos que se van a predecir, esto, con la finalidad de poder evaluar de mejor forma los resultados obtenidos. A continuación, se describen éstos.

2.1 Porosidad.

Existen diversas definiciones para el concepto de porosidad, una de ellas es la dada por **Jorden & Campbell (1984)** que dice; “La porosidad es la fracción del volumen de la roca ocupada por el poro”, es decir;

$$\frac{\text{Volumen de Poro}}{\text{volumen volumétrico}} = \frac{\text{volumen volumétrico} - \text{volumen de sólidos minerales}}{\text{volumen volumétrico}} \dots \text{Ec. 2. 1}$$

La porosidad es una propiedad fundamental volumétrica de la roca; describe el volumen potencial de almacenamiento de fluidos (gas, agua, aceite) e influye en la mayoría de las propiedades de las rocas (velocidad de la onda elástica, resistividad eléctrica, y densidad).

Entonces, la porosidad puede ser definida como la sumatoria del volumen de todos los poros, fracturas, fisuras, etc., o generalizando todos los fluidos (gas, agua, hidrocarburos) o de aquella parte no sólida de una muestra en relación al volumen total de la muestra (**Fig. 1**)

La porosidad es dada como una fracción del volumen (sin dimensión) o como un porcentaje. La definición de arriba describe a la “**porosidad total**”. Si la roca contiene una parte no conectada o poros separados (vúgulos, poros móldicos, etc.), entonces esta parte no contribuye a ningún fluido transportado dentro de la roca y es “**no efectivo**”. Así, la porosidad efectiva o interconectada es la razón entre el volumen de poros conectados y el volumen total de la roca.

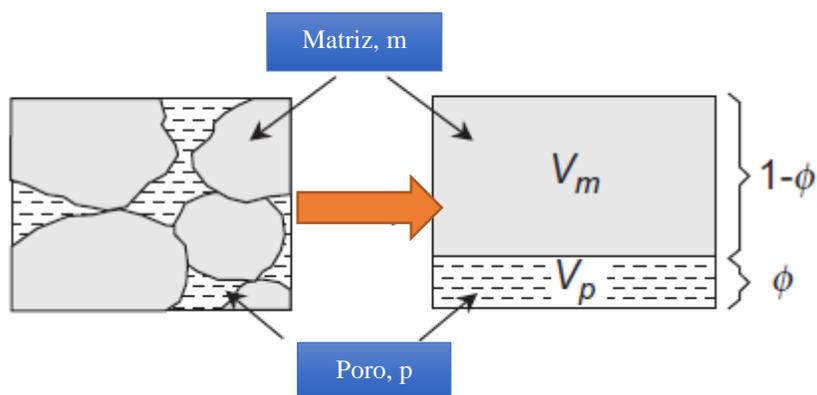
Para la descripción de reservas, es importante distinguir entre:

***Porosidad total** (ϕ_t), la fracción de volumen volumétrico ocupado por el espacio total del poro.

***Porosidad efectiva** (ϕ_e), la fracción de volumen volumétrico ocupado por espacio de poros interconectados.

Amyx et al. (1960) define porosidad total como “... la razón entre el espacio vacío total en la roca y el volumen total de ésta; la porosidad efectiva es la razón entre el espacio vacío interconectado en la roca y el volumen total de la ésta...”

Figura 1. Definición de porosidad.



Adaptado de J. H Schön. Physical Properties of Rocks. 2012 Elsevier B.V. p. 18

El tamaño de poro puede ser clasificado como se muestra en la **Tabla 1** después de una compilación de **Mann et al. (1977)** (citado en Schön, 2012, p.19)

La porosidad presente en las rocas clásticas no es de la misma naturaleza que la presentada en las rocas carbonatadas, lo que influirá que los mecanismos a través de los cuales los hidrocarburos se desplacen sean diferentes, y por ello generalmente se describe de manera detallada y por separado el tipo de porosidad presente en los dos principales tipos de roca de interés en la industria petrolera. No obstante, el área de estudio de este trabajo corresponde a una zona de rocas clásticas, es por ello, que solo se abordará lo correspondiente a este tipo de rocas.

<i>Tabla 1 Clasificación del poro en función de su tamaño.</i>	
Tipo de Poro	Tamaño (Diámetro)d
Poro grande	$d > 2\text{mm}$
Macro capilar	$2\text{mm} > d > 50\mu\text{m}$
Capilar	$50\mu\text{m} > d > 2\mu\text{m}$
Micro capilar	$2\mu\text{m} > d > 50\text{nm}$
Microporo	$50\text{nm} > d > 2\text{nm}$
Mesoporo	$2\text{nm} > d > 0.8\text{nm}$
Microporo	$0.8\text{nm} > d$

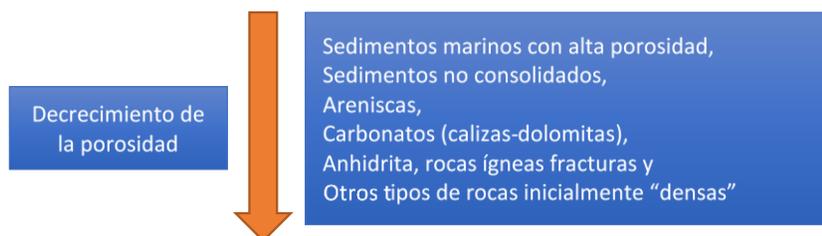
Adaptado Mann et al. (1977)

2.1.1 Porosidad de las rocas clásticas.

En los sedimentos clásticos, se presenta un tipo de porosidad controlado por los factores pre-diagenéticos, estos factores, así mismo controlan entre otros aspectos el tamaño y distribución de los tamaños de grano, el empacamiento de grano y la geometría de la partícula.

Para que se tenga idea del grado de porosidad que generalmente se presenta los diferentes tipos de roca, la **Fig. 2** muestra la tendencia de decremento de la porosidad relacionada a series típicas de rocas.

Figura 2. Tendencia del decremento de la porosidad relacionado a diferentes series de rocas.



Adaptado de J. H Schön. óp. cit., p. 20.

En áreas sedimentarias, la porosidad decrece de manera no lineal conforme a la profundidad como resultado de la compactación. La propiedad que controla la compactación en este proceso es el esfuerzo efectivo principal. La primera formulación empírica es probablemente la formulada por **Athy (1930)**:

$$\phi(z) = \phi_0 * e^{-b*z} \dots Ec. 2. 2$$

donde:

ϕ_0 ; es la porosidad a la profundidad de referencia ($z=0$).

z ; es la profundidad.

b ; es un parámetro de caracterización de compresibilidad del sedimento.

Así mismo, dado que la porosidad está relacionada directamente con otras propiedades de las rocas, ésta muchas veces es obtenida a partir de mediciones de registros geofísicos, específicamente el registro de densidad, el registro neutrón, y el registro sísmico. Es importante señalar, que los modelos empleados para establecer la relación entre las mediciones dadas por los registros y la obtención de ésta, y otras propiedades petrofísicas, es muy ideal, ya que se establecen relaciones lineales entre sus elementos, es por ello por lo que las estimaciones realizadas por los inversionistas sean mejores, al basarse éstos en modelos más robustos y realistas. Aunando a lo anterior, también el costo de la toma de registro es muy elevado, lo que, en la mayoría de las veces, no se encuentran éstos disponibles en gran cantidad, requiriendo el uso de técnicas matemáticas - estadísticas para caracterizar a los yacimientos.

2.2 Arcillas.

Si bien las arcillas no son propiamente un parámetro petrofísico como tal, la presencia de este tipo de litología en las formaciones influye importante en las propiedades físicas de las rocas, conllevando a complicaciones no sólo en la caracterización estática, sino también en la explotación de los yacimientos, de ahí la importancia de hablar sobre ellas.

En primer lugar, se debe mencionar que los términos lutita (shale) y arcilla (clay), son usados de manera indiferente en la práctica de toma de registro geofísicos. No obstante, de manera rigurosa, estos términos son diferentes. Para entender las propiedades físicas de las rocas sedimentarias es importante distinguir los diferentes términos denominados lutita y arcilla. A continuación, se discutirá este punto.

El término "**lutita**" describe un tipo de roca sedimentaria que es una mezcla de partículas del tamaño de la arcilla (principalmente minerales de arcilla), partículas del tamaño de limo (cuarzo, feldespato, calcita) y quizás algunas partículas del tamaño de arena como, por ejemplo, cuarzo, ocasionalmente feldespato, calcita (**Jorden y Campbell, 1984**). **Serra (2007)** escribió "de una recopilación de 10,000 análisis de lutitas hecha por Yaalon (1962), la composición promedio de las lutitas es del 59% de minerales de arcilla, predominantemente ilita; 20% de cuarzo; 8% de feldespato; 7% carbonatos; 3% de óxidos de hierro; 1% de material orgánico; y otros 2%". Muchas propiedades de las lutitas están controladas por los componentes de arcilla (por ejemplo, radiación gamma, propiedades eléctricas, capacidad de intercambio catiónico (CEC), respuesta neutrón, permeabilidad).

Mientras que el término "**arcilla**", hace referencia a un grupo de minerales, los cuales son aluminosilicatos con una estructura de lámina. Los principales elementos de construcción son dos tipos de hojas o unidades (**Schön, 2012**):

- * **Una unidad tetraédrica de un átomo central de Si y átomos de O circundantes;**
- * **Una unidad octaédrica de átomos de O y grupos OH alrededor de un átomo central de Al.**

Los minerales de arcilla (**caolinita, ilita, montmorillonita, clorita**) se caracterizan por diferentes combinaciones de apilamiento o "arquitectura" de los dos elementos de construcción. Las figuras individuales siempre representan un cristal. En la **Fig. 3**, se observan los diferentes minerales arcillosos que pueden ser encontrados.

Así, en resumen, se puede decir que:

- * **la arcilla describe un grupo de minerales.**
- * **la arcilla también define un tamaño de partícula (< 0.002 mm);**
- * **El término lutita describe un tipo de roca.**

En este trabajo, también se usa de manera indistinta los términos, arcillas y lutitas, pero se es consciente de las diferencias existentes entre éstos.

Figura 3. Minerales arcillosos, así como algunas de sus propiedades físicas.

<p>Estr. Moléc. de las arcillas</p> <p>Lámina tetraédrica de sílice</p> <p>Lámina octaédrica de aluminio</p>	
<p>Montmorillonita</p> <p>$(Ca,Na)_7(Al,Mg,Fe)_4(Si,Al)_8O_{20}(OH)_4$</p> <p>densidad=2.33 g/cm³</p>	 <p>Unicapa</p> <p>9.7 ... 17.2 * 10⁻¹⁰ m</p> <p>n-H₂O & Mg, Na,Ca</p> <p>n-H₂O & Mg, Na,Ca</p>
<p>Ilita</p> <p>$K_{1-1.5}Al_4(Si_{7.0-6.5}Al_{1-1.5}O_{20}(OH)_4$</p> <p>densidad=2.76 g/cm³</p>	 <p>Unicapa</p> <p>10 * 10⁻¹⁰ m</p> <p>K</p> <p>K</p>
<p>Clorita</p> <p>$(Mg,Al,Fe)_{12}(Si,Al)_8O_{20}(OH)_{16}$</p> <p>$(Mg,Al,Fe)_6(Si,Al)_4^o(OH)_8$</p> <p>densidad=2.77 g/cm³</p>	 <p>Unicapa</p> <p>14 * 10⁻¹⁰ m</p>
<p>Caolinita</p> <p>$Al_4(Si_4O_{10})(OH)_8$</p> <p>densidad=2.69 g/cm³</p>	 <p>Unicapa</p> <p>7.2 * 10⁻¹⁰ m</p>

Adaptado de J. H Schön. óp. cit., p. 9

La forma en la que la arcillosidad afecta a las lecturas de los registros, y consecuentemente a los resultados obtenidos después de la interpretación de estos, depende de la proporción presente de arcillas, y a sus propiedades físicas (**Schlumberger(comp), 1997**). Para muchas herramientas de registros geofísicos (resistivos, sónico, Sp, y registro nuclear magnético), también se ven afectados por la distribución que presenten las arcillas dentro de la formación. Las respuestas de los registros radiactivos (Gamma, Neutrón, densidad, etc.) no son afectados por la forma en la que las arcillas se encuentran distribuidas (**Schlumberger(comp), 1997**).

A través del análisis de núcleos, se ha observado que las arcillas presentan tres tipos de distribuciones:

- ✚ **Arcilla laminar:** Este tipo de arcilla se encuentra en forma de láminas entre los granos, o bien entre láminas de arenas. Este tipo de arcilla no afecta de manera

considerable a la permeabilidad y porosidad de las arenas, siempre y cuando no esté presente de manera cuantiosa.

$$\phi_e = \phi_t(1 - V_{cl}) \dots Ec. 2.3$$

- ✚ **Arcilla Estructural:** Este tipo de arcilla, sustituye granos de cuarzo durante su depositación, así que no afecta en la porosidad ni en la permeabilidad de la Formación. Las formaciones que presentan este tipo de distribución, su porosidad efectiva es igual a su porosidad total.

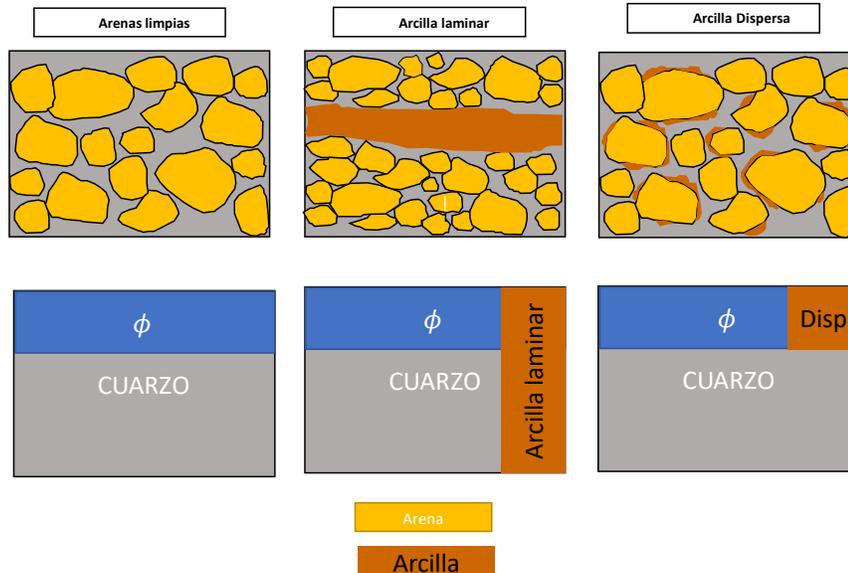
$$\phi_e = \phi_t \dots Ec. 2.4$$

- ✚ **Arcilla Dispersa;** Las arcillas dispersan, son las que más afectan en la porosidad y permeabilidad de las rocas, dado que se encuentran distribuidas alrededor de los granos, es decir, estas al depositarse, ocupan el espacio que antes ocupaban los poros.

$$\phi_e = \phi_t - V_{cl} \dots Ec. 2.5$$

En la **Fig. 4** se presenta en una forma esquemática los diferentes tipos de distribuciones que presentan las arcillas, y como la porosidad y permeabilidad se ven afectadas por la presencia de ésta.

Figura 4. Tipo de distribución que presentan las arcillas, y su influencia sobre la porosidad y permeabilidad de las rocas.



Adaptado de Schlumberger, (comp). *Principios y aplicaciones de la interpretación de registros*. p.18

En este trabajo se asume que las arcillas presentes en la Formación poseen una distribución mayormente laminar. Así que la porosidad efectiva calculada, se determinó una vez llevada a cabo la evaluación petrofísica donde se determinó entre otros parámetros, el volumen de arcilla.

Existen, varios modelos, tales como; el modelo de **Clavier (1984)**, el modelo de **Larionov (1974)**, etc., que permiten la estimación del volumen de arcilla con el empleo de registros geofísicos (generalmente usan los registros de Rayos Gamma y potencial espontáneo). Es importante mencionar, que en este trabajo se determinó el volumen de arcilla con varios modelos, y se tomó aquel que daba la menor cantidad de arcilla, esto es hecho comúnmente en la industria para obtener resultados de saturación de hidrocarburos más optimistas.

2.3 Velocidad de propagación de las ondas elásticas.

El movimiento ondulatorio puede considerarse como un transporte de energía y cantidad de movimiento desde un punto del espacio a otro, sin transporte de materia. Una onda elástica es una perturbación tensional que se propaga a lo largo de un medio elástico, el cual puede ser un sólido o un fluido ideal (**Loaiza, 2011**).

Las ecuaciones que gobiernan los problemas de los medios elásticos son las ecuaciones de movimiento, las de deformación y las constitutivas de la elasticidad lineal (**Billingham, J. King, A.C**), dichas ecuaciones no son nada simple de resolver, sin embargo, al hacer las consideraciones tal como que las ondas se propagan a través de un medio homogéneo e isotrópico, hacen que éstas se simplifiquen en gran medida, a tal grado de poder ser resueltas con métodos analíticos.

Existen diferentes clasificaciones de las ondas elásticas, dependiendo del criterio que se tome como base para llevarlo a cabo, en este trabajo solo nos interesa la clasificación que se hace de las ondas en función de la dirección del movimiento de las partículas.

Las ondas elásticas, de acuerdo con la dirección del movimiento de las partículas, se clasifican en (**Cheeke, 2012**):

- ✓ **Ondas longitudinales:** son aquellas que se caracterizan porque las partículas del medio se mueven o vibran paralelamente a la dirección de propagación de la onda. Por ejemplo, las ondas sísmicas P, las ondas sonoras y un muelle que se comprime dan lugar a una onda longitudinal.
- ✓ **Ondas transversales:** son aquellas que se caracterizan porque las partículas del medio vibran perpendicularmente a la dirección de propagación de la onda. Por ejemplo, las olas del mar, las ondas que se propagan en una cuerda y las ondas sísmicas S.

La Ec. de onda, una ecuación diferencial parcial lineal de segundo orden, es la ecuación que describe matemáticamente la propagación de las ondas.

Al igual que en el caso de la porosidad, se han creado modelos petrofísicos (**Willy, 1956; Raymer, 1980; Han, 1986**), a partir de los cuales se hace una estimación de la velocidad de propagación de la onda P, presentando los mismos inconvenientes. Aun así, estos modelos son usados para la determinación de la velocidad de propagación de las ondas, y se emplean

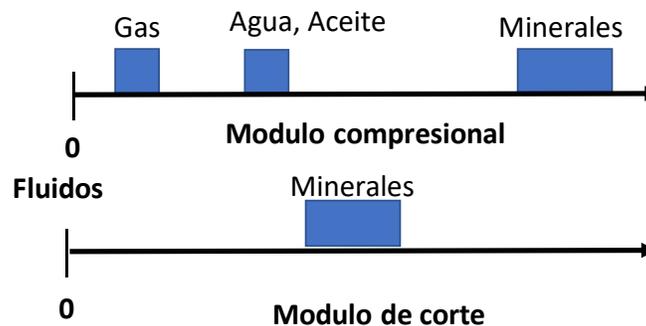
métodos geoestadísticos convencionales- variogramas, regresión lineal, regresión lineal multivariantes, etc.- para el geoplado, de ésta y otras propiedades petrofísicas de interés.

Dada la necesidad de una mayor precisión en la caracterización estática de los yacimientos, se ha tenido la necesidad de la introducción de nuevas herramientas de modelado no lineal, como el llevado en el presente trabajo, para una estimación más realista de las propiedades petrofísicas de toda un área. El empleo de herramientas de modelado no lineal es consecuencia de la alta heterogeneidad en las formaciones, y ésta, es a la vez producto de las diferentes propiedades físicas que poseen los diferentes constituyentes en las rocas, minerales-fluidos (La propagación de las ondas en las rocas está definida por las propiedades elásticas de los diferentes constituyentes de ésta). Una de las técnicas que ha tomado auge en las últimas décadas son las redes neuronales, la cual es la usada en la metodología que se propone en este trabajo, para la construcción de secciones de propiedades petrofísicas- V_{cl} , V_p y Rho .

Las propiedades elásticas de las rocas están dominadas por las propiedades de la roca sólida (esqueleto), incluyendo "defectos" como poros, fracturas y grietas (Schön, 2012). Estos defectos tienen, en la mayoría de los casos, dimensiones más pequeñas que la longitud de onda. La Fig. 5 ilustra las tendencias generales para los componentes básicos de la roca (Schön, 2012):

- ✓ **Minerales sólidos:** minerales, componentes de la matriz.
- ✓ **Fluidos:** líquidos (agua, aceite), aire, gas.

Figura 5. Módulo compresional y de corte para los elementos principales de las rocas.



Adaptado de J. H Schön. *óp. cit.*, p. 154

Para velocidades de onda compresionales, $V_{p,mineral} > V_{p,agua;aceite} > V_{p,gas}$ y para el módulo de compresión correspondiente, $k_{mineral} > k_{agua;aceite} > k_{gas}$. El módulo de corte muestra un comportamiento completamente diferente, porque por definición, los fluidos no tienen resistencia al corte ($\mu_{fl} = 0$). Es por ello, que la onda de corte muestra solo una menor dependencia de los fluidos de poros a través del término de densidad. Inmediatamente, se puede deducir lo siguiente:

- ✚ el aumento de la porosidad disminuye las velocidades de las ondas de compresión y de corte;
- ✚ la velocidad de la onda compresional está controlada por el tipo de fluido del poro (gas, líquido);
- ✚ la velocidad de la onda de corte no está fuertemente controlada por el tipo de fluido en el poro.

La **Tabla 3** es un fragmento de la mostrada por **J. H Schön, 2012**, donde se proporciona una orientación para la magnitud de las velocidades de propagación de las ondas y los parámetros elásticos de los principales minerales que conforman las rocas.

De esta tabla podemos observar que:

- ✓ Los componentes básicos de las rocas ígneas presentan módulos y velocidades de propagación mayores, que para aquellos componentes ácidos.
- ✓ Existe una diferencia significativa, en las velocidades y módulos elásticos, entre los 3 componentes principales, presentes en las rocas de los yacimientos; calita, dolomía y cuarzo.

En cuanto a los fluidos de los poros (gas, aceite y agua), generalmente tienen distintos intervalos de módulo de compresión o de volumen (el módulo de cizalla de los fluidos es cero). En la **Tabla 2** se muestra los rangos que se presentan generalmente, para el módulo de compresibilidad y la densidad, en los fluidos del yacimiento.

<i>Tabla 1. Rango de valores que presentan los fluidos de los yacimientos, en su módulo de compresibilidad, k, y su densidad</i>		
Gas	$k_{fl} \approx 0.01 - 0.4 \text{ Gpa}$	$\rho_{fl} \approx 0.1 - 0.5 \text{ kg m}^{-3}$
Aceite	$k_{fl} \approx 0.4 - 3.0 \text{ Gpa}$	$\rho_{fl} \approx 0.7 - 1.1 \text{ kg m}^{-3}$
Agua	$k_{fl} \approx 2.0 - 4.0 \text{ Gpa}$	$\rho_{fl} \approx 0.9 - 1.2 \text{ kg m}^{-3}$

Adaptada de **J. H Schön, óp. cit., p. 155**

Así mismo, en la **Tabla 4** se presentan los valores que presenta la onda compresional en los fluidos presentes en los yacimientos, y algunos otras más, que, si bien no están presentes en los yacimientos, sirven como referencia.

La **Fig. 6** muestra el rango promedio de las velocidades de las ondas compresionales y de corte para los tipos de rocas que son encontradas comúnmente e indica algunas tendencias y características típicas:

- Las rocas porosas sedimentarias muestran un amplio rango de velocidad, donde las altas velocidades son típicas para aquellas rocas que son relativamente densas (baja porosidad) y las bajas velocidades son típicas para los miembros porosos (portadores de gas)
- Las velocidades más bajas se presentan en las rocas no consolidadas (secas) como resultado del contacto específico grano-grano y la alta porosidad.

Como se puede observar en la figura, existe una amplia dispersión en los valores de V_p para las rocas porosas (resultado de la combinación de la influencia de la porosidad, fluido de poros y composición mineral) en particular y un rango pequeño de valores para una roca densa con una composición bien definida (anhidrita).

Dado que la mayoría de los yacimientos petrolíferos, se encuentran sobre rocas sedimentarias, se hará un análisis más detallado de este tipo de rocas.

Tabla 2. Densidad, módulos elásticos y velocidad de onda P de los minerales más comunes que conforman a las rocas.

Mineral	ρ $kg\ m^{-3}$	k GPa	μ GPa	V_p $m\ s^{-1}$	V_s $m\ s^{-1}$	ν	Referencia
Cuarzo	2650	38.2	43.4	6050	4090	0.08	G
	2650	37	44	6050	4090	0.08	C
	2650	36.5	45.6	6060	4150	0.06	M
Hornblenda	3124	87	43	6810	3720	0.29	G
Olivino	3320	130	80	8540	4910	0.24	C
Ortoclasas	2570	46.8	27.3	5690	3260		A
Biotita	3050	51	27	5350	3000	0.27	G
	3050	59.7	42.3	6170	3730	0.21	M
	3050	41.1	12.4	4350	2020	0.36	M
Moscovita	2790	52	32	5810	3370	0.25	G
	2790	42.9	22.2	5100	2820	0.28	M
	2790	52	30.9	5780	3330	0.25	M
	2790	61.5	41.1	6460	3840	0.23	M
Caolinita	1580	1.5	1.4	1440	930	0.14	M
Arcilla	2580	20.9	6.85				H
Arcillas del Golfo	2550	25	9	3810	1880	0.34	M
	2600	21	7	3410	1640	0.35	M
Clorita	95.3	11.4					K
Ilita	39.4	11.7					K
Calcita	2714	73	32	6540	3430	0.31	G
	2710	76.8	32	6640	3440	0.32	M
Dolomía	2860	94	46	7370	4000	0.29	G
	2870	94.9	45	7340	3960	0.30	M
Anhidrita	2970	55	30	5620	3140	0.27	G
	2980	56.1	29.1	5640	3130	0.28	M

Código de referencia: G, Gebrande et al. (1982); C, Carmichael (1989); M, datos de la compilación de Mavko et al. (1998), H, Helgerud et al. (1999). Adaptada de J. H Schön, óp. cit., pp. 156-157

<i>Tabla 3. Velocidad de onda compresional para algunos fluidos.</i>		
Fluidos	V_p en $m\ s^{-1}$	Observaciones, parámetros
Aire	263	$T = 173^\circ K$
	332	$T = 273^\circ K$
	387	$T = 373^\circ K$
Metano	488	$P = 0.103\ MPa$
Dióxido de carbono	259	
Aceite, natural	1035 – 1370; media 1280	
Parafina	1420	$T = 307^\circ K, \rho = 835\ kg\ m^{-3}$
Aceite para gasolina	1250	$T = 307^\circ K, \rho = 803\ kg\ m^{-3}$
Agua	1497	$T = 298^\circ K$, destilado
	1585	$p = 0.103\ MPa$; 1000ppm NaCl
	1638	$p = 0.103\ MPa$; 1500ppm NaCl
	1689	$p = 0.103\ MPa$; 2000ppm NaCl

Después de Ebert (1976) Bulatova et al. (1970), Gregory (1990), Serra (1984), Ellis (1987), Baker Atlas/Western Atlas (1992). Adaptada de J. H Schön, óp. cit., p. 158

Con respecto a la velocidad y su comportamiento, hay dos tipos de rocas sedimentarias (Schön, 2012):

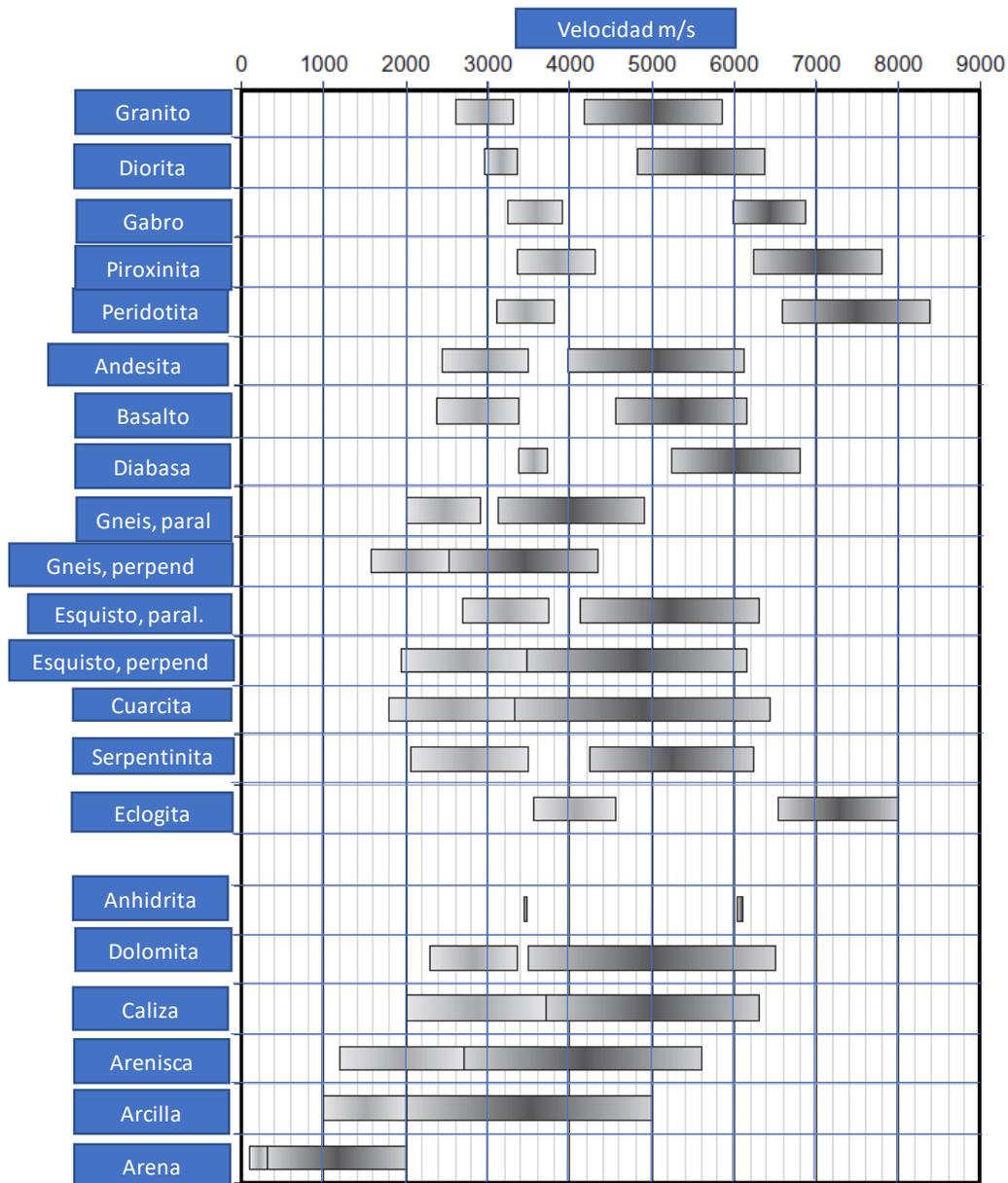
1. Rocas densas sin poros (anhidrita, sal) con una velocidad bien definida que está controlada por las propiedades minerales y la composición.
2. Rocas porosas con un rango de velocidad con fuerte influencia de la porosidad y el fluido de poro, pero también las propiedades de contacto de los componentes sólidos de las rocas (que van desde rocas cementadas hasta rocas no consolidadas) y la composición mineral (arenisca, dolomita, caliza, influencia de las arcillas).

Son de particular interés en la industria petrolera el segundo grupo; la diversidad de influencias origina la variedad de relaciones y correlaciones que existen en el área, y que se empleen técnicas de modelado no lineal para el geoplado de propiedades petrofísicas en general, como ya se mencionó anteriormente.

En la siguiente sección se verá cómo se ve afectada la velocidad de propagación debido a la composición mineralógica de la roca, y la porosidad.

Es importante mencionar, que las fracturas tienen gran influencia en la propagación de las ondas, en el caso de los carbonatos. No obstante, este tema está fuera del alcance de este trabajo.

Figura 6. Rango de valores que presentan las rocas, en la velocidad de propagación de las ondas compresionales (valores altos), y en las ondas de cizalla (valores bajos).



Adaptado de J. H Schön. óp. cit., p. 160.

2.3.1 Influencia de la litología (mineralogía) y la porosidad sobre la velocidad.

La influencia litológica o mineralógica sobre la velocidad se expresa principalmente por (Schön, 2012):

- ✓ la diferencia entre rocas clásticas (arenisca) y carbonatos (dolomita, piedra caliza);
- ✓ la diferencia entre rocas clásticas consolidadas (arenisca) y no consolidadas (arena);
- ✓ la influencia de las arcillas en la velocidad.

La disminución de la velocidad con el aumento de la porosidad es una característica general de todas las rocas porosas.

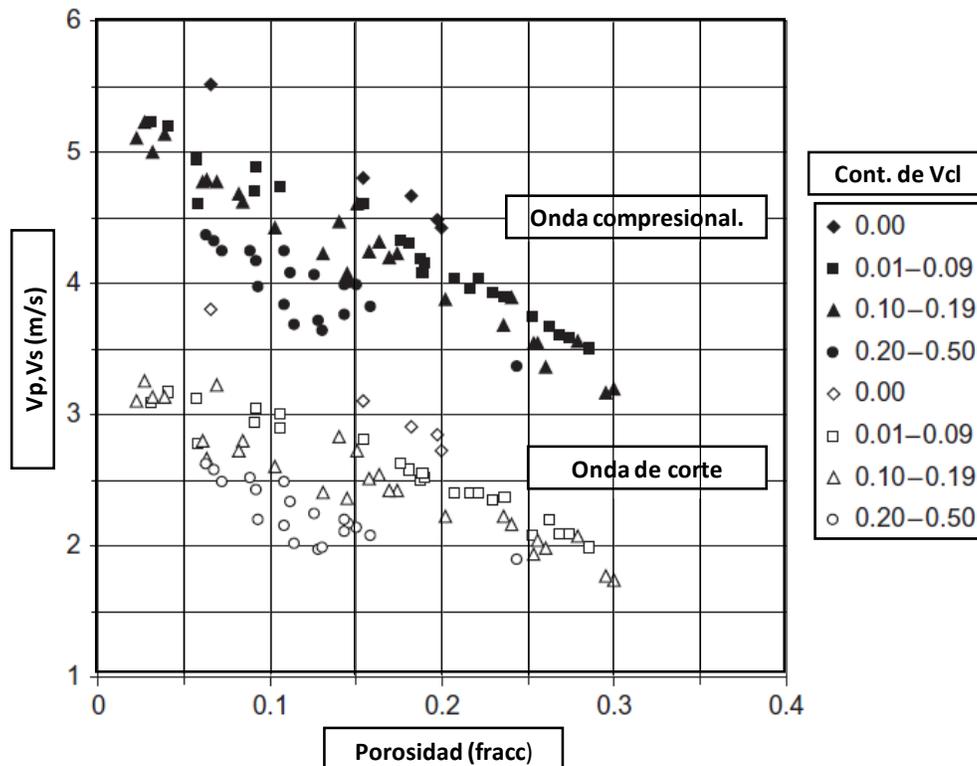
La influencia de la composición mineral en la velocidad se expresa mediante la llamada velocidad de la matriz, que es la velocidad extrapolada de un tipo de roca porosa para la porosidad cero. Para los tres componentes minerales principales de las rocas de yacimiento, las velocidades medias de la matriz son:

Cuarzo (Areniscas)	$V_{p,ma} = 5500 \text{ ms}^{-1}$	$V_{p,ma} = 18000 \text{ fts}^{-1}$
Calcita (Caliza)	$V_{p,ma} = 6400 \text{ ms}^{-1}$	$V_{p,ma} = 20400 \text{ fts}^{-1}$
Dolomita (Dolomía)	$V_{p,ma} = 7000 \text{ ms}^{-1}$	$V_{p,ma} = 23000 \text{ fts}^{-1}$

Adaptado de J. H Schön. óp. cit., p. 165

Para mostrar el efecto del contenido de arcilla y la porosidad, en rocas clásticas, específicamente areniscas, tomaremos los resultados de investigaciones experimentales llevadas a cabo por **Han et al, (1986)**, los cuales son mostrados en la **Fig. 7**

Figura 7. Velocidades de ondas compresivas y de corte versus porosidad (fracción) para 75 muestras de areniscas a una presión de confinamiento de 40 MPa y una presión de poro de 1.0 MPa; datos después de Han et al. (1986).



Adaptado de J. H Schön. óp. cit., p. 166

De este gráfico se puede observar que:

- ✓ El efecto de porosidad en las velocidades de compresión y de onda de corte es comparable;
- ✓ Tanto la velocidad de corte como la velocidad compresional se ve mermada ante la presencia de arcilla.
- ✓ El efecto de la arcilla es más importante para la velocidad de la onda de corte.

2.4 Densidad.

La densidad es una propiedad física de todo cuerpo, que se define como la relación de la masa entre el volumen:

$$\rho = m/V \dots Ec. 2. 6$$

Las unidades de la densidad en el Sistema Internacional son kg/m³.

La densidad de una roca estará en función de los principales componentes de las rocas, teniéndose la necesidad de hacer una distinción de las densidades de estos componentes. Entonces, sabiendo que las rocas son heterogéneas, se tienen las siguientes "tipos" de densidades en éstas (Schön, 2012);

- Densidad de volumen, ρ , es la densidad promedio de un volumen específico de roca (incluyendo poros, etc.); por ejemplo, la densidad de las areniscas.
- ρ_i , hace referencia a la densidad de cada uno de los minerales, i , que componen a la roca; por ejemplo, la densidad del cuarzo, de la calcita, etc.
- ρ_{ma} , es la densidad promedio del material sólido de la roca (mineral o mezcla de minerales), también se le llama densidad del grano; por ejemplo, se tiene la densidad de la matriz de carbonato (sin los fluidos de los poros).
- ρ_{fl} , hace referencia a la densidad promedio de los fluidos contenidos en los poros o fracturas; por ejemplo, densidad del agua, ρ_a

La densidad de volumen de un material compuesto (roca), que está compuesto por n componentes es entonces;

$$\rho = \sum_{i=1}^n \rho_i * V_i \dots \dots 2. 7$$

Donde V_i es la fracción volumen de cada componente dentro de la roca.

Con base a lo anterior, entonces para una roca porosa, donde la fracción de volumen de espacio ocupado está dado por la porosidad, ϕ , sería:

$$\rho = (1 - \phi)\rho_{ma} + \phi * \rho_{fl} \dots 2. 8$$

En el caso, que la roca contenga más de un fluido dentro de los poros, la ecuación anterior quedaría como;

$$\rho = (1 - \phi)\rho_{ma} + \phi(\rho_a * S_a + \rho_{ac} * S_{ac} + \rho_g * S_g) \dots 2.9$$

Donde los subíndices a,ac y g hacen referencia a los fluidos agua, aceite y gas, respectivamente.

Al igual que en las propiedades petrofísicas descritas anteriormente, la densidad en las rocas será función directa y proporcional de los componentes principales de éstas (minerales y fluidos), lo cual está expresado matemáticamente en la **Ec. 2.7**. A continuación, se aborda de manera separada la influencia de cada uno de ellos sobre la densidad promedio de las rocas.

La densidad de los minerales está controlada por su composición elemental y estructura y enlaces internos. **Clark (1966), Dortman (1976), Olhoeft y Johnson (1989), Rosler y Lange (1972), Serra (1984) y Wohlenberg (1982)** publican las compilaciones de las densidades de minerales (**citados en Schön, 2012, p.98**). Algunas densidades de minerales se presentan en la sección anterior en la **Tabla. 3**

El rango de densidad en los minerales es (**Schön, 2012**):

- Los minerales más abundantes en la conformación de las rocas presentan un rango de densidades entre 2.2-3.5 E3 kg/m³.
- Los minerales ore presentan un rango de densidades entre 4-8 E3 kg/m³.

En cuanto a la densidad de los líquidos y gases, ésta está controlada por:

- ✓ **Composición química (incluidos los componentes solubles);**
- ✓ **Temperatura y presión (la densidad aumenta al aumentar la presión y decrece al aumentar la temperatura).**

Para sustancias líquidas, **Schlumberger (2000)** da el siguiente rango de valores de densidad:

- ✓ **Agua fresca, 1 E3 kg/m³.**
- ✓ **Salmuera (200,000 ppm), 1.146 E3 kg/m³.**
- ✓ **Aceite, 0.85 E3 kg/m³.**

En el caso de los gases, su densidad está controlada muy fuertemente por la presión. La densidad de la mezcla de gases se especifica frecuentemente por una gravedad específica, que es la relación entre la densidad del gas de interés y la densidad del aire a 15.6 ° C (60 ° F) y a la presión atmosférica. La **Fig. 8** muestra la densidad del metano (gas seco) en función de la temperatura y la presión.

La densidad aparente de las rocas está en función de los constituyentes de las rocas tratados anteriormente, la cual sigue exactamente la Ec. 2.7 y depende de:

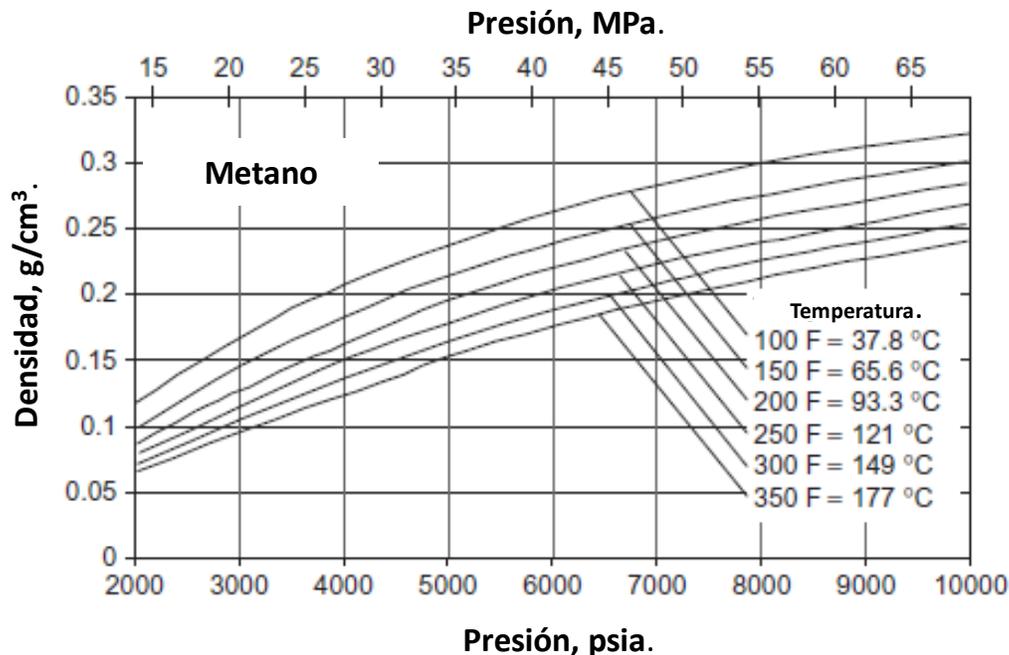
- * la composición mineral (densidades minerales y fracciones volumétricas);
- * porosidad (poros, fracturas) y densidad de fluidos de poro.

Esto explica la regla general de la variación de densidad:

- * Las rocas ígneas muestran un aumento de la densidad de félsicos (ácido) a máficos.
- * Las rocas porosas muestran una disminución de la densidad con el aumento de la porosidad y disminución de la saturación de agua.

Igualmente, solo nos concentraremos en las rocas sedimentarias, las cuales se caracterizan por ser rocas porosas, con fluidos.

Figura 8. Densidad del gas metano a diferentes presiones y temperaturas.



Adaptado de J. H Schön. óp. cit., p. 102

La Fig. 9 brinda una visión general sobre la densidad de las rocas sedimentarias. En las rocas sedimentarias, como ya se mencionó en el apartado anterior cuando se habló de la propagación de las ondas elásticas, hay dos grupos principales con respecto a los factores de control de la densidad (Schön, 2012):

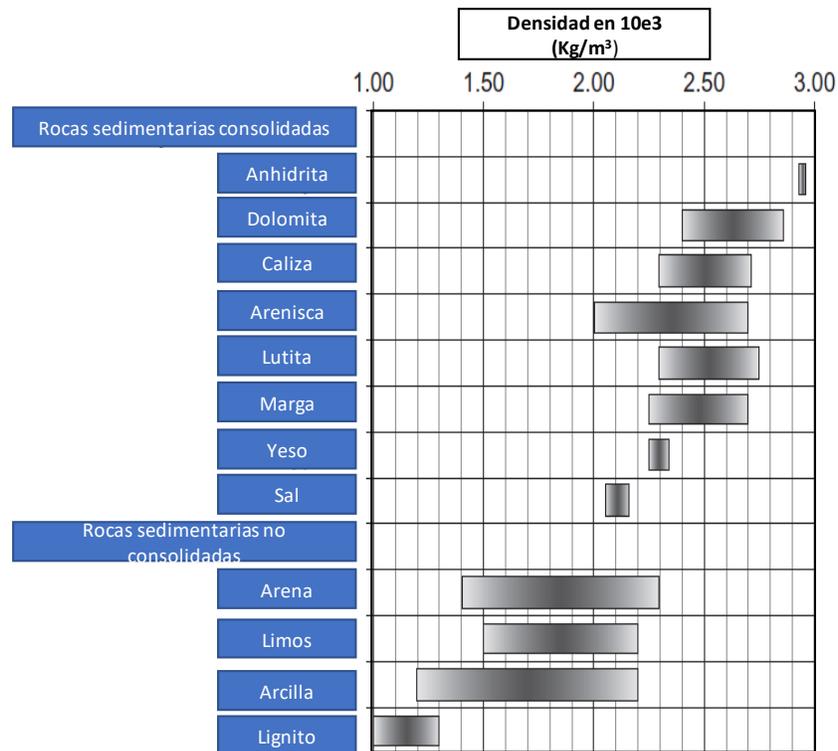
1. Sedimentos libres de poros (sal, anhidrita, carbonato denso)
2. Sedimentos porosos (el grupo más abundante, la roca sedimentaria típica).

La densidad de los sedimentos libres de poros está determinada por su composición mineral; algunos de ellos son rocas mono-minerales (especialmente sales). Como resultado de una composición casi constante, este grupo se caracteriza por rangos estrechos de valores de densidad.

La densidad de los sedimentos porosos está controlada por (Schön, 2012):

- * la composición mineral (densidad de la matriz, también llamada densidad de grano).
- * la porosidad.
- * la composición del contenido de poro (saturación).

Figura 9. Rango de valores promedio de densidad para rocas sedimentarias más comunes.



Adaptado de J. H Schön. óp. cit., p. 104

Como resultado de la clara diferencia entre el rango de densidad media de la matriz (alrededor de $2.5- 2.9 \cdot 10^3 \text{ kg/m}^3$) y el rango de densidad media del fluido poroso (líquidos alrededor de $0.8- 1.2 \cdot 10^3 \text{ kg/m}^3$); gases, $<0.4 \cdot 10^3 \text{ kg/m}^3$), hay una fuerte correlación entre la densidad aparente de rocas ρ y la porosidad ϕ , dada en la Ec. 2.8.

Para un tipo de roca sedimentaria, generalmente hay una amplia gama de valores de densidad que refleja las variaciones de porosidad y saturación. Los valores más altos representan miembros densos (baja porosidad) del tipo, mientras que los valores más bajos resultan del aumento de la porosidad y/o del aumento del contenido de gas en los poros.

Esta propiedad, generalmente se estima en campo mediante el empleo de registro geofísicos, que no miden propiamente la densidad de las rocas, sino la densidad eléctrica en éstas, el cual es transformado posteriormente a valores de densidad. La importancia de contar con secciones de esta propiedad, que se obtienen extrapolando los valores generalmente usando

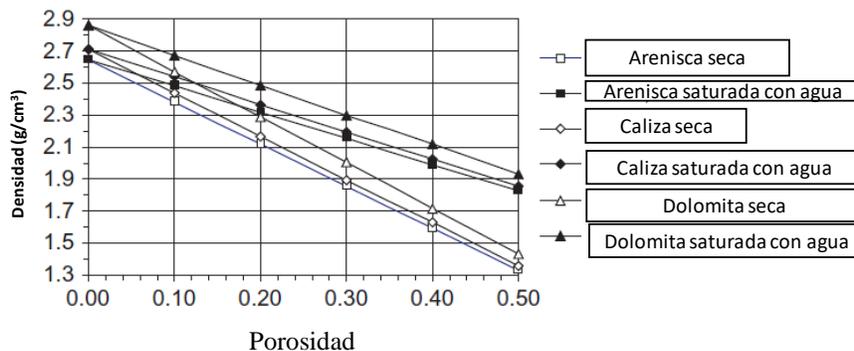
métodos geostadísticos tradicionales, es que se aplica esta medición para el cálculo de porosidades totales. La reorganización de la **Ec. 2.8** da como resultado:

$$\phi = \frac{\rho_{ma} - \rho}{\rho_{ma} - \rho_{fl}} \dots 2.5$$

Para la determinación de la porosidad, la densidad aparente medida ρ (de registros), la densidad de la matriz, ρ_{ma} , y la densidad del fluido, ρ_{fl} , son entradas necesarias. La **Fig. 10** muestra la relación entre la densidad aparente y la porosidad de la roca arenisca, caliza y dolomita para la roca seca y saturada de agua.

La fuerte correlación entre la porosidad y la densidad aparente da como resultado un aumento de la densidad con la presión o la profundidad, que es el resultado de una disminución de la porosidad con la profundidad

Figura 10. Relación entre la densidad volumétrica y la porosidad para areniscas ($\rho_{ma} = 2.65 \text{ g/cm}^3$), caliza ($\rho_{ma} = 2.71 \text{ g/cm}^3$), dolomita ($\rho_{ma} = 2.86 \text{ g/cm}^3$). Los caracteres no rellenos indican que las rocas están secas ($\rho_{fl} = 0$); y los rellenos indican que las rocas están saturadas con agua ($\rho_{fl} = 1 \text{ g/cm}^3$).



Adaptado de J. H Schön. óp. cit., p. 105

CAPITULO III

REDES NEURONALES ARTIFICIALES

3.1 Fundamentos.

La inteligencia artificial generalmente se divide en dos categorías básicas: sistemas basados en reglas (expertos) y sistemas adaptativos (neuronales) (Mohaghegh et al., 1995). Este documento se concentrará en *las redes neuronales*.

Dado que una red neuronal artificial, es un esquema de cálculo inspirado biológicamente, específicamente en el cerebro humano, pues al igual que éste, es un sistema cognitivo que da solución a problemas de manera adaptativa, distributiva y paralela. Este esquema de solución de problemas ha sido ampliamente usado en la comunidad financiera, y aproximadamente desde hace tres décadas se empezó a emplear en la solución de los problemas propios de la industria petrolera, uno de los cuales será tratado en este documento; **caracterización petrofísica de yacimientos**.

El interés principal en las redes neuronales tiene sus raíces en el reconocimiento de que el cerebro procesa la información de una manera diferente a las computadoras digitales convencionales. Es por ello por lo que iniciaremos este capítulo hablando del órgano que fue la inspiración para el desarrollo de este tipo de inteligencia artificial; el cerebro humano.

El término cerebro, deriva del latín "cerebrum", con su raíz indoeuropea «ker», *cabeza, en lo alto de la cabeza* y «brum», *llevar*; teniendo el significado arcaico de *lo que se lleva en la cabeza*. Éste, definido desde el punto de vista anatómico, es un órgano complejo que forma parte del Sistema Nervioso Central (SNC) y que constituye la parte más voluminosa y conocida del encéfalo. Está situado en la parte anterior y superior de la cavidad craneal y está presente en todos los vertebrados. Dentro del cráneo, el cerebro flota en un líquido transparente, llamado líquido cefalorraquídeo, que cumple funciones de protección, tanto físicas como inmunológicas (Changeux, 2016). Sin embargo, también puede ser concebido como un potente ordenador capaz de resolver problemas de alta complejidad, de manera simultánea en corto tiempo. Éste presenta la capacidad de organizar sus millones de células constituyentes, las neuronas, para llevar a cabo ciertas actividades, o cálculos, tales como: reconocimiento de patrones, la percepción, y el control motriz, por mencionar algunos. Considere, por ejemplo, el sentido del gusto del hombre, el cual, como alguna otra de nuestras capacidades sensoriales, puede ser considerada como una herramienta procesadora de información (Marr, 1982; Levine, 1985, citados por Haykin, S., 1999). Una de las tareas de este sistema es proveer una representación de nuestro alrededor, y más importante aún, suministrar información que necesitamos para poder interactuar con el exterior. Siendo más específicos, el cerebro ejecuta tareas de reconocimiento perceptivo (por ejemplo, reconocimiento de un sabor, aun cuando tengamos los ojos vendados.) en aproximadamente 300-400 milisegundos (Churchland, 1992), mientras que tareas de mucho menor

complejidad le tomarían días a un ordenador llevarlas a cabo. Es importante mencionar, que el proceso de identificación de un platillo en particular es resultado gracias de la información que es percibida por los sentidos de la vista y el olfato.

Pero entonces, he ahí la gran pregunta, ¿Cómo es que el cerebro lo hace? Al nacer, el cerebro humano presenta una gran estructura y la habilidad de construir sus propias reglas a través de lo que conocemos como la “*experiencia*”. De hecho, esta se construye a lo largo del tiempo, siendo el tiempo más crucial para su desarrollo, durante sus dos primeros años de existencia; no obstante, su desarrollo continúa hasta la muerte del sujeto.

El “desarrollo neuronal” puede ser visto como la capacidad maleable del cerebro (**Haykin, S., 1999**). Esta maleabilidad permite el desarrollo del sistema nervioso y poder adaptarse al entorno. Tal como la maleabilidad es esencial para el funcionamiento de las neuronas como unidades de procesamiento de información en el cerebro humano, lo es también para las redes neuronales constituidas por neuronas artificiales.

El sistema nervioso del hombre puede ser visto como un sistema de tres etapas, como se muestra el diagrama de bloque de la **Fig. 11 (Arbib, 1987)**. El elemento central del sistema es el cerebro, representado por la red neuronal, el cual continuamente recibe y percibe información, y toma las decisiones apropiadas con base a ella. Dos sets de flechas se muestran en la figura. Aquellas que apuntan desde la izquierda hacia la derecha indica la transmisión de información (señales) hacia delante a través del sistema. Las flechas apuntando desde la derecha hacia la izquierda indican la presencia de retroalimentación en el sistema. Los *receptores* convierten un estímulo desde el cuerpo humano o desde el entorno en impulsos eléctricos que transmiten la información a la red neuronal (el cerebro). Los *ejecutores* convierten los impulsos eléctricos generados por la red neuronal en respuestas distinguibles.

Figura 11. Representación en diagrama de bloques del sistema nervioso.



Fuente: Modificado de Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, New Jersey: Prentice Hall.p.6

El problema de entender al cerebro ha sido más fácil gracias al trabajo de **Ramón y Cajal (1911)**, quien introdujo la idea de las neuronas como constituyentes estructurales del cerebro. Típicamente, las neuronas son alrededor de 5 o 6 órdenes de magnitud más lentas que las compuertas lógicas de silicón; sin embargo, la diferencia de rapidez de procesamiento entre éstos es que éste está constituido por alrededor de 10 billones de neuronas, y 60 trillones de conexiones o sinapsis, lo que hace que éste sea una estructura enormemente eficiente. Específicamente, la eficiencia energética del cerebro es de alrededor de 10 E-16 Joules (J) por operación por segundo, mientras que valor correspondiente de los mejores ordenadores actuales es de alrededor de 10 E-6 (J). (**Shepherd y Koch, 1990, citado por Haykin, S., 1999**)

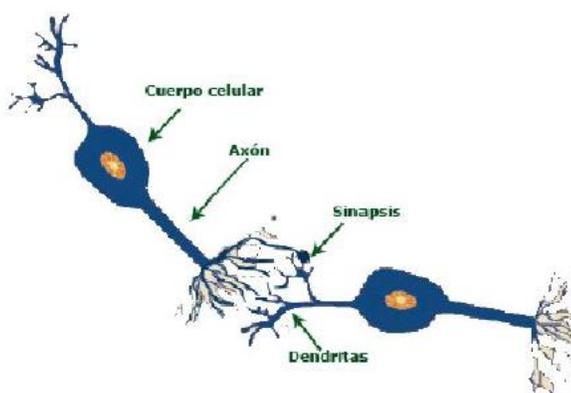
Un concepto importante que nos ayudará a entender el proceso cognitivo que ocurre en nuestro cerebro, y posteriormente el llevado a cabo en las redes neuronales artificiales, es la **sinapsis**. Las sinapsis son unidades estructurales y funcionales elementales que son el medio a través de los cuales se ejecuta la interacción entre las neuronas. El tipo más común de sinapsis es el tipo químico, que funciona de la siguiente manera. Un proceso presináptico libera una sustancia transmisora que se difunde a través de la unión entre neuronas, y posteriormente actúa sobre un proceso postsináptico (**Shepherd y Koch, 1990, citado por Haykin, S., 1999**). Así que se puede decir que una sinapsis es un proceso a través del cual una señal eléctrica presináptica es convertida en una señal química, que posteriormente esta se transforma nuevamente en otra señal eléctrica, pero ahora postsináptica. Es importante mencionar, que este proceso solo se realiza en una sola dirección. Con base a lo anterior podemos decir que la sinapsis es el medio a través del cual las neuronas se transmiten la información, mediante la excitación o inhibición de una a la otra.

Como se mencionó anteriormente, la “plasticidad” en el sistema nervioso, es lo que le permite adaptarse a las condiciones del medio. En el cerebro de un hombre adulto, la plasticidad se refleja en éste a través de dos mecanismos (**Haykin, S., 1999**):

- ✓ **La creación de nuevas conexiones sinápticas entre neuronas.**
- ✓ **La modificación de las conexiones sinápticas preexistentes.**

Los axones, las líneas de transmisión, y las dendritas, las zonas receptoras, constituyen dos tipos de filamentos celulares, que se distinguen entre ellos por su morfología. Un axón presenta una superficie más suave, con menos ramificaciones, y cuya longitud es menor; mientras que una dendrita (así llamada por su semejanza a un árbol) tiene una superficie irregular y más ramificaciones (**Freeman, 1975**). La **Fig. 12** muestra de manera simplificada la representación de dos neuronas y sus elementos.

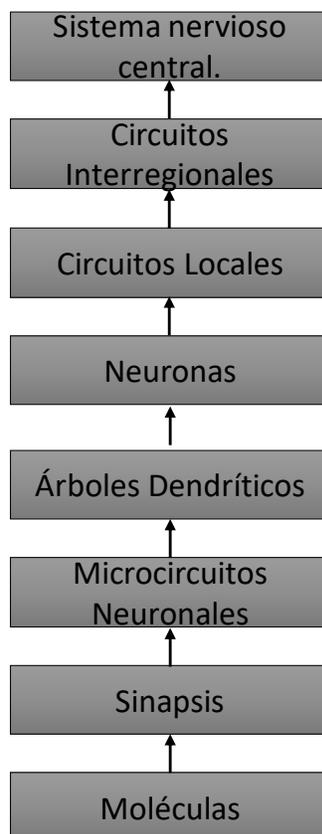
Figura 12. Modelo biológico de una neurona.



En el cerebro del hombre se pueden encontrar tanto organizaciones anatómicas a pequeña escala, como a una escala grande, así que diferentes funciones se llevan cabo tanto a niveles inferiores como superiores. La **Fig. 13** muestra la jerarquía de los niveles de organización intracelular que son resultado del trabajo extensivo sobre el análisis de las regiones locales en el cerebro (**Shepherd y Koch, 1990; Churchland y Sejnowski, 1992, citados en**

Haykin, S., 1999, p.29). La **sinapsis** representa el nivel mas fundamental, el cual depende de la acción de las moléculas e iones. A niveles superiores, encontramos los **microcircuitos neuronales**, los **árboles dendríticos**, y posteriormente las **neuronas**. Un **microcircuito neuronal** consiste en un ensamble de conexiones sinápticas, las cuales se encuentran organizadas en patrones de conectividad para producir una operación de interés. Se podría comparar un microcircuito neuronal con un chip de silicon, contituido por un ensamble de transmisores. Los circuitos neuronales estan agrupados para formar **subunidades dentriticas** dentro de los árboles dendríticos de neuronas individuales. Una **neurona** completa, de tamaño de 100 μm , contiene numerosas subunidades dentriticas. Al siguiente nivel de complejidad tenemos los **circuitos locales** (con un tamaño de alrededor de 1 mm), los cuales están constituidos de neuronas con propiedades similares o diferentes; estos ensambles de neuronas ejecutan operaciones características de una región localizada en el cerebro. Éstos son seguidos por los **circuitos interregionales** constituídos de columnas, rutas y mapas topográficos, los cuales involucran múltiples regiones localizados en diferentes partes del cerebro. (**Haykin, S., 1999**)

Figura 13. Niveles de organización estructural en el cerebro.



Modificado de Haykin, S. (1999). óp. cit., p. 9

Es importante reconocer que los niveles estructurales de organización descritos aquí son características únicas del cerebro. Estas no pueden ser encontradas actualmente en una

computadora digital, y ni si quiera estamos cercanos de poder recrearlas con las redes artificiales neuronales. Todos los esfuerzos están orientados a recrear dichos niveles cognitivos en el diseño de nuestra inteligencia artificial, no obstante, siendo objetivos estamos en niveles muy incipientes, primitivos, en comparación al funcionamiento del cerebro humano. Las redes neuronales que hoy en día somos capaces de diseñar son tan primitivas tanto como los circuitos locales e interregionales encontrados en el cerebro. Lo que es remarcable, sin embargo, es el progreso que se ha hecho en las últimas dos décadas. Tomando como base de inspiración a las neurociencias, y aunado a los avances tecnológicos, no hay duda de que en una década más el entendimiento de las redes neuronales artificiales será mucho más sofisticado que lo que es hoy en día.

3.2 Redes Neuronales.

3.2.1 Definición.

En la literatura, no existen definiciones realmente claras y definitivas de lo que son las Redes Neuronales Artificiales (RNA). Muchas de las definiciones explican las RNA's haciendo referencia solo de la analogía de éstas con los mecanismos de cálculo que se llevan a cabo en las neuronas reales. Las definiciones más ampliamente utilizadas, sin embargo, involucran dos aspectos: las unidades de procesamiento y el algoritmo de aprendizaje. El aprendizaje se define como la modificación de pesos sinápticos para capturar información (**Haykin, S.,1999**). En algunas otras definiciones también se agrega que las RNA's pueden modificar su propia topología mientras aprenden.

Una buena definición de RNA, es la dada por **Haykin, S (1999)** que describe una RNA como una combinación de unidad de procesamiento masivamente paralela simple que puede adquirir conocimiento del entorno a través de un proceso de aprendizaje y almacenar el conocimiento en sus conexiones.

“Una red neuronal es un procesador masivo de distribución paralela, constituido por unidades simples de cálculo, el cual tiene la habilidad natural de almacenar conocimiento adquirido a través de la experiencia, el cual lo hace disponible para su uso.”

Esta definición hace recordar al cerebro humano en dos aspectos:

- 1-. El conocimiento es adquirido por la red desde sus alrededores, ambiente, a través de un proceso de aprendizaje.
- 2-. La fuerza de interconexión entre las neuronas, conocida por muchos autores como **pesos sinápticos**, son usadas para almacenar el conocimiento adquirido.

Sin embargo, esta definición es algo limitada al no considerar el algoritmo de aprendizaje y otros aspectos esenciales de una red, considerando más apropiada y completa, la dada por **R. Rojas (1996)**:

" Las RNA's son máquinas de aprendizaje distribuidas, adaptativas y generalmente no lineales construidas a partir de muchos elementos de procesamiento (EP) diferentes. Cada

EP recibe conexiones de otros EP y/o de sí mismo. La interconectividad define la topología. Las señales que fluyen en las conexiones se escalan mediante parámetros ajustables denominados pesos, w_{ij} . Los EP suman todas estas contribuciones y producen un resultado que es una función no lineal (estática) de la suma. Las salidas de EP se convierten en salidas del sistema o se envían al mismo u otros EP "

3.2.2 Modelo artificial de una neurona

Primeramente empecemos a definir el elemento básico de toda red neuronal computacional, la neurona (**Haykin, S., 1999**). El diagrama de bloques de la **Fig. 14** muestra el modelo de una neurona, que forma la base para el diseño de redes neuronales (artificiales) y lo que sería su analogía con una neurona real. Aquí identificamos tres elementos básicos del modelo neuronal:

1.-Un conjunto de sinapsis o enlaces de conexión, cada uno de los cuales se caracteriza por un peso propio. Específicamente, una señal x_j a la entrada de la sinapsis j conectada a la neurona k es multiplicada por el peso sináptico w_{kj} . "Es importante notar la manera en que son escritos los subíndices del peso sináptico w_{kj} . El primer subíndice se refiere a la neurona en cuestión y el segundo subíndice se refiere al extremo de entrada de la sinapsis a la que se refiere el peso. A diferencia de una sinapsis en el cerebro, el peso de la sinapsis de una neurona artificial puede estar en un rango que incluye tantos valores negativos, así como también valores positivos.

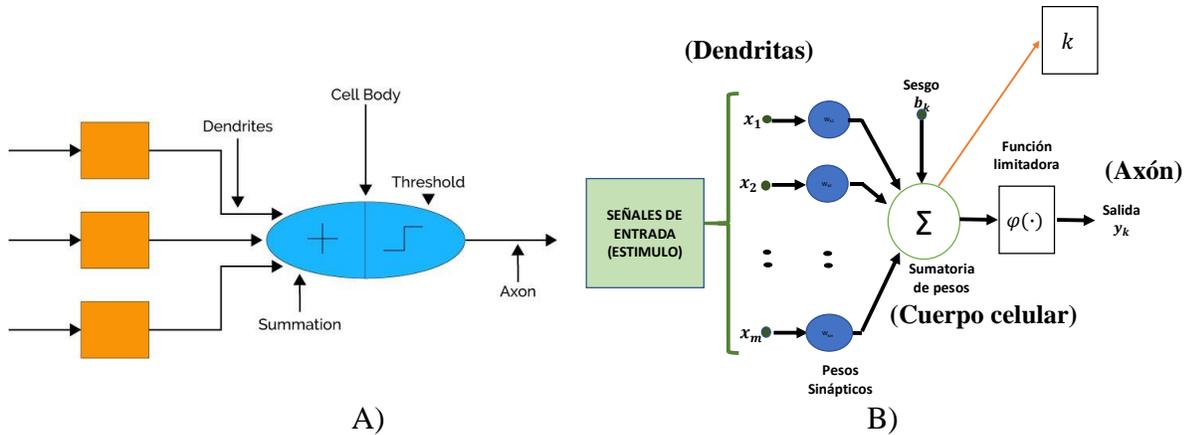
2.-Un operador aditivo para sumar las señales de entrada, ponderado por los respectivos pesos sinápticos de la neurona; las operaciones descritas aquí constituyen un combinador lineal.

3.-Una función de activación para limitar la amplitud de la salida de una neurona. La función de activación también se conoce como función de "aplastamiento" en la medida en que aplasta (limita) el rango de amplitud permisible de la señal de salida a algún valor finito.

Típicamente, el rango de amplitud normalizado de la salida de una neurona se escribe como el intervalo cerrado unitario $[0,1]$ o alternativamente $[-1,1]$.

El modelo neuronal de la **Fig. 16** también incluye un sesgo aplicado externamente, denotado por b_k . El sesgo b_k tiene el efecto de aumentar o disminuir la entrada neta de la función de activación, dependiendo de si es positivo o negativo, respectivamente.

Figura 14. Modelo no lineal de una neurona real y artificial. A) Neurona real. B) Neurona artificial.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 11

En términos matemáticos, podemos describir una neurona k escribiendo el siguiente par de ecuaciones:

$$u_k = \sum_{j=1}^m w_{kj} x_j \dots 3.1$$

$$y_k = \varphi(u_k + b_k) \dots 3.2$$

donde x_1, x_2, \dots, x_m son las señales de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ son los pesos sinápticos de la neurona k ; u_k es la salida del combinador lineal debido a las señales de entrada; b_k es el sesgo; $\varphi(\cdot)$ es la función de activación, y y_k es la señal de salida de la neurona. El uso del sesgo b_k tiene el efecto de aplicar una transformación afín a la salida u_k del combinador lineal en el modelo de la **Fig. 16**, como se muestra en:

$$v_k = u_k + b_k \dots 3.3$$

En particular, dependiendo de si el sesgo b_k es positivo o negativo, la relación entre el **campo local inducido** o el **potencial de activación** v_k de la neurona k y la salida del combinador lineal u_k se modifica de la manera ilustrada en la **Fig. 15**; en lo sucesivo, se usa el término "**campo local inducido**". Tengase en cuenta que, como resultado de esta transformación afín, la gráfica de v_k versus u_k ya no pasa por el origen.

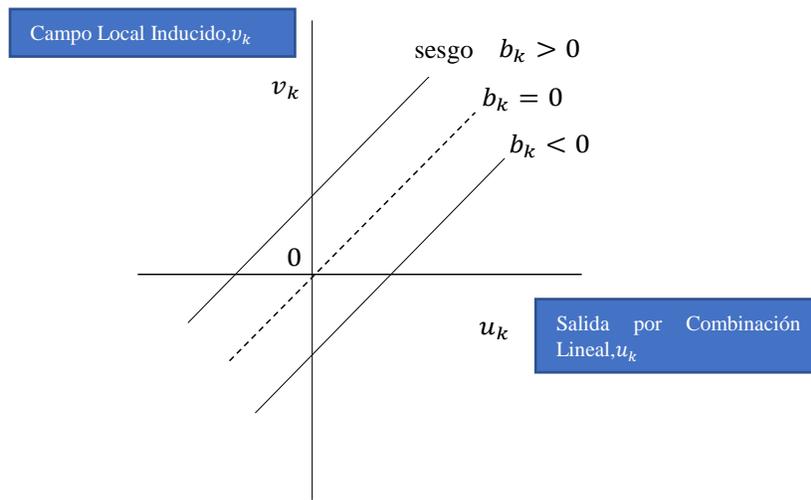
El sesgo b_k es un parámetro externo de la neurona artificial k . Se puede dar cuenta de su presencia en la Ec. 3.2. De manera equivalente, se puede formular la combinación de Ecs. 3.1 a 3.3 de la siguiente manera:

$$v_k = \sum_{j=0}^m w_{kj} x_j \dots \dots 3.4$$

y

$$y_k = \varphi(v_k) \dots \dots 3.5$$

Figura 15. Transformación llevada a cabo al incluir el sesgo en el diseño de la red neuronal. Nótese que $v_k = b_k$ en $u_k = 0$



Modificado de Haykin, S. (1999). *óp. cit.*, p. 12

En la Ec. 3.4 se ha agregado un nuevo peso sináptico. Su entrada es:

$$x_0 = +1 \dots 3.6$$

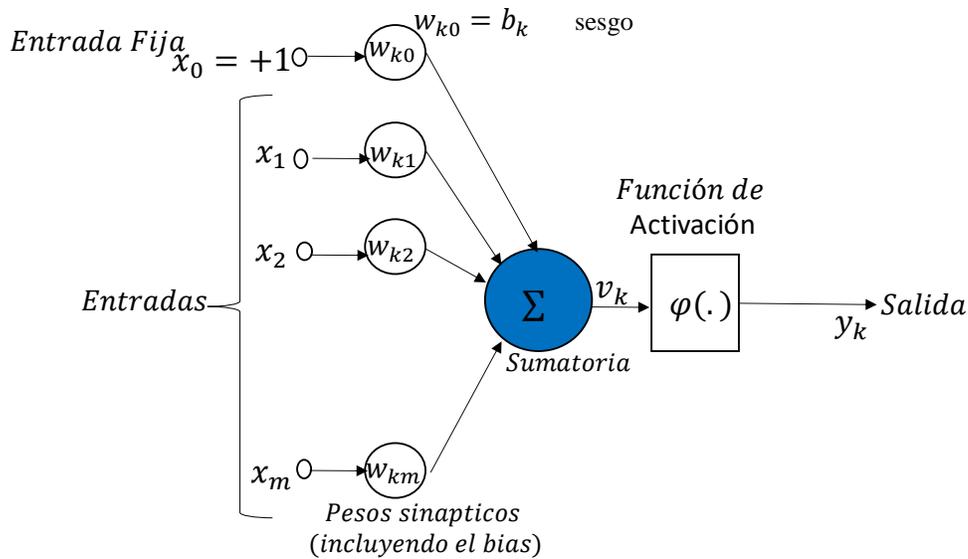
y su peso es:

$$w_{k0} = b_k \dots \dots 3.7$$

Por lo tanto, se puede reformular el modelo de neurona k como en la **Fig. 16**. En esta figura el efecto del sesgo se aplica haciendo dos cosas:

- (1) **agregar una nueva señal de entrada fija en +1 y,**
- (2) **agregar un nuevo peso sináptico igual al sesgo b_k .**

Figura 16. Modelo no lineal de una neurona.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 13

Aunque los modelos de las Figs. 14 y 16 son diferentes en apariencia, son matemáticamente equivalentes.

3.2.3 Tipos de funciones de activación.

La función de activación, denotado por $\varphi(v)$, define la salida de una neurona k en términos del campo local inducido, v_k . Aquí sólo se mostrarán 3 tipos básicos de funciones de activación (Haykin, S., 1999):

- 1- **Función de umbral.** Para este tipo de función de activación, descrita en la Fig. 17.a, tenemos:

$$\varphi(v) = \begin{cases} 1 & \text{si } v \geq 0 \\ 0 & \text{si } v < 0 \end{cases} \dots 3.8$$

En textos de ingeniería, este tipo de función es comúnmente referida como función Heaviside. Correspondientemente, la salida de la neurona k empleando esta función sería:

$$y_k = \begin{cases} 1 & \text{si } v_k \geq 0 \\ 0 & \text{si } v_k < 0 \end{cases} \dots 3.9$$

donde v_k es el campo local inducido de la neurona, esto es:

$$v_k = \sum_{j=1}^m w_{kj}x_j + b_k \dots 3.10$$

Tal neurona se menciona en la literatura como el modelo de **McCulloch-Pitts**, en reconocimiento del trabajo pionero realizado por **McCulloch y Pitts (1943)**. En este modelo, la salida de una neurona toma el valor de 1 si el campo local inducido de esa neurona es no negativo, y 0 en caso contrario. Esta declaración describe la propiedad de todo o nada del modelo de McCulloch-Pitts.

2- Función lineal por partes. Para la función lineal por partes descrita en la **Fig. 17 b**, tenemos,

$$\varphi(v) = \begin{cases} 1, & v \geq 0 \\ v, & 1/2 > v > -1/2 \dots\dots 3.11 \\ 0, & v \leq -1/2 \end{cases}$$

donde el factor de amplificación dentro de la región lineal de operación es la unidad. Esta forma de función de activación puede ser vista como una aproximación a un amplificador no lineal (**Haykin, S., 1999**). Las siguientes dos situaciones pueden verse como formas especiales de la función lineal por partes:

- Un combinador lineal surge si la región de operación se mantiene, sin llegar a valores de saturación.
- La función lineal por partes se reduce a una función de umbral si el factor de amplificación de la región lineal se hace infinitamente grande.

3-. Función Sigmoidal . La función sigmoidal, cuya gráfica figura una s, es por mucho la forma más común de función de activación utilizada en la construcción de redes neurales artificiales. Se define como una función estrictamente creciente que exhibe un equilibrio armónico entre el comportamiento lineal y el no lineal. Un ejemplo de la función sigmoidal es la función logística, definida por:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \dots\dots 3.12$$

donde **a** es el parámetro de la pendiente de la función sigmoideal. Al variar el parámetro a, se obtienen funciones sigmoidales de diferentes pendientes, como se ilustra en la **Fig. 17 c**. De hecho, la pendiente en el origen es igual a a / 4. En el límite, cuando el parámetro de la pendiente se aproxima al infinito, la función sigmoidal se convierte simplemente en una función de umbral (**Haykin, S., 1999**). Mientras que una función umbral asume el valor de 0 ó 1, una función sigmoidal asume un rango continuo de valores de 0 a 1. Téngase en cuenta también, que la función sigmoidal es diferenciable, mientras que la función umbral no lo es. (La diferenciable es una característica importante de la teoría de redes neuronales, como se describirá más adelante.)

Las funciones de activación definidas en las Ecs. 3.8, 3.11 y 3.12 presentan rango de valores de 0 a +1. A veces es deseable que el rango de valores en la salida de la función de activación sea de -1 a +1, en cuyo caso la función de activación asume una forma antisimétrica con

respecto a el origen; es decir, la función de activación es una función impar del campo local inducido. Específicamente, la función de umbral de la Ec. 3. 8 ahora se define como:

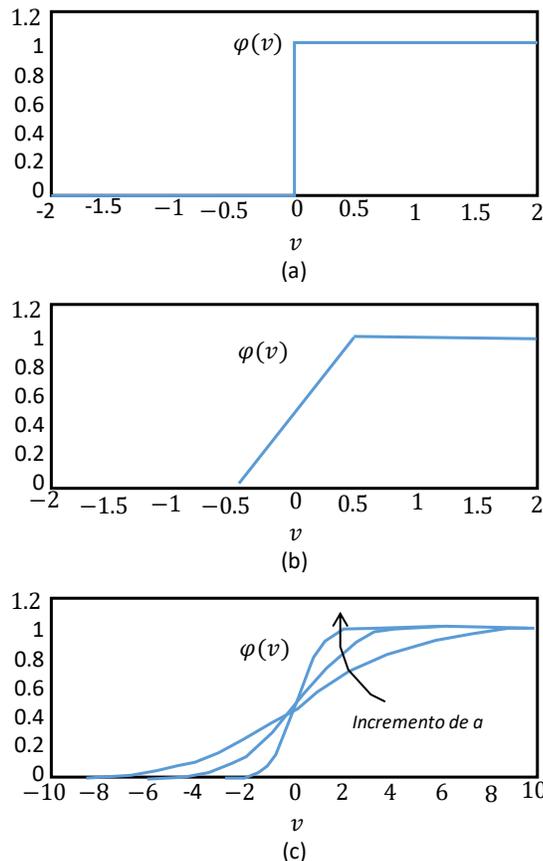
$$\varphi(v) = \begin{cases} 1 & \text{si } v > 0 \\ 0 & \text{si } v = 0 \\ -1 & \text{si } v < 0 \end{cases} \dots 3.13$$

la cual comúnmente es referida como **función signo**. Para la forma correspondiente de una función signo, se puede usar la función tangente hiperbólica, definida por:

$$\varphi(v) = \tanh(v).. 3.14$$

Permitiendo que una función de activación del tipo sigmoidal asuma valores negativos, como se muestra en la Ec. 3.14 tiene sus ventajas analíticas este hecho (como se mostrará más adelante)

Figura 17. A) *Función de umbral.* B) *Función lineal por partes.* C) *Función Sigmoidal, variando la pendiente a.*



Modificado de Haykin, S. (1999). óp. cit., p. 13

3.2.4 Diferentes representaciones de las redes neuronales.

El diagrama de bloques de la **Fig. 14** o el de la **Fig. 16** proporciona una descripción funcional de los diversos elementos que constituyen el modelo de una neurona artificial. Se puede

simplificar la apariencia del modelo utilizando la idea de **gráficos de flujo de señales** sin sacrificar ninguno de los detalles funcionales del modelo. **Los gráficos de flujo de señal** con un conjunto bien definido de reglas fueron desarrollados originalmente por **Mason (1953, 1956)** para redes lineales. La presencia de no linealidad en el modelo de una neurona limita el alcance de su aplicación a las redes neuronales. Sin embargo, los gráficos de flujo de señal proporcionan un método claro para la representación del flujo de señales en una red neuronal. Así mismo, es importante que el lector este familiarizado con cualquier de las formas de representación de las redes neuronales, puesto que en este trabajo se usarán todas éstas.

Un gráfico de flujo de señal es una red de enlaces dirigidos (ramas) que están interconectados en ciertos puntos llamados nodos (**Mason, 1953**). Un nodo típico j tiene una señal de nodo x_j . Un enlace dirigido típico se origina en el nodo j y termina en el nodo k ; tiene una función de transferencia asociada que especifica la manera en la cual la señal y_k en el nodo k depende de la señal x_j en el nodo j . El flujo de las señales en varias partes del gráfico esta dictado por 3 reglas básicas (**Mason, 1953, 1956**):

REGLA 1. Una señal fluye a lo largo de un enlace solamente en la dirección definida por la flecha del enlace.

Dos diferentes tipos de enlace pueden ser distinguidos:

- ✓ **Enlaces sinápticos**, cuyo comportamiento se rige por una relación de entrada-salida lineal. Específicamente, la señal del nodo x_j se multiplica por el peso sináptico w_{kj} para producir el nodo señal y_k , como se ilustra en la **Fig. 18a**.
- ✓ **Enlaces de activación**, cuyo comportamiento se rige en general por una relación de entrada-salida no lineal. Esta forma de relación se ilustra en la **Fig. 18b**, donde $\varphi(\cdot)$ es la función de activación no lineal.

REGLA 2. La señal de un nodo es igual a la suma algebraica de todas las señales entrando a dicho nodo a través de los enlaces de entrada.

Esta segunda regla es ilustrada en la **Fig. 18c** para el caso de la convergencia sináptica.

REGLA 3. La señal en un nodo se transmite a cada enlace saliente que proviene de ese nodo, siendo la transmisión completamente independiente de las funciones de transferencia de los enlaces de salida.

Esta tercera regla se ilustra en la **Fig. 18d** para el caso de divergencia sináptica o abanico abierto.

Por ejemplo, usando estas reglas se puede construir el gráfico de flujo de señal de la **Fig. 19** del modelo de una neurona, correspondiente al diagrama de bloques de la **Fig. 16**. La representación que se muestra en la **Fig. 19** es claramente más simple en apariencia que la de la **Fig. 16**, pero contiene todos los detalles funcionales representados en el último diagrama. Tenga en cuenta que en ambas figuras la entrada es $x_0 = +1$, junto con su peso sináptico asociado $w_{k0} = b_k$, donde b_k es el sesgo aplicado a la neurona k .

Figura 18. Esquema ilustrativo de las reglas básicas para la construcción de gráficos de flujo de señal.

$$x_j \circ \xrightarrow{w_{kj}} \circ y_k = w_{kj}x_j$$

(a)

$$x_j \circ \xrightarrow{\varphi(\cdot)} \circ y_k = \varphi(x_j)$$

(b)

(c)

(d)

Modificado de Haykin, S. (1999). *óp. cit.*, p. 16.

Un gráfico dirigido así definido, se dice que es un **gráfico completo** en el sentido de que describe no solo el flujo de la señal de neurona a neurona, sino también el flujo de la señal dentro de cada neurona. Sin embargo, cuando el foco de atención se restringe al flujo de la señal de neurona a neurona, podemos usar una forma reducida de este gráfico omitiendo los detalles del flujo de la señal dentro de las neuronas individuales. Tal gráfico dirigido se dice que está **parcialmente completo**. Se caracteriza de la siguiente manera (Mason, 1956):

1. Los nodos fuente proporcionan señales de entrada al gráfico.
2. Cada neurona está representada por un único nodo llamado **nodo de cálculo**.
3. Los enlaces de comunicación que interconectan la fuente y los nodos de cálculo del gráfico no tienen peso; simplemente proporcionan direcciones de flujo de señal en el gráfico.

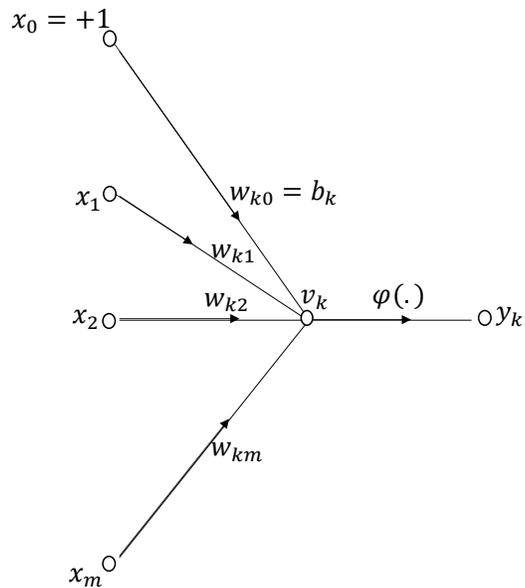
Un *gráfico parcialmente dirigido* definido de esta manera se conoce como un **gráfico arquitectónico**, y describe sólo el diseño de la red neuronal. En la **Fig. 20** se ilustra el caso simple de una sola neurona con m nodos fuente y un solo nodo con valor fijo $+1$ para el sesgo. Tenga en cuenta que el nodo de cálculo que representa la neurona se muestra sombreado, y el nodo fuente se muestra como un cuadrado pequeño.

En resumen, se manejarán en este trabajo tres representaciones gráficas de una red neuronal:

- **Diagrama de bloques**, que proporciona una descripción funcional de la red.
- **Gráfico de flujo de señal**, que proporciona una descripción completa del flujo de señal en la red.

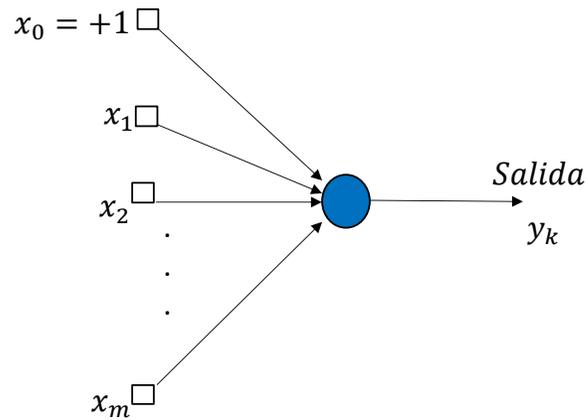
- **Gráfico arquitectónico**, que describe el diseño de la red.

Figura 19. Gráfico de señal de flujo del modelo de una neurona.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 18.

Figura 20. Gráfico arquitectónico de una neurona.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 17

3.2.5 Clasificación de redes neuronales.

Las redes neuronales pueden ser clasificadas de dos diferentes formas; dependiendo de su arquitectura, y en función de su aprendizaje (**Fig. 21**). En la **Fig. 22** se muestra la gran diversidad de tipo de redes que existen. Como se menciona en el resumen, se usó redes neuronales de flujo hacia delante, que como se puede observar en el diagrama de la **Fig.22** son uno de los tipos de redes más simples que hay. La razón por la cual se usó este tipo de redes es por limitación que trae consigo la poca cantidad de datos disponibles para la construcción del modelo. (Como se observará más adelante, el número de datos para el entrenamiento de las redes deberá ser siempre mayor al número de pesos sinápticos para que se tenga un sistema subestimado, así, el uso de redes más complejas requerirá de mayor número de datos para el entrenamiento.)

3.2.5.1 Tipos de red de acuerdo con su arquitectura.

La manera en que se estructuran las neuronas de una red neuronal está íntimamente relacionada con el algoritmo de aprendizaje utilizado para entrenar a la red.

En general, se puede identificar fundamentalmente tres clases de arquitecturas en las redes, que a continuación se describen:

Figura 21. Clasificación de las redes neuronales.

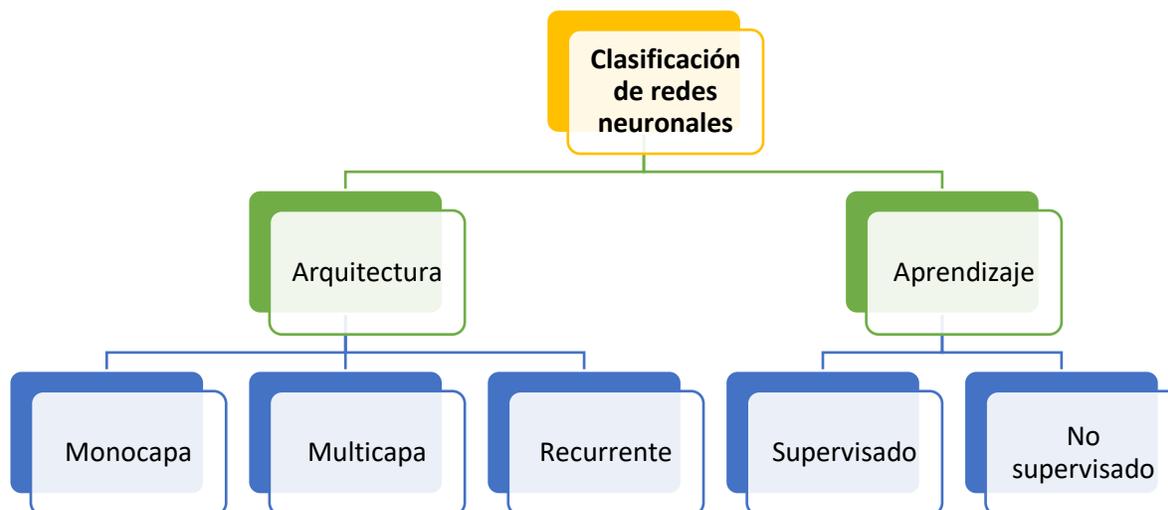
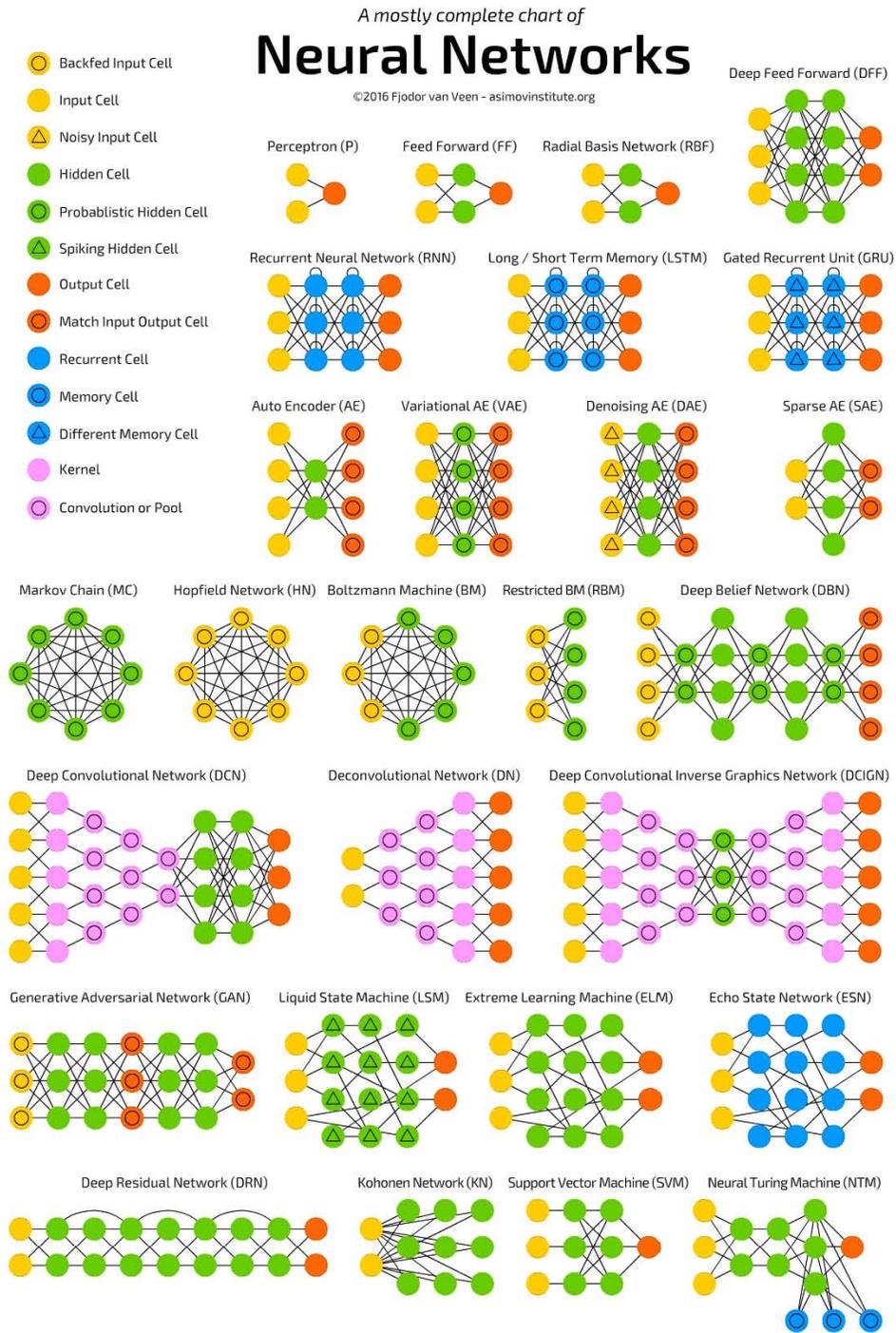


Figura 22. Diagrama que muestra los diferentes tipos de redes neuronales que existe.

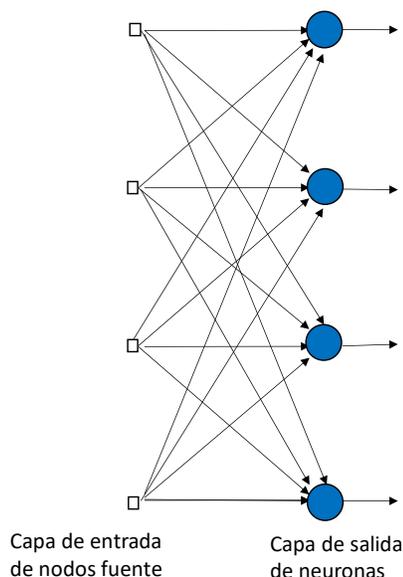


Recuperado de <http://www.asimovinstitute.org/neural-network-zoo/>

1. Redes monocapa de flujo hacia delante

En una red neuronal en capas, las neuronas están organizadas en forma de capas. En la forma más simple de una red en capas, se tiene una capa de entrada de nodos fuente que se proyecta en una capa de salida de neuronas (nodos de cálculo), pero no viceversa. En otras palabras, esta red es estrictamente del tipo de flujo hacia delante o acíclico. Se ilustra en la **Fig. 23** para el caso de cuatro nodos tanto en la capa de entrada como en la de salida. Dicha red se denomina red de una sola capa, con la designación "monocapa" que hace referencia a la capa de salida de los nodos de cálculo (neuronas). No se cuenta la capa de entrada de los nodos fuente porque no se realiza ningún cálculo allí.

Figura 23. Red de flujo hacia delante con una sola capa de neuronas.

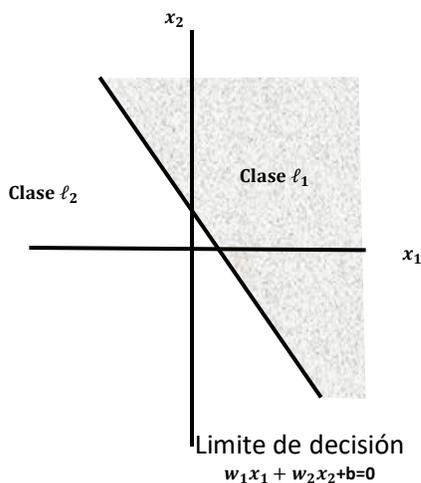


Modificado de Haykin, S. (1999). óp. cit., p. 21.

El modelo neuronal monocapa más reconocido, es la máquina de aprendizaje creada por **Rosenblatt**, la cual la llamó **perceptrón**. El perceptrón es la forma más simple de una red neuronal utilizada para la clasificación de patrones que se dice son linealmente separables (**Hagan et. al, 2004**) (es decir, patrones que se encuentran en lados opuestos de un hiperplano (**Fig. 24**)). El perceptrón se construye alrededor de una neurona no lineal, es decir, el modelo de una **neurona de McCulloch-Pitts**, como se recordará dicho modelo neuronal consiste en un combinador lineal seguido de un limitador (que realiza la función signo), como se representa en la **Fig. 25**. El algoritmo utilizado para ajustar los parámetros libres de esta red neuronal apareció por primera vez en un procedimiento de aprendizaje desarrollado por **Rosenblatt (1958, 1962)** para su modelo de cerebro perceptrónico. De hecho, Rosenblatt demostró que si los patrones (vectores) utilizados para entrenar el perceptrón se dibujan a partir de dos clases linealmente separables, el **algoritmo perceptrón** converge y posiciona la superficie de decisión en forma de un hiperplano entre las dos clases. La prueba de la

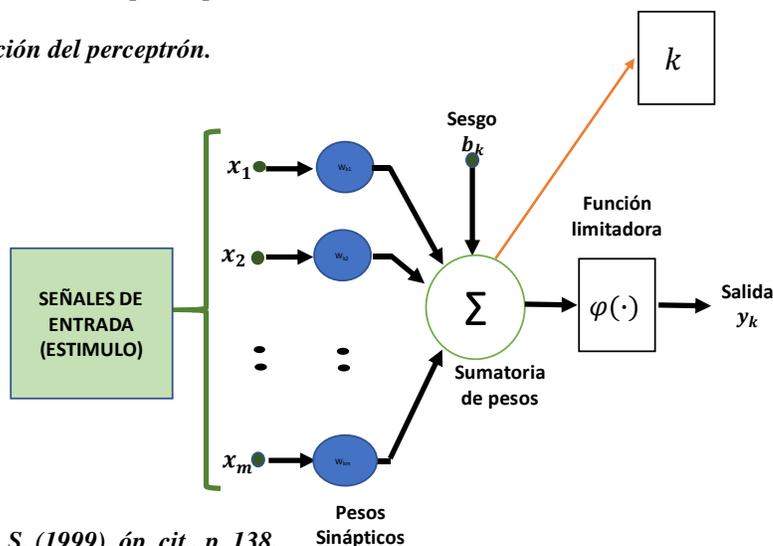
convergencia del algoritmo se conoce como *el teorema de convergencia del perceptrón* (Haykin, S., 1999. pp 137-142). El perceptrón construido alrededor de una sola neurona se limita a realizar una clasificación de patrones con solo dos clases (hipótesis). Al expandir la capa de salida (capa de cálculo) del perceptrón para incluir más de una neurona, podemos formar una clasificación correspondiente con más de dos clases. Sin embargo, las clases tienen que ser linealmente separables para que el perceptrón funcione correctamente. El punto importante aquí es que, en lo que se refiere a la teoría básica del perceptrón como clasificador de patrones, solo se debe considerar el caso de una sola neurona. La extensión de la teoría al caso de más de una neurona es trivial.

Figura 24. Ilustración del hiperplano (en este ejemplo, la línea recta) como límite de decisión para un problema de clasificación de patrones en dos diferentes clases.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 137.

Figura 25. Representación del perceptrón.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 138.

2. Redes multicapa de flujo hacia delante.

La segunda clase de una red neuronal de flujo hacia delante se distingue por la presencia de una o más capas ocultas, cuyos nodos de cálculo se denominan neuronas ocultas o unidades ocultas. La función de las neuronas ocultas es intervenir entre la entrada externa y la salida de la red en alguna manera útil. Agregando uno o más capas ocultas, la red está habilitada para extraer estadísticas de orden superior (**Churchland, 1989**). En un sentido bastante amplio, la red adquiere una perspectiva global a pesar de su conectividad local debido al conjunto adicional de conexiones sinápticas y la dimensión adicional de las interacciones neuronales (**Churchland y Sejnowski, 1988**). La capacidad de las neuronas ocultas para extraer estadísticas de orden superior es particularmente valiosa cuando el tamaño de la capa de entrada es grande (**Haykin, S., 1999**).

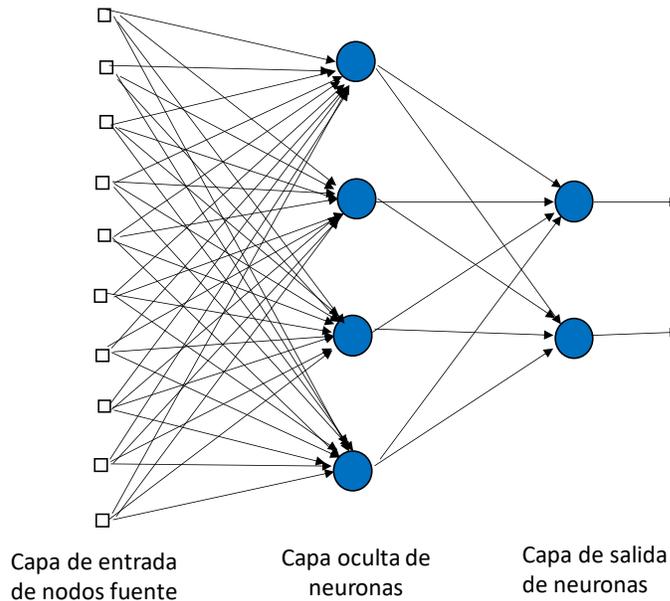
Los nodos fuente en la capa de entrada de la red suministran elementos respectivos del patrón de activación (vector de entrada), que constituyen las señales de entrada aplicadas a las neuronas (nodos de cálculo) en la segunda capa (es decir, la primera capa oculta).

Las señales de salida de la segunda capa se utilizan como entradas a la tercera capa, y así sucesivamente para el resto de la red. Normalmente, las neuronas en cada capa de la red tienen como entradas las señales de salida de la capa anterior solamente. El conjunto de señales de salida de las neuronas en la capa de salida (final) de la red constituye la respuesta global de la red al patrón de activación suministrado por los nodos fuente en la capa de entrada (primera). El gráfico arquitectónico de la **Fig. 26** ilustra el diseño de una red neuronal de flujo hacia delante multicapa para el caso de una sola capa oculta. Para abreviar, la red en la **Fig. 26** se conoce como una red 10-4-2 porque tiene 10 nodos fuente, 4 neuronas ocultas y 2 neuronas de salida. Como otro ejemplo, una red de flujo hacia delante con m nodos fuente, h_1 neuronas en la primera capa oculta, h_2 neuronas en la segunda capa oculta y q neuronas en la capa de salida se conoce como una red $m-h_1-h_2-q$.

Se dice que la red neuronal en la **Fig. 26** está *completamente conectada* en el sentido de que cada nodo en cada capa de la red está conectado a cada otro nodo de la capa siguiente adyacente. Sin embargo, si faltan algunos de los enlaces de comunicación (conexiones sinápticas) de la red, decimos que la red está *parcialmente conectada*.

Este tipo de redes es el que se usará para la construcción de las secciones de propiedades, así que se retomarán en el siguiente capítulo de una manera más extendida.

Figura 26. Red de flujo hacia delante completamente conectada con una capa oculta y una capa de salida.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 29.

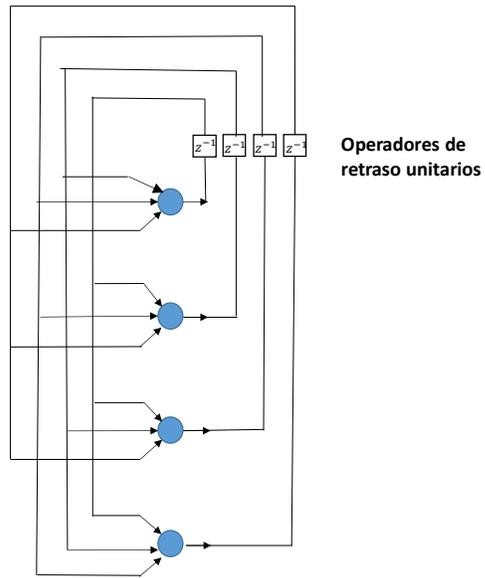
3. Redes recurrentes.

Una **red neuronal recurrente** se distingue de una red neuronal de flujo hacia delante porque tiene al menos un bucle de retroalimentación. Por ejemplo, una red recurrente puede consistir en una sola capa de neuronas con cada neurona alimentando su señal de salida a las entradas de todas las otras neuronas, como se ilustra en el gráfico arquitectónico de la **Fig. 27**. En la estructura representada en esta figura, no hay bucles de auto retroalimentación en la red; el término de auto alimentación se refiere a una situación en la que la salida de una neurona se retroalimenta a su propia entrada.

La red recurrente ilustrada en la **Fig. 27** tampoco tiene neuronas ocultas. En la **Fig. 28** se ilustra otra clase de redes recurrentes con neuronas ocultas. Las conexiones de retroalimentación que se muestran en la **Fig. 28** se originan tanto en las neuronas ocultas como en las neuronas de salida.

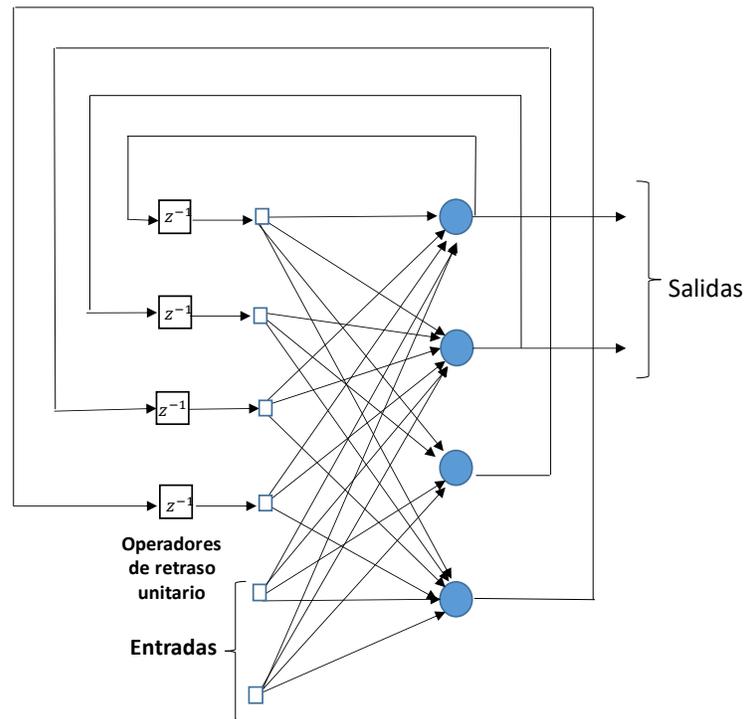
La presencia de bucles de retroalimentación ya sea en la estructura recurrente de la **Fig. 27** o en la de la **Fig. 28** tiene un profundo impacto en la capacidad de aprendizaje de la red y en su rendimiento (**Haykin, S., 1999**). Además, los bucles de retroalimentación implican el uso de ramas particulares compuestas de elementos de retardo unitarios (denotados por Z^{-1}), que dan como resultado un comportamiento dinámico no lineal, suponiendo que la red neuronal contiene unidades no lineales.

Figura 27. Red recurrente sin loops de retroalimentación, ni con neuronas ocultas.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 23.

Figura 28. Red recurrente con neuronas ocultas.



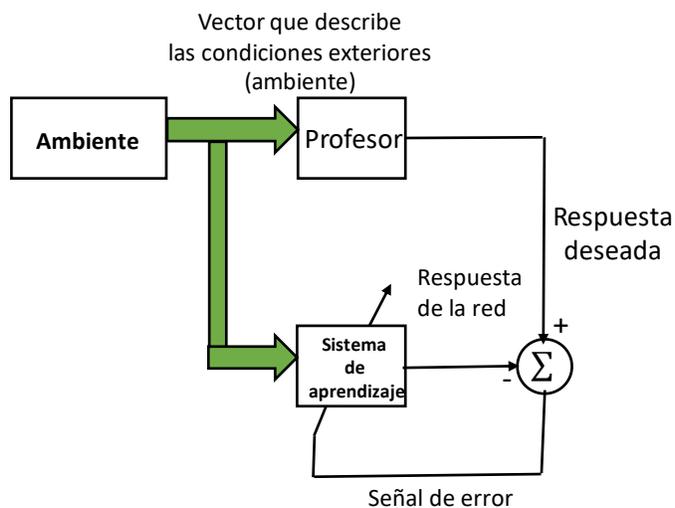
Modificado de Haykin, S. (1999). *óp. cit.*, p. 24.

3.2.5.2 Tipos de red de acuerdo con el aprendizaje.

Aprendizaje Supervisado

Comencemos considerando el aprendizaje con un guía que también se conoce como *aprendizaje supervisado*. La **Fig. 29** muestra un diagrama de bloques que ilustra esta forma de aprendizaje. En términos conceptuales, se puede pensar que el guía tiene conocimiento del entorno, y que ese conocimiento está representado por un conjunto de ejemplos de entrada y salida. Sin embargo, el entorno es desconocido para la red neuronal de interés. Supóngase ahora, que el guía y la red neuronal están expuestos a un vector de entrenamiento (es decir, un ejemplo) extraído del entorno. En virtud del conocimiento incorporado, el guía puede proporcionar a la red neuronal una respuesta deseada para ese vector de entrenamiento. De hecho, la respuesta deseada representa la acción óptima que debe realizar la red neuronal. Los parámetros de la red se ajustan bajo la influencia combinada del vector de entrenamiento y la señal de error. La *señal de error* se define como la diferencia entre la respuesta deseada y la respuesta real de la red. Este ajuste se lleva a cabo de forma iterativa paso a paso con el objetivo de finalmente hacer que la red neuronal emule al guía; se presume que la emulación es óptima en algún conjunto estadístico (**Haykin, S., 1999**). De esta forma, el conocimiento del entorno disponible para el guía se transfiere a la red neuronal a través de la información lo más completa posible. Cuando se alcanza esta condición, podemos prescindir del guía y dejar que la red neuronal se ocupe completamente del entorno por sí misma. Este el tipo de aprendizaje empleado por las redes neuronales usadas en este trabajo.

Figura 29. Diagrama de bloque del aprendizaje supervisado.



Modificado de Haykin, S. (1999). óp. cit., p. .63

Aprendizaje no supervisado.

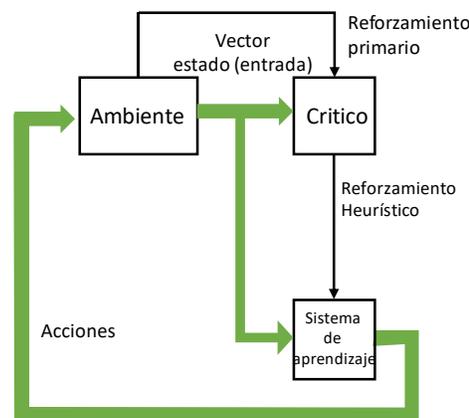
En el aprendizaje supervisado, el proceso de aprendizaje se lleva a cabo bajo la tutela de un guía. Sin embargo, en el paradigma conocido como aprendizaje sin guía, como su nombre lo indica, no hay un “maestro” que supervise el proceso de aprendizaje. Es decir, no hay

ejemplos etiquetados de la función que la red deba aprender. Bajo este segundo paradigma, se identifican dos subdivisiones:

1. Aprendizaje por refuerzo / Programación neurodinámica.

En el aprendizaje por refuerzo, el aprendizaje de un mapeo de entrada-salida se realiza a través de una interacción continua con el entorno para minimizar un índice de rendimiento escalar. La **Fig. 30** muestra el diagrama de bloques de una forma de un sistema de aprendizaje de refuerzo construido alrededor de un crítico que convierte una señal de refuerzo primaria recibida del ambiente en una señal de refuerzo de mayor calidad llamada señal de refuerzo heurístico, ambas son entradas escalares (**Barto et al., 1983**).

Figura 30. Diagrama de bloque del aprendizaje reforzado.



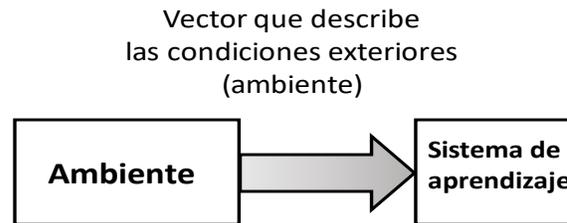
Modificado de Haykin, S. (1999). óp. cit., p. 63.

2. Aprendizaje no supervisado.

En el aprendizaje no supervisado o autoorganizado, no existe un profesor o crítico externo para supervisar el proceso de aprendizaje, como se indica en la **Fig. 31**. Más bien, se prevén medidas para una medida dependiente de la calidad de la representación que la red debe aprender, y los parámetros libres de la red están optimizados con respecto a esa medida. Una vez que la red se ha sintonizado a las regularidades estadísticas de los datos de entrada, desarrolla la capacidad de formar representaciones internas para las funciones de codificación de la entrada y, de ese modo, crear nuevas clases automáticamente (**Becker, 1992**).

Para realizar un aprendizaje sin supervisión, podemos usar una regla de aprendizaje competitivo. Por ejemplo, podemos usar una red neuronal que consta de dos capas: una capa de entrada y una capa competitiva. La capa de entrada recibe los datos disponibles. La capa competitiva consiste en neuronas que compiten entre sí (de acuerdo con una regla de aprendizaje) por la "oportunidad" de responder a las características contenidas en los datos de entrada. En su forma más simple, la red opera de acuerdo con una estrategia de "el ganador se lo lleva todo".

Figura 31. Diagrama de bloque del aprendizaje no supervisado.



Modificado de Haykin, S. (1999). *óp. cit.*, p. 65.

3.2.6 Beneficios de las redes neuronales.

La aparente superioridad del uso de las redes neuronales sobre otros métodos de solución de problemas no lineales, se atribuyen a dos características que presentan éstas:

- 1- **Su estructura paralelamente distribuida.**
- 2- **Su habilidad de aprender y por ende de generalizar.**

Su habilidad de generalización se refiere a la capacidad que tienen éstas de deducir, es decir, de producir soluciones, salidas, razonables para entradas, problemas, no encontrados durante el entrenamiento (aprendizaje). Estas dos capacidades de procesamiento de información, hacen posible para las redes neuronales resolver problemas complejos a gran escala, que hace pocos años eran imposible pensar su tratamiento. No obstante, es importante mencionar que las redes neuronales no pueden proporcionar una solución trabajando individualmente, sino que se requiere que estas estén integradas en un enfoque sistemático ingenieril. Es decir, un problema complejo es resuelto, descomponiéndolo en un número de tareas relativamente simple, de las cuales un set de éstas son asignadas a redes neuronales, de acuerdo a sus inherentes capacidades. Lo anterior, deja entrever que todavía falta mucho para poder construir una herramienta tan eficiente como el cerebro humano.

El uso de redes neuronales ofrecen las siguientes propiedades y capacidades útiles (**Haykin, S., 1999**):

- ✚ **No linealidad:** Una neurona artificial puede ser lineal o no lineal. Una red neuronal, constituida por una interconexión de neuronas no lineales, es por si misma no lineal. La no linealidad es una propiedad muy importante, sobre todo si el mecanismo físico responsable para la generación de la señal de entrada (por ejemplo la señal de la voz) es inherentemente no lineal.
- ✚ **Mapeado de entradas-salidas.** Como se mencionó anteriormente, la red neuronal aprende a través de un proceso de entrenamiento, el cual consiste en proporcionarle varios ejemplos, el cual cada uno consiste en una única señal de entrada y su correspondiente señal de respuesta deseada, e ir modificando los pesos sinápticos para minimizar la diferencia entre la respuesta deseada y la respuesta producida por la red.

Este entrenamiento se repite para muchos ejemplos hasta que la red alcance un estado estable donde no haya cambios significativos en los pesos. Los ejemplos de entrenamiento aplicados anteriormente se pueden volver aplicar pero en diferente orden durante la sesión de entrenamiento. Es así como la red aprende a partir de los ejemplos, mediante la construcción de un mapa de entradas-salidas del problema. Dicho enfoque trae a la mente el estudio de la **inferencia no estadística**, la cual es una rama de la estadística que trata de la estimación de modelos libres; el término **no paramétrico** se emplea para resaltar el hecho de que no se requieren asunciones previas a un modelo estadístico para los datos de entrada (**Haykin, S., 1999**).

- ✚ **Adaptabilidad.** Las redes neuronales tienen la capacidad de adaptar sus pesos sinápticos a los cambios de su entorno. En particular, una red neuronal entrenada para operar en un ambiente específico puede ser fácilmente reentrenada para tratar con cambios menores en las condiciones de operación del entorno. Además, cuando se opera en un entorno no estacionario (es decir, uno donde estadísticamente existen cambios con respecto al tiempo), una red neuronal puede ser diseñada para que cambie sus pesos sinápticos en tiempo real. La arquitectura natural de una red neuronal para la clasificación de patrones, procesamiento de señales, y aplicaciones de control, se acopla con la capacidad adaptativa de la red, haciéndola una herramienta muy útil para este tipo de tareas. Como regla general, se puede decir que entre más adaptativo se haga al sistema, asegurando que el sistema permanezca estable, más robusto probablemente su funcionamiento será, cuando el sistema sea requerido bajo condiciones no estacionarias. Es importante decir, que la adaptabilidad no asegura la robustez del sistema, de hecho, generalmente estos dos aspectos son inversos.

- ✚ **Tolerancia a fallas.** Una red neuronal tiene el potencial de ser inherentemente tolerante a fallas, dado que su rendimiento decrece de manera gradual bajo condiciones de operación adversas. Esto gracias a que la información se encuentra almacenada de manera distribuida en la red. Para asegurar que la red neuronal en verdad sea tolerante a fallas, podría ser necesario tomar medidas correctivas durante el diseño del algoritmo usado para el entrenamiento de la red (**Kerlirzin y Vallet, 1993**).

- ✚ **Uniformidad de análisis y diseño.** Las redes neuronales son conocidas de manera universal como procesadores de información. En ese aspecto se puede decir que la misma anotación es usada en todos los dominios donde se aplican las redes neuronales. Esta característica se manifiesta por sí misma, en diferentes formas:
 - ✓ **Las neuronas en una u otra forma, representan el ingrediente común en todas las redes neuronales.**

- ✓ Este hecho común hace posible que se puedan compartir las teorías y los algoritmos de aprendizaje en las diferentes aplicaciones donde se usan las redes neuronales.

3.2.7 Algoritmos de aprendizaje.

La propiedad que es de primordial importancia para una red neuronal es la capacidad de la red para aprender de su entorno, y la de mejorar su rendimiento a través del aprendizaje. La mejora en el rendimiento se lleva a cabo a lo largo del tiempo de acuerdo con alguna medida prescrita. Una red neuronal aprende sobre su entorno a través de un proceso interactivo de ajustes aplicados a sus pesos sinápticos y niveles de sesgo (Haykin, S., 1999). Idealmente, la red se vuelve más conocedora de su entorno después de cada iteración del proceso de aprendizaje.

Hay demasiadas actividades asociadas con el concepto de "aprendizaje" para justificar su definición de manera precisa. Además, el proceso de aprendizaje es una cuestión de punto de vista, lo que hace que sea aún más difícil ponerse de acuerdo sobre una definición precisa del término. Por ejemplo, el concepto de aprendizaje desde el punto de vista de un psicólogo es bastante diferente al proceso de aprendizaje en el sentido del aula. Reconociendo que el interés particular del presente documento está en las redes neuronales, utilizamos una definición de aprendizaje adaptada de Mendel y McClaren (1970).

Se define el aprendizaje en el contexto de las redes neuronales como:

“El aprendizaje es un proceso mediante el cual los parámetros libres de una red neuronal se adaptan a través de un proceso de estimulación por el entorno en el que está inserta la red. El tipo de aprendizaje está determinado por la manera en que se producen los cambios de los parámetros”.

Esta definición del proceso de aprendizaje implica la siguiente secuencia de eventos:

1. La red neuronal es estimulada por un entorno.
2. La red neuronal sufre cambios en sus parámetros libres como resultado de esta estimulación.
3. La red neuronal responde de una nueva manera al entorno debido a los cambios que se han producido en su estructura interna.

Un conjunto prescrito de reglas bien definidas para la solución de un problema de aprendizaje se denomina *algoritmo de aprendizaje*¹. Como es de esperar, no existe un algoritmo de aprendizaje único para el diseño de redes neuronales. Más bien, se tiene un "kit de herramientas" representado por una variedad diversa de algoritmos de aprendizaje, cada uno de los cuales ofrece ventajas propias. Básicamente, los algoritmos de aprendizaje difieren

¹ Haykin, S., 1999. pp 51-61

entre sí en la forma en que se formula el ajuste al peso sináptico de una neurona, lo cual dependerá del método de optimización (algunos de estos serán abordados más adelante de manera breve) a usar para minimizar la **función de costo** (error medio de la red, ϵ_{prom}). Otro factor a considerar es la forma en que una red neuronal (máquina de aprendizaje) compuesta por un conjunto de neuronas interconectadas, se relaciona con su entorno. En este último contexto, se habla de un paradigma de aprendizaje que se refiere a un modelo del entorno en el que la red neuronal opera.

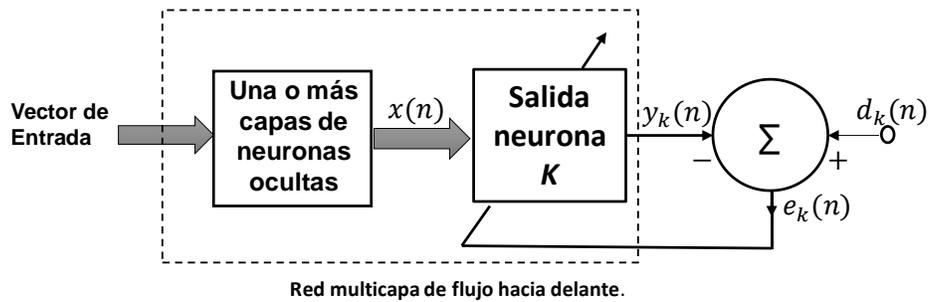
Por mencionar solo algunos algoritmos de aprendizajes, está el algoritmo de Hebb, el cual es la más famosa y antigua de todas las reglas de aprendizaje, puesto que constituye la primera regla para el aprendizaje auto-organizado. La manera en la que se ajustan los pesos aquí es emulando el aprendizaje asociativo en el cerebro humano, es decir, considera ambos lados de la sinapsis. O bien el aprendizaje competitivo, el cual como su nombre hace alusión, las neuronas de salida de una red neuronal compiten entre sí para activarse mientras que en una red neuronal basada en el aprendizaje de Hebbian, varias neuronas de salida pueden estar activas simultáneamente, en el aprendizaje competitivo, solo una neurona de salida está activa en un momento dado. Esta característica hace que el aprendizaje competitivo sea muy adecuado para descubrir características estadísticamente destacadas que pueden usarse para clasificar un conjunto de patrones de entradas. Este tipo de aprendizaje es muy usado en la clasificación de patrones en redes no supervisadas, un ejemplo útil de éste, *podría ser la clasificación de facies en un yacimiento*.

No obstante, dado el tipo de red que se utilizará para lograr los objetivos de este trabajo, redes supervisadas multicapa de flujo hacia delante, se empleará *el algoritmo de aprendizaje por corrección de error*. A continuación, se detalla éste.

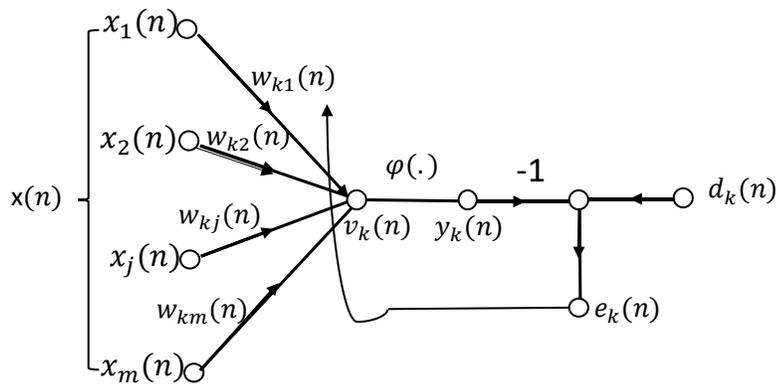
Para ilustrar esta regla de aprendizaje, considérese el caso simple de una neurona k que constituye el único nodo de cálculo en la capa de salida de una red neuronal de flujo hacia delante como se representa en la **Fig. 32a**. La neurona k está alimentada por un vector de señal $x(n)$ producido por una o más capas de neuronas ocultas, que a su vez son alimentadas por un vector de entrada (estímulo) aplicado a los nodos fuente (es decir, capa de entrada) de la red neuronal. El argumento n denota un tiempo discreto, o más precisamente, el paso de tiempo de un proceso iterativo involucrado en el ajuste de los pesos sinápticos de la neurona k . La señal de salida de la neurona k se denota por $y_k(n)$. Esta señal de salida, que representa la única salida de la red neuronal, se compara con una respuesta deseada o salida objetivo, denotada por $d_k(n)$. En consecuencia, se produce una señal de error, indicada por $e_k(n)$. Por definición, así tenemos;

$$e_k(n) = d_k(n) - y_k(n) \dots 3.15$$

Figura 32. Ilustración del aprendizaje por corrección de error.



(a) Diagrama de bloque de una red neuronal, resaltando la única neurona de la capa de salida.



(b) Gráfico de flujo de señal de una neurona de salida.

Modificado de Haykin, S. (1999). *óp. cit.*, p. 52.

La señal de error $e_k(n)$ activa un mecanismo de control, cuyo objetivo es aplicar una secuencia de ajustes correctivos a los pesos sinápticos de la neurona k . Los ajustes correctivos están diseñados para hacer que la señal de salida $y_k(n)$ se acerque a la respuesta deseada $d_k(n)$ en una forma progresiva, paso a paso. Este objetivo es alcanzado a través de la minimización de una **función de costo** o **índice de rendimiento**, $\xi(n)$, definida en términos de $e_k(n)$ como:

$$\xi(n) = \frac{1}{2} e_k^2(n) \dots 3.16$$

Esto es, $\xi(n)$ es el *valor instantáneo de la energía de error*. Los ajustes paso a paso de los pesos sinápticos de la neurona k continúan hasta que el sistema alcanza un estado estable (es decir, los pesos sinápticos están esencialmente estabilizados). En ese punto, el proceso de aprendizaje finaliza.

El proceso de aprendizaje descrito previamente obviamente se denomina **aprendizaje por corrección de error**. En particular, la minimización de la función de costo $\xi(n)$, en la siguiente sección se describirán varias técnicas de optimización, conduce a un dominio de aprendizaje comúnmente denominado **regla delta** o **regla Widrow-Hoff** nombrado en honor de sus creadores (**Widrow y Hoff. 1960**). Sea que w_{kj} denote el valor del peso sináptico w_{kj} de la neurona k excitado por el elemento $x_j(n)$ del vector señal $x(n)$ en el paso de tiempo n .

De acuerdo con la regla delta, el ajuste $\Delta w_{kj}(n)$ aplicado al peso sináptico w_{kj} en el paso de tiempo n está definido por:

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) \dots 3.17$$

donde η es una constante positiva que determina la *rapidez de aprendizaje* a medida que avanzamos de un paso en el proceso de aprendizaje a otro. Por lo tanto, es natural que nos refiramos a η como el *parámetro de rapidez de aprendizaje*. En otras palabras, la regla delta puede establecerse como:

El ajuste hecho a los pesos sinápticos de una neurona es proporcional al producto de la señal de error por la señal de entrada de la sinapsis en cuestión.

Tenga en cuenta que la regla delta, como se indica aquí, supone que la señal de error se puede medir directamente. Para que esta medición sea factible, claramente se necesita un suministro de la respuesta deseada de alguna fuente externa, que sea directamente accesible para la neurona k . En otras palabras, la neurona k es visible para el mundo exterior, como se muestra en la **Fig. 32a**, es por ello por lo que esta regla delta es la que se aplica en el **perceptrón**. De esta figura también se observa que el aprendizaje de corrección de errores es de hecho de naturaleza local. Esto simplemente dice que los ajustes sinápticos hechos por la regla delta se localizan alrededor de la neurona k .

Habiendo calculado el ajuste sináptico, $\Delta w_{kj}(n)$, el valor actualizado del peso sináptico w_{kj} está determinado por:

$$w_{kj}(n + 1) = w_{kj}(n) + \Delta w_{kj}(n) \dots 3.18$$

En efecto, $w_{kj}(n)$ y $w_{kj}(n+1)$ pueden ser vistos como el valor pasado y actual del peso sináptico, $w_{kj}(n)$, respectivamente. En términos computacionales, se puede escribir esto de la siguiente manera también,

$$w_{kj}(n) = z^{-1}[w_{kj}(n + 1)] \dots 3.19$$

donde z^{-1} es el operador de retardo unitario. Es decir, z^{-1} representa un elemento de almacenamiento.

La **Fig. 32b** muestra una representación gráfica del proceso de aprendizaje por corrección de error, centrándose en la actividad que rodea a la neurona k . La señal de entrada x_j y el campo local inducido v_k de la neurona k son conocidas como señales presináptica y postsináptica de la sinapsis j -ésima de la neurona k , respectivamente. De la **Fig. 32b** vemos que el aprendizaje por corrección de error es un ejemplo de un sistema de retroalimentación de bucle cerrado. De la teoría de control se sabe que la estabilidad de tal sistema está determinada por aquellos parámetros que constituyen los bucles de retroalimentación del sistema. En nuestro caso, se tiene un solo bucle de retroalimentación, y uno de esos parámetros de particular interés es el parámetro de rapidez de aprendizaje η . Por lo tanto, es importante que η se seleccione cuidadosamente para garantizar que se logre la estabilidad o la convergencia del

proceso de aprendizaje iterativo. La elección de η también tiene una profunda influencia en la precisión y otros aspectos del proceso de aprendizaje. En resumen, el parámetro de velocidad de aprendizaje η desempeña un papel clave en la determinación del rendimiento del aprendizaje por corrección de error en la práctica.

La manera en que se usa la señal de error $e(i)$ para controlar los ajustes a los pesos sinápticos de la neurona se determina por la función de costo utilizada. Este problema está estrechamente relacionado con el de la optimización. Por lo tanto, se presenta una revisión breve de los métodos de optimización sin restricciones en el **apéndice A**.

3.2.7 Tareas de aprendizaje.

En secciones anteriores de este capítulo, se han discutido diferentes algoritmos de aprendizaje y paradigmas de aprendizaje. En esta sección, se describen algunas tareas básicas de aprendizaje que pueden ser llevadas a cabo por las redes neuronales. La elección de un algoritmo de aprendizaje particular está influenciada por la tarea de aprendizaje que una red neuronal debe realizar. En este contexto, se ha identificado tres tareas de aprendizaje que se aplican al uso de redes neuronales de una forma u otra y que tienen varias aplicaciones en el campo de las geociencias; **Asociación y reconocimiento de patrones y aproximación de funciones**.

Asociación de Patrones.

Una memoria asociativa es una memoria distribuida parecida a un cerebro que aprende por asociación. Se sabe que la asociación es una característica prominente de la memoria humana desde Aristóteles, y todos los modelos de cognición usan la asociación de una forma u otra como la operación básica (**Anderson, 1995**).

La asociación toma una de dos formas: **autoasociación** o **heteroasociación** (**Anderson, 1995**). En la *autoasociación* se requiere una red neuronal para almacenar un conjunto de patrones (vectores) presentándolos repetidamente en la red. Subsecuentemente, se presenta a la red una descripción parcial o una versión distorsionada (ruidosa) de un patrón original almacenado en ella, y la tarea es recuperar ese patrón en particular. La heteroasociación difiere de la autoasociación en que un conjunto arbitrario de patrones de entrada se combina con otro conjunto arbitrario de patrones de salida. La autoasociación implica el uso de aprendizaje no supervisado, mientras que el tipo de aprendizaje involucrado en la heteroasociación es supervisado.

El número q de patrones almacenados en una memoria asociativa proporciona una medida directa de la capacidad de almacenamiento de la red. Al diseñar una memoria asociativa, el desafío es hacer que la capacidad de almacenamiento q (expresada como un porcentaje del número total N de neuronas utilizadas para construir la red) sea lo más grande posible e insistir en que una gran fracción de los patrones memorizados se recupere correctamente.

Reconocimiento de patrones.

Los humanos somos buenos en el reconocimiento de patrones. Recibimos datos del mundo que nos rodea a través de nuestros sentidos y somos capaces de reconocer la fuente de los datos. A menudo podemos hacerlo casi de inmediato y prácticamente sin ningún esfuerzo. Por ejemplo, podemos reconocer el rostro familiar de una persona a pesar de que esa persona ha envejecido desde nuestro último encuentro, identificar a una persona conocida por su voz en el teléfono a pesar de una mala conexión, y distinguir un huevo hervido que es bueno de uno malo oliéndolo. Los humanos realizamos reconocimiento de patrones a través de un proceso de aprendizaje; lo mismo ocurre con las redes neuronales.

El reconocimiento de patrones se define formalmente como el proceso mediante el cual un patrón / señal recibida se asigna a una de un número prescrito de clases (categorías) (Haykin, S., 1999). Una red neuronal realiza el reconocimiento de patrones, experimentando primero una sesión de entrenamiento, durante la cual a la red se le presenta repetidamente un conjunto de patrones de entrada junto con la categoría a la que pertenece cada patrón particular. Más tarde, se presenta un nuevo patrón a la red que no se ha visto antes, pero que pertenece a la misma población de patrones utilizados para entrenar a la red. La red puede identificar la clase de ese patrón particular debido a la información que ha extraído de los datos de entrenamiento. El reconocimiento de patrones realizado por una red neuronal es de naturaleza estadística, con los patrones representados por puntos en un espacio de decisión multidimensional (Haykin, S., 1999). El espacio de decisión se divide en regiones, cada una de las cuales está asociada a una clase. Los límites de decisión están determinados por el proceso de entrenamiento. La construcción de estos límites se hace estadísticamente por la variabilidad inherente que existe dentro y entre las clases.

Aproximación de la función.

La tercera tarea de aprendizaje, la cual es en la que más se está interesado por el objetivo del trabajo de investigación, es la **aproximación a una función**. Considere un mapeo de entrada-salida no lineal descrito por la relación funcional,

$$\mathbf{d} = f(\mathbf{x}) \dots 3.20$$

donde el vector \mathbf{x} es la entrada y el vector \mathbf{d} es la salida. Se supone que la función de valor vectorial $f(\cdot)$ es desconocida. Para compensar la falta de conocimiento sobre la función $f(\cdot)$, se nos proporciona el conjunto de ejemplos etiquetados:

$$\mathcal{F} = \{(\mathbf{x}_i, \mathbf{d}_i) \mid_{i=1}^N \dots 3.21$$

El requisito es diseñar una red neuronal que se aproxime a la función desconocida $f(\cdot)$ de modo que la función $F(\cdot)$ que describe el mapeo de entrada-salida realmente realizada por la red sea lo suficientemente cercana a $f(\cdot)$ en un sentido euclidiano sobre todas las entradas (Haykin, S., 1999), como se muestra por:

$$\|F(\mathbf{x}) - f(\mathbf{x})\| < \epsilon \text{ para todo } \mathbf{x} \dots 3.22$$

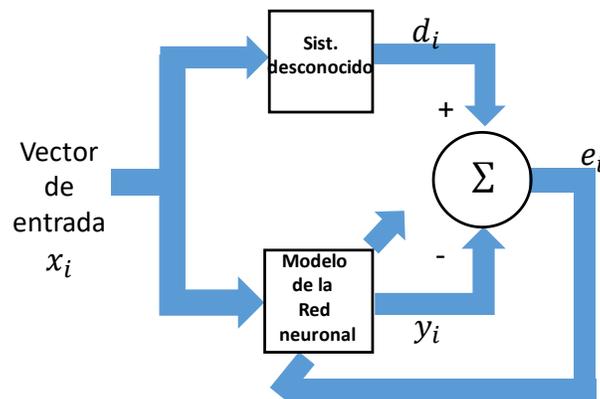
donde ε es un pequeño número positivo. Siempre que el tamaño N del conjunto de entrenamiento sea lo suficientemente grande y la red esté equipada con un número adecuado de parámetros libres, entonces el error de aproximación ε puede hacerse lo suficientemente pequeño para la tarea.

El problema de aproximación descrito aquí es un candidato perfecto para el aprendizaje supervisado con \mathbf{x}_i jugando el papel de vector de entrada y \mathbf{d}_i tomando el rol de la respuesta deseada.

La capacidad de una red neuronal para aproximarse a un mapeo de entrada-salida desconocido puede explotarse de dos maneras importantes (Haykin, S., 1999):

- **Identificación del sistema.** Dado que la Ec. 3.20 describe la relación entrada-salida de un sistema desconocido de múltiples entradas y salidas sin memoria; por un sistema "sin memoria" nos referimos a un sistema que es invariante en el tiempo. Entonces se puede usar el conjunto de ejemplos etiquetados en la Ec. 3.21 para entrenar una red neuronal como modelo del sistema. Así, \mathbf{y}_i denota la salida de la red neuronal producida en respuesta a un vector de entrada \mathbf{x}_i . La diferencia entre \mathbf{d}_i (asociada con \mathbf{x}_i) y la salida de la red \mathbf{y}_i proporciona el vector de señal de error \mathbf{e}_i , como se representa en la Fig. 33. Esta señal de error a su vez se usa para ajustar los parámetros libres de la red para minimizar la diferencia al cuadrado entre las salidas del sistema desconocido y la red neuronal en un sentido estadístico, y se calcula sobre el conjunto completo de entrenamiento (Haykin, S., 1999).

Figura 33. Diagrama de bloque de un sistema de identificación.



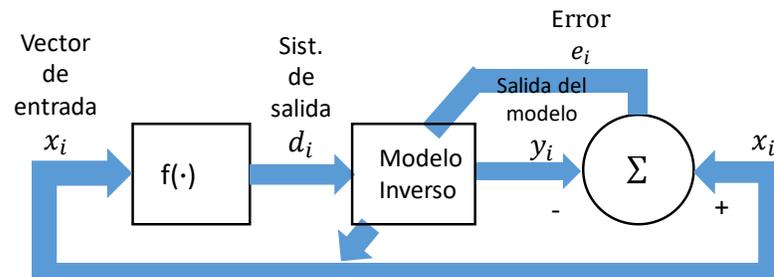
Modificado de Haykin, S. (1999). *óp. cit.*, p. 69.

Sistema inverso. Supóngase que luego se proporciona un sistema conocido de entradas-salidas múltiples sin memoria cuya relación entrada-salida esta descrita por la Ec. 3.23. El requisito en este caso es construir un sistema inverso que produzca el vector \mathbf{x} en respuesta al vector \mathbf{d} . El sistema inverso puede así describirse por:

$$\mathbf{x} = f^{-1}(\mathbf{d}) \dots 3.23$$

donde la función de valor vectorial $f^{-1}(\cdot)$ denota el inverso de $f(\cdot)$. Nótese que, sin embargo, $f^{-1}(\cdot)$ no es el recíproco de $f(\cdot)$; más bien, el uso de superíndice - 1 es simplemente una etiqueta para indicar un inverso. En muchas situaciones que son encontradas en la práctica, la función de valor vectorial $f(\cdot)$ es demasiado compleja e inhibe una formulación directa de la función inversa $f^{-1}(\cdot)$. Dado el conjunto de ejemplos etiquetados en la Ec. 3.23, se puede construir una red neuronal de aproximación para $f^{-1}(\cdot)$ utilizando el esquema que se muestra en la **Fig. 34**. En la situación descrita aquí, los roles de \mathbf{x}_i y \mathbf{d}_i se intercambian: el vector \mathbf{d}_i se usa como la entrada y \mathbf{x}_i es tratado como la respuesta deseada. Considere que el vector de señal de error \mathbf{e}_i denote la diferencia entre \mathbf{x}_i y la salida real \mathbf{y}_i de la red neuronal producida en respuesta a \mathbf{d}_i . Al igual que con el problema de identificación del sistema, este vector de señal de error se usa para ajustar los parámetros libres de la red neuronal para minimizar la diferencia al cuadrado entre las salidas del sistema inverso desconocido y la red neuronal en un sentido estadístico, y se calcula sobre el conjunto de entrenamiento.

Figura 34. Diagrama de bloque de un sistema inverso de modelado.



Modificado de Haykin, S. (1999). óp. cit., p. 70.

CAPITULO IV

REDES NEURONALES MULTICAPA.

4.1 Generalidades.

En este capítulo se hablará de las redes neuronales multicapa de flujo hacia adelante, una clase importante de redes neuronales, las cuales serán usadas en este trabajo para la geopoblado de parámetros petrofísicos, específicamente V_{cl} , ϕ_{hie} , Rho y V_p . Típicamente, la red consta de un conjunto de unidades sensoriales (nodos fuente) que constituyen la capa de entrada, una o más capas ocultas de nodos de cálculo y una capa de nodos de salida. La señal de entrada se propaga a través de la red en una dirección hacia adelante, capa por capa. Estas redes neuronales se conocen comúnmente como perceptrones multicapa, por su acrónimo en inglés, MLP, (Multilayer Perceptron), que representan una generalización del perceptrón de una sola capa.

Los perceptrones multicapa se han aplicado con éxito para resolver algunos problemas complicados, entrenándolos de forma supervisada con un algoritmo muy popular conocido como *algoritmo de retropropagación de error*. Este algoritmo se basa en la regla de aprendizaje por corrección de error; éste es un caso particular del método de optimización *descenso de gradiente* en el espacio de pesos para una red de flujo hacia adelante de múltiples capas. La idea básica es calcular eficientemente las derivadas parciales de una función aproximada $F(\mathbf{w}, \mathbf{x})$ realizada por la red con respecto a todos los elementos del vector de pesos ajustables \mathbf{w} para un valor dado del vector de entrada \mathbf{x} . Esta función (\mathbf{w}, \mathbf{x}) , se puede visualizar como una superficie de error-rendimiento multidimensional o simplemente una superficie de error, con los parámetros libres como coordenadas (**Haykin, S., 1999**). La superficie de error verdadero se promedia sobre todos los posibles ejemplos de entrada. Cualquier operación dada del sistema bajo la supervisión del guía se representa como un punto en la superficie del error. Para que el sistema mejore el rendimiento a lo largo del tiempo y, por lo tanto, aprenda del supervisor, el punto de operación debe descender sucesivamente hacia un punto mínimo de la superficie de error; el punto mínimo puede ser un mínimo local o un mínimo global (**Hagan et al., 2004**). Un sistema de aprendizaje supervisado puede hacer esto con la información útil que tiene sobre el gradiente de la superficie de error correspondiente al comportamiento actual del sistema. El gradiente de una superficie de error en cualquier punto es un vector que apunta en la dirección del descenso más inclinado. De hecho, en el caso del aprendizaje supervisado a partir de ejemplos, el sistema puede usar una estimación instantánea del vector gradiente, con los índices de ejemplo que se presume son los de tiempo. El uso de tal estimación da como resultado un movimiento del punto de operación en la superficie de error que típicamente tiene la forma de una "caminata aleatoria".

El aprendizaje por retropropagación de error consta de dos flujos o etapas a través de las diferentes capas de la red (**Haykin, S., 1999**): un flujo hacia adelante y un flujo hacia atrás. En el flujo hacia adelante, se aplica un patrón de actividad (vector de entrada) a los nodos sensoriales de la red, y su efecto se propaga a través de la red capa por capa. Finalmente, se produce un conjunto de resultados que viene siendo la respuesta real de la red. Durante el flujo hacia adelante, los pesos sinápticos de la red son fijos todos. Durante el flujo hacia atrás, por otro lado, los pesos sinápticos se ajustan de acuerdo con una regla de corrección de errores. Específicamente, la respuesta real de la red se resta de una respuesta deseada (objetivo) para producir una señal de error. Esta señal de error luego se propaga hacia atrás a través de la red, en contra de la dirección de las conexiones sinápticas, de ahí el nombre "**retropropagación de error**". Los pesos sinápticos se ajustan para hacer que la respuesta real de la red se acerque más a la respuesta deseada en un sentido estadístico. El algoritmo de retropropagación de error también se menciona en la literatura como algoritmo de propagación inversa. A partir de ahora se referirá a él como algoritmo de retropropagación. El proceso de aprendizaje realizado con el algoritmo se llama aprendizaje de retropropagación.

Un perceptrón multicapa tiene tres características distintivas (**Haykin, S., 1999**):

1-. El modelo de cada neurona en la red incluye una función de activación no lineal. El punto importante que destacar aquí es que la no linealidad es suave (es decir, diferenciable en todas partes), a diferencia de la limitación estricta utilizada en el perceptrón de Rosenblatt. Una forma de no linealidad comúnmente utilizada que satisface este requisito es una no linealidad sigmoideal definida por la función logística:

$$y_j = \frac{1}{1 + \exp(-v_j)}$$

donde v_j es el *campo local inducido* (es decir, la suma ponderada de todas las entradas sinápticas más el sesgo) de la neurona j , y y_j es la salida de la neurona. La presencia de no linealidades es importante porque, de lo contrario, la relación entrada-salida de la red podría reducirse a la de un perceptrón de una sola capa. Además, el uso de la función logística tiene una motivación biológica, ya que intenta dar cuenta de la fase refractaria de las neuronas reales (**Haykin, S., 1999**).

2. La red contiene una o más capas de neuronas ocultas que no son parte de la entrada o salida de la red. Estas neuronas ocultas permiten que la red aprenda tareas complejas extrayendo progresivamente características más significativas de los patrones de entrada (vectores).

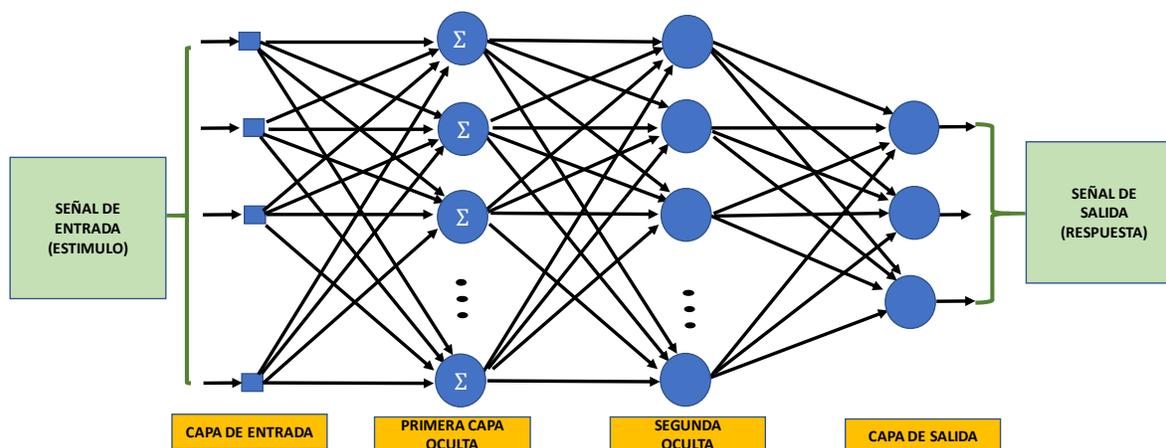
3. La red exhibe un alto grado de conectividad, determinada por las sinapsis de la red. Un cambio en la conectividad de la red requiere un cambio en la población de conexiones sinápticas o sus pesos.

Es a través de la combinación de estas características junto con la capacidad de aprendizaje a través de la experiencia (por un proceso de entrenamiento) que el perceptrón multicapa deriva su poder de cálculo. Estas mismas características, sin embargo, también son

responsables de las deficiencias del estado actual de conocimiento sobre el comportamiento de la red. En primer lugar, la presencia de una forma distribuida de no linealidad y la alta conectividad de la red dificultan el análisis teórico de un perceptrón multicapa. En segundo lugar, el uso de neuronas ocultas hace que el proceso de aprendizaje sea más difícil de visualizar. En un sentido implícito, el proceso de aprendizaje debe decidir qué características del patrón de entrada deberían representarse mediante las neuronas ocultas. Por lo tanto, el proceso de aprendizaje se hace más difícil porque la búsqueda debe realizarse en un espacio mucho más grande de funciones posibles, y debe hacerse una elección entre representaciones alternativas del patrón de entrada (**Hinton, 1989**).

La **Fig. 35** muestra la representación arquitectónica de un perceptrón multicapa con dos capas ocultas y una capa de salida, cabe mencionar que esta estructura es la mayormente usada. Para establecer el escenario para una descripción del perceptrón multicapa en su forma general, la red que se muestra aquí está completamente conectada. Esto significa que una neurona en cualquier capa de la red está conectada a todos los nodos / neuronas de la capa anterior. El flujo de señal a través de la red progresa en dirección hacia adelante, de izquierda a derecha y capa por capa.

Figura 35. Representación arquitectónica de una red neuronal multicapa, con dos capas ocultas.



Modificado de Haykin, S. (1999). óp. cit., p. 181.

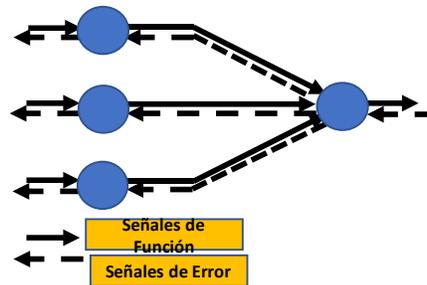
La **Fig. 36** representa una parte del perceptrón multicapa. En esta red se identifican dos tipos de señales (**Parker, 1987**):

1. **Señales de función.** Una señal de función es una señal de entrada (estímulo) que entra en el extremo de entrada de la red, se propaga hacia adelante (neurona por neurona) a través de la red y emerge en el extremo de salida de la red como una señal de salida. Generalmente, tal señal es referida como una "señal de función" por dos razones (**Parker, 1987**). En primer lugar, se presume que realiza una función útil en la salida de la red. En segundo lugar, en cada neurona de la red a través de la cual pasa una señal de función, la señal es calculada como una función de las entradas y los pesos.

asociados aplicados a esa neurona. La señal de función también se conoce como la señal de entrada.

2. **Señales de error.** Una señal de error se origina en una neurona de salida de la red y se propaga hacia atrás (capa por capa) a través de la red. Nos referimos a ella como una "señal de error" porque su cálculo por cada neurona de la red implica una función dependiente del error de una forma u otra.

Figura 36. Ilustración de las direcciones de dos flujos de señales básicas en un perceptrón multicapa: propagación hacia delante de las señales función y retropropagación de la señal error.



Modificado de Haykin, S. (1999). óp. cit., p. 181.

Las neuronas de salida (nodos de cálculo) constituyen las capas de salida de la red. Las neuronas restantes (nodos de cálculo) constituyen las capas ocultas de la red. Por lo tanto, las unidades ocultas no son parte de la salida o entrada de la red, es por ello su designación como "ocultas". La primera capa oculta se alimenta desde la capa de entrada formada por las unidades sensoriales (nodos fuente); las salidas resultantes de la primera capa oculta se aplican a su vez a la siguiente capa oculta; y así sucesivamente para el resto de la red.

Cada neurona oculta o de salida de un perceptrón multicapa está diseñada para realizar dos cálculos (**Haykin, S., 1999**):

1. **El cálculo de la señal de función que aparece en la salida de una neurona, que se expresa como una función no lineal continua de la señal de entrada y los pesos sinápticos asociados con esa neurona.**
2. **El cálculo de una estimación del vector gradiente (es decir, los gradientes de la superficie de error con respecto a los pesos conectados a las entradas de una neurona), que se necesita para la etapa hacia atrás a través de la red.**

4.2 Algoritmo de retropropagación.

A continuación, se presenta la derivación del algoritmo de retropropagación, el cual es el algoritmo base de aprendizaje en las redes usadas en el presente trabajo.

Cabe mencionar que esta fuera de alcance del presente trabajo los métodos de optimización que son como operan los diferentes algoritmos de aprendizaje por corrección de error, por lo

que no se programó ni el algoritmo de retropropagación ni alguno otro algoritmo usado para el entrenamiento de las redes, sino que se usó el **Toolbox** de Redes Neuronales de **Matlab** para dicho propósito. No obstante, se incluye en este capítulo no sólo la derivación del algoritmo de retropropagación, sino también heurísticas para un mejor diseño de una red neuronal, lo cual sin duda alguna contribuirá no sólo a una mejor comprensión del tema por parte del lector, sino también será de gran utilidad para todos aquellos interesados en el desarrollo de algoritmos que mejoren el desempeño del funcionamiento de las redes neuronales.

La derivación del algoritmo de propagación inversa es bastante complicada. Para aliviar la carga matemática implicada en esta derivación, primero se presenta un resumen de las notaciones utilizadas en su derivación.

- Los índices i , j y k se refieren a diferentes neuronas en la red; con señales que se propagan a través de la red de izquierda a derecha, la neurona j se encuentra en una capa a la derecha de la neurona i , y la neurona k se encuentra en una capa a la derecha de la neurona j cuando la neurona j es una unidad oculta.
- En la iteración (paso de tiempo) n , el n -ésimo patrón de entrenamiento (ejemplo) se presenta a la red. El símbolo $\xi(n)$ se refiere a la suma instantánea de cuadrados de error o energía de error en la iteración n . El promedio de $\xi(n)$ sobre todos los valores de n (es decir, todo el conjunto de entrenamiento) produce la energía de error promedio ξ_{prom} .
- El símbolo $e_j(n)$ se refiere a la señal de error en la salida de la neurona j para la iteración n .
- El símbolo $d_j(n)$ se refiere a la respuesta deseada para la neurona j y se usa para calcular $e_j(n)$.
- El símbolo $y_j(n)$ se refiere a la señal de función que aparece en la salida de la neurona j en la iteración n .
- El símbolo $w_{ji}(n)$ denota el peso sináptico que conecta la salida de la neurona i con la entrada de la neurona j en la iteración n . La corrección aplicada a este peso en la iteración n se denota por $\Delta w_{ji}(n)$.
- El campo local inducido (es decir, la suma ponderada de todas las entradas sinápticas más el sesgo) de la neurona j en la iteración n se denota por $v_j(n)$; constituye la señal aplicada a la función de activación asociada a la neurona j .
- La función de activación que describe la relación funcional entrada-salida de la no linealidad asociada con la neurona j se denota por $\varphi_j(\cdot)$.
- El sesgo aplicado a la neurona j se denota por b_j ; su efecto está representado por una sinapsis de peso $w_{j0} = b_j$ conectada a una entrada fija igual a $+1$.
- El i -ésimo elemento del vector de entrada (patrón) se denota por $x_i(n)$.
- El elemento k -ésimo del vector de salida general (patrón) se denota por $O_k(n)$.
- El parámetro de rapidez de aprendizaje se denota por η .
- El símbolo m_l denota el tamaño (es decir, el número de nodos) en la capa l del perceptrón multicapa; $l = 0, 1, \dots, L$, donde L es la "profundidad" de la red. Entonces

m_0 denota el tamaño de la capa de entrada, m_1 denota el tamaño de la primera capa oculta, y m_L , denota el tamaño de la capa de salida. La notación $m_L = M$ también es usada.

La señal de error en la salida de la neurona j en la iteración n (es decir, la presentación del n -ésimo ejemplo de entrenamiento) está definida por:

$$e_j(n) = d_j(n) - y_j(n) \quad \text{la neurona } j \text{ es una nodo de salida .. (4.1)}$$

Definiendo el valor instantáneo de la energía de error para la neurona j como $\frac{1}{2}e_j^2(n)$. Correspondientemente, el valor instantáneo $\xi(n)$ de la energía de error total se obtiene sumando $\frac{1}{2}e_j^2(n)$ sobre todas las neuronas en la capa de salida; estas son las únicas neuronas "visibles" para las cuales las señales de error se pueden calcular directamente. Por lo tanto, podemos escribir:

$$\xi(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \dots \dots (4.2)$$

donde el conjunto C incluye todas las neuronas en la capa de salida de la red. **La energía de error cuadrática promedio** se obtiene sumando $\xi(n)$ sobre todo n y luego normalizando con respecto al tamaño establecido N (N denota el número total de patrones (ejemplos) contenidos en el conjunto de entrenamiento), como se muestra,

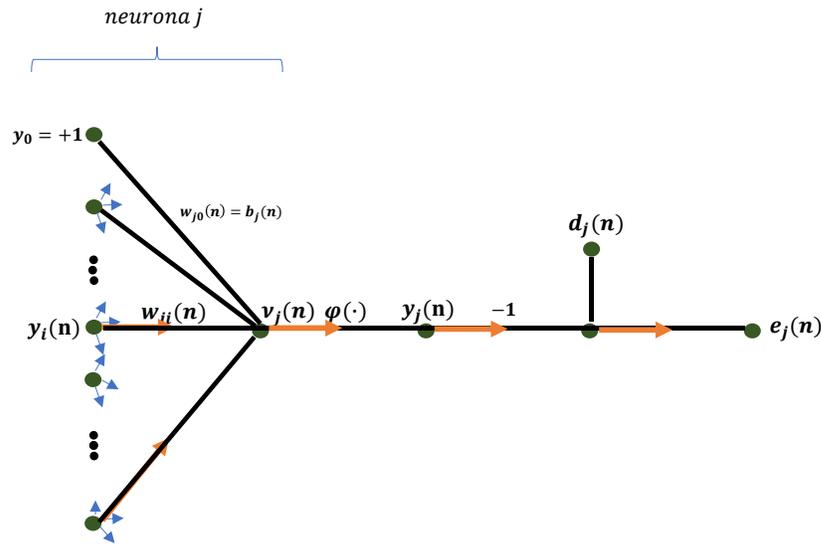
$$\xi_{prom} = \frac{1}{N} \sum_{n=1}^N \xi(n) \dots (4.3)$$

La energía de error instantánea $\xi(n)$, y por lo tanto la energía de error promedio, ξ_{prom} , es una función de todos los parámetros libres (es decir, pesos sinápticos y sesgos) de la red. Para un conjunto de entrenamiento dado, ξ_{prom} representa la **función de costo** como una medida del desempeño del aprendizaje (Haykin, S., 1999). El objetivo del proceso de aprendizaje es ajustar los parámetros libres de la red para minimizar ξ_{prom} . Para hacer esta minimización, se usa una aproximación similar a la utilizada para la derivación del algoritmo **LMS**.² Específicamente, consideramos un método simple de entrenamiento en el que los pesos se actualizan patrón por patrón hasta que se haya tratado una época, es decir, una presentación completa de todo el conjunto de entrenamiento. Los ajustes de los pesos se hacen de acuerdo con los errores respectivos calculados para cada patrón presentado a la red.

El promedio aritmético de estos cambios de peso individuales en el conjunto de entrenamiento es, por lo tanto, una estimación del cambio real que resultaría de la modificación de los pesos basadas en la minimización de la función de costo ξ_{prom} en todo el conjunto de entrenamiento (Haykin, S., 1999).

² Haykin, S.,1999,pp 128-132

Figura 37. Gráfico de señal de flujo que realza los detalles de la salida para la neurona j .



Modificado de Haykin, S. (1999). *óp. cit.*, p.162.

Considérese entonces la **Fig. 37**, que representa a la neurona j alimentada por un conjunto de señales de función producidas por una capa de neuronas a su izquierda. El **campo local inducido** $v_j(n)$ producido a la entrada de la función de activación asociada con la neurona es, por lo tanto,

$$v_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n) \dots (4.4)$$

donde m es el número total de entradas (excluyendo el sesgo) aplicado a la neurona j . El peso sináptico w_{j0} (que corresponde a la entrada fija $y_0 = +1$) es igual al sesgo b_j aplicado a la neurona j . Por lo tanto, la señal de función $y_j(n)$ que aparece en la salida de la neurona j en la iteración n es:

$$y_j(n) = \varphi_j(v_j(n)) \dots (4.5)$$

De una manera similar al algoritmo LMS, el algoritmo de retropropagación aplica una corrección $\Delta w_{ji}(n)$ al peso sináptico $w_{ji}(n)$, que es proporcional a la derivada parcial de $\frac{\partial \xi(n)}{\partial w_{ji}(n)}$. De acuerdo con la regla de la cadena de cálculo, podemos expresar este gradiente como:

$$\frac{\partial \xi(n)}{\partial w_{ji}(n)} = \frac{\partial \xi(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \dots (4.6)$$

La derivada parcial $\frac{\partial \xi(n)}{\partial w_{ji}(n)}$ representa un *factor de sensibilidad*, que determina la dirección de búsqueda en el espacio de los pesos para el peso sináptico w_{ji} .

Diferenciando ambos lados de la Ec. 4.2 con respecto a $e_j(n)$, se tiene,

$$\frac{\partial \xi(n)}{\partial e_j(n)} = e_j(n) \dots (4.7)$$

Diferenciando ambos lados de la Ec. 4.1 con respecto a $y_j(n)$, se tiene,

$$\frac{\partial y_j(n)}{\partial e_j(n)} = -1 \dots (4.8)$$

Posteriormente, diferenciando la Ec. 4.5 con respecto a $v_j(n)$, se obtiene,

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n)) \dots (4.9)$$

donde se emplea el apóstrofe para hacer referencia que se diferencia con respecto al argumento. Finalmente, diferenciando la Ec. 4.4 con respecto a $w_{ji}(n)$, conduce a;

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \dots (4.10)$$

Sustituyendo ahora las Ecs. 4.7 a la 4.10 en la Ec. 4.6, se obtiene:

$$\frac{\partial \xi(n)}{\partial w_{ji}(n)} = -e_j(n) \varphi_j'(v_j(n)) y_i(n) \dots (4.11)$$

La corrección $\Delta w_{ji}(n)$ aplicada a $w_{ji}(n)$ está definida por la *regla delta*;

$$\Delta w_{ji}(n) = -\eta \frac{\partial \xi(n)}{\partial w_{ji}(n)} \dots (4.12)$$

donde η es el *parámetro de rapidez de aprendizaje* del algoritmo de retropropagación. El uso del signo menos en la Ec. 4.12 representa el descenso del gradiente en el espacio de peso (es decir, se busca una dirección tal que el cambio de pesos reduzca el valor de $\xi(n)$). En consecuencia, el uso de la Ec. 4.11 en 4.12 produce:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \dots (4.13)$$

donde *el gradiente local* $\delta_j(n)$ está definido por:

$$\begin{aligned} \delta_j(n) &= -\frac{\partial \xi(n)}{\partial v_j(n)} \\ &= -e_j(n) \varphi_j'(v_j(n)) \dots (4.14) \end{aligned}$$

El gradiente local señala los cambios requeridos en los pesos sinápticos. De acuerdo con Ec. 4.14, el gradiente local $\delta_j(n)$ para la salida neurona j es igual al producto de la correspondiente señal de error $e_j(n)$ para esa neurona y la derivada $\varphi'_j(v_j(n))$ de la función de activación asociada.

De las Ecs. 4.13 y 4.14 se observa que un factor clave involucrado en el cálculo del ajuste de peso $\Delta w_{ji}(n)$ es la señal de error $e_j(n)$ a la salida de la neurona j . En este contexto, se puede identificar dos casos distintos dependiendo de en qué parte de la red está ubicada la neurona j . En el caso 1 la neurona j es un nodo de salida. Este caso es fácil de manejar porque cada nodo de salida de la red se suministra con una respuesta propia deseada, por lo que es una cuestión sencilla calcular la señal de error asociada. En el caso 2, la neurona j es un nodo oculto. Aunque las neuronas ocultas no son accesibles directamente, comparten la responsabilidad por cualquier error cometido en la salida de la red. La gran pregunta sería entonces, saber cómo penalizar o recompensar a las neuronas ocultas por su parte de la responsabilidad. Esto es resuelto de una manera ingeniosa mediante la retropropagación de las señales de error a través de la red.

Caso 1 Neurona j es un nodo de salida.

Cuando la neurona j se encuentra en la capa de salida de la red, se suministra con la respuesta deseada propia. Se puede usar la Ec. 4.1 para calcular la señal de error $e_j(n)$ asociada con esta neurona; véase la **Fig. 37**. Habiendo determinado $e_j(n)$, es una cuestión simple calcular el gradiente local $\delta_j(n)$ usando Ec. 4.14.

Caso 2 Neurona j es un nodo oculto.

Cuando la neurona j está ubicada en una capa oculta de la red no hay una respuesta deseada especificada para esa neurona. En consecuencia, la señal de error para una neurona oculta tendría que determinarse recursivamente en términos de las señales de error de todas las neuronas a las que está conectada directamente esa neurona oculta; aquí es donde el desarrollo del algoritmo de retropropagación se complica. Considérese la situación representada en la **Fig. 38**, que representa la neurona j como un nodo oculto de la red. De acuerdo con Ec. 4.14, se puede redefinir el gradiente local $\delta_j(n)$ para neurona j oculta como:

$$\begin{aligned} \delta_j(n) &= - \frac{\partial \xi(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \dots \dots (4.15) \\ &= - \frac{\partial \xi(n)}{\partial y_j(n)} \varphi'_j(v_j(n)) \quad \text{para una neurona oculta} \end{aligned}$$

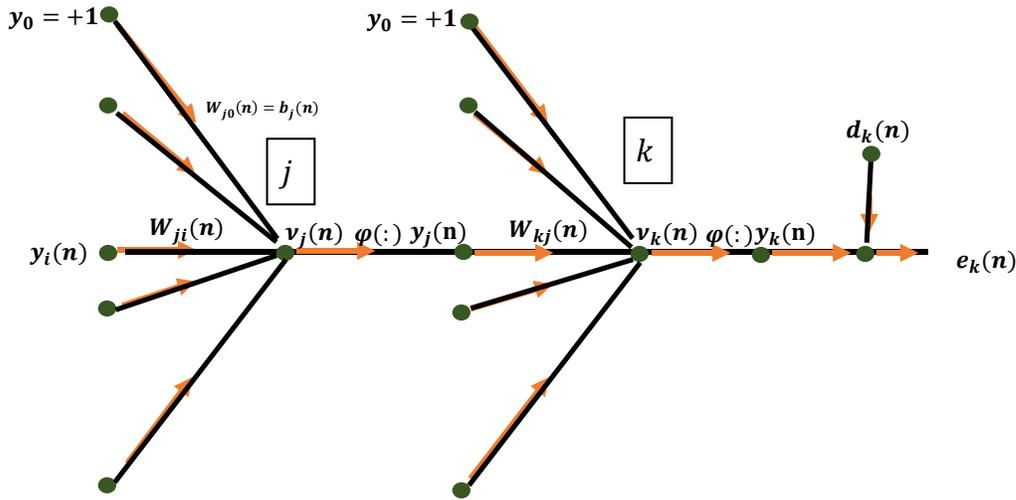
donde en la segunda línea se ha empleado la Ec. 4.9. Para calcular la derivada parcial $\frac{\partial \xi(n)}{\partial y_j(n)}$, se procede de la manera siguiente. A partir de la **Fig. 38** se observa que:

$$\xi(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \dots \dots (4.16) \quad \text{la neurona } k \text{ es un nodo de salida}$$

Como se recordará esta ecuación, es la Ec. 4.2 con la diferencia que se emplea el índice k en lugar de j . Esto se ha llevado a cabo para evitar la confusión con el uso del índice j que hace referencia a una neurona oculta para el caso 2. Diferenciando la Ec. 4.16 con respecto a la función de salida $y_j(n)$, se obtiene,

$$\frac{\partial \xi(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} \dots (4.17)$$

Figura 38. Gráfico de flujo de señal que resalta los del detalle de la neurona k conectada la neurona oculta j .



Modificado de Haykin, S. (1999). *óp. cit.*, p.165.

Posteriormente, se usa la regla de la cadena para la derivada parcial $\frac{\delta e_k(n)}{\delta y_j(n)}$, y reescribiendo la Ec. 4.17;

$$\frac{\partial \xi(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \dots (4.18)$$

Sin embargo, de la Fig. 39 se puede notar que,

$$\begin{aligned} e_k(n) &= d_k(n) - y_k(n) \\ &= d_k(n) - \varphi_k(v_k(n)) \text{ la neurona } k \text{ es un nodo de salida} \end{aligned} \quad (4.19)$$

Por lo tanto,

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'_j(v_j(n)) \dots (4.20)$$

También se puede observar de la Fig. 38 que para la neurona k el campo local inducido es:

$$\mathbf{v}_k(\mathbf{n}) = \sum_{j=0}^m \mathbf{w}_{kj}(\mathbf{n}) \mathbf{y}_j(\mathbf{n}) \dots \dots (4.21)$$

Donde m es el número total de entradas (excluyendo el sesgo) aplicadas a la neurona k . Aquí, una vez más el peso sináptico $w_{k0}(\mathbf{n})$ es igual al sesgo $b_k(\mathbf{n})$ aplicado a la neurona k , y la entrada correspondiente tiene un valor fijo de $+1$. Diferenciando Ec. 4.21 con respecto a $y_j(\mathbf{n})$, se obtiene;

$$\frac{\partial \mathbf{v}_k(\mathbf{n})}{\partial \mathbf{y}_j(\mathbf{n})} = \mathbf{w}_{kj}(\mathbf{n}) \dots \dots (4.22)$$

Usando las Ecs. 4.20 y 4.22 en 4.18, se obtiene la derivada parcial deseada:

$$\begin{aligned} \frac{\partial \xi(\mathbf{n})}{\partial \mathbf{y}_j(\mathbf{n})} &= - \sum_k \mathbf{e}_k(\mathbf{n}) \varphi'_j(\mathbf{v}_j(\mathbf{n})) \mathbf{w}_{kj}(\mathbf{n}) \\ &= - \sum_k \delta_k(\mathbf{n}) \mathbf{w}_{kj}(\mathbf{n}) \dots \dots (4.23) \end{aligned}$$

donde en la segunda línea se ha usado la definición del gradiente local $\delta_k(\mathbf{n})$ dado en la Ec. 4.14 con el índice k reemplazando para j .

Finalmente, usando la Ec. 4.23 en la 4.15, se obtiene la **fórmula de retropropagación** para el gradiente local $\delta_j(\mathbf{n})$:

$$\delta_j(\mathbf{n}) = \varphi'_j(\mathbf{v}_j(\mathbf{n})) \sum_k \delta_k(\mathbf{n}) \mathbf{w}_{kj}(\mathbf{n}) \text{ la neurona } j \text{ esta oculta} \dots \dots (4.24)$$

La **Fig. 39** muestra el gráfico de flujo de señal de la Ec. 4.24, suponiendo que la capa de salida consiste en m_l de neuronas.

El factor $\varphi'_j(\mathbf{v}_j(\mathbf{n}))$ involucrado en el cálculo del gradiente local $\delta_j(\mathbf{n})$ en la Ec. 4.24 depende únicamente de la función de activación asociada con la neurona oculta j . El factor restante involucrado en este cálculo, es decir la suma sobre k , depende de dos conjuntos de términos. El primer conjunto de términos, el $\delta_k(\mathbf{n})$ requiere el conocimiento de las señales de error $e_k(\mathbf{n})$, para todas las neuronas que se encuentran en la capa a la derecha inmediata de la neurona oculta j , y que están directamente conectados a la neurona j : ver **Fig. 39**. El segundo conjunto de términos, la $w_{kj}(\mathbf{n})$, consiste en los pesos sinápticos asociados con estas conexiones.

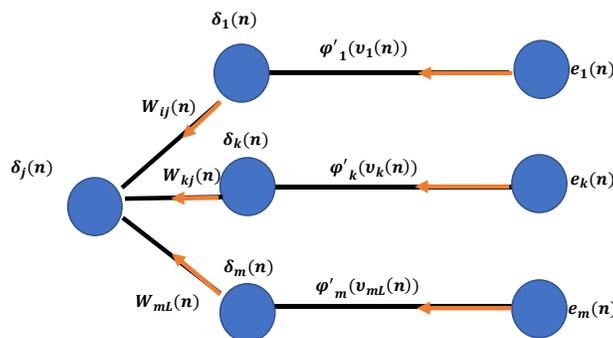
Ahora resumiendo las relaciones que se han obtenido para el algoritmo de retropropagación; primero, la corrección $\Delta w_{ji}(\mathbf{n})$, aplicada al peso sináptico que conecta la neurona i con la neurona j se define mediante la regla delta:

$$\left(\begin{array}{c} \text{corrección del peso sináptico} \\ \Delta w_{ji}(\mathbf{n}) \end{array} \right) = \left(\begin{array}{c} \text{parám. veloc. de aprendizaje} \\ \eta \end{array} \right) \cdot \left(\begin{array}{c} \text{Grad. local} \\ \delta_j(\mathbf{n}) \end{array} \right) \cdot \left(\begin{array}{c} \text{Señal de entrada de} \\ \text{la neurona } j \\ \mathbf{y}_i(\mathbf{n}) \end{array} \right) \dots \dots (4.25)$$

En segundo lugar, el gradiente local $\delta_j(n)$ depende de si la neurona j es un nodo de salida o un nodo oculto:

- 1-. Si la neurona j es un nodo de salida, $\delta_j(n)$ es igual al producto de la derivada $\varphi'_j(v_j(n))$ y la señal de error $e_j(n)$, ambas asociadas con neurona j ; ver Ec. 4.14.
- 2-. Si neurona j es un nodo oculto, $\delta_j(n)$ es igual al producto de la derivada asociada $\varphi'_j(v_j(n))$ y la suma ponderada del δ_s calculada para las neuronas en la siguiente capa oculta o de salida que están conectado a neurona j ; ver Ec. (4.24).

Figura 39. Gráfico de señal de flujo de una parte del sistema adjunto perteneciente a la retropropagación de la señal error.



Modificado de Haykin, S. (1999). *óp. cit.*, p.166.

4.2.1 Función de activación.

Cómo se pudo observar en las Ecs. anteriores, es necesario para el cálculo de δ para cada neurona del perceptrón multicapa el conocimiento de la derivada de la función de activación $\varphi(\cdot)$ asociada con esa neurona. Para que exista esta derivada, se requiere que la función sea $\varphi(\cdot)$ continua. En términos básicos, la diferenciabilidad es el único requisito que una función de activación debe satisfacer. Un ejemplo de una función de activación no lineal continuamente diferenciable usada comúnmente en los perceptrones multicapa, es la no linealidad sigmoideal; dos formas son descritas a continuación (Stewart, 2008; Haykin, S., 1999):

1-. Función logística. Esta forma de no linealidad sigmoideal en su forma general está definida por:

$$\varphi_j(v_j(n)) = \frac{1}{1 + \exp(-av_j(n))} \quad a > 0 \text{ y } -\infty < v_j(n) < \infty \dots \dots (4.26)$$

donde $v_j(n)$ es el campo local inducido de la neurona j . De acuerdo con esta no linealidad, la amplitud de la salida caerá en el rango $0 \leq y_j \leq 1$. Diferenciando la Ec. 4.26 con respecto a $v_j(n)$, se obtiene:

$$\varphi'_j(v_j(n)) = \frac{a \exp(-av_j(n))}{[1 + \exp(-av_j(n))]^2} \dots \dots (4.27)$$

Con $y_j(n) = \varphi_j(v_j(n))$, se podría eliminar el término del exponencial $\exp(-av_j(n))$ de la Ec. 4.27, y así expresar la derivada $\varphi'_j(v_j(n))$ como:

$$\varphi'_j(v_j(n)) = ay_j(n)[1 - y_j(n)] \dots (4.28)$$

Para la neurona j localizada en la capa de salida, $y_j(n) = o_j(n)$. Por lo tanto, se puede expresar el gradiente local para la neurona j como:

$$\begin{aligned} \delta_j(n) &= e_j(n)\varphi'_j(v_j(n)) \\ &= a[d_j(n) - o_j(n)]o_j(n)[1 - o_j(n)], \quad \text{neurona } j \text{ es un nodo de salida } \dots \quad (4.29) \end{aligned}$$

donde $o_j(n)$ es la señal función en la salida de la neurona j , y $d_j(n)$ es la respuesta deseada. Por otro lado, para una neurona arbitraria oculta j , se puede expresar el gradiente local como:

$$\begin{aligned} \delta_j(n) &= \varphi'_j(v_j(n)) \sum_k \delta_k(n)w_{kj}(n) \\ &= ay_j[1 - y_j(n)] \sum_k \delta_k(n)w_{kj}(n), \quad \text{neurona } j \text{ está oculta } \dots \quad (4.30) \end{aligned}$$

Nótese de la Ec. 4.28 que la derivada $\varphi'_j(v_j(n))$ alcanza su valor máximo en $y_j(n) = 0.5$, y su valor mínimo (cero) en $y_j(n) = 0$, o $y_j(n) = 1.0$. Como la cantidad de cambio en un peso sináptico de la red es proporcional a la derivada $\varphi'_j(v_j(n))$ se sigue que para una función de activación sigmoideal los pesos sinápticos que más cambian son los pertenecientes a las neuronas de la red donde las señales de función están en su rango medio. De acuerdo con **Rumelhart et al. (1986a)**, es esta característica del aprendizaje de retropropagación lo que contribuye a su estabilidad como algoritmo de aprendizaje.

2. Función tangente hiperbólica. Otra forma comúnmente utilizada de la no linealidad sigmoideal es la función tangente hiperbólica, que en su forma más general se define como:

$$\varphi_j(v_j(n)) = \text{atanh}(bv_j(n)) \quad (a, b) > 0 \dots \dots (4.31)$$

donde a y b son constantes. En realidad, la función tangente hiperbólica es solo la función logística re-escalada y con sesgo. Su derivada con respecto a $v_j(n)$ es dado por:

$$\begin{aligned}
\varphi'_j(v_j(n)) &= ab \operatorname{sech}^2(bv_j(n)) \\
&= ab \left[1 - \tanh^2(bv_j(n)) \right] \dots \dots \dots (4.32) \\
&= \frac{b}{a} [a - y_j(n)][a + y_j(n)]
\end{aligned}$$

Para una neurona j localizada en la capa de salida, el gradiente local es:

$$\begin{aligned}
\delta_j(n) &= e_j(n) \varphi'_j(v_j(n)) \\
&= \frac{b}{a} [d_j(n) - o_j(n)][a - o_j(n)][a + o_j(n)] \dots \dots (4.33)
\end{aligned}$$

Para una neurona j en una capa oculta, se tiene;

$$\begin{aligned}
\delta_j(n) &= \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \\
&= \frac{b}{a} [a - y_j(n)][a + y_j(n)] \sum_k \delta_k(n) w_{kj}(n) \text{ la neurona } j \text{ esta oculta } \dots (4.34)
\end{aligned}$$

Usando las Ecs. 4.29 y 4.30 para la función logística y las Ecs. 4.33 y 4.34 para la función tangente hiperbólica, se puede calcular el gradiente local δ_j sin tener conocimiento explícito de la función de activación.

4.2.2 Rapidez de aprendizaje.

El algoritmo de retropropagación proporciona una "aproximación" a la trayectoria en el espacio de peso calculado por el método de descenso más empinado. Un parámetro sin duda de gran importancia es el de rapidez de aprendizaje, η , pues de este depende que el método converja o no. Cuanto más pequeño se haga el parámetro de rapidez de aprendizaje η , menores serán los cambios en los pesos sinápticos en la red de una iteración a la siguiente, y más suave será la trayectoria en el espacio de pesos. Sin embargo, esta mejora se logra a costa de una rapidez de aprendizaje más lenta. Si, por otro lado, se hace que el parámetro de rapidez de aprendizaje η sea demasiado grande para acelerar la velocidad de aprendizaje, los grandes cambios resultantes en los pesos sinápticos suponen tal forma que la red puede volverse inestable (es decir, oscilatoria). Un método simple para aumentar este parámetro y evitar el peligro de inestabilidad es modificar la regla delta de la Ec. 4.13 incluyendo un término de momentum. Como se muestra en (Rumelhart et al, 1986),

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n) \dots \dots (4.35)$$

donde α es generalmente un número positivo llamado constante de momentum. Controla el bucle de realimentación que actúa alrededor de $\Delta w_{ji}(n)$, c. La Ec. 4.35 se llama **regla delta generalizada**; incluye la regla delta de Ec. 4.13 como un caso especial (es decir, $\alpha=0$).

Para observar el efecto de la secuencia de presentaciones de patrones en los pesos sinápticos debido a la constante de momentum α , se reescribe la Ec. (4.35) como una serie de tiempo con índice t . El índice t va desde el tiempo inicial 0 hasta el tiempo actual n . La Ec. 4.35 puede verse como una ecuación de diferencias de primer orden en la corrección del peso. Resolviendo esta ecuación para $\Delta w_{ji}(n)$ tenemos;

$$\Delta w_{ji}(n) = \eta \sum_{t=0}^n \alpha^{n-t} \delta_j(t) y_i(t) \dots (4.36)$$

que representa una serie temporal de longitud $n+1$. De las Ecs. 4.11 y 4.14 observamos que el producto $\delta_j(n)$ y $y_i(n)$ es igual a $-\frac{\partial \xi(n)}{\partial w_{ji}(t)}$ (n). En consecuencia, se puede reescribir la Ec. (4.36) en la forma equivalente:

$$\Delta w_{ji}(n) = -\eta \sum_{t=0}^n \alpha^{n-t} \frac{\partial \xi(t)}{\partial w_{ji}(t)} \dots \dots (4.37)$$

Con base en esta relación, se pueden hacer las siguientes observaciones perspicaces (**Watrous, 1987; Jacobs, 1988**):

1. El ajuste actual $\Delta w_{ji}(n)$ representa la suma de una serie exponencial en tiempo ponderada. Para que la serie de tiempo *sea convergente*, la constante de momentum debe restringirse al rango $0 \leq |\alpha| \leq 1$. Cuando α es cero, el algoritmo retropropagación funciona sin momentum. También la constante de momentum α , puede ser positiva o negativa, aunque es poco probable que se use una negativa en la práctica.
2. Cuando la derivada parcial $\frac{\partial \xi(n)}{\partial w_{ji}(t)}$ tiene el mismo signo algebraico en iteraciones consecutivas, la suma exponencial de pesos $\Delta w_{ji}(n)$ crece en magnitud, y entonces el peso $w_{ji}(n)$ se ajusta en una gran cantidad. La inclusión del momentum en el algoritmo de retropropagación tiende a acelerar el descenso en direcciones constantes hacia abajo.
3. Cuando la derivada parcial $\frac{\partial \xi(n)}{\partial w_{ji}(t)}$ tiene signos opuestos en iteraciones consecutivas, la suma exponencial de pesos $\Delta w_{ji}(n)$ disminuye en magnitud, entonces el peso $w_{ji}(n)$ se ajusta en una pequeña cantidad. La inclusión del momentum en el algoritmo de retropropagación tiene un efecto estabilizador en las direcciones que oscilan en el signo.

La incorporación de momentum en el algoritmo de retropropagación representa una modificación menor a la actualización de peso, aunque puede tener algunos efectos beneficiosos sobre el comportamiento de aprendizaje del algoritmo (**Haykin, S., 1999**). El término de momentum también puede tener el beneficio de evitar que el proceso de aprendizaje termine en un mínimo local superficial en la superficie del error. Se implementó este algoritmo de aprendizaje durante el entrenamiento de las redes para la predicción de las propiedades petrofísicas de interés.

Al derivar el algoritmo de retropropagación, se supuso que el parámetro de rapidez de aprendizaje fuera constante, denotado por η . Sin embargo, en realidad debe definirse como η_{ji} ; es decir, el parámetro de rapidez de aprendizaje debe depender de las conexiones. De hecho, se pueden hacer muchas cosas interesantes haciendo que el parámetro de rapidez de aprendizaje sea diferente para diferentes partes de la red, no obstante, esta fuera del alcance del presente trabajo.

También es digno de mencionarse que, en la aplicación del algoritmo de retropropagación, se puede elegir que todos los pesos sinápticos de la red sean ajustables, o se puede restringir cualquier cantidad de pesos en la red para que permanezcan fijos durante el proceso de adaptación. En este último caso, las señales de error se propagan por la red de la manera habitual; sin embargo, los pesos sinápticos fijos se dejan inalterados. Esto se puede hacer simplemente haciendo que el parámetro de rapidez de aprendizaje η_{ji} para el peso sináptico $w_{ji}(n)$ sea igual a cero.

4.2.3 Criterios para detener el proceso de aprendizaje de la red (entrenamiento).

En general, no se puede mostrar que el algoritmo de retropropagación converja y no existen criterios bien definidos para detener su funcionamiento (es por ello por lo que en este trabajo se propone el empleo de la prueba Gamma para dar una solución a éste y otros problemas que se presentan en las redes neuronales). Por el contrario, existen algunos criterios razonables, cada uno con su propio mérito práctico, que pueden utilizarse para finalizar los ajustes de pesos. Para formular tal criterio, es lógico pensar en términos de las propiedades únicas de un **mínimo local** o **global** de la superficie de error. Considérese que el vector de ponderación \mathbf{w}^* denote un mínimo, ya sea local o global. Una condición necesaria para que \mathbf{w}^* sea mínimo es que el vector de gradiente $\mathbf{g}(\mathbf{w})$ (es decir, la derivada parcial de primer orden) de la superficie de error con respecto al vector de peso \mathbf{w} sea cero en $\mathbf{w} = \mathbf{w}^*$. En consecuencia, se puede formular un criterio de convergencia sensible para el aprendizaje de retropropagación de la siguiente manera (**Kramer y Sangiovanni-Vincentelli, 1989**):

Se considera que el algoritmo de retropropagación ha convergido cuando la norma euclidiana del vector gradiente alcanza un valor suficientemente pequeño.

El inconveniente de este criterio de convergencia es que, para ensayos exitosos, los tiempos de aprendizaje pueden ser largos. Además, requiere el cálculo del vector del gradiente $\mathbf{g}(\mathbf{w})$.

Otra propiedad única de un mínimo que podemos usar es el hecho de que la función de costo o el error medido $\xi_{prom}(w)$ es estacionario en el punto $\mathbf{w} = \mathbf{w}^*$. Por lo tanto, **Haykin, S., 1999** sugiere un criterio diferente de convergencia:

Se considera que el algoritmo de retropropagación ha convergido cuando la rapidez absoluta de cambio en el error cuadrado medio por época es suficientemente pequeña.

La rapidez de cambio en el error promedio cuadrado se considera típicamente lo suficientemente pequeña si se encuentra en el rango de 0.1 a 1 por ciento por época. Algunas

veces se usa un valor tan pequeño como 0.01 por ciento por época. Desafortunadamente, este criterio puede resultar en una finalización prematura del proceso de aprendizaje.

Existe otro criterio útil y teóricamente compatible para la convergencia. Después de cada iteración de aprendizaje, la red se prueba para su rendimiento de generalización. El proceso de aprendizaje se detiene cuando el rendimiento de generalización es adecuado o cuando es aparente que el rendimiento de generalización ha alcanzado su máximo, sin embargo, éste criterio tiene el inconveniente de que se tiene que dividir los datos en entradas, unos para el entrenamiento, y otros más para la validación, lo cual resulta muy malo cuando se cuenta con pocos datos.

El software **Matlab** permite establecer varios criterios para llevar a cabo la terminación del entrenamiento de las redes de manera simultánea. Éstos son: Número de épocas, el error mínimo a alcanzar y cuando la rapidez de cambio en el error es mínima. Dado que la prueba Gamma, proporciona el error mínimo a alcanzar, este criterio será el más preponderante en el momento del entrenamiento de las redes, en el caso de no alcanzar dicho error, en un número considerable de épocas, se considerará aumentar el número de neuronas en las capas ocultas, siempre y cuando el número de pesos sea menor al número de datos de entrada, esto para conservar un sistema subestimado.

4.2.4 Heurísticas para mejorar el desempeño del algoritmo de retropropagación.

A menudo se dice que el diseño de una red neuronal utilizando el algoritmo de retropropagación es más un arte que una ciencia en el sentido de que muchos de los factores que intervienen en el diseño son el resultado de la propia experiencia personal. Hay algo de verdad en esta declaración. Sin embargo, existen métodos que mejorarán significativamente el rendimiento del algoritmo de retropropagación, o bien se pueden emplear herramientas de análisis estadístico que nos ayuden a mejorar tanto el diseño como el desempeño de las redes neuronales, lo cual es el propósito del presente trabajo. A continuación, se describe los enfoques clásicos comúnmente utilizados para optimizar el desempeño de las redes, algunos de los cuales, fueron empleados en el presente trabajo (**Haykin, S., 1999**).

1. Actualización de los pesos de manera secuencial contra por lotes, el modo secuencial del aprendizaje de retropropagación (que implica la actualización patrón por patrón) es computacionalmente más rápido que el modo por lotes. Esto es especialmente cierto cuando el conjunto de datos de entrenamiento es grande y altamente redundante. (Los datos altamente redundantes plantean problemas de cálculo para la estimación del Jacobiano requerido para la actualización por lotes).

2. Maximizar el contenido de la información. Como regla general, cada ejemplo de entrenamiento presentado al algoritmo de retropropagación debe elegirse sobre la base de que su contenido de información es el más amplio posible para la tarea en cuestión (**LeCun, 1993**). Existen dos formas de lograr este objetivo, las cuales son (**Haykin, S., 1999**):

- **El uso de un ejemplo que resulta en el mayor error de entrenamiento.**

- **El uso de un ejemplo que es radicalmente diferente de todos los utilizados anteriormente.**

Estas dos heurísticas están motivadas por un deseo de buscar más en el espacio de pesos. En las tareas de clasificación de patrones que utilizan el aprendizaje de retropropagación secuencial, una técnica simple que se usa comúnmente es aleatorizar (es decir, mezclar) el orden en el que se presentan los ejemplos al perceptrón de varias capas de una época a la siguiente. Idealmente, la aleatorización asegura que los ejemplos sucesivos en una época presentados a la red rara vez pertenezcan a la misma clase (este enfoque se empleó durante el entrenamiento de las redes en el presente trabajo).

Para una técnica más refinada, se puede utilizar un esquema de énfasis, que implica la presentación a la red de patrones más difíciles en lugar de fáciles (LeCun, 1993). Si un patrón particular es fácil o difícil se puede identificar al examinar el error que produce, en comparación con las iteraciones previas del algoritmo.

Sin embargo, hay dos problemas con el uso de un esquema de énfasis que debe examinarse cuidadosamente:

- **La distribución de ejemplos dentro de una época presentada a la red está distorsionada.**
- **La presencia de un ejemplo atípico o mal etiquetado puede tener una consecuencia catastrófica en el rendimiento del algoritmo; el aprendizaje de tales valores atípicos compromete la capacidad de generalización de la red sobre las regiones más probables del espacio de entrada.**

3. Función de activación. Un perceptrón multicapa entrenado con el algoritmo retropropagación puede, en general, aprender más rápido (en términos del número de iteraciones de entrenamiento requeridas) cuando la función de activación sigmoideal incorporada en el modelo neuronal de la red es antisimétrica que cuando no es simétrica (Haykin, S.,1999). Se dice que una función de activación $\varphi(v)$ es antisimétrica (es decir, una función impar de su argumento) si,

$$\varphi(-v) = -\varphi(v)$$

cómo se representa en la **Fig. 40 a**. Esta condición no está satisfecha con la función logística estándar indicada en la **Fig. 40 b**.

Un ejemplo popular de una función de activación antisimétrica es una no linealidad sigmoideal en forma de una tangente hiperbólica, definida por:

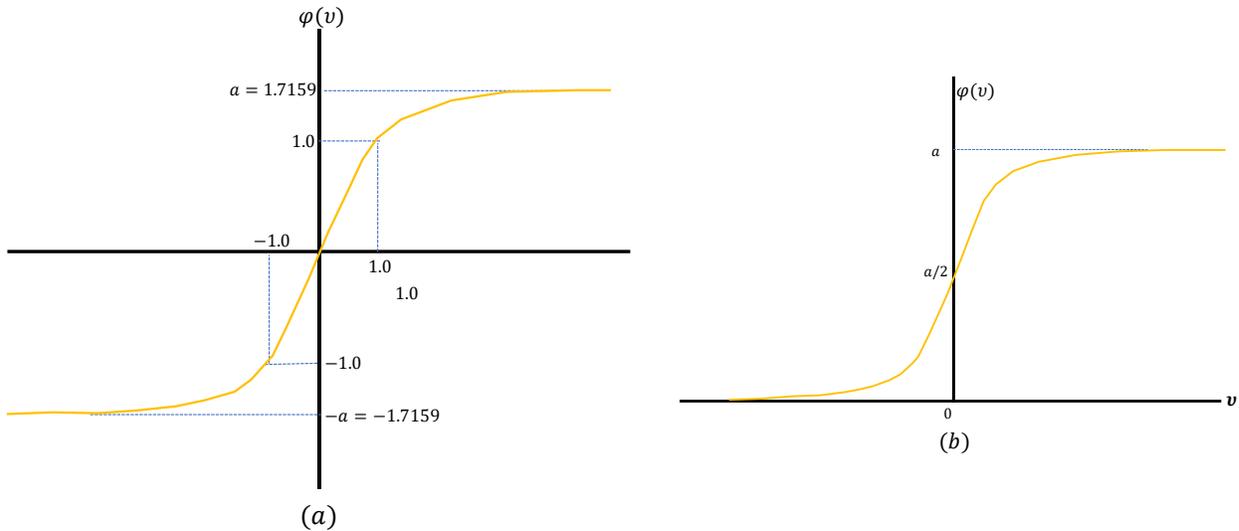
$$\varphi(v) = \operatorname{atanh}(bv)$$

donde a y b son constantes. Los valores adecuados para las constantes a y b son (LeCun, 1989, 1993):

$$a = 1.7159$$

$$b = 2/3$$

Figura 40. a) Función de activación antisimétrica. b) Función de activación no simétrica.



Modificado de Haykin, S. (1999). *óp. cit.*, p.180.

La función tangente hiperbólica así definida tiene las siguientes propiedades útiles:

- $\varphi(1) = 1$ y $\varphi(-1) = -1$
- En el origen, la pendiente (es decir, ganancia efectiva) de la función de activación está cerca de la unidad, como es mostrado por:

$$\begin{aligned} \varphi(0) &= a \\ &= 1.7159 * 2/3 \\ &= 1.1424 \end{aligned}$$

- La segunda derivada de $\varphi(v)$ adquiere su máximo valor en $v = 1$

4. Valores objetivo. Es importante que los valores objetivo (respuesta deseada) se elijan dentro del rango de la función de activación sigmoideal. Más específicamente, la respuesta deseada d_j para la neurona j en la capa de salida del perceptrón multicapa debe ajustarse en cierta cantidad ε lejos del valor límite de la función de activación sigmoideal, dependiendo de si el valor límite es positivo o negativo. De lo contrario, el algoritmo de retropropagación tiende a conducir los parámetros libres de la red al infinito, y por lo tanto ralentiza el proceso de aprendizaje al conducir las neuronas ocultas a la saturación. Para ser específico, considere la función de activación antisimétrica de la **Fig. 40 a**. Para el valor límite $+a$, se establece;

$$d_j = a - \varepsilon$$

donde ε es una constante positiva apropiada. Para la elección de un $a=1.7159$ referido antes, podemos establecer $\varepsilon=0.7159$, en cuyo caso el valor objetivo (respuesta deseada) d_j puede convenientemente ser elegido como ± 1 , como se indica en la **Fig. 40a**.

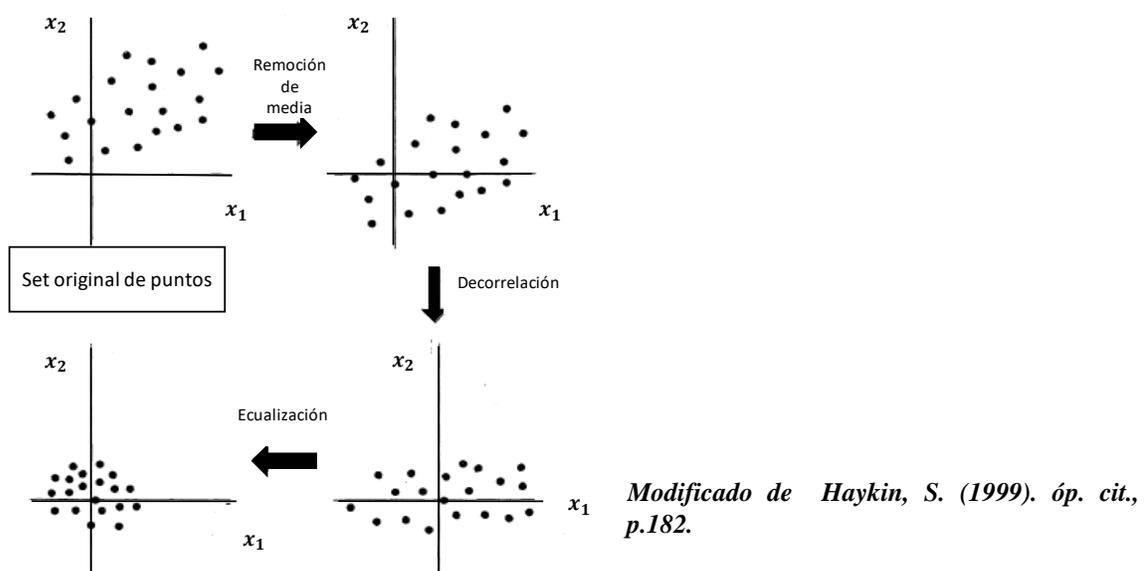
5. Normalizar las entradas. Cada variable de entrada debe preprocesarse para que el valor medio, promediado sobre todo el conjunto de entrenamiento, sea cercano a cero, o si no que sea pequeño en comparación con su desviación estándar (**LeCun, 1993**). Para apreciar la importancia práctica de esta regla, considere el caso extremo donde las variables de entrada son consistentemente positivas. En esta situación, los pesos sinápticos de una neurona de la primera capa oculta pueden solo aumentar o disminuir juntos. En consecuencia, si el vector de peso de esa neurona es cambiar de dirección, solo puede hacerlo zigzagueando su trayectoria por la superficie error, lo cual es normalmente lento y, por lo tanto, debe evitarse.

A fin de acelerar el proceso de aprendizaje de retropropagación, a parte de la normalización de los datos, se deben también incluir otras dos medidas (**LeCun, 1993**):

- Las variables de entrada contenidas en el conjunto de entrenamiento no deben estar correlacionadas; esto puede hacerse usando el análisis de componentes principales u otro método geoestadística, en nuestro caso se usó la **función estadística Gamma**. (véase **apéndice D**)
- Las variables de entrada no correlacionadas deben escalarse para que sus covarianzas sean aproximadamente igual, asegurando así que los diferentes pesos sinápticos en la red aprendan aproximadamente a la misma velocidad.

La **Fig. 41** ilustra los resultados de tres pasos de normalización: eliminación de la media, decorrelación y ecuilización de covarianza, aplicados en ese orden.

Figura 41. Ilustración de las operaciones de normalizado: remoción de la media, decorrelación y ecuilización de la covarianza para un espacio de entrada 2D.



6. Inicialización. Una buena opción para los valores iniciales de los pesos sinápticos y los umbrales de la red pueden ser de gran ayuda en un diseño de red exitoso. La pregunta clave es: ¿cuál es una buena opción?

Cuando a los pesos sinápticos se les asignan valores iniciales grandes, es muy probable que las neuronas en la red sean conducidas a la saturación. Si esto sucede, los gradientes locales en el algoritmo de retropropagación asumen valores pequeños, lo que a su vez hará que el proceso de aprendizaje se ralentice. Sin embargo, si a los pesos sinápticos se les asignan valores iniciales pequeños, el algoritmo de retropropagación puede operar en un área muy plana alrededor del origen de la superficie de error; esto es particularmente cierto en el caso de las funciones de activación antisimétrica, como la función de tangente hiperbólica. Por estas razones, se debe evitar el uso de valores grandes y pequeños para inicializar los pesos sinápticos. La elección correcta de inicialización se encuentra en algún lugar entre estos dos casos extremos.

7. Rapidez de aprendizaje. Todas las neuronas en el perceptrón multicapa idealmente deberían aprender a la misma velocidad. Las últimas capas generalmente tienen gradientes locales más grandes que las capas en el frente de la red. Por lo tanto, debe asignarse el parámetro " η " de rapidez de aprendizaje un valor más pequeño en las últimas capas que en las capas frontales. La neurona con muchas entradas debería tener un parámetro de rapidez de aprendizaje más grande que las neuronas con pocas entradas a fin de mantener un tiempo de aprendizaje similar para todas las neuronas en la red. En **LeCun (1993)**, sugirió que, para una neurona dada, la tasa de aprendizaje debería ser inversamente proporcional a la raíz cuadrada de las conexiones sinápticas hechas a esa neurona.

Es importante mencionar que algunas de estas heurísticas no fueron usadas para el diseño de las redes para la estimación de los parámetros petrofísicos, dado que el software empleado no permite llevarlas a cabo, y también porque esta fuera del alcance de este trabajo. No obstante, es importante que el lector tenga conocimiento de ellas, por si en algún momento éste quiera llevar algún trabajo empleando redes neurales.

4.2.4 Capacidad de generalización.

En el aprendizaje de retropropagación, normalmente se comienza con una muestra de entrenamiento y se usa el algoritmo de retropropagación para calcular los pesos sinápticos de un perceptrón multicapa cargando (codificando) la mayor cantidad posible de ejemplos de entrenamiento en la red. La esperanza es que la red neuronal así diseñada generalice. Se dice que una red **generaliza** bien cuando la asignación de entrada-salida calculada por la red es correcta (o casi) para los datos de prueba nunca utilizados en el entrenamiento de la red (**Haykin, S., 1999**); el término "**generalización**" se toma prestado de la psicología. Aquí se supone que los datos de prueba se obtienen de la misma población utilizada para generar los datos de entrenamiento.

El proceso de aprendizaje (es decir, el entrenamiento de una red neuronal) se puede ver como un problema de "ajuste de curvas". La red en sí misma puede considerarse simplemente como

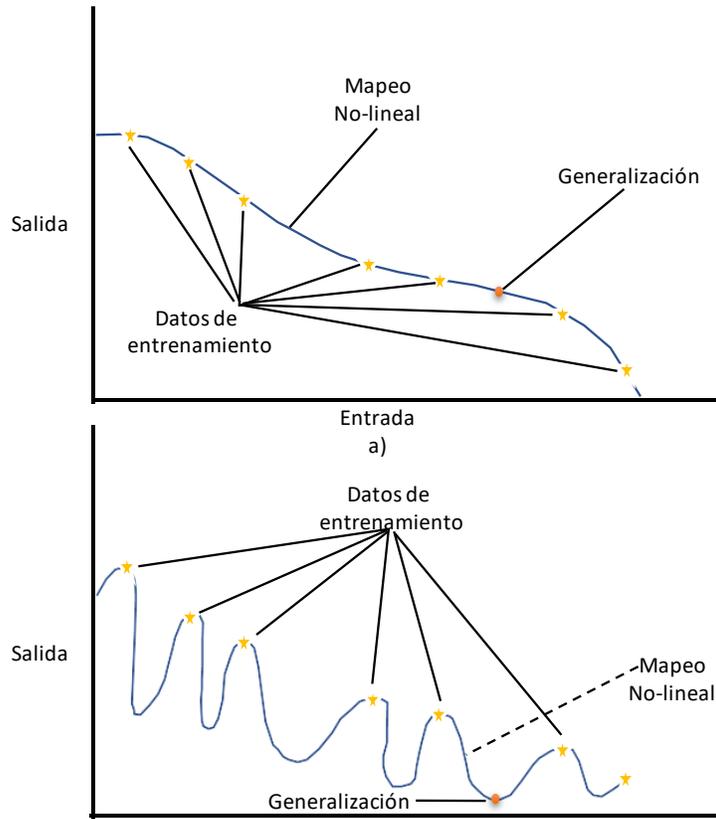
un mapeo de entrada-salida no lineal. Desde ese punto de vista, permite ver a la generalización no como una propiedad mística de las redes neuronales sino simplemente como el efecto de una buena interpolación no lineal de los datos de entrada (**Wieland y Leighton, 1987**). La red realiza una interpolación útil principalmente porque los perceptrones multicapa con funciones de activación continuas conducen a funciones de salida que también son continuas.

La **Fig. 42 a** ilustra cómo puede ocurrir la generalización en una red hipotética. El mapeo no lineal de entrada-salida representado por la curva representada en esta figura es calculado por la red como resultado del aprendizaje de los puntos etiquetados como "datos de entrenamiento". El punto marcado en la curva como "generalización" se ve, así como el resultado de la interpolación realizado por la red.

Una red neuronal que está diseñada para generalizar bien producirá un mapeo correcto de entrada-salida incluso cuando la entrada sea ligeramente diferente al de los ejemplos utilizados para entrenar la red, como se ilustra en la figura. Sin embargo, cuando una red neuronal aprende demasiados ejemplos de entrada-salida, la red puede terminar memorizando los datos de entrenamiento. Esto puede ser debido a que la red encuentre una característica (debido al ruido, por ejemplo) que está presente en los datos de entrenamiento, pero que no es cierta para la función subyacente que se va a modelar.

Normalmente, cargar datos en un perceptrón multicapa de esta manera requiere el uso de más neuronas ocultas de las que realmente se necesitan, resultando que las contribuciones no deseadas en el espacio de entrada debido al ruido se almacenen en los pesos sinápticos de la red. Un ejemplo de cómo puede ocurrir una pobre generalización debido a la memorización en una red neuronal se ilustra en la **Fig. 42 b** para los mismos datos representados en la **Fig. 42 a**. La "memorización" es esencialmente una "tabla de búsqueda", lo que implica que la asignación de entrada-salida calculada por la red neuronal no es uniforme. Como se señala en **Poggio y Girosi (1990 a)**, la uniformidad del mapeo entrada-salida está estrechamente relacionada con los criterios de selección del modelo, cuya esencia es seleccionar la función "más simple" en ausencia de cualquier conocimiento previo. En el contexto de la discusión actual, la función más simple significa la función más suave que se aproxima al mapeo para un criterio de error dado, porque tal elección generalmente requiere la menor cantidad de recursos computacionales. La suavidad también es natural en muchas aplicaciones, dependiendo de la escala del fenómeno que se estudia. Por lo tanto, es importante buscar un mapeo no lineal suave para las relaciones de entrada y salida mal planteadas, de modo que la red sea capaz de clasificar los patrones nuevos correctamente con respecto a los patrones de entrenamiento (**Wieland y Leighton, 1987**).

Figura 42. a) Datos propiamente ajustados (buena generalización). b) Datos sobre ajustados (generalización pobre).



Modificado de Haykin, S. (1999). *óp. cit.*, p.207.

Tamaño suficiente del conjunto de entrenamiento para una generalización válida

La generalización está influenciada por tres factores (Haykin, S., 1999):

- (1) el tamaño del conjunto de entrenamiento y cuán representativo es el entorno de interés,
- (2) la arquitectura de la red neuronal y
- (3) la complejidad física del problema en cuestión.

Claramente, no se tiene control sobre este último. En el contexto de los otros dos factores podemos ver el problema de la generalización desde dos perspectivas diferentes (Hush y Horne, 1993):

- La arquitectura de la red es fija (con suerte de acuerdo con la complejidad física del problema subyacente), y el problema a resolver es el de determinar el tamaño del conjunto de entrenamiento necesario para que ocurra una buena generalización.

- El tamaño del conjunto de entrenamiento es fijo, y el tema de interés es determinar la mejor arquitectura de red para lograr una buena generalización.

Ambos puntos de vista son válidos en sus formas individuales. En la presente discusión, se enfocará en el primer punto.

En particular, se tienen fórmulas en el peor de los casos para estimar el tamaño de la muestra de entrenamiento que son suficiente para un buen rendimiento de generalización. Desafortunadamente, a menudo se encuentra que existe una gran brecha numérica entre el tamaño de la muestra de entrenamiento realmente necesaria y la predicha por estas fórmulas. Es esta brecha la que ha convertido el problema de la complejidad de la muestra en un área de investigación abierta continua. En la práctica, parece que todo lo que realmente se necesita para una buena generalización es que el tamaño del conjunto de entrenamiento, N , satisfaga la condición:

$$N = O \left[\frac{w}{\epsilon} \right] \dots (4.40)$$

donde w es el número total de parámetros libres (es decir, pesos y sesgos sinápticos) en la red, y ϵ indica la fracción de errores de clasificación permitida en los datos de prueba (como en la clasificación de patrones) y $O(\cdot)$ indica el orden de la cantidad encerrada adentro. Por ejemplo, con un error del 10 por ciento, el número de ejemplos de entrenamiento necesarios debería ser aproximadamente 10 veces mayor al número de parámetros libres en la red.

En el presente trabajo, se aborda este problema mediante el uso de la prueba Gamma, la cual también nos permite conocer en número mínimo de datos necesarios, para poder llevar a cabo la construcción del modelo no lineal, con base al análisis llevado a cabo de los valores del estadístico Gamma, Γ , mediante la prueba M .

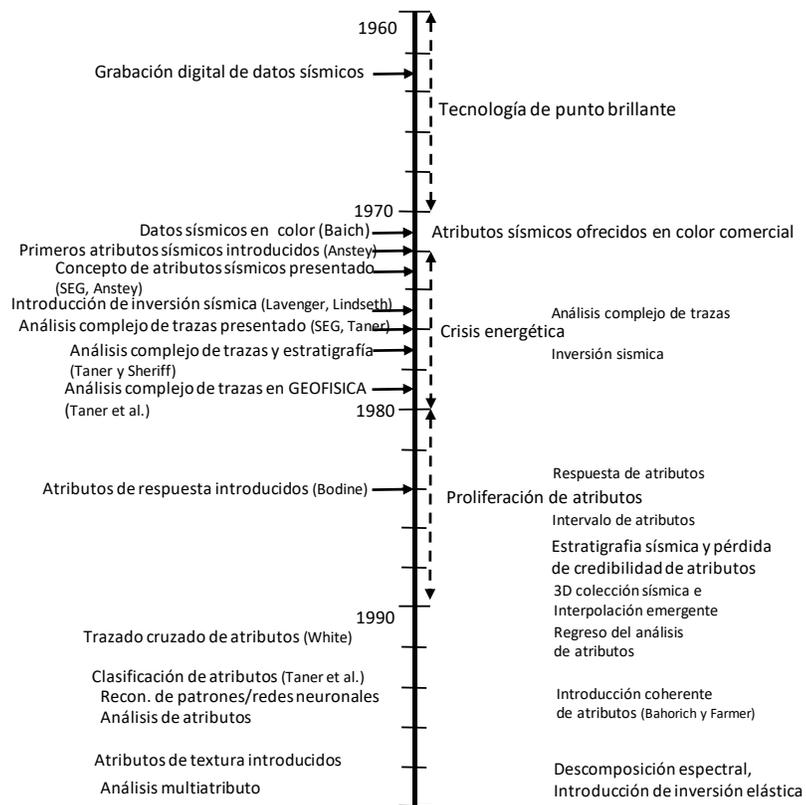
CAPITULO V

ATRIBUTOS SÍSMICOS

5.1 Perspectiva Histórica.

La evolución de los atributos sísmicos en la industria petrolera ha estado determinada por el desarrollo tecnológico. En la **Fig. 43** se muestra la evolución de éstos en el siglo XX. Desde su descubrimiento durante los primeros años de los 60's, el cual fue meramente fortuito, asociado al descubrimiento de los puntos brillantes (Bright Spots); su proliferación durante los años 70's; la pérdida de credibilidad de éstos por parte de las geociencias entre otras muchas cosas por la poca comprensión de éstos; y finalmente su resurgimiento durante la última década del siglo XX, a la par del uso extendido de la adquisición sísmica 3D.

Figura 43. Línea de tiempo del desarrollo de los atributos sísmicos, y su relación con los avances claves en la tecnología de exploración sísmica.



Adaptado de Satinder Chopra and Kurt J. Marfurt. Seismic attributes — A historical perspective.p.3

No hay duda de que la interpretación sísmica es una mezcla de arte y ciencia. Aunque los intérpretes, generalmente geofísicos, llevan a cabo la identificación de características del yacimiento de interés asociados a ciertos patrones sísmicos, con base a principios geológicos y matemáticos, muchas veces éste lo lleva a cabo ya de manera empírica y creativa. De hecho, esta peculiar característica es lo que diferencia un intérprete experimentado de uno amateur que suele actuar fundamentalmente con sustento a su formación académica. Y es precisamente este hecho, es decir, la necesidad de poder capturar esa manera intuitiva de los intérpretes del reconocimiento de patrones, a través de la cuantificación de la amplitud y de las características morfológicas presentes en los datos sísmicos, que surge el análisis multiatributo.

Desde 1975, 3 atributos principales - Envolvente, fase y frecuencia, habían sido establecidos (**Chopra, 2003**):

- ✓ **Envolvente instantánea (magnitud de la reflexión).**
- ✓ **Fase instantánea.**
- ✓ **Frecuencia instantánea.**

La **envolvente** es la envolvente de la señal sísmica. Tiene una apariencia de baja frecuencia y solo amplitudes positivas. A menudo destaca las principales características sísmicas. La envolvente representa la energía instantánea de la señal y es proporcional en su magnitud al coeficiente de reflexión (**Subrahmanyam et al., 2008**).

La envolvente es útil para resaltar discontinuidades, cambios en la litología, fallas, cambios en la deposición y límites de secuencia. También es proporcional a la reflectividad y, por lo tanto, es útil para analizar las anomalías de AVO (**Chopra, 2007**). Si hay dos volúmenes que difieren solo por el desplazamiento de fase constante, sus envolventes serán las mismas.

Este atributo representa principalmente el contraste de impedancia acústica, por lo tanto, la reflectividad. Este atributo es principalmente útil para identificar (**Subrahmanyam et al., 2008**):

- ✓ Puntos brillantes.
- ✓ Acumulación de gas.
- ✓ Límites de secuencia, cambios importantes o ambientes deposicionales.
- ✓ Efectos de ajuste de capa fina.
- ✓ Discordancias.
- ✓ Principales cambios de litología.
- ✓ Cambios locales que indican fallas.
- ✓ Correlación espacial con la porosidad y otras variaciones litológicas.

La **fase instantánea** se mide en grados ($-\pi, \pi$). Es independiente de la amplitud y muestra continuidad y discontinuidad de eventos (**Subrahmanyam et al., 2008**). Muestra la estratificación muy bien.

La fase a lo largo del horizonte no debería cambiar en principio, pueden surgir cambios si existiera algún problema durante su picado, o si la capa cambia lateralmente debido a "agujeros de hundimiento" u otros fenómenos.

Este atributo es útil como (**Subrahmanyam et al., 2008**):

- ✓ Mejor indicador de continuidad lateral.
- ✓ Se relaciona con el componente de fase de la propagación de la onda.
- ✓ Se puede usar para calcular la velocidad de fase.
- ✓ No contiene información relacionada con la amplitud, por lo tanto, todos los eventos están representados.
- ✓ Muestra discontinuidades, pero puede no ser el mejor. Es mejor para mostrar continuidad y límites de secuencia.
- ✓ Visualización detallada de configuraciones de las capas.
- ✓ Se utiliza en el cálculo de frecuencias instantáneas y aceleración.

La **frecuencia instantánea** es la derivada en el tiempo de la fase, es decir, la tasa de cambio de la fase.

La frecuencia instantánea representa la amplitud media de la ondícula. Este atributo es empleado ampliamente para (**Subrahmanyam et al., 2008**):

- ✓ Indicar el espesor de la capa y también los parámetros de litología.
- ✓ Corresponde a la frecuencia promedio (centroide) del espectro de amplitud de la ondícula sísmica.
- ✓ Indica los límites de las capas delgadas de baja impedancia,
- ✓ Indicador de hidrocarburos por anomalía de baja frecuencia. Este efecto es algunas veces acentuado por las arenas no consolidadas debido al contenido de aceite de los poros.
- ✓ Indicador de zona de fractura, aparecen como zonas de frecuencia más baja.
- ✓ Indicador de zona de reflexión caótica,
- ✓ Indicador de espesor de capa. Las frecuencias más altas indican capas delgadas de arcilla, mientras que las frecuencias más bajas indican capas con arena abundante.

Estos atributos son los conocidos también como atributos de traza compleja. El concepto de trazas compleja fue descrito por primera vez por **Taner, 1979**. La traza compleja se define como:

$$CT(t) = T(t) + iH(t) \dots 5.1$$

donde:

CT(t)= Complemento de traza.

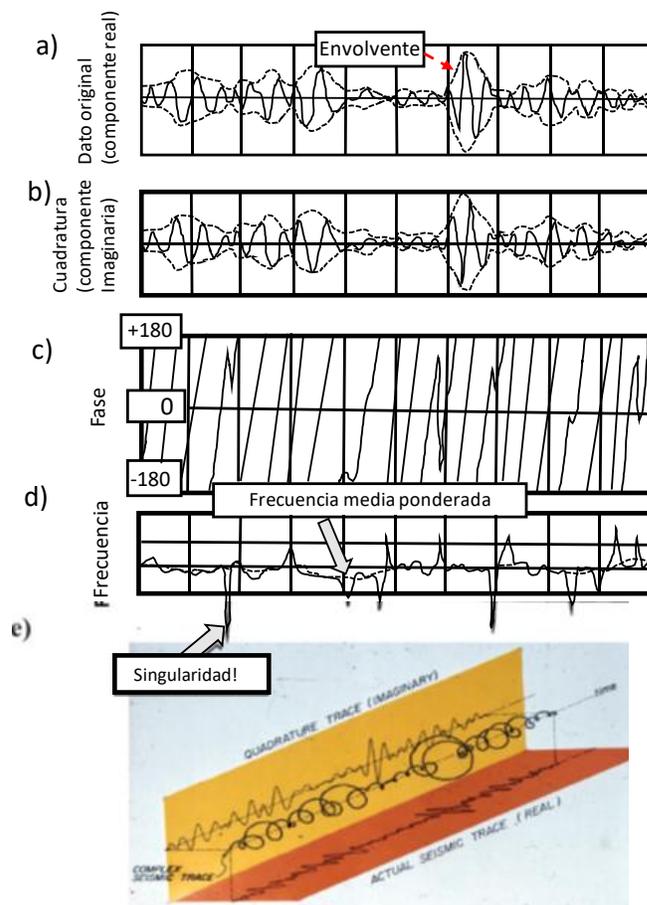
T(t)= Traza sísmica.

H(t)= Transformada de Hilbert T(t)

H(t) el cambio de fase es a 90° de T(t)

La amplitud de la traza sísmica se trata como la parte real de la señal analítica (compleja), mientras que la parte imaginaria de la señal se calcula tomando su transformada de Hilbert. En la **Fig. 44** se ilustra gráficamente el concepto de traza compleja propuesto por **Taner**.

Figura 44. a) Parte real de la traza sísmica, b) cuadratura, c) fase instantánea, d) frecuencia instantánea de acuerdo con lo desarrollado por **Taner et al. (1979)**. Nótese la frecuencia promedio ponderada indicada por la línea punteada para el inciso d. También nótese las singularidades que se presentan en la frecuencia instantánea debido a las interferencias de la forma de onda. E) Una copia escaneada de la diapositiva usada por **Taner** en su presentación hecha durante 1970 para explicar el análisis de la traza compleja.



Adaptado de Satinder Chopra and Kurt J. Marfurt. óp. cit., p. 8

Hoy en día, los atributos sísmicos agregan valor en el análisis estructural, estratigráfico, y así como también, en la predicción de facies y propiedades de los fluidos del yacimiento.

El uso ampliamente extendido de los atributos sísmicos en la actualidad se debe a dos factores principalmente (**Chopra, 2007**):

- 1- A los desarrollos contemporáneos sobre estudios en Física de rocas, lo cual proporcionó la base cuantitativa de como las propiedades de las rocas afectan los

datos sísmicos, permitiendo así relacionar a los atributos directamente con las propiedades de las rocas en una manera mucho más confiable, en comparación a la que se daba durante los años 80's.

- 2- Al desarrollo tecnológico, no solo de visualización, almacenamiento, sino también de procesamiento de los datos sísmicos.

5.2 Definición.

En el sentido más general, un atributo sísmico abarca todas las cantidades derivadas de los datos sísmicos. La definición dada a continuación, es una combinación de los conceptos presentados primeramente por **Tener (1977; 2000; 2001)**, **Barnes (2001)**, **Chopra y Marfurt (2005)**;

Un atributo sísmico puede ser definido como una transformación, lineal o no lineal de la traza sísmica, la cual puede o no incorporar otras fuentes de datos; Los atributos sísmicos son medidas específicas de la geometría, cinemática, de las características dinámicas o estadísticas de los datos sísmicos.

Con base a la definición anterior, es importante que se reconozca que estas estimaciones están de alguna manera contaminadas por errores y, por lo tanto, requieren de calibración con los datos de pozo mediante geoestadística u otras técnicas de integración de datos.

5.3 Clasificación.

El crecimiento descomunal, no solo en cantidad sino también en variedad, de los atributos sísmicos en las última tres décadas, justifica una clasificación sistemática, así como también, se necesita un enfoque sistemático para comprender el uso de cada uno de estos atributos y también sus limitaciones bajo diferentes circunstancias. Es por ello por lo que muchos autores han intentado clasificarlos en familias. Para poner este crecimiento en perspectiva, las ediciones de 1984, 1991 y 2002 de su diccionario enciclopédico de geofísica de Bob Sheriff contienen las siguientes entradas sobre los atributos: La edición de 1984 dedicó 26 líneas a atributos, análisis de trazas complejas e indicadores de hidrocarburos, además de una figura completa de página de análisis de trazas complejas. La edición de 1991 cubrió las mismas tres categorías y una tabla de indicadores de hidrocarburos (Fig. H-7), para un total de 48 líneas más la misma imagen de página completa de la edición de 1984. La edición de 2002 contiene 153 líneas en total (69 líneas de texto, una tabla de 34 líneas, 9 líneas sobre coherencia, 25 líneas en indicadores de hidrocarburos y 16 líneas en análisis de trazas complejas, más 6 figuras en atributos) (**Chopra, 2005**).

Taner et al. (1994) fue el primero en proponer una clasificación real y coherente de los atributos sísmicos. Éste divide los atributos en dos categorías generales: *geométricos* y *físicos*.

Los *atributos físicos* se definen como aquellos atributos que están directamente relacionados con la propagación de la onda, la litología y otros parámetros (**Chopra, 2005**). Estos atributos físicos se pueden clasificar además como atributos **pre-apilamiento** y **post-apilamiento**.

Los *atributos geométricos* son echado, azimut y discontinuidad. El atributo de echado o de la amplitud de los datos corresponde a los echados que presentan los eventos sísmicos. El echado es útil porque hace que las fallas sean más perceptibles. La amplitud de los datos en el atributo de azimut corresponde al azimut en la dirección del máximo echado del evento sísmico (**Chopra, 2005**).

Brown (1996 b, 2004) clasificó los atributos usando una estructura de árbol con el tiempo, la amplitud, la frecuencia y la atenuación como las ramas principales, que se subdividen aún más en categorías de post-apilamiento y pre-apilamiento. Los atributos de tiempo proporcionan información en la estructura, mientras que los atributos de amplitud proporcionan información en estratigrafía y del yacimiento (**Fig. 45**).

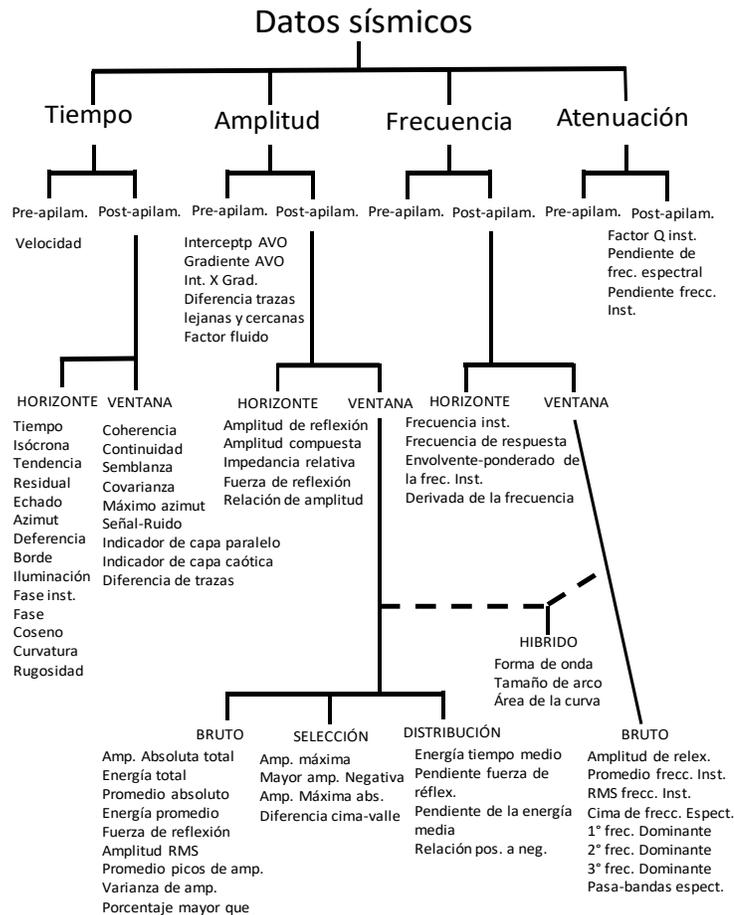
Otra clasificación, es la propuesta por **Chen y Sidney (1997)**, en la cual dividen los atributos sísmicos en dos categorías:

1) *atributos basados en el horizonte*, que vienen representando las propiedades promedio de la traza sísmica entre dos límites, generalmente definidos por horizontes picados por los intérpretes, los cuales son muy útiles cuando las formaciones son muy heterogéneas.

2) *atributos basados en muestras*, son transformaciones de la traza de entrada de tal forma que se produzca otra traza de salida con el mismo número de muestras que la entrada.

Una de las clasificaciones, que muchos prefieren es la clasificación de **Liner et al. (2004)**, en la cual establecen *categorías generales y específicas*. Los *atributos generales* de **Liner et al.** son medidas de características geométricas, cinemáticas, dinámicas o estadísticas derivadas de datos sísmicos. Incluyen la amplitud del reflector, tiempo del reflector, echado del reflector y azimut, amplitud y frecuencia compleja, atributos generalizados de Hilbert, iluminación, detección de límites/ coherencia, AVO y descomposición espectral. Estos atributos generales se basan en el carácter físico o morfológico de los datos vinculados a la litología o geología y por lo tanto son generalmente aplicables de una cuenca a otra en todo el mundo (**Chopra, 2005**). Por el contrario, los *atributos específicos* tienen una base menos definida en aspectos físicos o geológicos. Mientras un atributo específico dado puede estar bien correlacionado con una característica geológica o a la productividad del yacimiento dentro de una cuenca dada, estas correlaciones, en general, no se transfieren a una cuenca diferente (**Chopra, 2005**). Hay literalmente cientos de atributos específicos. Para que esta clasificación sea aún mejor, y contemple las nuevas tendencias de atributos desarrollados actualmente, podemos agregar una tercera categoría; la de *atributos "compuestos"*, también llamados *atributos meta* por **Meldahl et al. (2001)**. Muchos de los atributos específicos citados en la literatura son sumas, productos u otras combinaciones de atributos generales más fundamentales.

Figura 45. Clasificación de los atributos sísmicos según Brown (1996 b, 2004).



Modificado de Brown, 2001.

En este documento, consideramos solo los atributos basados en muestras y éstos se combinarán mediante la red neuronal (atributos compuestos) para la predicción de valores de las propiedades petrofísicas ya mencionadas. En el **apéndice B** se incluyen la parte matemática de la mayor parte de los atributos usados en el entrenamiento de la red. (Una clasificación completa, así como la base matemática de los atributos sísmicos, puede ser encontrado en *Seismic Attributes for Prospect Identification and Reservoir Characterization* por Satinder Chopra).

5.3.1 Criterios para la selección de atributos.

Entre más atributos sísmicos disponibles, hay más confusión seleccionando los apropiados. Una práctica común para identificar qué atributos sísmicos usar, es basándose únicamente en la fuerza de sus correlaciones observadas con las propiedades del yacimiento. Como con cualquier parámetro, cuando el tamaño de muestra es pequeño, la incertidumbre sobre el valor de la verdadera correlación puede ser bastante grande. Un problema menos conocido es lo que se nombra en inglés **experiment-wise error rate** (Parra et al, 2015). Esto sólo

significa que a medida que generamos un aumento constante del número de atributos, mayor es la posibilidad de observar al menos una correlación espuria en una muestra grande. Una correlación espuria es una correlación de muestra que es grande en valor absoluto, meramente por casualidad. Esto ha sido señalado en **Kalkomey (1997)**. La selección de los atributos apropiados no es una tarea nada fácil, de hecho, en la actualidad se usan herramientas de análisis estadístico para llevarlo a cabo, como es nuestro caso (Prueba Gamma), o bien se pueden seguir ciertos criterios, los cuales que, si bien son muy útiles, no cuentan con ningún sustento matemático, sino que más bien son meramente empíricos, es decir, son heurísticas.

A continuación, se mencionarán algunos de ellos (**Barnes, 2006**):

- ✓ Los atributos sísmicos deben ser únicos. Solo necesita un atributo para medir una propiedad sísmica determinada. Descartar atributos duplicados. Cuando múltiples atributos miden la misma propiedad, elija la que mejor funcione. Si no puede decir cuál funciona mejor, no importa cuál elija.
- ✓ Los atributos sísmicos deben tener significados claros y útiles. Si no sabe lo que significa un atributo, no lo use. Si se sabe lo que significa, pero no es útil, deséchalo. Prefiera atributos con significado geológico o geofísico; evitar los atributos con un significado puramente matemático.
- ✓ Los atributos sísmicos representan subconjuntos de la información en los datos sísmicos. Las cantidades que no son subconjuntos de los datos no son atributos y no deben usarse como atributos.
- ✓ Los atributos que difieren solo en resolución son el mismo atributo; se deberá tratarlos de esa manera.
- ✓ Los atributos sísmicos no deberían variar mucho en respuesta a pequeños cambios en los datos. Evite atributos demasiado sensibles.
- ✓ No todos los atributos sísmicos son creados de igual manera. Evite atributos mal diseñados.

Muchos de estos criterios fueron los usados durante revisión bibliográfica exhaustiva llevada a cabo, para preseleccionar los atributos para la predicción de las propiedades petrofísicas.

CAPITULO VI

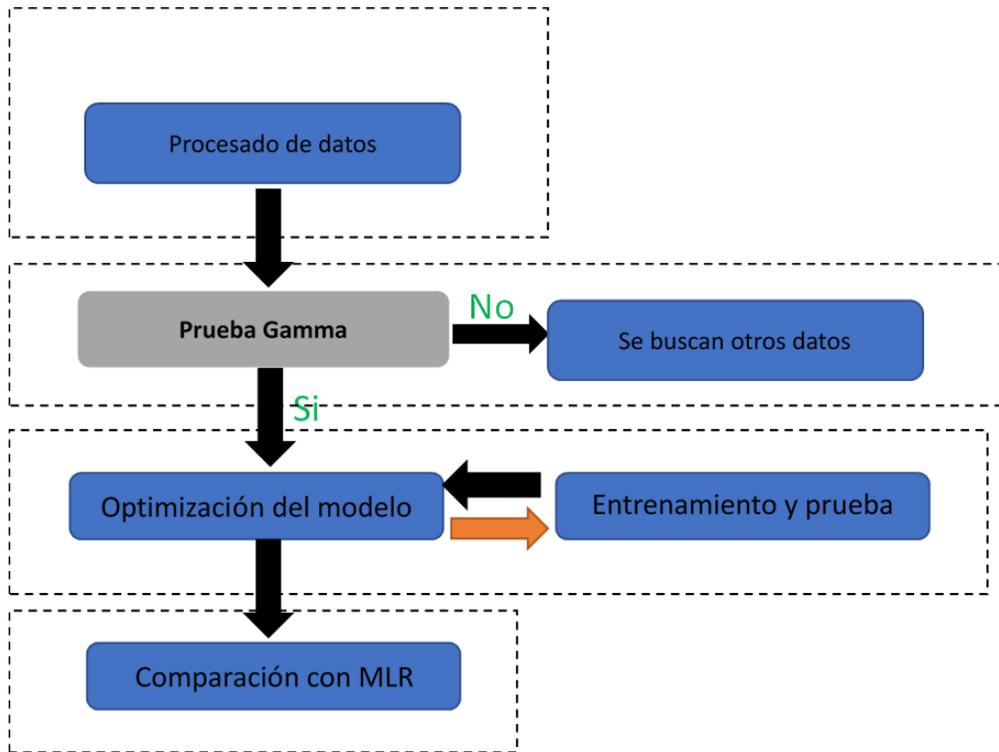
METODOLOGÍA

El objetivo general de desarrollar cualquier modelo de predicción es construir un modelo que resuelva el problema bajo investigación al nivel de precisión requerido (**Al Bulushi et al., 2009**). En la presente tesis, se desarrolla un flujo de trabajo sistemático para optimizar no sólo la construcción de una red neuronal sino también su desempeño, cuyo elemento principal de esta metodología es el empleo de la prueba Gamma. Esta metodología se basa en la propuesta por **Al Bulushi et al. (2010)**, y el análisis estadístico de los datos que se propone sigue el enfoque dado en **Parra et al. (2015)**.

La metodología propuesta se muestra en la **Fig. 46**. Como se puede observar en el diagrama de flujo de trabajo, ésta está constituida por 3 bloques principales, los cuales serán descritos de manera general y simple a continuación. El primer bloque básicamente consiste en el procesado de los datos, el cual consiste en la determinación y acondicionamientos de los datos que se usarán para la construcción del(os) modelo(s). En cuanto al segundo bloque, consiste en el análisis estadístico de los datos a través del empleo de la prueba Gamma a partir del cual se evaluará la factibilidad de la construcción del o de los modelos con los datos que se definieron en la etapa anterior (cantidad y calidad), en caso de ser viable, esta herramienta, también, proporcionará una forma de reducir la dimensionalidad del espacio de entrada mediante la selección de la mejor combinación de datos (atributos). Finalmente, el tercer bloque trata del entrenamiento-optimización de las redes neuronales, así como también, de la comparación de los resultados obtenidos con éstas y los obtenidos con otras técnicas lineales más simples, en este caso solo se compararán con los resultados obtenidos de la regresión lineal convolucional.

Una vez detalladas las bases teóricas, la metodología será aplicada en un campo real, donde se obtendrán secciones de las propiedades petrofísicas; Vol. de arcilla (Vcl), porosidad efectiva (ϕ_{ie}), velocidad de propagación de la onda P (V_p) y densidad de la roca (Rho).

Figura 46. Diagrama de la metodología propuesta para la obtención de modelos no lineales usando redes neurales y la prueba Gamma. Cada recuadro delineado con líneas punteadas constituye un bloque del flujo de trabajo de la metodología propuesta.



6.1 Procesado de los datos.

Esta etapa comprende la determinación de los datos a usar, así como su acondicionamiento. Así, el procesado de los datos incluirá las siguientes operaciones:

- ✓ **Definición de la propiedad y evaluación de los datos.**
- ✓ **Cálculo de los atributos sísmicos.**
- ✓ **Normalización y escalado de atributos**
- ✓ **Escalamiento de datos de registro de pozo.**

A continuación, se describe de manera detallada cada uno de los puntos anteriores.

Definición de la propiedad y evaluación de los datos.

Esta etapa comprende a la selección de aquellos atributos que responden a rasgos litológicos, lo cual se llevó a cabo bajo un enfoque completamente empírico, el cual se basa en la comprensión del problema bajo investigación.

Una vez que se definen los datos de entrada para entrenar el modelo, es importante evaluar la incertidumbre en los datos que serán usados en la salida de la red, como datos objetivo. Esto proporciona un límite de cuánto más debe optimizarse el modelo entrenado, en el sentido

que se tiene conocimiento de las incertidumbres propias de la herramienta y/o modelo usado para su determinación.

Cálculo de los atributos sísmicos.

Con base a los atributos matemáticos (traza compleja) y AVO descritos en el **apéndice B**, se calcula para la traza sísmica más cercana (Traza A536) al único pozo por la cual pasa la línea sísmica de estudio el conjunto de atributos.

Normalización y escalado de atributos.

Las estadísticas de los datos son un aspecto importante en el desarrollo de la red neuronal artificial. Es importante que los diferentes conjuntos de datos (entrenamiento y predicción) tengan características comparables. En caso de que los datos de operación presenten estadísticas muy diferentes, será necesario por llamarlo de alguna forma de una regularización de los datos de entrada (atributos en este caso en específico), la cual consiste básicamente en la normalización y escalado de estos. En este caso, los datos fueron escalados utilizando el método de media y desviación estándar (con una media de cero y desviación estándar de la unidad). Por cuestiones de tiempo, no se pudo investigar el impacto en el desempeño de las redes al considerar otros tipos de escalado de los datos.

Escalamiento de datos de registro de pozo.

Debido a que la densidad de las muestras por traza en un levantamiento sísmico 3-D es menor que en un registro de pozo, donde se toma una medida cada 15.24 cm, es necesario llevar un sobre escalamiento de los datos de registro de pozos a utilizar, tanto para el análisis estadístico llevado a cabo con la prueba Gamma, así como para el entrenamiento de las redes neuronales. Esto se realiza para que sea posible asociar de manera efectiva la propiedad petrofísica de interés obtenida de los registros o pseudo registros, a un evento específico en la traza sísmica.

6.2 Análisis estadístico de los datos a través de la prueba Gamma.

El análisis estadístico de los datos inicia con la determinación de la factibilidad de construcción de los modelos con la cantidad y calidad de los datos disponibles. Esto es posible llevarse a cabo con la prueba Gamma dada la naturaleza de ésta. La prueba Gamma funciona con la suposición de que, si dos puntos x' y x están muy juntos en el espacio de entrada, entonces sus salidas correspondientes y' y y deben estar cerca en el espacio de salida (**Parra et al,2015**). Si las salidas no están muy juntas, se considera que esta diferencia es por el ruido presente en los datos. Los vectores de entrada son los atributos sísmicos y la salida es cualquier de los parámetros petrofísicos que se quiera predecir, es decir, V_{cl} , ϕ_{hie} , V_p o ρ . Se supone que los vectores de entrada contienen factores que son útiles para influir en la salida. Una segunda suposición es que la relación subyacente del sistema bajo investigación es de la siguiente forma:

$$[V_{cl}, \rho_{hie}, V_p \text{ ó } \rho] = f(\text{atributos}) + n \dots 6.1$$

donde f es una función suave y n es una variable aleatoria que representa el ruido. El dominio de los posibles modelos está restringido a la clase de funciones suaves que tienen primeras derivadas parciales. Esta conjetura fue probada para una amplia clase de situaciones **Evans y Jones (2002)**, **Evans et al. (2002)** y **Evans y Jones (2008)**. Así también, se supone que el no determinismo en un modelo uniforme de entradas a salidas se debe a la presencia de ruido estadístico en las salidas, por lo que no todos los fenómenos que se podrían tratar de modelar caen en esta categoría (**Jones, 1998**). Por ejemplo, si el resultado que se intenta predecir a partir de observaciones es altamente probabilístico, entonces el modelo producido no será satisfactorio como herramienta de predicción.

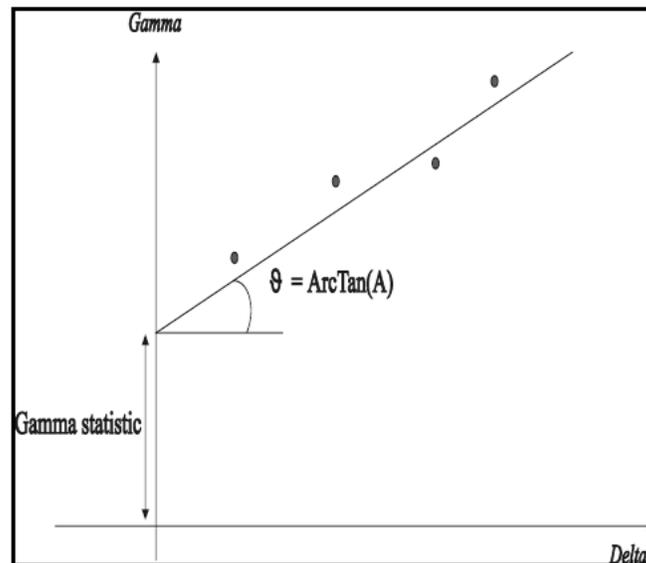
Con base a lo anterior, se puede decir que la prueba Gamma estima la varianza del ruido en cada salida. El parámetro que representa dicha varianza del ruido, el cual es el principal parámetro arrojado por esta prueba, es **la estadística Gamma Γ** . Así entonces, si la **estadística Gamma Γ** es pequeña, existe una fuerte relación predictiva entre las variables de entrada y la salida. Si Γ es grande, no existe una relación predictiva uniforme entre las entradas y la salida; es decir, las entradas son irrelevantes para la salida (**Jones, 1998**). De esta manera, el Γ es similar al **error cuadrático medio (por su acrónimo en inglés MSE)** alrededor de una línea de regresión. Sin embargo, la regresión lineal mide la precisión de ajuste de una línea recta, mientras que Γ mide la precisión de ajuste de cualquier y todas las curvas suaves. Dado que el valor de la estadística Gamma, Γ , es una estimación de la varianza del ruido en cada salida (aquella parte que no puede ser contabilizada por un modelo de datos uniforme), entonces, este parámetro puede ser empleado también como criterio para detener el entrenamiento de la red, antes que este empiece a modelar el ruido, y no la función que describe la relación de los datos de entrada-salida. Este criterio, es sumamente útil, lo cual es más plausible cuando se cuenta con pocos datos para la construcción del modelo (entrenamiento de la red), puesto que evita que los datos sean divididos en subconjuntos, entrenamiento-validación, que es una de las técnicas más empleadas para evitar el sobre entrenamiento de la red (**Amari S et al., 1997**), permitiendo con ello que se pueda extraer el máximo de información de los datos, haciendo posible con ello la construcción de un mejor modelo predictivo. Es importante señalar para este punto que se supone que la variación de ruido en cada una de las salidas está limitada y es independiente de los valores de entrada. Si la condición de independencia es falsa, esto no es necesariamente fatal, la prueba Gamma devolverá una varianza de ruido promedio en todo el espacio de entrada (**Jones, 1998**).

La estadística de Gamma se calcula al determinar una línea de regresión basada en las estadísticas del vecino derivado de los datos (ver **Fig. 47**). El diagrama de dispersión muestra pares de puntos (δ, γ), donde δ es la distancia al cuadrado de una entrada (x) desde uno de sus vecinos cercanos y γ es la mitad de la distancia cuadrada entre los dos valores correspondientes de salida escalar (y). Los puntos a los que se ajusta la línea de regresión se calculan al encontrar la media de $\delta(p)$ de δ y la de $\gamma(p)$ de γ , donde p se refiere al primer

vecino más cercano ($p = 1$), el segundo vecino más cercano ($p = 2$) y así sucesivamente hasta el número máximo de vecinos cercanos (p_{max}) que ha sido establecido por el usuario (Jones, 1998). Es importante mencionar, que para que los resultados arrojados por la prueba Gamma sean precisos se requiere determinar el número de vecinos más cercanos, primeramente, puesto que la prueba Gamma se basa en la teoría de los k - vecinos más cercanos. (En general, en un experimento de prueba Gamma se debe mantener el número de vecinos cercanos a menos de 30. Por lo general, 10-20 es una buena opción para iniciar el análisis estadístico.)

Una buena línea de regresión con puntos $(\delta(p), \gamma(p))$ aproximándose a $(\delta, \gamma) = (0, 0)$ indica que los valores de salida escalar de los vecinos cercanos de entrada están cerca. Si la línea de regresión tiene una pendiente pronunciada, esto indica que la función de modelado f que se busca aproximar es probable que sea bastante difícil de construir y se requerirá una gran cantidad de puntos de datos M . Si la línea es casi horizontal, la función es bastante sencilla (Jones, 1998).

Figura 47. Representación gráfica del estadístico Gamma, y su gradiente.



The WinGamma User Guide. Jones,1998. Cardiff: Universidad de Wales.p. 15

Como se puede ver en la Fig. 47, la estadística Gamma es en realidad la intersección vertical de la línea de regresión en la figura. Esta es la varianza estimada de los errores para cualquier modelo construido a partir de datos.

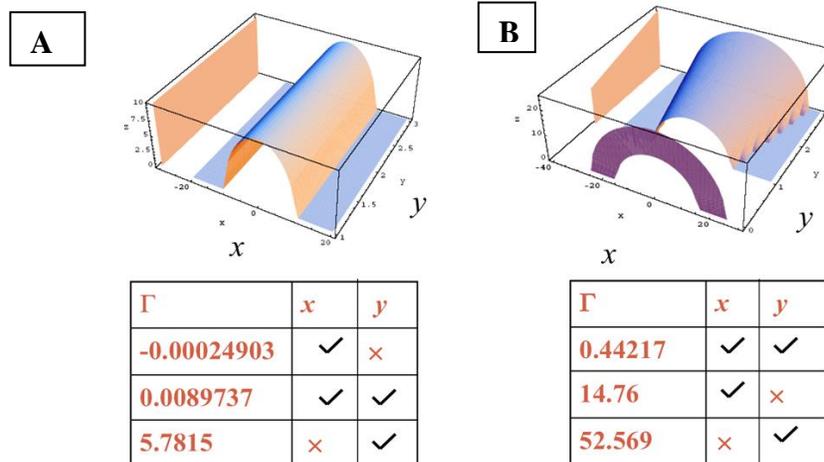
Para que quede aún más claro la funcionalidad de la prueba Gamma, imagínese que se tiene dos funciones desconocidas que representan las figuras A y B de la Fig. 48, respectivamente. Como se puede observar, ambas funciones son dependientes de las variables x y y ; no obstante, no es posible determinar a simple vista, que variable es la que define de manera predominante a las dos funciones. Sin embargo, esto es posible de acuerdo con lo dicho anteriormente de la prueba Gamma. Dado que el valor absoluto del estadístico Γ es menor solo empleando la variable y , ésta es la que predominantemente define a la función

desconocida que representa la figura A. Mientras que para la figura B, ambas variables son importantes para la construcción de dicha función.

En caso de que el resultado de este análisis de factibilidad sea que la varianza del ruido en el resultado que se intenta predecir es insatisfactoria, entonces se debe considerar, alguno de los siguientes puntos:

- ✓ **Aumentar la precisión de las medidas tanto de las entradas como de las salidas. La variación de ruido efectiva en la salida puede ser el resultado del error de medición en las entradas.**
- ✓ **Preguntarse si se ha incluido todas las variables principales de entrada causal que pueden afectar la salida. Si se ha pasado por alto algún factor obviamente importante, esto bien podría explicar por qué actualmente no se tenga éxito en predecir la variable de salida.**

Figura 48. Ejemplo del empleo de la prueba Gamma para la selección de los datos de entrada para la construcción de un modelo no lineal. Tanto la figura A como B están representados por dos funciones desconocidas dependientes de las variables x y, a través del valor absoluto del estadístico Γ obtenido al aplicar la prueba Gamma, se observa que la función representada por la figura A esta definida predominantemente por la variable y, mientras que la función que representa la figura B está definida por ambas variables.



Una vez que se ha comprobado que puede ser establecida una relación de los datos entrada-salida, es necesario determinar si se cuenta con la cantidad necesaria de datos para llevarlo a cabo. La teoría nos asegura que a medida que $M \rightarrow \infty$ el resultado Γ del algoritmo de prueba Gamma convergerá a la varianza del ruido verdadero con la probabilidad uno (**Evans y Jones, 2002**). Sin embargo, esto en sí mismo no nos dice qué tan grande M debe ser (tamaño del conjunto de datos de entrenamiento) para dar una estimación precisa de la verdadera variación de ruido. Realmente se necesita saber cuán rápido se estabilizará la estimación devuelta por el algoritmo a una aproximación cercana a la verdadera varianza del ruido. Una manera simple de lograr esto es calcular la estadística Gamma, Γ , aumentando M y

examinando el gráfico resultante para determinar si el gráfico parece acercarse a una asíntota estable, este procedimiento se denomina **prueba M (Končar, 1997)**.

Finalmente, el análisis de los datos mediante el empleo de la prueba Gamma termina con el acotamiento del espacio de entrada, es decir, seleccionando aquella combinación de datos, que permita la mejor estimación de la salida. La selección de las variables de entrada apropiadas es un problema importante no sólo en el modelado no lineal mediante el empleo de redes neuronales, sino en general para cualquier método empleado para su generación. Cuando se seleccionan más entradas que las requeridas, esto dará como resultado un gran tamaño de red y, en consecuencia, disminuirá la velocidad de aprendizaje y la eficiencia del método, y reducirá la capacidad de generalización. Por otro lado, seleccionar sólo algunos pocos datos de entrada podría no ser suficiente para modelar el problema bajo investigación. Hay varios enfoques, que se pueden emplear para dar solución a este problema de reducción del espacio de entradas, algunos de estos son; la **regresión inteligente**, **análisis de componentes principales**, etc. No obstante, si bien proporcionan una solución a dicho problema, no quiere decir que sea la mejor solución. La solución proporcionada por la prueba Gamma es superior a muchas de las que se aplican actualmente, en el sentido que permite probar cada combinación de entradas para determinar qué combinación produce el valor del estadístico Gamma absoluto más pequeño, esto es, aquella combinación de entradas que presentan una mayor relación con las salidas. Así, por ejemplo, si se tiene m entradas escalares, entonces habrá $2^m - 1$ posibles combinaciones para la construcción del modelo, que serán evaluadas con esta prueba para elegir la mejor (**Γ menor**).

La **Fig. 49** nos muestra el flujo de trabajo para llevar a cabo el análisis estadístico mediante la prueba Gamma, y que fue descrito ya anteriormente.

En el **apéndice D** se proporcionan más detalles de la prueba Gamma.

Figura 49. Flujo de trabajo para llevar a cabo el análisis de datos con la prueba Gamma.



6.3 Estructura de la red.

Desarrollar una estructura de red neuronal es el paso más difícil en el modelado de las redes. Los siguientes cuatro principales parámetros deben determinarse:

- **La cantidad de neuronas en las capas de entrada y salida.**
- **La cantidad de capas ocultas y el número de neuronas en estas capas.**
- **Seleccionar los criterios de detenimiento del entrenamiento.**
- **Seleccionar el algoritmo de aprendizaje de optimización.**
- **El tipo de activación de funciones ocultas y de salida capas.**

El desarrollo de una estructura de red depende completamente del problema ya que los diferentes problemas requieren estructuras diferentes. La cantidad de neuronas en las capas de entrada y salida está determinada por la naturaleza del problema. Por lo tanto, se pueden determinar a partir del número de variables (datos) de entrada y salida. La determinación de la cantidad de capas ocultas y sus neuronas es la tarea principal en la estructura de la red. Se pueden usar una capa o dos capas ocultas, dependiendo de la complejidad del problema y la cantidad de datos disponibles para construir el modelo.

Si bien es cierto que existen diferentes enfoques para determinar el número de neuronas en las capas ocultas, es un hecho que todavía hay cabida para optimizar este peldaño para la construcción de redes neuronales.

La regla esencial y que es respetada en todos los enfoques usados en la determinación del número de neuronas en las capas ocultas es que el número de parámetros libres sea menor que el número de datos usados para el entrenamiento (**Huang, 1991**). El número de parámetros libres (pesos) es fácilmente determinable (**Ec. 6.1**). En la práctica el número de datos a usar para el entrenamiento (m) de las redes es $m \sim 10n_t$ (**Huang, 1991**), esto para conservar el sistema subestimado.

$$n_t = (n_i + 1) * n_{h1} + (n_{h1} + 1) * n_0 \dots \dots 6.1$$

donde n_i es el número de datos de entrada, n_{h1} número de neuronas en la i -ésima capa oculta.

El número óptimo de neuronas ocultas se puede obtener mediante un método de prueba y error, el cual termina al alcanzar un error aceptable durante el entrenamiento. Pero la gran pregunta es ¿Cuál es un error aceptable? Bueno, la prueba Gamma una vez más proporciona la respuesta a esta interrogante que no es nada trivial. La prueba Gamma al proporcionar, el ruido de la función (la cual se desconoce) modelada a partir de nuestros datos de entrada-salida, a través del valor del estadístico Gamma, permite determinar el error mínimo alcanzable, lo cual con base a ello se determina el número de neuronas en las capas ocultas, es decir, se colocarán tantas neuronas sean necesarias en las capas ocultas, para alcanzar dicho error, o al menos a cercarse a él, siempre y cuando el número de parámetros libres sea menor que el número de datos.

En cuanto a la determinación del número de capas, este dependerá esencialmente de dos aspectos esencialmente:

- **Número de datos disponibles.**
- **Complejidad de la relación entrada-salida, lo cual está definido parcialmente por el gradiente arrojado por la prueba Gamma (Tsui, 1999 a).**

Los ciclos de entrenamiento (épocas) junto con el error mínimo a alcanzar (valor del estadístico Gamma para la mejor combinación de datos) se usarán como criterio de detención, por las razones ya mencionadas en secciones anteriores.

En cuanto a la selección de las funciones de activación, se tomará la función sigmoideal asimétrica, que ha demostrado ser más eficiente. (Haykin, S., 1999)

6.4 Entrenamiento y validación de la red.

Una vez que se determina la arquitectura de la red, la red está lista para entrenarse y validarse. Los datos operativos se utilizan para entrenar el modelo con la cantidad mínima predeterminada de neuronas ocultas. El entrenamiento se realiza varias veces, cada una con una inicialización de pesos diferentes. Esto es para garantizar el inicio en un punto diferente en la superficie de error para minimizar el efecto de los mínimos locales. Dada la poca información con la que se cuenta, y a la usencia de homogeneidad lateral en el área que nos hubiera permitido proyectar otro pozo cercano y validar así la capacidad de generalización de las redes de una manera más rigurosa, ésta se lleva acabo meramente de forma matemática como se explica a continuación. Primeramente, cada uno de los modelos se valida a través de la *validación cruzada*³, donde se toma el 85% de los datos para entrenar, y el resto para validar. Estos puntos de validación se toman de manera aleatoria. Una vez llevada a cabo esta validación en cada una de las redes para cada uno de los parámetros petrofísicos a predecir, se selecciona aquella red cuyo desempeño es el peor, para realizar un segundo entrenamiento, y validándolo una vez más con la técnica anterior, pero en esta ocasión, tomando los puntos ciegos de manera que correspondan a aquel intervalo de tiempo (profundidad) donde se presentan los mayores errores durante el primer entrenamiento. Esto se lleva a cabo bajo la premisa de que, si se presentan nuevamente errores bajos en los datos de validación en este nuevo entrenamiento, la capacidad de generalización de esta red estará comprobada (al menos matemáticamente), y consecuentemente, la capacidad de generalización del resto de las redes, al ser ésta la que tuvo el peor desempeño durante el primer entrenamiento.

³ Es una técnica estándar, que consiste en repetir el entrenamiento n veces. En cada entrenamiento, el conjunto de datos de entrenamiento se divide en p subconjunto de datos de manera aleatoria, reservando uno de esos p subconjuntos para la validación, y el resto p-1 para el entrenamiento de la red (Iturrarán, 2011).

La selección del algoritmo de aprendizaje usado para el entrenamiento de la red dependerá tanto de la cantidad de datos disponibles, complejidad de la relación que se intenta establecer, hardware disponible, primordialmente.

Los resultados se analizan utilizando el error cuadrático medio (MSE) y el coeficiente de correlación (r).

6.5 Optimización del modelo.

Es importante ejecutar diferentes análisis de sensibilidad para investigar si el modelo se puede optimizar aún más. El análisis de sensibilidad incluye lo siguiente (Al Bulushi et al., 2009):

- 1) **Probar la cantidad óptima de neuronas ocultas.**
- 2) **Probar diferentes tipos de escalado.**
- 3) **Prueba con diferentes algoritmos de aprendizaje.**

En este trabajo, solo se pusieron en práctica los puntos 1 y 3.

6.6 Comparación de la red neuronal con modelo de regresión estadística convencional; Regresión lineal multivariable convolucional.

Finalmente, el flujo de trabajo termina comparando los resultados obtenidos con el empleo de las redes neuronales artificiales con otras técnicas de modelado, esto para demostrar la superioridad de los métodos no lineales frente a métodos lineales más simples. En este trabajo solo se comparará con la regresión múltiple lineal convolucional. Este tipo de regresión se programó en Matlab, y se presentan los fundamentos matemáticos en el **apéndice C**.

Cabe mencionarse, que se usó el mismo tipo de entradas que el de las redes neuronales, las cuales fueron determinadas por la prueba Gamma.

CAPITULO VII

RESULTADOS

Como se pudo observar el flujo de trabajo es demasiado extenso, por ello se dividirán los resultados y su correspondiente análisis en dos partes; una correspondiente a la prueba Gamma, y la segunda lo correspondiente a la construcción del modelo no lineal (entre atributos sísmicos y datos de pozo) mediante el empleo de redes neuronales, que permitieron la construcción de las secciones de propiedades petrofísicas, definidas ya anteriormente.

Para la construcción del modelo se utilizaron como datos de entrada **19 atributos instantáneos** derivados de datos sísmicos sensibles a rasgos litológicos, enumerados en la **Tabla 5** y descritos en el **apéndice B**. Estos atributos conforman una clase llamada **atributos basados en muestras**. Estos corresponden a transformaciones de la traza de entrada en tal manera de producir otra traza de salida con el mismo número de muestras que la entrada. Es importante mencionar, que se optó analizar no los datos sísmicos en sí, sino los atributos extraídos de éstos, dado que la naturaleza de muchos de éstos es no lineal, lo que aumenta el poder predictivo de la técnica. Una segunda razón, es que a menudo se obtienen beneficios al desglosar los datos de entrada en componentes. Este proceso se denomina preprocesamiento o extracción de características y, a menudo, puede mejorar enormemente el rendimiento de un sistema de reconocimiento de patrones al reducir la dimensionalidad de los datos antes de usarlos para entrenar el sistema (**Hampson et., al, 2001**). Cabe mencionar también, que los atributos usados para el entrenamiento de la red corresponden a la traza más cercana al pozo, traza 532, y corresponden al intervalo 1398-2860 m (1000-1864 ms), siendo igual a 433 muestras de datos por cada atributo.

Tabla 4. Conjunto de atributos sísmicos considerados para construir los modelos.

Atributos Sísmicos
1.Tiempo
2.Amplitud traza
3.Varianza
4.Tatenuation
5.Sweetness
6.StructuralS
7.RMS Amplitudes
8.InstantanusB
9.LocalS
10.Localflatness
11.Isofrequency
12.Echado Iluminado
13.Dip_Deviation
14.Chaos
15.Amplitud
16.Edge
17.Envelope
18. AVO
19-First_Derivative

En la **Fig. 50** se muestran gráficos cruzados de algunos de los atributos que fueron tomados inicialmente. En esta se puede observar que algunos de estos están relacionados entre sí, no

obstante, no es algo que nos deba preocupar, en el sentido que la prueba Gamma descarta aquellos atributos que son dependientes durante el proceso de selección de la mejor combinación de atributos para la construcción de los modelos. Mientras que en la **Tabla 6** se presentan las medidas de dispersión más comunes, tanto de los datos usados para el entrenamiento (traza 532), así como para los atributos del resto de las trazas (los cuales se identifican por el sufijo one). Comparando las medidas de dispersión, se puede observar que difieren mucho éstas entre los diferentes tipos de atributos, es por ello de la necesidad e importancia de llevar a cabo el preprocesado de los datos, señalado en el diagrama de trabajo.

Los datos de salida que fueron usados para guiar a las redes corresponden a las propiedades petrofísicas medidas directamente de los registros geofísicos, para el caso de Vp y Rhob, mientras que los valores de phie y Vcl, se obtuvieron de la evaluación petrofísica.

Figura 50. Gráfico cruzado de algunos de los atributos que aparecen en la Tabla 5. Se puede observar en una primera instancia, que algunos de ellos son redundantes, es decir, responden de la misma forma.

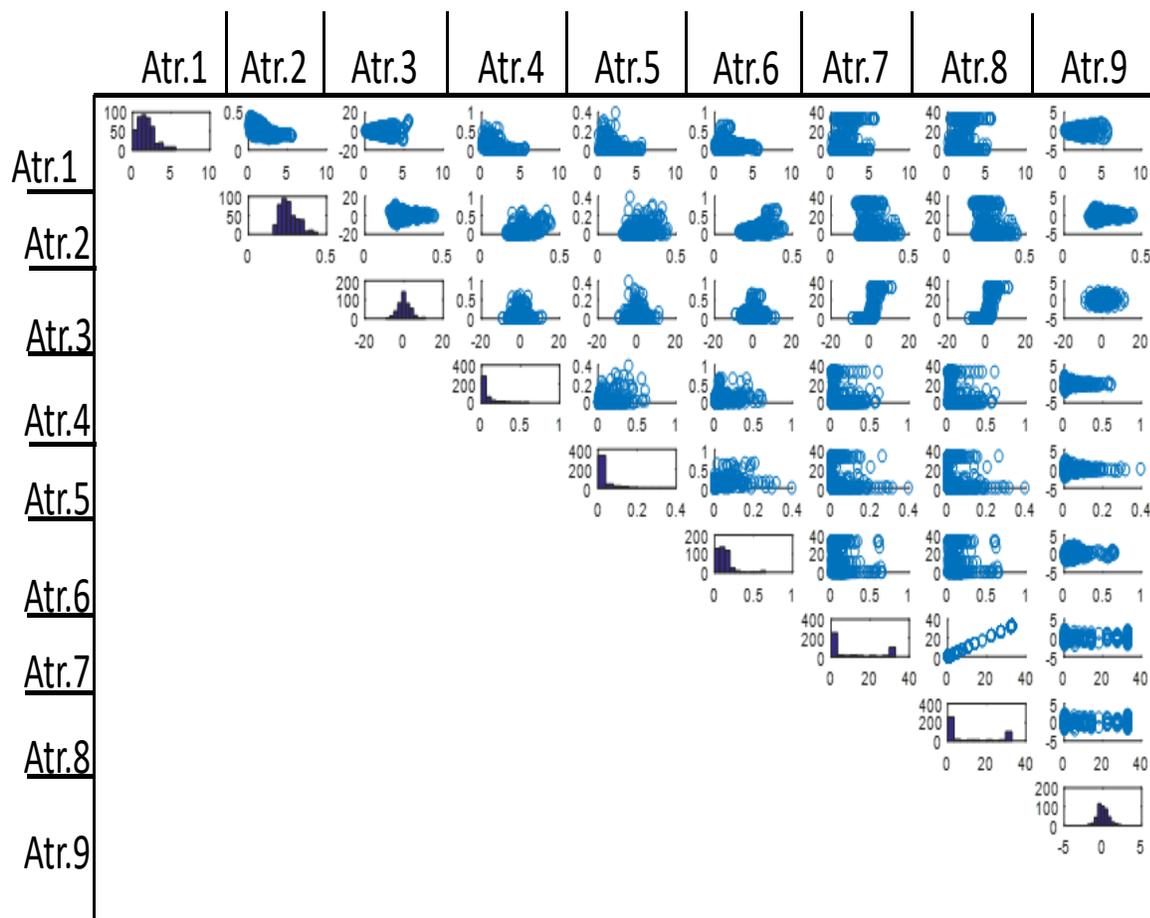


Tabla 5. Estadísticos descriptivos de los atributos a la traza más cercana al pozo, los cuales fueron los usados durante el entrenamiento de las redes, seguido de los estadísticos de la totalidad de los atributos en el resto de las trazas.

Variable	Media	Desv. Est.	Varianza	Mínimo	Máximo	Rango
1.CHAOS	0.08547	0.11397	0.01299	0.00086	0.60654	0.60569
CHAOS_ONE	0.14252	0.22447	0.05039	0.000000	1.00000	1.00000
2.DIP_DEVIATION	0.02862	0.05590	0.00312	0.00000	0.39313	0.39313
DIP_DEVIATIONONE	0.032881	0.075113	0.005642	0.000000	0.834810	0.834810
3.DIP_ILUMINATION	0.12732	0.11347	0.01288	0.00001	0.65452	0.65451
DIP_ILUMINATIONONE	0.22056	0.20812	0.04331	0.000000	3.99990	3.99990
4.EDGE	10.010	13.711	187.993	0.000	32.904	32.904
EDGEONE	7.2981	11.5953	134.4509	0.00000	32.9180	32.9180
5.FIRST_DERIVATIVE	-0.0003	0.6891	0.4748	-1.9587	2.3596	4.3183
FIRST_DERIVATIVEONE	0.000009	0.61116	0.37351	-5.21740	5.08000	10.29740
6.INSTANTANUSB	13.173	15.729	247.416	0.046	99.128	99.082
INSTANTANUSBONE	10.53	13.31	177.08	0.00000	1004.70	1004.70
7.ISOFREQUENCY	0.77888	0.17127	0.02933	0.18272	0.99855	0.81582
ISOFREQUENCYONE	0.74539	0.24456	0.05981	-0.55794	0.99999	1.55793
8.LOCALS	164.41	88.74	7875.17	90.00	270.00	180.00
LOCALSONE	152.68	85.75	7353.63	90.00	270.00	180.00
9.RMS_AMPLITUD	1.3091	0.7090	0.5026	0.2185	3.7923	3.5738
RMS_AMPLITUDONE	1.0961	0.6592	0.4345	0.000000	6.8114	6.8114
10.STRUCTURALS	-0.0026	1.2159	1.4784	-3.6328	4.6883	8.3211
STRUCTURALSONE	0.000086	1.01831	1.03696	-7.91560	6.66920	14.58480
11.TATENUATION	-0.00275	0.05460	0.00298	-0.12226	0.12006	0.24231
TATENUATIONONE	0.000394	0.054737	0.002996	-0.311730	0.237730	0.549460
12.VARIANCE	0.02037	0.02630	0.00069	0.00021	0.17246	0.17225
VARIANCEONE	0.03576	0.06358	0.00404	0.000000	1.00000	1.00000
13.AMPL_TRAZA	-0.0014	1.4916	2.2250	-4.6836	5.5198	10.2035
AMPL_TRAZAONE	0.000086	1.2790	1.6358	-10.0340	9.3210	19.3550

7.1 Resultados del análisis estadístico.

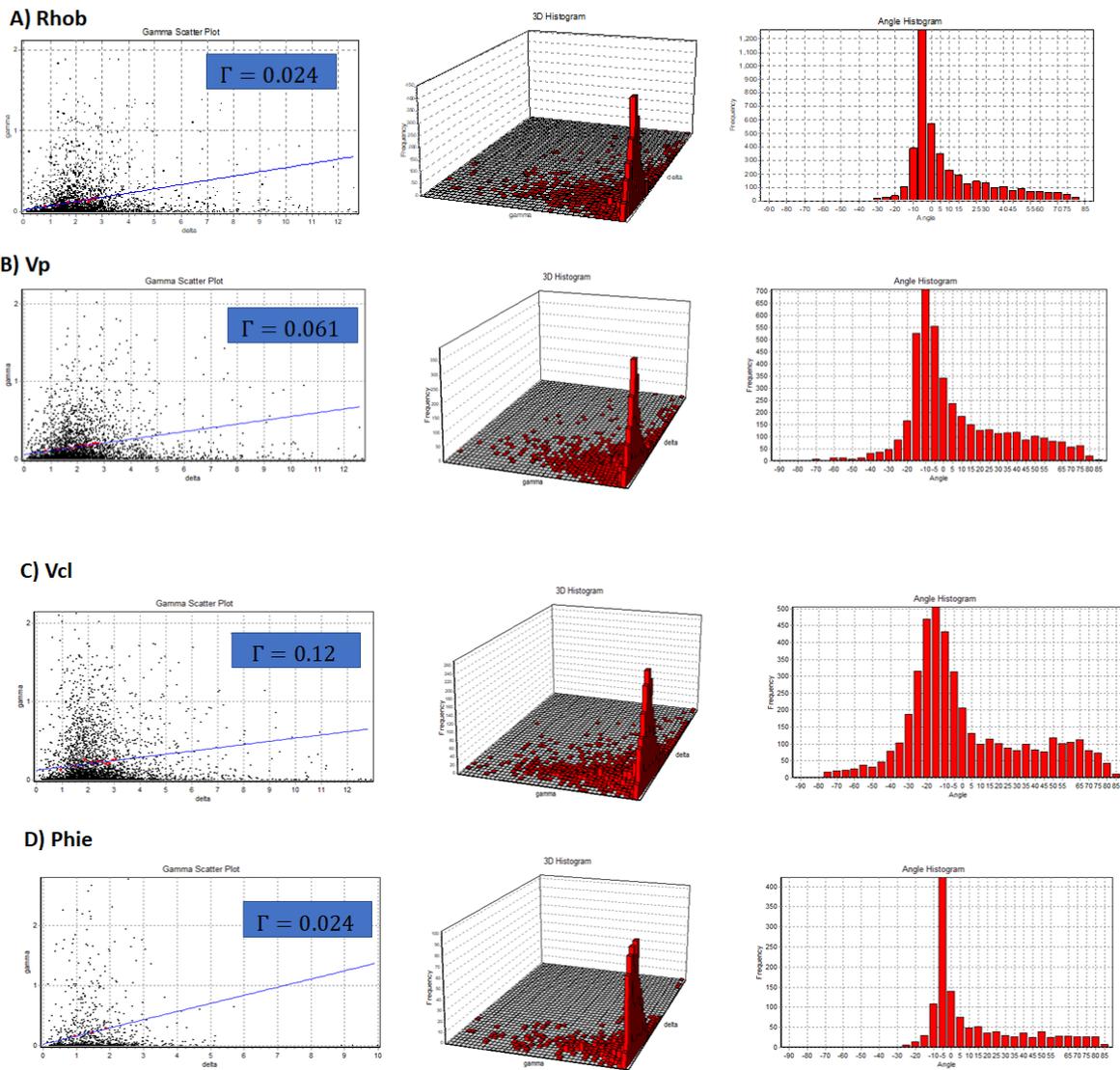
La **Fig. 51** muestra los resultados de la prueba Gamma preliminar llevada a cabo para cada uno de los modelos que se pretende construir - Vcl, Vp, Rhob y phie. Esta primera prueba, permite de una manera rápida determinar si es posible establecer un modelo con los datos de entrada y salida.

Analizando el gráfico de dispersión de cada una de las variables, se observa que aparece un espacio en blanco en forma de cuña en la esquina superior izquierda en todos ellos, esto es característico para los datos con poco ruido, dado que lo que se grafica son los pares $\left((x_{N(i,k)} - x_i)^2, \frac{1}{2} (y_{N(i,k)} - y_i)^2 \right)$, esto quiere decir, que los valores de salida, y , correspondientes a puntos de entrada del conjunto de datos original que se encuentran cercanos entre sí, están distanciados igualmente cerca, a partir de lo cual se puede concluir que existe una relación fuerte entre los datos de entrada-salida y/o que estos presentan poco ruido. Esto, puede ser constatado cuantitativamente con el valor del estadístico Gamma, Γ , el cual presenta valores bajos comparado con el rango de valores que presenta cada uno de los parámetros petrofísicos, con excepción del Vol. de arcilla cuyo valor es relativamente alto, no obstante, como se señaló anteriormente, éste es un primer análisis, donde se están considerando la totalidad de atributos, y tampoco se ha optimizado el número de vecinos cercanos. Así mismo, se puede observar en este tipo de gráfico que la pendiente de la línea de regresión de cada una de las variables no es muy pronunciada por lo que las relaciones de las variables entrada-salida (atributos sísmicos-parámetros petrofísicos) es relativamente compleja, así que la arquitectura de la red será relativamente grande (**Tsui, 1999 a**).

La poca cantidad de ruido también puede observarse en el gráfico de histograma de ángulos, el cual se construye, al unir la intercepción gamma en el eje vertical del gráfico de línea de regresión con el punto de dispersión. El ángulo que forma la línea resultante con el eje horizontal positivo es entonces calculado. Este ángulo se encuentra entre $\left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$. La característica por buscar en este histograma es la frecuencia de ángulos cercanos al extremo derecho, es decir, cerca de $\frac{\pi}{2}$. Si no hay puntos cercanos a $\frac{\pi}{2}$ (= 90 grados), este es un buen indicador para un modelado suave. Si hay muchos puntos cerca de a $\frac{\pi}{2}$ este es una señal muy mala. La importancia de la distribución cercana a $\frac{\pi}{2}$ en el histograma de ángulo es otra forma de visualizar la cuña superior izquierda de la gráfica de dispersión.

Con base a todo lo anterior, se concluye que es posible llevar a cabo la construcción de modelos suaves con los datos disponibles.

Figura 51. Resultados preliminares de la prueba Gamma usando un número de vecinos igual a 10 para cada una de las propiedades petrofísicas, a partir de las cuales se pretende construir un modelo. Se presentan las 3 diferentes presentaciones que proporciona el software WinGamma.



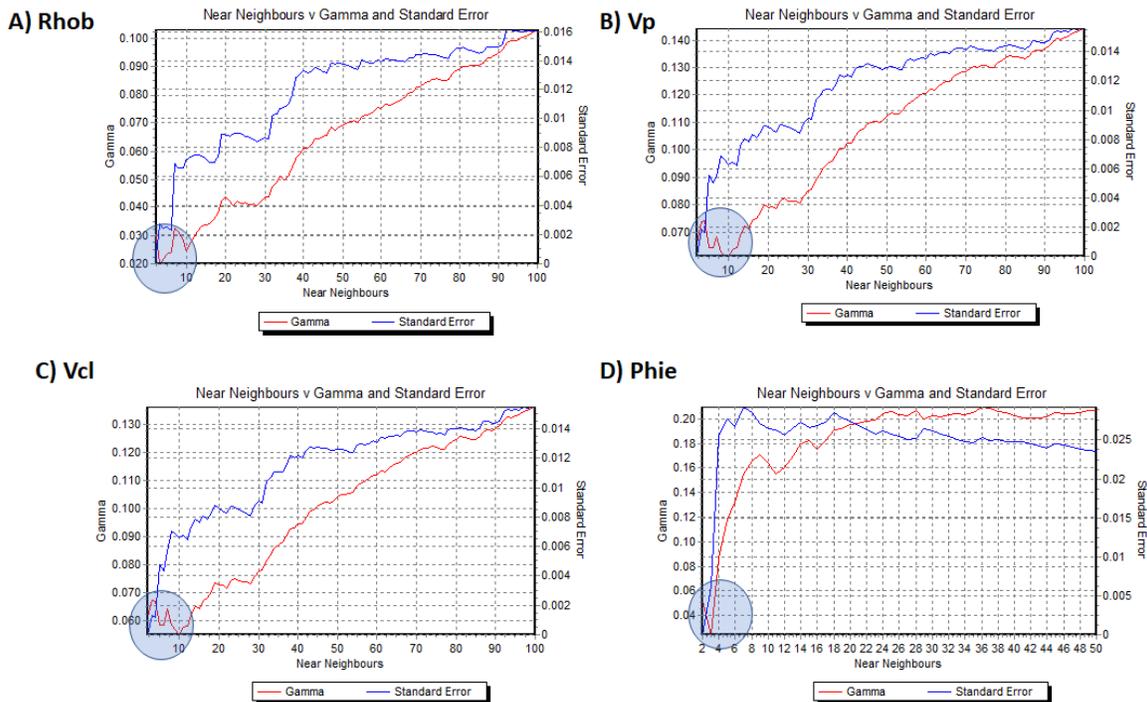
Una vez determinada la factibilidad de construir los modelos con los datos disponibles, y dado que la efectividad y precisión de la estimación del ruido en los datos a través de la prueba Gamma depende de establecer el número de los k-vecinos más cercanos, se procedió a determinar el número óptimo de vecinos más cercanos. En la **Tabla 6** se presenta el número de vecinos más cercanos óptimo obtenidos una vez llevada a cabo la prueba de incremento de número de vecinos más cercanos (**Fig. 52**). Esta prueba consiste básicamente en estimar con diferente número de vecinos más cercanos, el valor del estadístico Gamma. La forma ideal de saber el número óptimo de vecinos más cercanos es esperar a que se establezca el estadístico Gamma, y ver con cuántos vecinos se tiene el error más pequeño, no obstante, se puede observar que en ninguno de los casos se estabiliza el valor del estadístico Gamma, por

lo que se tomaron varios valores de vecinos más cercanos que proporcionaban los errores más bajos, esto con la finalidad de evitar mínimos locales. (Sólo se presentan aquel número de vecinos con el que se obtuvieron mejores resultados en cada uno de los casos.)

Tabla 6. Número de vecinos óptimo para cada uno de los parámetros petrofísicos.

Variable	Núm. de vecinos
Rhob.	2
Vp	8
Vcl	3
Phie	3

Figura 52. Prueba de incremento del número de vecinos más cercanos. Se puede observar que en ninguno de los casos se estabiliza el valor del estadístico Gamma, por lo que se tomó diferentes valores de vecinos más cercanos y posteriormente se llevaron a cabo el resto de las pruebas, esto con la finalidad de evitar mínimos locales. A) Para la variable Rhob, el número de vecinos más cercanos óptimo es 2. B) Para la variable Vp, el número de vecinos más cercanos óptimo es 8. C) El número de vecinos más cercanos óptimo para la variable Vcl es 3. D) El número de vecinos más cercanos óptimo para la variable Phie es 3.



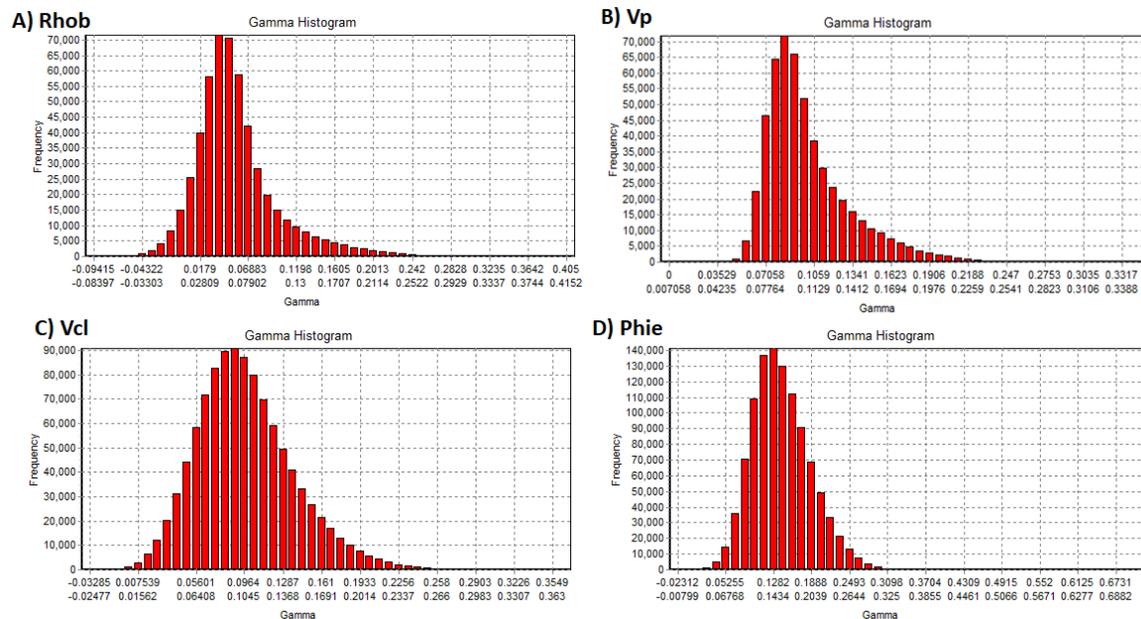
Selección de un subconjunto adecuado de entradas.

En la **Fig. 53** se puede observar que para cada uno de los modelos que se intenta construir existen un gran número de combinaciones posibles de entradas que minimizan el valor del

estadístico Gamma, Γ . Esta minimización del estadístico Gamma se debe a la depuración de aquellos atributos que son, ya sea redundantes y/o tienen poca relación con la variable de la que se está tratando de modelar.

Antes de presentar los resultados de la selección de datos, es imprescindible explicar el concepto de máscara, puesto que es cómo funciona el software WinGamma. En WinGamma una incrustación se designa con una cadena de "1" y "0" llamada máscara. Por lo tanto, si hay cinco entradas, la máscara 10111 indica que se deben usar todas las entradas en la inserción, excepto la segunda. Una incrustación completa prueba cada combinación de entradas para determinar qué combinación produce el valor de Gamma absoluto más pequeño. Por lo que, en este caso en particular, la máscara tendrá una extensión de 19 elementos (la misma que la cantidad de atributos que se está considerando). El concepto de una máscara ayuda a indicar claramente cualquier combinación particular de atributos. Para comenzar con el análisis de datos se necesita establecer el orden en que se considerarán los atributos en cualquier máscara. Por lo tanto, el orden en que se etiquetó cada atributo se da en **Tabla 5**. Estableciendo el orden en que los atributos aparecen en cualquier máscara se puede interpretar cualquier máscara dada.

Figura 53. Histograma de los valores de Gamma para la búsqueda de los mejores atributos para cada uno de los parámetros petrofísicos, se incluyen todas las posibles combinaciones. Se puede observar que menos de un cuarto del total de las combinaciones son las que producen los valores más pequeños del estadístico $|\Gamma|$ para cada una de las variables.



En las **Tablas 8, 9, 10 y 11** se presenta sólo aquella combinación de entradas (máscara) para la predicción de cada una de las propiedades petrofísicas de interés; Vcl, phie, Vp y Rhob, respectivamente, cuyo valor de la estadística Gamma, $|\Gamma|$ fue el menor. Es importante señalar, que este análisis se llevó a cabo con el número de vecinos cercanos óptimo obtenido anteriormente.

Tabla 7. Mejor combinación de atributos para la predicción de Vcl. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario

Máscara	tiempo	Amplitud Traza	Varianza	Tatenuation	Sweetness	StructuralS	RMS Amplitud	Inst.B	LocalS	Localflatness	Isofrequency	Dip_Illumination	Dip_Deviation	Chaos	Amplitud	Edge	Envelope	AVO	First_Derivative	MSE
Máscara 1	1	0	0	0	1	0	1	0	1	1	1	1	0	1	0	1	0	1	1	3.72E-05

Tabla 8. Mejor combinación de atributos para la predicción de porosidad efectiva. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario.

Máscara	tiempo	Amplitud Traza	Varianza	Tatenuation	Sweetness	StructuralS	RMS Amplitud	Inst.B	LocalS	Localflatness	Isofrequency	Dip_Illumination	Dip_Deviation	Chaos	Amplitud	Edge	Envelope	AVO	First_Derivative	MSE
Máscara 1	0	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	0	1	0.00035

Tabla 9. Mejor combinación de atributos para la predicción de Vp. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario.

Máscara	tiempo	Amplitud Traza	Varianza	Tatenuation	Sweetness	StructuralS	RMS Amplitud	Inst.B	LocalS	Localflatness	Isofrequency	Dip_Illumination	Dip_Deviation	Chaos	Amplitud	Edge	Envelope	AVO	First_Derivative	MSE
Máscara 1	0	0	0	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0.039

Tabla 10. Mejor combinación de atributos para la predicción de Rhob. Se escribe 0 si la entrada correspondiente no es considerada, 1 para el caso contrario.

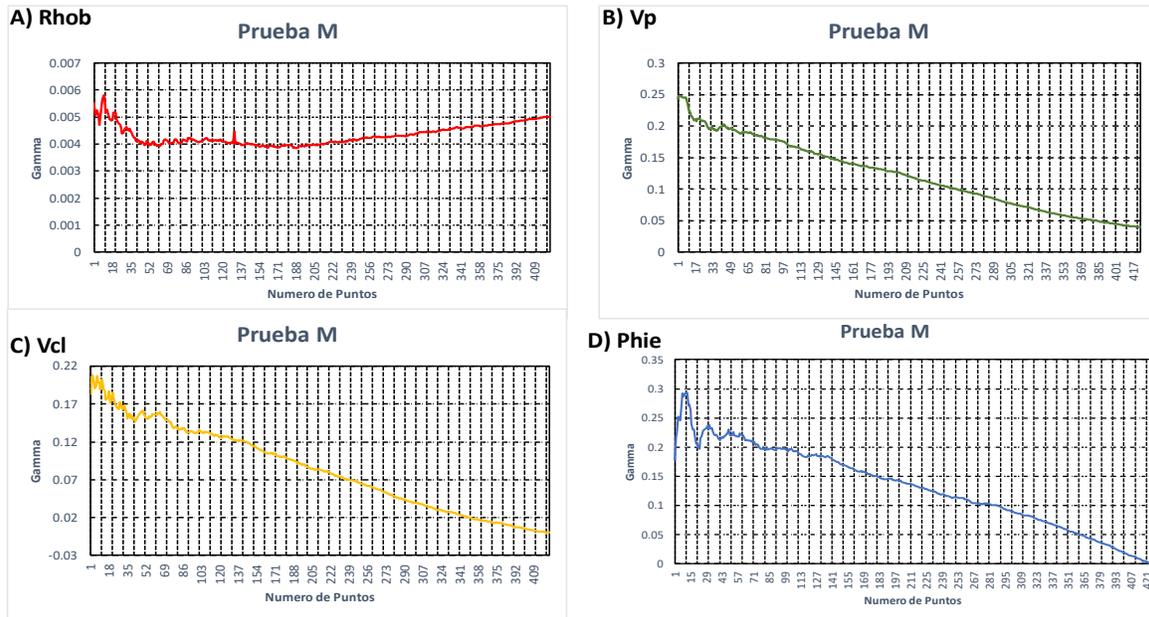
Máscara	tiempo	Amplitud Traza	Varianza	Tatenuation	Sweetness	StructuralS	RMS Amplitud	Inst.B	LocalS	Localflatness	Isofrequency	Dip_Illumination	Dip_Deviation	Chaos	Amplitud	Edge	Envelope	AVO	First_Derivative	MSE
Máscara 1	0	0	0	1	0	1	0	1	0	0	1	0	1	0	1	1	1	0	0	3.32E-07

Se puede observar claramente que, para todas las predicciones, el valor del estadístico Gamma disminuyó considerablemente en comparación a los resultados del análisis preliminar, donde se consideró todos los elementos de las **Tabla 5** y un número de vecinos más cercanos igual a 10. Estos últimos resultados permiten establecer que el ruido en los datos es poco, lo que significaría que se pueden construir buenos modelos entre las variables de entrada y salida con los datos que se tienen. Así mismo, dado que la capacidad de generalización de una red está limitada principalmente por el ruido en los datos, estos resultados implican, que uno puede esperar que el desempeño de las redes neuronales sea bueno para la construcción de los modelos, siempre y cuando el resto de las trazas contengan información similar.

Finalmente, se lleva a cabo la prueba M, para conocer la confiabilidad del valor del estadístico Gamma arrojado por la prueba. La **Fig. 54** representa el promedio de los resultados de 40 pruebas M para cada una de las variables, esto dado que el resultado de esta prueba depende de la forma en que uno selecciona cada conjunto de datos (esto se hace al azar).

Como se puede observar para ninguno de los parámetros petrofísicos se alcanza el valor asintótico del estadístico Gamma, sino más bien se observa que parece ser que apenas inicia este comportamiento, con excepción de phie y Rhob. Aun cuando para phie y Rhob, todavía no se vislumbra el comportamiento asintótico de gamma, se puede confiar que el valor del estadístico Gamma arrojado por la prueba Gamma para cada uno de los casos representa la variación de ruido real en la salida, dado que como se observa el rango de variación de ésta es muy pequeño en ambos casos.

Figura 54. Prueba M para los 4 modelos. La prueba M se usa para mostrar cómo varía la estadística Gamma a medida que se utilizan más datos para calcularla. Eventualmente, si se utilizan suficientes datos, la estadística Gamma debe sugerir una variación de ruido real en la salida para la cual se ha calculado. Esta prueba también nos puede decir cuántos datos es probable que se necesiten para obtener un modelo de una calidad determinada, en el sentido de predecir con un MSE alrededor del nivel de ruido. Se tomó la máscara que produce el MSE más bajo para cada uno de los parámetros petrofísicos para ejecutar la prueba M.



Hasta este punto, ya se sabe cuál es la mejor combinación de atributos sísmicos (para los datos dados), cuántos puntos de datos mínimo se necesitan para el entrenamiento y el error mínimo a alcanzar durante el entrenamiento de cada una de las redes para la construcción de cada uno de los modelos (**valor de $|\Gamma|$**), antes de que se sobreentrene la red.

La siguiente sección proporciona detalles sobre la red neuronal artificial que se utilizó para la construcción de los modelos no lineales de cada uno de los parámetros petrofísicos

7. 2 Redes neuronales: diseño, entrenamiento y resultados.

En la **Tabla 12** se presenta la configuración, el algoritmo de optimización, y el error alcanzado, de cada una de las redes usadas para la estimación de cada parámetro petrofísico.

La arquitectura de las redes dependió de la complejidad del modelo a construir, dado que todos son relativamente complejos se usaron dos capas ocultas, cuya función de activación fue sigmoideal para ambas capas.

En lo que respecta al algoritmo de aprendizaje empleado en cada una de las redes neuronales fueron dos los que se probaron durante el entrenamiento de éstas: **Retropropagación con momentum**, muy ampliamente usado en la estimación de propiedades petrofísicas, y el de **Broyden-Fletcher-Goldfarb-Shanno (BFGS)** que proporciona un progresivo ajuste de los

pesos de la red neuronal por descenso del gradiente, utilizando aproximaciones mejoradas al Hessiano inverso, en lugar del verdadero inverso.

El proceso de entrenamiento fue costoso hablando en términos de tiempo y requerimientos computacionales. En algunos casos se entrenó a las RNA's durante 4 o 5 días para permitir que la MSE alcanzará un nivel cercano al objetivo MSE (o igual a éste) dictado por la estadística Gamma.

Tabla 11. Arquitectura, algoritmo de entrenamiento y MSE alcanzado para cada una de las redes, según la propiedad petrofísica a estimar. La arquitectura corresponde al número de neuronas en cada capa oculta.

<i>Propiedad Petrofísica</i>	<i>Mejor arquitectura</i>	<i>Algoritmo de entrenamiento</i>	<i>MSE alcanzado</i>
<i>Vcl</i>	<i>21*15</i>	<i>BFGS</i>	<i>.0178</i>
<i>Rho</i>	<i>11*11</i>	<i>Bpm</i>	<i>0.0130</i>
<i>Phie</i>	<i>17*25</i>	<i>BFGS</i>	<i>3.5 E-4</i>
<i>Vp</i>	<i>8*8</i>	<i>BFGS</i>	<i>0.039</i>

***Bpm: Back Propagation Momentum.**

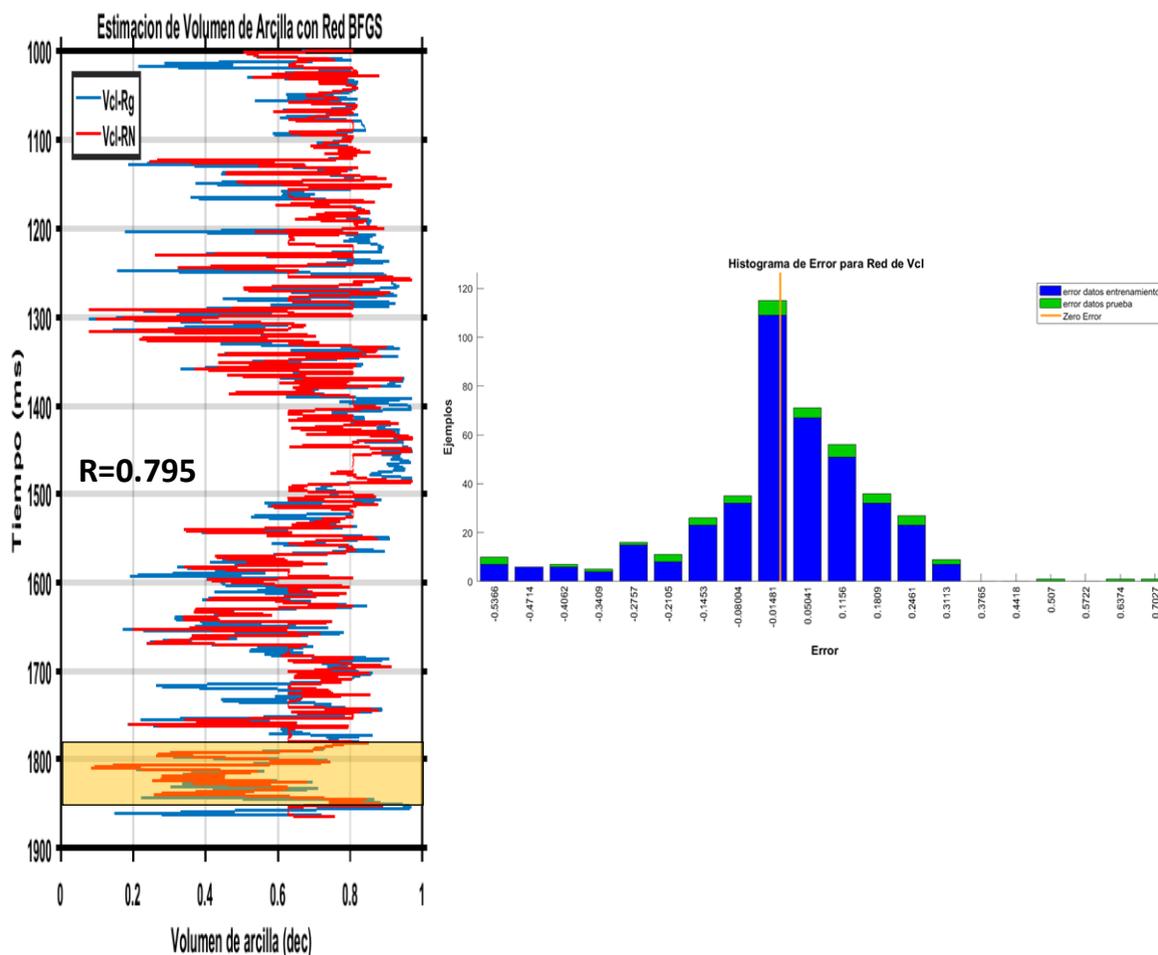
***BFGS: Broyden-Fletcher-Goldfarb-Shanno.**

Un punto que llama la atención de esta tabla es el MSE alcanzado por cada una de las redes. Se observa que solo dos alcanzaron el MSE establecido en la prueba Gamma, cuyo valor está asociado al ruido verdadero, así que, dos de los modelos obtenidos (para Vcl y Rhob) son susceptibles a ser mejorados durante el entrenamiento, al no alcanzar el valor asociado al ruido. Es importante declarar, que distintas configuraciones de valores de los pesos pueden dar un mismo error, con esto, se puede considerar que las redes neuronales funcionan como un proceso de inversión, donde no existe una solución única.

En la red usada para la estimación del volumen de arcilla (Vcl), se obtuvo un factor de correlación (R) de 0.795 entre los datos obtenidos por la red y los datos del pseudo registro. Cuando se usó todo el set de datos, como era de esperarse se obtuvo un factor de correlación menor de 0.62. En la **Fig. 55** se muestran dos gráficos; el primero, corresponde a un gráfico donde se comparan los valores estimados por la red y los valores del pseudo registro para el conjunto completo de datos durante el entrenamiento. El segundo gráfico, corresponde a un histograma, donde se observa el error que se presentan en los datos usados para el entrenamiento, y el error en los datos usados para la validación.

Lo remarcable en el histograma de errores, aparte del hecho que presenta casi una distribución normal, es que los errores más bajos, tienen las frecuencias mal altas, tanto para los datos de entrenamiento, así como para los datos de validación, siendo, éstos últimos de mayor interés, puesto que avalan en un cierto sentido, la capacidad de generalización de la red. Así mismo, dado los bajos errores presentes en los datos de entrenamiento, es de esperarse que exista una buena correlación (0.795) entre la curva del pseudo registro de Vcl y la curva de Vcl arrojada por la red.

Figura 55. Se presentan dos tipos de gráficos para Vcl. El primero corresponde a un gráfico donde se compara los valores estimados por la red y los valores del pseudoregistro. La parte sombreada en amarillo corresponde a las arenas productoras. El segundo, es un histograma donde se observa el error durante el entrenamiento, las barras en color azul, y el error al aplicarse a los puntos ciegos para la validación de la red, barras en color verde.



El conjunto usado para evaluar el desempeño de la red entrenada para la estimación de ϕ_{ie} , ρ y V_p , presentan un factor de correlación (R) de 0.78, 0.95 y 0.93, respectivamente. Cuando se usó todos los puntos de los registros de pozo, este factor disminuyó naturalmente en todos los casos, siendo éstos; 0.71, 0.88 y 0.8. Las **Figs. 56-58** son análogas a la presentada para el volumen de arcilla.

Se obtienen resultados bastante buenos en cada uno de los modelos para predecir los parámetros petrofísicos, como puede ser observado en cada una de las figuras. No obstante, se puede observar que, en el caso de la red usada para la predicción de la V_p , se presentan los errores mayores en los datos de validación, en comparación de los modelos (redes) para las otras propiedades.

Figura 56. Se presentan dos tipos de gráficos para ϕ_{ie} . El primero corresponde a un gráfico donde se compara los valores estimados por la red y los valores del registro. La parte sombreada en amarillo corresponde a las arenas productoras. El segundo, es un histograma donde se observa el error durante el entrenamiento, las barras en color azul, y el error al aplicarse a los puntos ciegos para la validación de la red, barras en color verde.

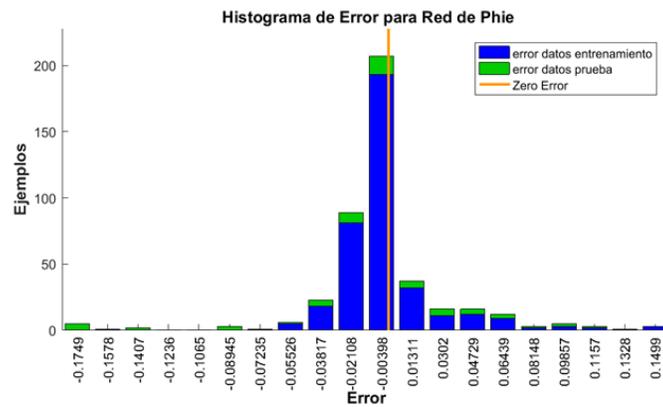
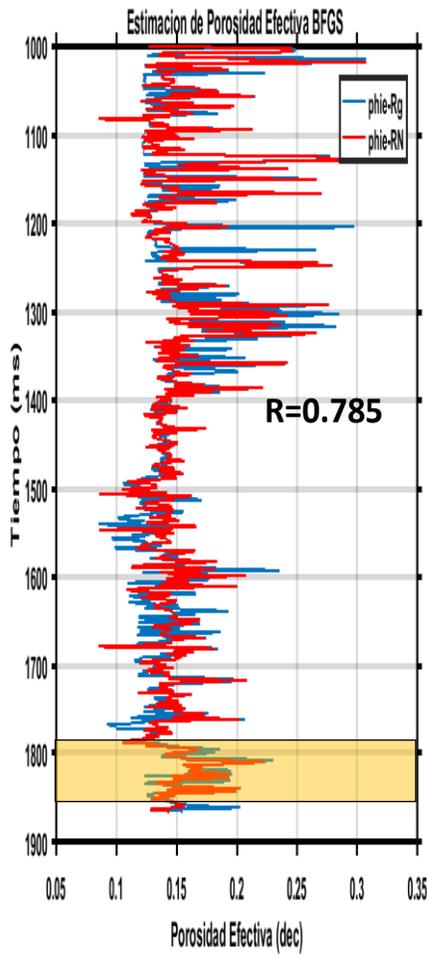


Figura 57. Se presentan dos tipos de gráficos para densidad o Rhob. El primero corresponde a un gráfico donde se compara los valores estimados por la red y los valores del registro. La parte sombreada en amarillo corresponde a las arenas productoras. El segundo, es un histograma donde se observa el error durante el entrenamiento, las barras en color azul, y el error al aplicarse a los puntos ciegos para la validación de la red, barras en color verde.

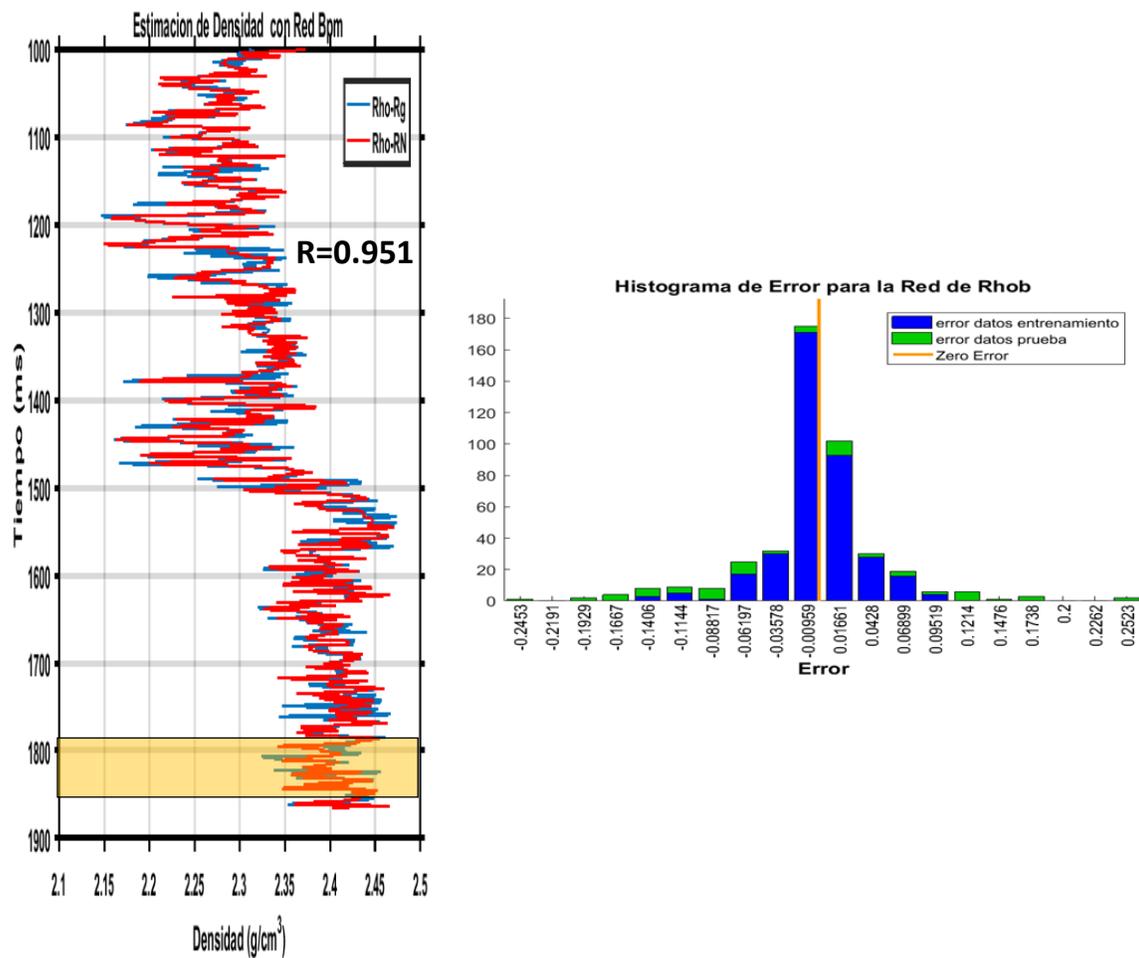
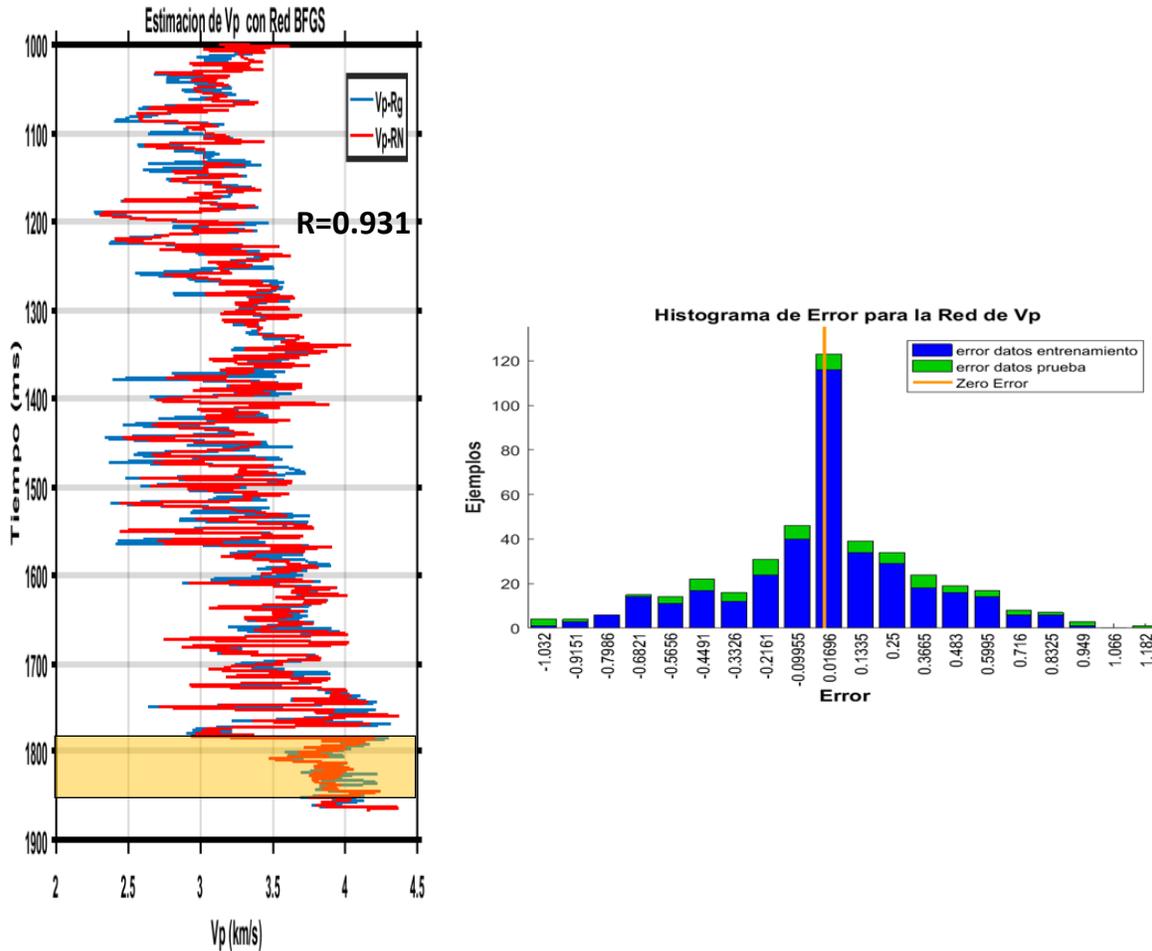


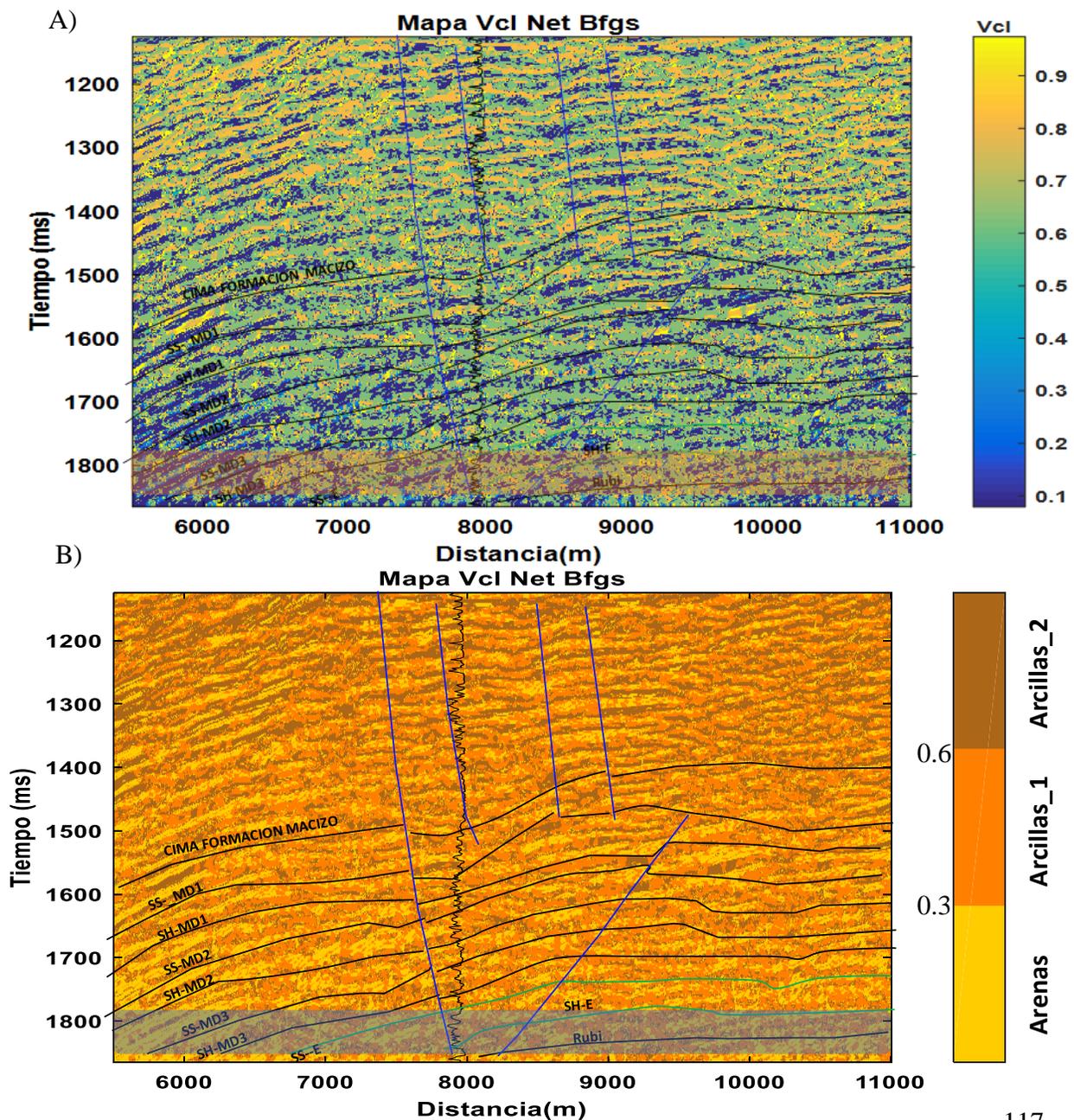
Figura 58. Se presentan dos tipos de gráficos para Vp. El primero corresponde a un gráfico donde se compara los valores estimados por la red y los valores del registro. La parte sombreada en amarillo corresponde a las arenas productoras. El segundo, es un histograma donde se observa el error durante el entrenamiento, las barras en color azul, y el error al aplicarse a los puntos ciegos para la validación de la red, barras en color verde.



Las estimaciones para cada uno de los parámetros petrofísicos a escala sísmica se obtienen aplicando las RNA's entrenadas a las 903 trazas sísmicas restantes que comprenden a las secciones. La **Fig. 59a** muestra la estimación continua del volumen de arcilla (Vcl). En este mapa, se puede distinguir claramente la continuidad de las capas arenosas, así como de aquellas capas arcillosas. En el gráfico cruzado de **Vcl vs Rhob (Fig. 60)**, se puede identificar claramente tres diferentes familias de datos de acuerdo con su volumen de arcilla, que pueden ser asociados de manera arbitraria a tres diferentes litofacies. Por lo que se procedió a construir una sección de 3 litofacies (**Fig. 59 b**), en donde para valores de Vcl entre 0.0-0.35 corresponden a arenas, valores de Vcl 0.35-0.60 corresponde a lutitas tipo 1, y para valores de Vcl >0.60 estarían asociados a lutitas tipo 2. Los diferentes volúmenes de arcilla pueden ser asociados a tipos específicos de ambiente de deposición, que, en este caso en particular, corresponderían a depósitos relacionados a los diferentes elementos morfológicos de los sistemas meándricos. Así, los volúmenes de arcilla grandes estarían asociados a sistemas de baja energía, que corresponderían a depósitos de planicie de inundación y/o depósitos de lago

lateral. Los volúmenes intermedios de Vcl, pueden ser asociados a depósitos de desborde o a depósitos de barra de punto. Finalmente, los volúmenes bajos de Vcl, estarían asociados a sistemas de alta energía, tal como, depósitos de canal. Es importante señalar, que lo anteriormente mencionado es solo una interpretación incipiente, que deber ser validada y sustentada con sus respectivos estudios sedimentológicos, estratigráficos y demás.

Figura 59. Secciones de Vcl que abarcan la ventana de tiempo 1126-1863 ms. A) Estimación continua del volumen de arcilla, usando una red neuronal 21 por 15 con el algoritmo de aprendizaje BFGS. B) Sección donde se incluye tres litofacies: Las arenas están representadas en color amarillo, mientras que las lutitas tipo 1 en color naranja y en color marrón las lutitas tipo 2. Se incluye en ambas secciones, tanto el pseudoregistro de Vcl, el cual su valor incrementa hacia la derecha, así como la interpretación estructural-estratigráfica (las fallas están en color azul). La parte sombreada corresponde a las arenas productoras ubicadas aproximadamente en el intervalo en tiempo de 1783-1843 ms.



El gráfico cruzado de la **Fig. 60a**, también, permite diferenciar dos tipos de minerales arcillosos, uno altamente denso, que podría ser asociado a minerales de ilita, y uno, poco denso, que podría ser asociado a minerales de montmorillonita. En la **Fig. 60b**, es posible establecer que en general las litofacias asociadas a arenas presentan una mayor porosidad efectiva en comparación con las litofacias de lutita. Los casos donde esto no se cumple, una de las causas podría ser que las arcillas presenten un tipo de distribución diferente (dispersa) que esté produciendo que los valores de porosidad sean más bajos de los esperados. Lo anterior, puede ser constatado en la evaluación petrofísica del área mostrada en la **Fig. 61**.

Las **Figs. 62 a, b, c**, corresponde a las secciones de las propiedades petrofísicas Rho, phie y Vp, respectivamente, obtenidas con las mejores redes neuronales entrenadas, para cada caso. En cada una de ellas, se comprueba que las redes están presentando una capacidad de generalización de buena a muy buena, en el sentido que se puede observar que se conserva el echado de los estratos, así también, que existe correspondencia entre los diferentes mapas, es decir, en aquellas áreas donde se presentan velocidades de onda P elevadas, corresponden a áreas de alta densidad, y baja porosidad efectiva dado que dichas áreas estarían asociadas en su mayoría a las lutitas de alta densidad. Dicha correspondencia puede ser vista con mayor claridad en la **Fig. 63**, en donde se sobrepone el registro de cada propiedad, sobre una paleta de colores, que representa los valores arrojados por cada una de las redes neuronales entrenadas sobre el intervalo de arenas productoras, 1783-1843 ms. En esta figura, también, se puede observar más claramente el buen entrenamiento que se llevó a cabo en cada una de las redes. Las redes que presentan mejores resultados son las correspondientes a la estimación del volumen de arcilla y porosidad efectiva. Mientras que las redes entrenadas para la estimación de Rho y Vp presentan en general, un comportamiento de sobreestimación en algunos de los valores para el caso de Rho, y una subestimación para el caso de Vp. Los tiempos donde se presentan los mayores errores son; 1794,1806,1812,1824,1826 y 1841 ms, y 1788,1818,1832 y 1834 ms, para la red de Rho y Vp respectivamente. Para el caso de la red que modela el volumen de arcilla, se puede observar claramente que está distinguiendo las 3 litofacias propuestas.

En la **Fig. 64**, se presentan las mismas secciones, pero ahora se le sobrepone los datos sísmicos, donde se observa que las altas amplitudes coinciden a altos valores de Vp, altos valores de Rho, y a bajos valores de phie, lo cual confirma nuevamente lo dicho en el párrafo anterior.

Aunado a lo dicho anteriormente a cuanto, a la comprobación de la capacidad de generalización de cada una de las redes, y siguiendo la metodología presentada en el capítulo anterior para dicho punto (comprobación de la generalización de las redes), se seleccionó a la red que estima los valores de Vp, dado que es la que presentó un mayor error para los puntos de validación; es importante mencionar sin embargo que todas las redes presentaron un buen desempeño en general en su validación durante el primer entrenamiento. La **Fig. 65a** muestra un histograma donde se muestra tanto las salidas de la red (color rojo) y los datos del registro Vp (color azul), y se usó para detectar en que parte se tuvo el peor desempeño, esto para la selección de los nuevos puntos de validación. En este se puede observar que uno de los intervalos donde se presentó mayor problema durante el primer entrenamiento fue en

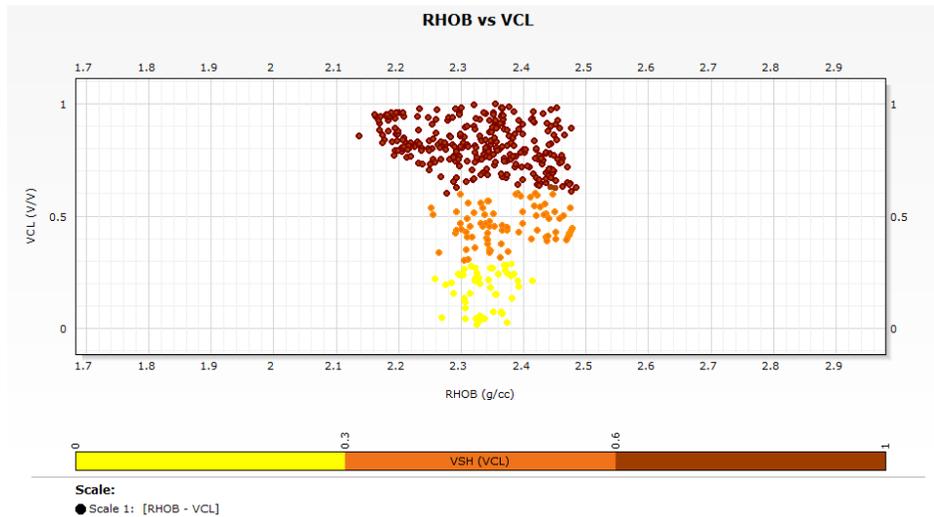
la ventana de tiempo de 1000-1126 ms. La Fig. **65b** es un histograma de errores, tanto de los datos de entrenamiento como los de validación para el segundo entrenamiento llevado a cabo, donde se observa que una vez más se obtuvieron errores pequeños, lo cual nos permite confiar en la capacidad de generalización de las redes usadas para la predicción de las propiedades petrofísicas.

Finalmente, para dar por concluido este trabajo, y siguiendo el flujo de trabajo presentado en la sección anterior, en la **Fig. 66** se presentan los resultados obtenidos al usar regresión lineal convolucional (7 número de operadores) para la predicción de las mismas propiedades petrofísicas usando las mismas entradas y salidas. Cabe mencionar, que solo se presenta las secciones de Rhob y Vcl, dado que no fue posible aplicar el método en las otras dos propiedades, ya que la matriz inversa de los datos entrada (mejor combinación de atributos dada por la prueba Gamma), la cual su cálculo es necesario y esencial para la aplicación de esta técnica, no pudo efectuarse por el mal condicionamiento de los datos.

Se observa claramente la superioridad de las secciones obtenidas por las redes neuronales sobre las obtenidas por la regresión lineal. En la sección de volumen arcilla obtenida a través de este método es imposible poder discretizar las tres litofacies, inclusive observar la continuidad de las arenas. Para el caso de Rhob, se observa que la sección obtenida está muy suavizada, aparte de que no se distinguen con claridad las litofacies más densas. No obstante, se conservan los echados de los estratos. Es importante mencionar que, con este método se obtuvieron factores de correlación menores de 0.10 y 0.15 para Vcl y Rhob respectivamente.

Figura 60. A) Gráfico cruzado Rhob vs Vcl. Se puede distinguir tres familias de datos claramente de acuerdo con el volumen de arcillas. Los círculos en amarillo representan las litofacies más limpias (arenas) cuya densidad va de 2.25 a 2.4 g/cc. Los círculos naranjas representarían las arcillas tipo 1, cuya densidad es para la mitad de los datos aproximadamente, más alta que las areniscas. Los círculos marrones representan las lutitas tipo 2, y cuyos valores de densidad es muy variable, éstos van de 2.15-2.44 g/cc. B) Mismo gráfico cruzado, sólo que ahora la escala de colores hace referencia a la porosidad.

A)



B)

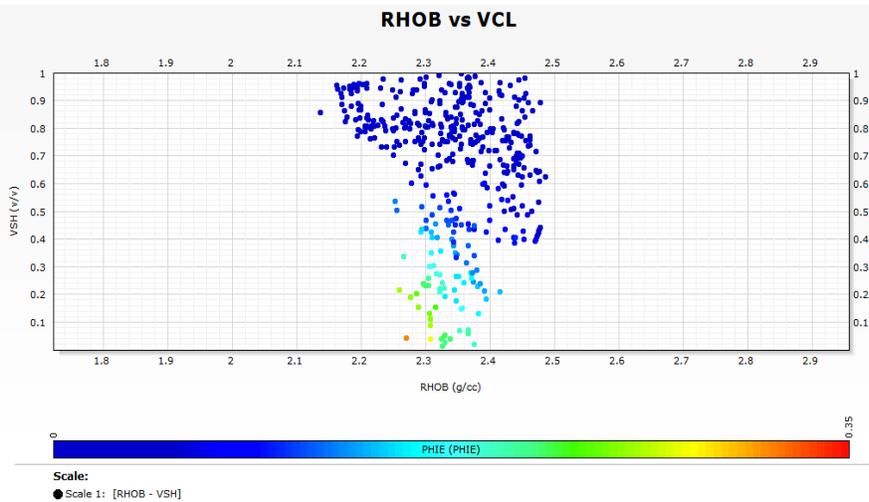
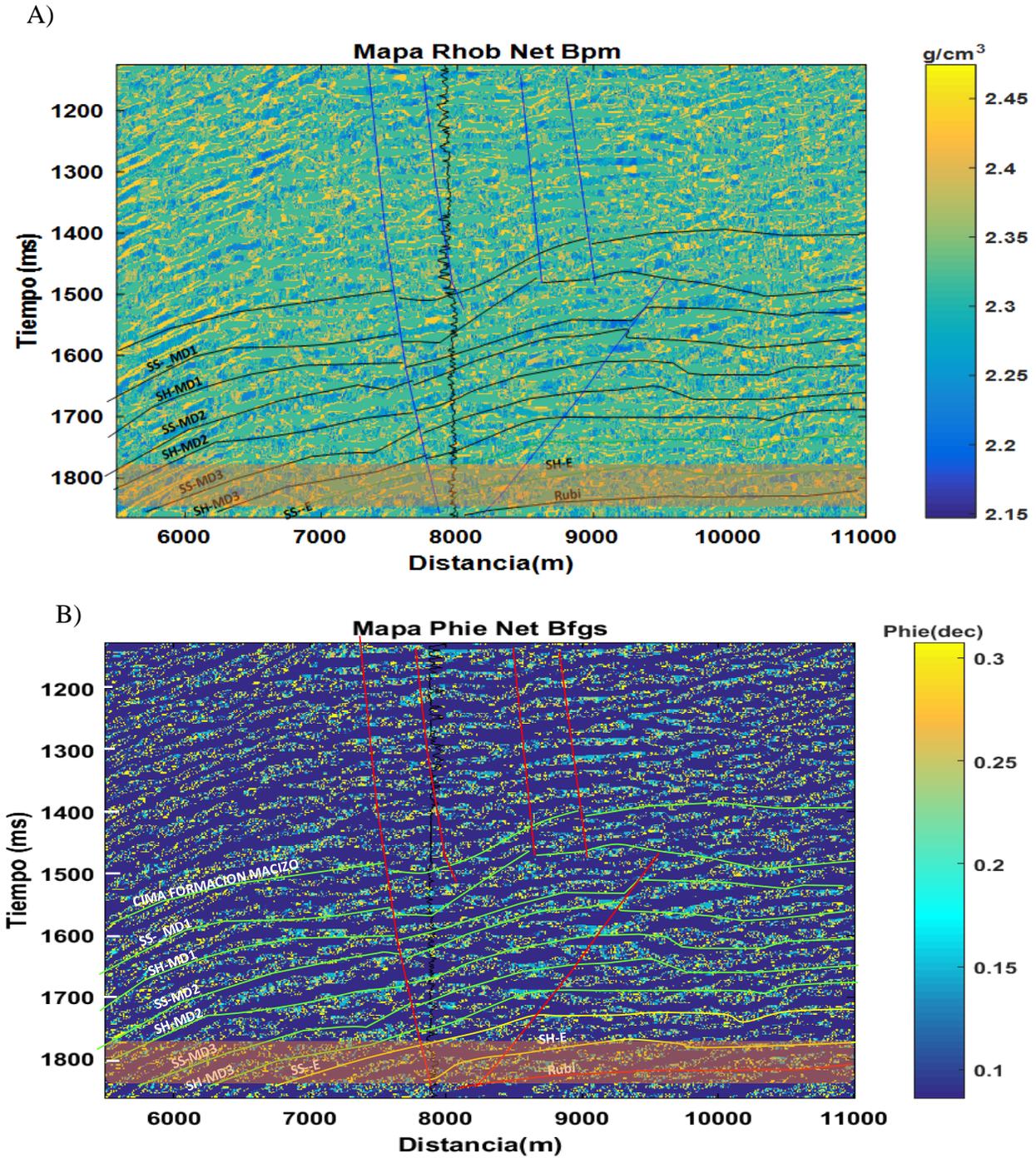


Figura 62. Secciones de las propiedades petrofísicas Rho, Phie y Vp, obtenidas con las redes con mejor desempeño. A) Mapa de Rho. B) Mapa correspondiente a phie. C) Mapa de Vp. En cada una de las secciones se incluye su interpretación estructural-estratigráfica, así como el registro correspondiente a cada propiedad, cuyo valor de las curvas incrementa hacia la derecha. La parte sombreada corresponde a las arenas productoras ubicadas aproximadamente en el intervalo en tiempo de 1783-1843 ms.



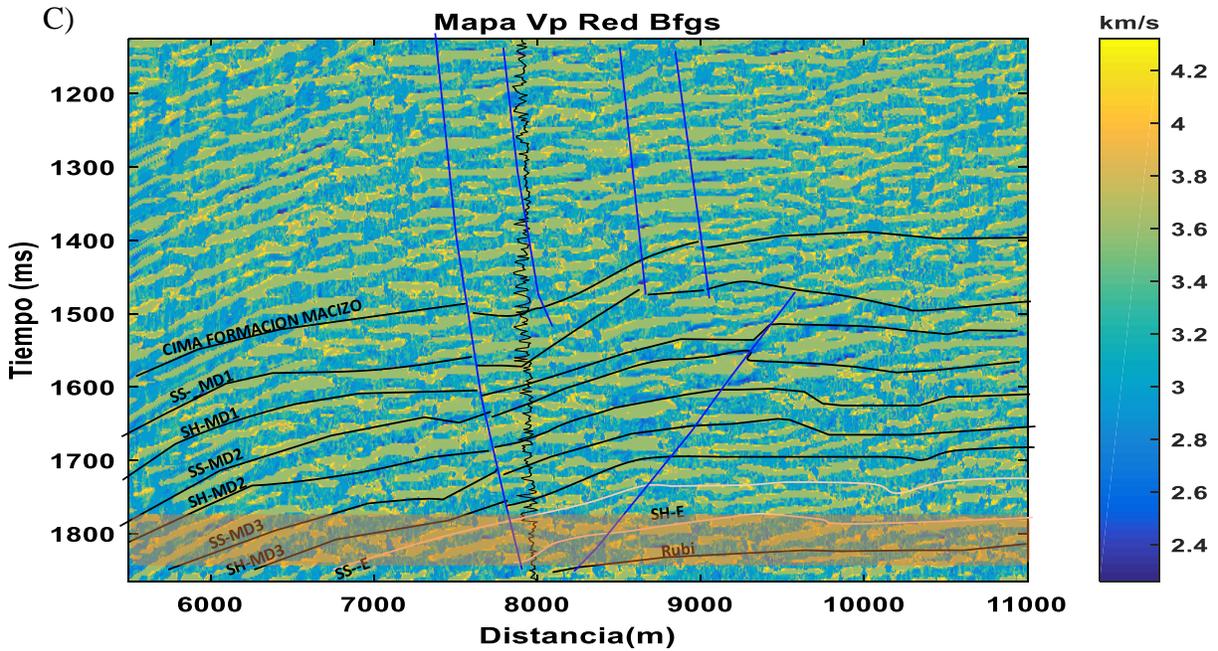


Figura 63. Sobreposición del valor de registro de cada una de las propiedades petrofísicas, sobre paleta de colores que representa el valor de salida de su respectiva red, durante el entrenamiento. Las líneas verticales negras representan los límites asociados a la escala de colores de la paleta.

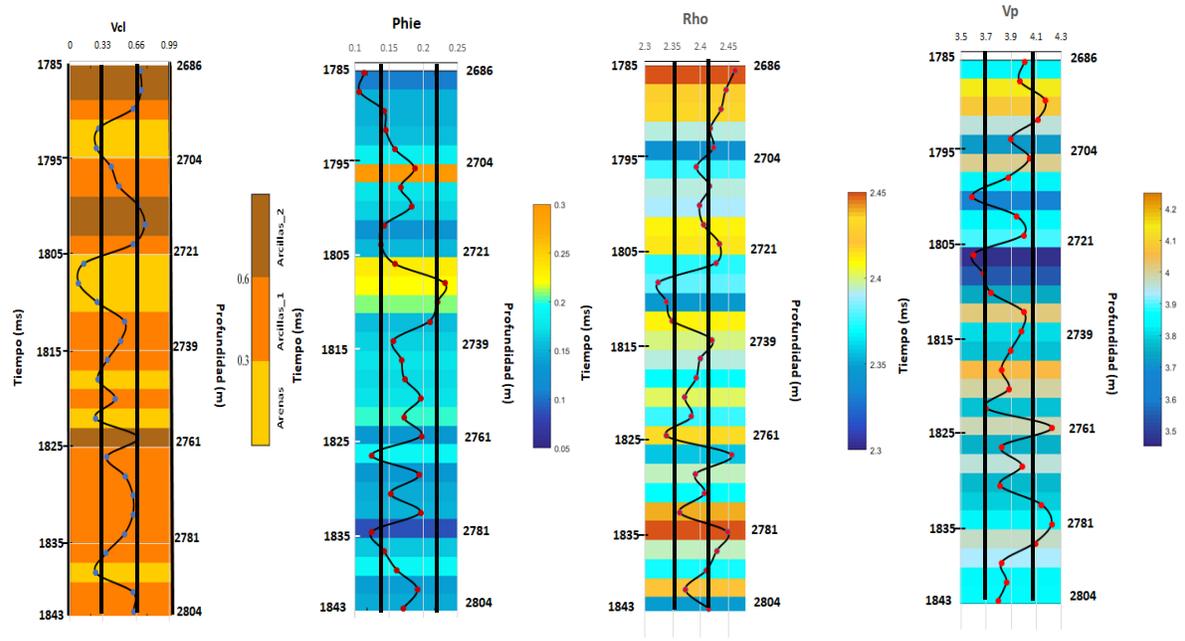
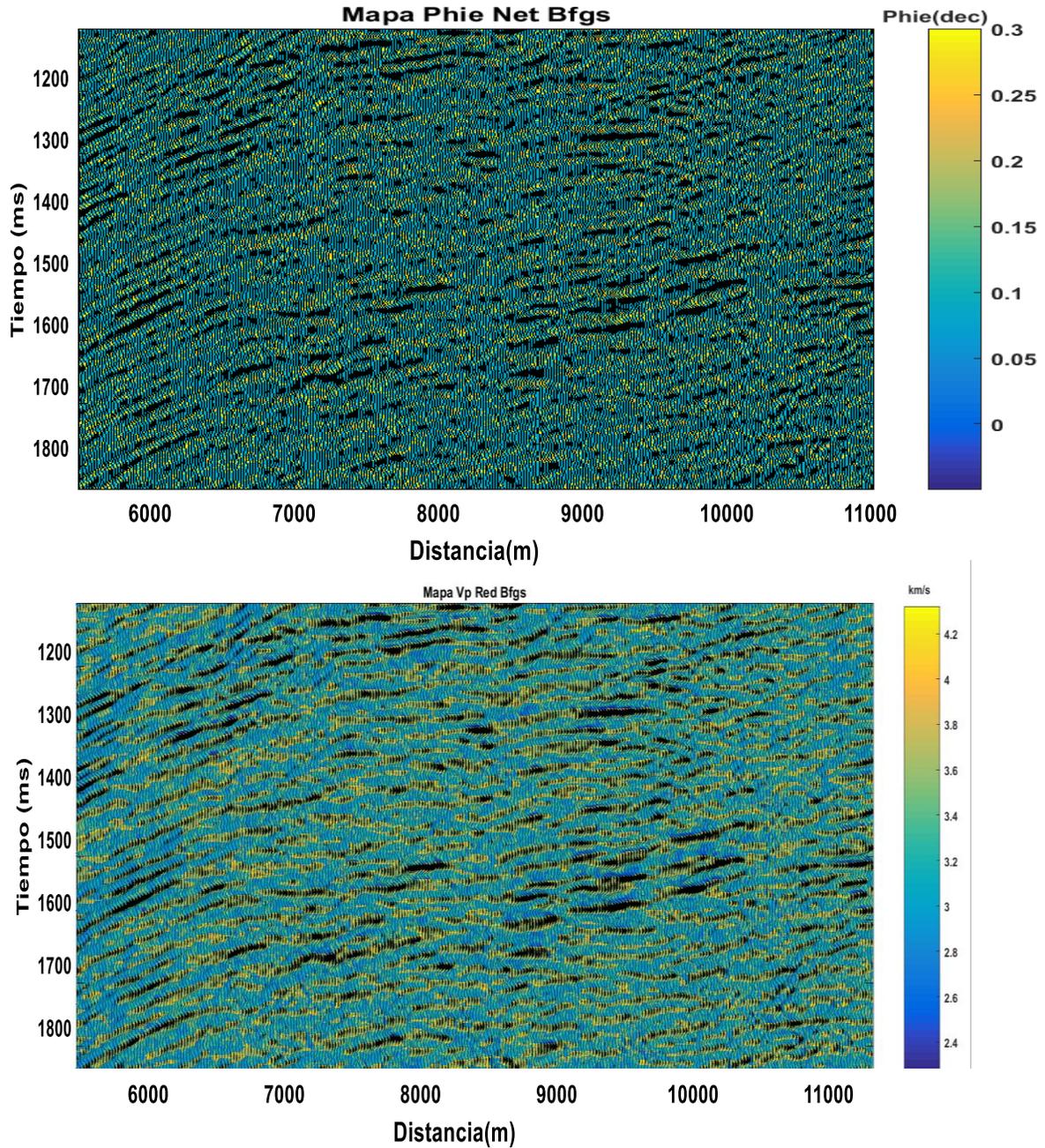


Figura 64. Secciones de R_{hob} , ϕ_{hie} y V_p , respectivamente, donde se sobreponen los datos sísmicos. Se observa que existe correspondencia entre todas ellas, es decir, aquellas áreas más densas que coinciden con las amplitudes altas de las ondículas, y que podrían ser asociadas a lutitas, corresponden a valores bajos de ϕ_{hie} y a valores altos de V_p .



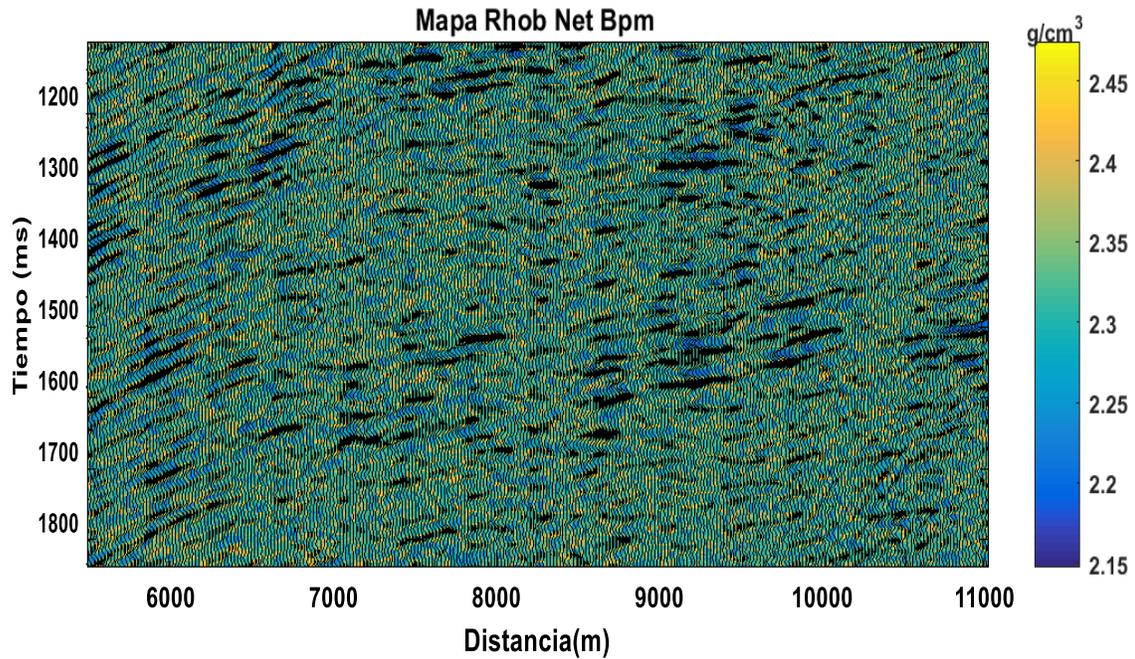
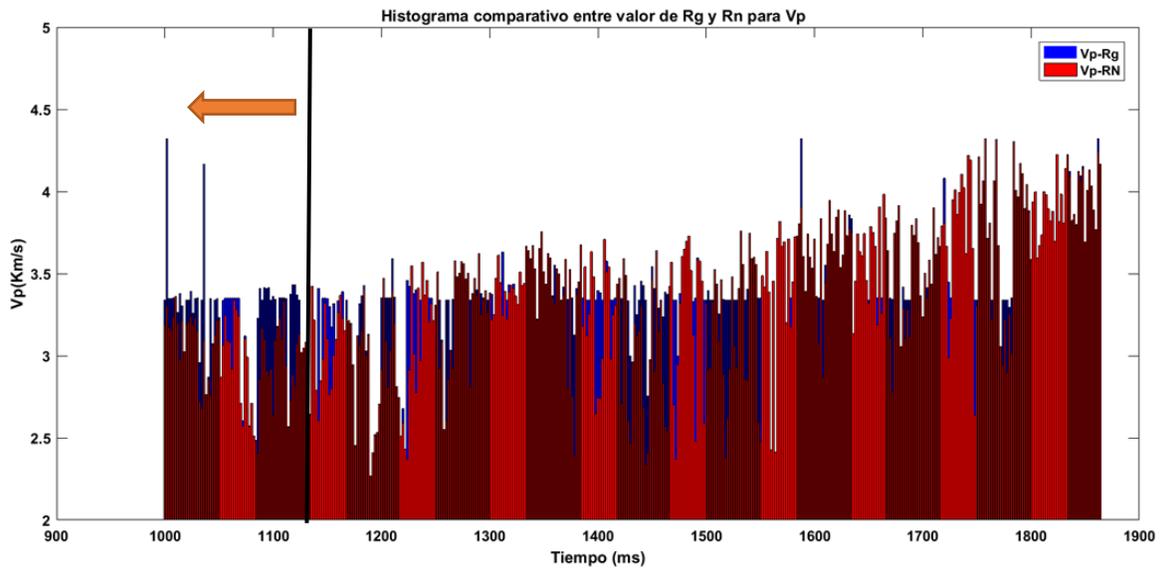


Figura 65. A) Histograma donde se compara la salida de la red para V_p (color rojo), y los valores del registro V_p (color azul). Se puede observar que una de las zonas donde se presentan mayores problemas (mayor diferencia entre ambos valores), es la correspondiente al intervalo 1000-1126 ms. B) Histograma de errores, de la red para predecir V_p , durante el segundo entrenamiento, donde los puntos de validación corresponden a unos de los intervalos más problemáticos, durante el primer entrenamiento.

A)



B)

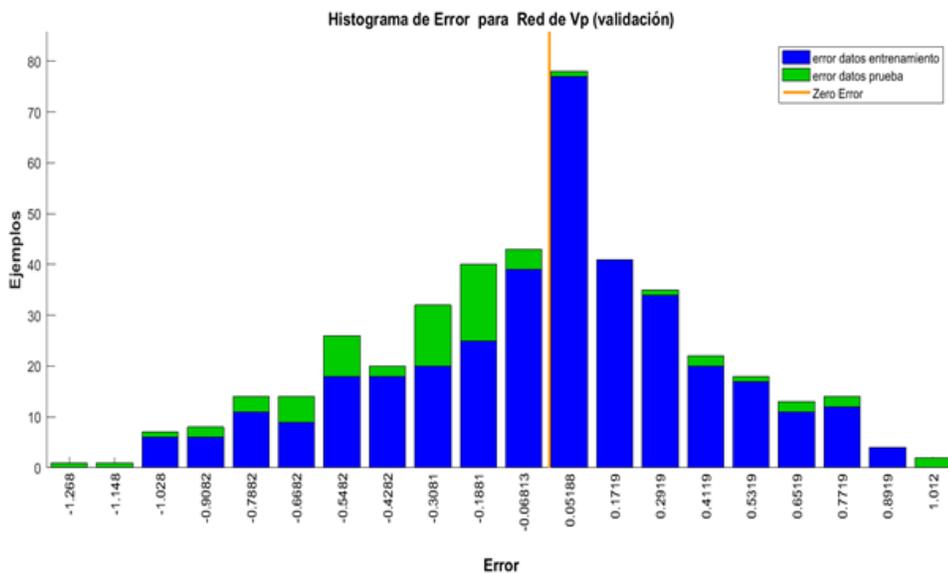
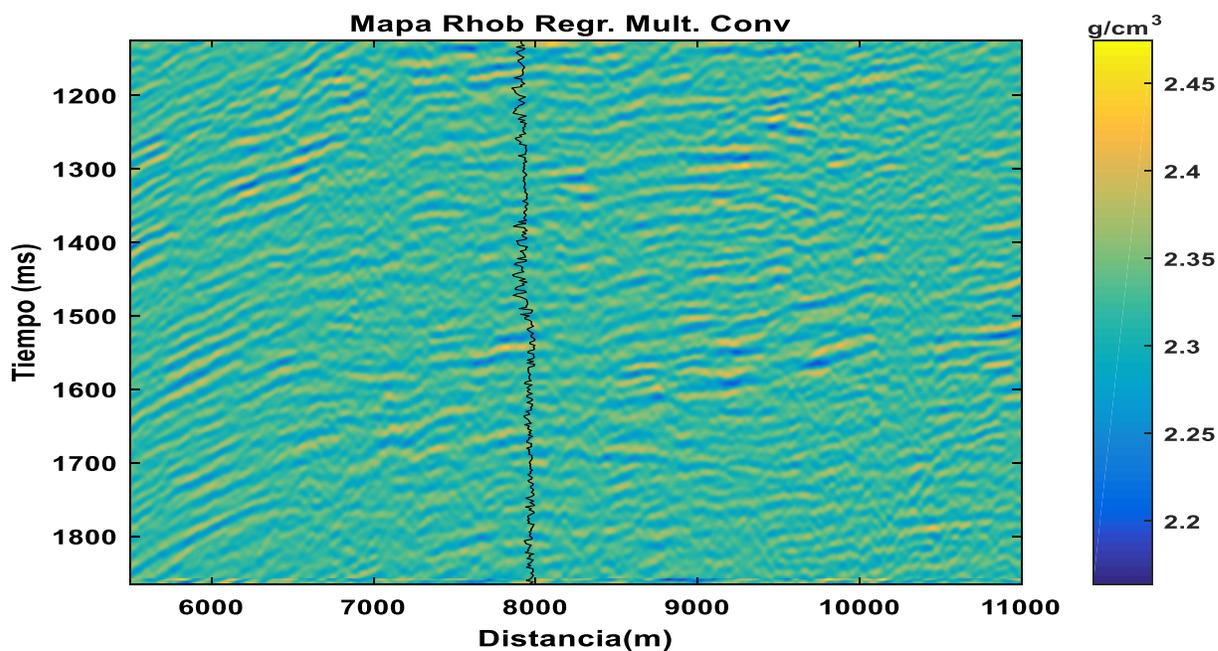
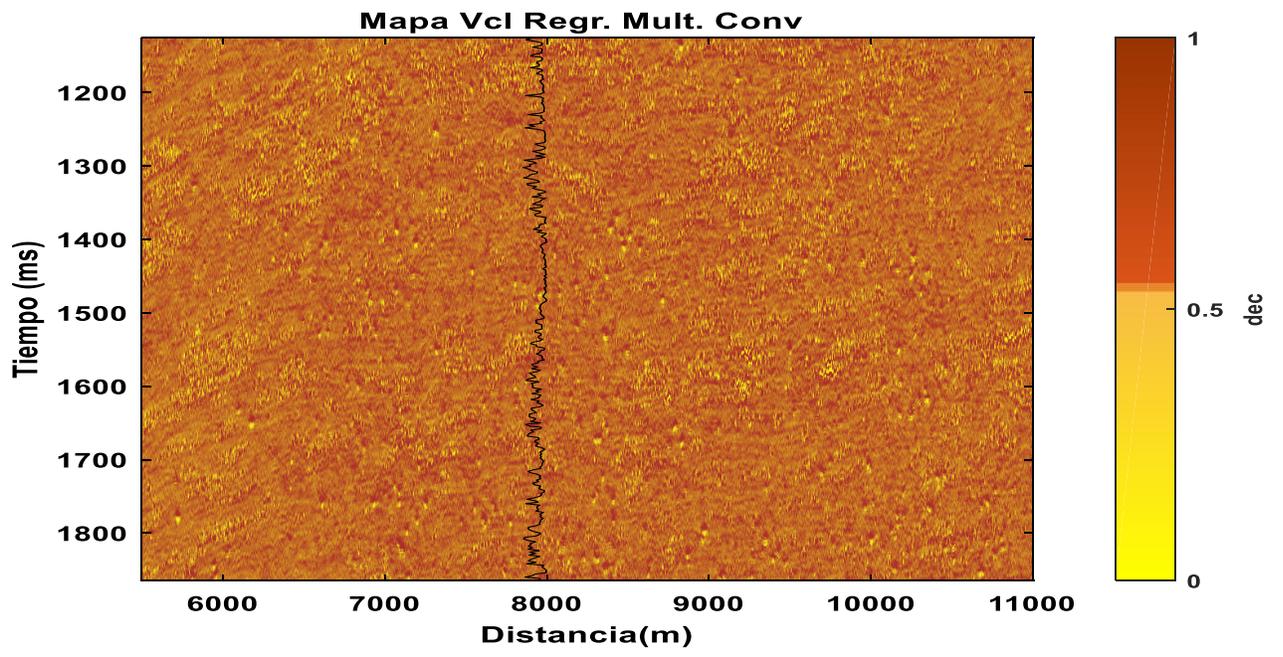


Figura 66. Secciones de R_{hob} y V_{cl} , respectivamente, obtenidas por el método de regresión lineal convolucional, con un número de operadores igual a 7. En la sección de R_{hob} , se conserva el echado de los estratos, no obstante, es complicado distinguir, las litofacies de alta densidad. Para el caso de la sección de V_{cl} , es imposible distinguir más de dos litofacies, inclusive. Además, no se conservan los echados de los estratos.





CONCLUSIONES Y RECOMENDACIONES

- El elemento de la prueba Gamma, como elemento de análisis de los datos de entrada en la metodología aquí propuesta, permite establecer de manera más rápida en comparación de los análisis de factibilidad clásicos, no sólo si es posible la construcción del modelo, sino también la calidad de éste. Lo anterior es posible, siempre y cuando se cuente con la cantidad de datos necesarios para poder estimar el valor real del estadístico Gamma (representativo del ruido contenido en los datos), y que la función a modelar sea suave y continua.
- La prueba Gamma permite llevar a cabo la reducción de la dimensión del espacio de entrada, el cual no es un problema exclusivo del método de redes neuronales para la construcción de modelos no lineales, de una manera más eficiente y precisa, en el sentido que permite evaluar las $2^n - 1$ posibles combinaciones de los datos de entrada en un tiempo relativamente corto, lo que no es posible llevar a cabo con la técnica de regresión inteligente ("stepwise") por poner un ejemplo. Así mismo, la naturaleza de la prueba Gamma es netamente no lineal, lo que los resultados arrojados serán probablemente más precisos como se dijo ya anteriormente, que los obtenidos con técnicas tales como análisis de componentes principales.
- La prueba Gamma permite solucionar el problema de ajuste excesivo que se presenta algunas veces durante el entrenamiento de las redes neurales, lo cual repercute gravemente la capacidad de generalización de las redes, de manera tal que no se sacrifica información (datos) como se hace generalmente, esto es sumamente valioso cuando se cuenta con pocos datos y que quiere construir un buen modelo.
- El entrenamiento de una red neuronal supervisada consiste básicamente en estimar aquella configuración de pesos (w) que minimiza una función de costo (error promedio). Esto es lo que se llama comúnmente problema de optimización o de inversión, en este sentido se podrá tener varias configuraciones de pesos que proporcionen una solución, es por ello que es imprescindible que se tenga una comprensión del fenómeno que se quiere modelar, para la selección de modelo correcto.
- La capacidad de generalización de una red depende principalmente de la cantidad de ruido presente en los datos con los que se entrena la red, así como de la similitud que presentes los datos a los cuales se va a aplicar la red ya entrenada.
- La buena capacidad de generalización de las redes entrenadas y usadas para la construcción de las secciones de propiedades petrofísicas en este documento, queda comprobado tanto de manera cualitativa, en tanto que se puede observar claramente que en cada una de las secciones se conserva el echado de los estratos; así mismo, se

observa que las áreas donde se tiene altas amplitudes corresponden a áreas de alta densidad, alta velocidad y baja porosidad, como cuantitativamente, al obtener errores bajos en los puntos de validación, inclusive durante el segundo entrenamiento de aquella red de la que tuvo el peor desempeño durante el primer entrenamiento (red que se usó para la estimación de V_p).

- Las secciones de las propiedades petrofísicas obtenidas a partir de los modelos no lineales fueron mejores que las obtenidas por el método de regresión lineal convolucional con 7 operadores. Los factores de correlación para la red con mejor desempeño de cada una de las propiedades son: 0.795, 0.785, 0.951 y 0.931 para V_{cl} , ϕ_{hie} , ρ y V_p respectivamente. Mientras que los factores de correlación obtenidos con la regresión lineal convolucional fueron de 0.15 y 0.30 para V_{cl} y ρ respectivamente.

RECOMENDACIONES

- Aplicar la metodología propuesta en este documento, en alguna otra área donde se cuente con una mayor cantidad de información que permita la evaluación de la capacidad de generalización de las redes no solo de manera matemática.
- Implementar alguna metodología de sobre escalamiento de datos al momento de tratar de ajustar la escala de los registros de pozo con la empleada con la sísmica, esto con el objeto de poder asociar mejor los eventos, y así poder construir mejores modelos.

Fuentes de Consulta

- AL-Bulushi, N., King, P., & Blunt, M. (2010). Artificial neural networks workflow and its application.
- Ali, J. K. (s.f.). Neural network:a new tool for petroleum industry. *SPE*.
- Amari, S., Murata, N., & Muller, K. (1997). Asymptotic statistical theory of overtraining and cross-validation . *IEEE*.
- Amix, J., Bass, D., & Whiting, R. (1960). *Petroleum Reservoir Engineering Physical*. New York: McGraw Hill.
- Anderson. (1995). *Learning and memory; An integrated approach* (segunda ed.).
- Arbib, M. (s.f.). *Brains,machines, and the mathematics*. New York: Springer-Verlag.
- Athy, L. (1930). Density,porosity and compactation of sedimentary rocks. *Bull Am. Ass. Pet Geol.*, 1-24.
- Azevedo, L., & Raposo, P. (2009). *Seismic Attributes in Hydrocarbon Reservoirs (M.E thesis)*. Aveiro University.
- Barnes, A. (2001). Seismic attributes in your facies. *CSEG*, 41-47.
- Barnes, A. (2006). Too many seismic attributes? *Landmark Graphics Corporation*.
- Barto, A., Sutton, R., & Anderson, C. (1983). Neuron like adaptive elements than can solve difficult learning control problems. *IEEE*.
- Becker, S., & Hinton, G. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 161-163.
- Bhatt, A., & Helle, H. (2002). Determination of facies from well logs using modular neural networks. *Norwegian University of Science and Technology*.
- Billingham, J., & King, A. (2000). *Wave motion*. Cambridge : Cambridge University Press.
- Brown, A. (1986). Interpretation of 3D seismic data. *AAPG*.
- Brown, A. (1996a). Interpretation of 3D seismic data. *AAPG*.
- Brown, A. (1996b). Seismic attributes and their classification. *The leading edge*, 1090.
- Brown, A. (2004). Interpretation of 3D seismic data. *AAPG*.
- Calderon-Macias, C., Sen, M., & Stoffa, P. (1997). Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation. *Geophysics*, 992-1002.
- Cediel, F., Shaw, R., & Caceres, C. (2003). The Circum-gulf of Mexico and the Caribbean: Hydrocarbon Habitats, Basin Formation, and Plate Tectonics. *AAPG*, 815-848.

- Changeux, J. (2016). *El cerebro y la complejidad*. Medellin.
- Cheeke, J. (2002). *Fundamentals and applications of ultrasonic waves*. Florida: CRC.
- Chen, Q., & Sidney, S. (1997). Seismic attribute technology for reservoir forecasting and monitoring. *The leading edge*.
- Chopra, S., & Marfurt, J. (2003). Seismic attributes- a historical perspective. *Geophysics*.
- Chopra, S., & Marfurt, J. (2007). Seismic attributes for prospect identification and reservoir characterization. *SEG*.
- Churchland, P. (1989). *A neurocomputational perspective: the nature of mind and the structure of Science*. Cambridge: MIT Press.
- Cooper, M., Addison, F., Alvarez, R., Coral, M. G., & R.H., H. A. (1995). Basin development and tectonic history of the Llanos basin, eastern Cordillera, and Middle Magdalena valley, Colombia. . *American Association of Petroleum Geologists*, 1421-1443.
- Dengo, C., & Covey, M. (1993). Structure of the Eastern Cordillera of Colombia; Implications for Trap Styles and Regional Tectonics. *AAPG*, 1315-1337.
- Etayo-Serna, F. (1994.). Estudios geologicos del Valle Superior del Magdalena.
- Evans, D. (2002). *The Gamma test. Data derived estimates of noise for unknown smooth models using near neighbour asymptotics (Ph.D Thesis)*. U.K: Cardiff University.
- Evans, D., & Jones, A. (2002). A proof of the Gamma test. . *Proceedings of the Royal Society*, 2759-2799.
- Evans, D., & Jones, A. (2008). Non-parametric estimation of residual moments and covariance. *Proceedings of the Royal Society*, 2831-2846.
- Evans, D., Jones, A., & Schmidt, W. (2002). Asymptotic moments of near neighbour distance distributions. *Proceedings of the Royal Society*, 2839-2849.
- Freeman, W. (1975/2004). *Mass action in the nervous system*. New York: Academic Press.
- Gomez, E. (2001.). *Tectonic controls on the Late Cretaceous to Cenozoic sedimentary Fill of the Middle Magdalena Valley Basin, Eastern Cordillera and Llanos Basin, Colombia. (Ph.D Thesis)*. New York, Cornell University.
- Gomez, E., Jordan, T., Allmendinger, R., Hegarty, K., & Kelley, S. (2005.). Syntectonic Cenozoic sedimentation in the northern middle Magdalena Valley Basin of Colombia and implications for exhumation of the Northern Andes. *Geological Society of America*, 547-569.
- Grossberg, S. (1976). Adaptive pattern classification and universal recording. *Biological cybernetics*, 23, 121-134.
- Hagan, M., Demuth, H., & Beale, M. (2004). *Neural Network Design (Segunda ed.)*.

- Hampson, D., Schuelke, J., & Quirein, J. A. (2001). Use of multiattribute transforms to predict log properties. *Geophysics*, 220-236.
- Han, D., Nur, A., & Morgan, D. (1986). Effects of porosity and clay content on wave velocities in sandstone. *Geophysics*, 2093-2107.
- Haykin, S. (1999). *Neural networks: A comprehensive foundation* (segunda ed.). Prentice Hall International.
- Haythorn, W. (abril de 2011). *Smooth Regression*. Obtenido de <http://www.smoothregression.com>.
- Hebb, D. O. (1949). The organization of behaviour.
- Hérault, J., & Jutten, C. (s.f.). *Réseaux neuronaux et traitement de signal*. (Hérmes, Ed.)
- Hernández, O. (2017). *Aplicación de redes neuronales en la ingeniería petrolera* (Tesis de maestría). Ciudad de México: UNAM.
- Hopfield, J. (1986). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 1.
<http://www.asimovinstitute.org/neural-network-zoo/>. (s.f.). Obtenido de <http://www.asimovinstitute.org/neural-network-zoo/>
- Huang, S. C., & Huang, Y. F. (s.f.). Bounds on the number of hidden neurons in multilayer perceptrons. *IEEE Trans Neur. Networks.*, 47-55.
- Huang, Z., Shimeld, J., Williamson, M., & Katsube, J. (1996). Permeability prediction with artificial neural network modeling in the Ventura gas field, offshore eastern Canada. *Geophysics*, 422-436.
- Hush. (1989). Classification with neural networks: A performance analysis . *Department of Electrical Engineering and Computer Engineering: University of New Mexico*.
- Hush, D., & Horne, B. (1993). Progress in supervised neural networks. *IEEE*, 8-39.
- Iturrarán, Ú. (2011). Smooth regression to estimate effective porosity using seismic attributes. *Journal of Applied Geophysics*.
- Jones. (1998). *The WinGamma user guide*. Cardiff: Universidad de Wales.
- Jorden, J., & Comberll, F. (1984). Well logging I-rock properties, borehole environment, mud and temperature logging. *SPE*.
- Karayiannis, N., & Venetsanopoulos, A. (s.f.). *Artificial Neural Networks; Learning algorithms, performance evaluation and applications*.
- Kavzoglu, T. (1999). Determining Optimum Structure for Artificial. *25° Annual Technical Conference and Exhibition of the Remote Sensing Society*, 675-682.

- Kerlirzin, P., & Vallet, F. (1993). Robustness in Multilayer Perceptrons. *Neural computation*, 473-482.
- Kingston, D. R., Dishroon, C. P., & Williams, P. (1983). Global Basin Classification System. *The American Association of Petroleum Geologist Bulletin*, 2175-2193.
- Kohonen, T. (1972). Correlation matrix memories. *IEEE transactions of computers*, 353-359.
- Kohonen, T. (1989). Self-organization and associate memory.
- Konkar, N. (1997). *Optimization methodologies dor direct inverse neurocontrol (Ph.D thesis)*. London: Imperial college of Science, Technologie and Medicine, University of London.
- Kromer, A., & Sangiovanni, A. (1989). Optimization techniques for neural networks. *University of California*, 19.
- Langer, H., Nunnari, G., & Occhipinti, L. (1996). Estimation of seismic waveform governing parameters with neural networks. *Journal of Geophysics*, 20109-20118.
- Lecun, Y. (1986). A theoretical framework for backpropagation. *University of Toronto*.
- Lecun, Y. (1993). Pattern recognition and neural networks.
- Lecun, Y., Boser, B., Denker, J., & Henderson, D. (1989). Backpropagation applied to handwritten zip code recognition, neural computation. 541-551.
- Liner, C., Gersztenkorn, A., & Smythe, J. (2004). SPICE; A new general seismic attribute. *SEG*, 433-436.
- Loaiza, M. (2011). *Modelado y Simulación de la Propagación de Ondas Elásticas en Medios Heterogéneos con Métodos Espectrales de Chebyshev (Tesis Licenciatura)*. Universidad Nacional Autónoma de México.
- Ma, Z., & Gómez, E. (2017). Integration of seismic and well-log data using statistical and neural network methods. *The leading edge*.
- Mason, S. (1953). *Feedback theory: some properties of signal flow graphs*.
- Mason, S. (1956). *Feedback theory: Further properties of signal flow graphs*.
- McCormack, M., Zaucha, D., & Dushek, D. W. (1993). First break refraction event picking and seismic data trace editing using neural networks. *Geophysics*, 67-78.
- McCulloch, W. S., & Pitts, W. (s.f.). A logical calculus of ideas immanent in nervous activity. 115-133.
- Meldahl, P., Heggland, R., & Bril, B. (2001). Identifying faults and gas chimneys using multiattributes and neural networks. *The leading edge*, 474-478.

- Mendel, J., & McLaren, R. (1970). 8 reinforcement-learning control and pattern recognition systems. 287-318.
- Mesa, A. (1995.). Diagénesis y calidad del reservorio Campo La Cira, Formaciones Mugrosa y Colorado, Valle Medio del Magdalena. *Tech. rept. Ecopetrol-Instituto Colombiano del Petróleo*.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge.
- Mohaghegh, S., & Ameri, S. (1995). Artificial Neural Network As A Valuable Tool For Petroleum Engineers. *SPE*.
- Murat, M., & Rudman, A. (1992). Automated first arrival picking; A neural network approach. *Geophysics Prosp.*, 587-604.
- Murata, N., & Amori, S. (1994). Network information criterion- determining the number of hidden units for an artificial neural network model. *IEEE*, 865-872.
- Parker, D. (1987). Optimal algorithms for adaptive networks; second order direct propagation, and second order hebbian learning. *IEEE*.
- Parra, J., & Iturrarán, U. (2015). Artificial Neural Networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. *Journal of Applied Geophysics*, 45-54.
- Pindell, J., & Tabbutt, K. (1995.). Mesozoic-Cenozoic Andean paleogeography and regional controls on hydrocarbon systems. (A. Tankard, R. Suarez, & H. Welsink, Edits.) *AAPG*, 101-128.
- Raymer, D., Hunt, E., & Gardner, J. (1980). An improved sonic transit time to porosity transform. *SPWLA*.
- Restrepo, P. (1995). *Late Precambrian to early Mesozoic tectonic evolution of the Colombian Andes, based on new geochronological, geochemical and isotopic data(Ph.D. thesis)*. Tucson, University of Arizona.
- Rogers, L., & Dowla, F. (1994). Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water resources*, 457-481.
- Rogers, S., Fang, J., Farr, C., & Stanley, D. (1992). Determination of lithology from well logs using a neural network. *AAPG*, 731-739.
- Rojas, R. (1996). *Neural Networks*. Berlin: Springer-Verlag.
- Rosenblatt, F. (1958). The perceptron; a probabilistic model for information storage and organization in the brain. *65*, 316-408.
- Roth, G., & Tarantola, A. (1994). Neural networks and inversion of seismic data. *Journal of Geophysics*, 6753-6768.

- Rumelhart, D. E., & McClelland, J. L. (1986). *Explorations in the micrstructure of cognition*. Cambridge: Parallel distributed proccessing.
- Saggaf, M., & Nebrija. (2000). Estimation of lithologies and depositional facies from wireline logs. *AAPG*, 1633-1646.
- Sarmiento-Rojas, L. (2001). *Mesozoic rifting and Cenozoic basin inversion history of the Eastern Cordillera, Colombian Andes. Inferences from tectonic models. (Ph.D. thesis)*. Amsterdam, The Netherlands.
- Schamel, S. (1991). Middle and Upper Magdalena basins, Colombia. *AAPG*, 283-301.
- Schlumberger. (1997). Log interpretation-Principles. 1-2. (Schlumberger, Recopilador)
- Schon, J. (2012). *Physical Properties of Rocks*. ELSEVIER.
- Serra, O. (2007). *Well logging*. Paris: Technip .
- Singh, S. (2005). Permeability prediction using artificial neural network: A case study of Uinta Basin. *SPE*.
- Soubotcheva, N., & Stewart, R. (2004). Predicting porosity logs from seismic attributes using geostatistics.
- Stewart, J. (2008). *Cálculo de una variable (sexta ed.)*. Cengage learning.
- Suarez, M. (1997). *Facies analysis of the Upper Eocene La Paz Formation, and regional evaluation of the Postmiddle Eocene stratigraphy, Northern Middle Magdalena Valley Basin(M.Phil Thesis)*. University of Colorado.
- Subrahmanyam, D., & Rao, P. (2008). Seismic attributes: a review. *International Conference and Exposition on Petroleum Geophysics*, 7.
- T, P., & F, G. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. 978-982.
- Taner, M. (2001). Seismic attributes: Canadian Society of Exploration. *Geophysics*, 48-56.
- Taner, M. (22 de junio de 2018). *Attributes revisited*:<http://www.rock-solidimages/pdf/attrib.htm>.
- Taner, M., & Sheriff, R. (1977). Application of amplitude,frequency and other attributes to stratigraphic. *SEG*.
- Taner, M., Schuelke, J., & Donerty, R. (1994). Seismic attributes. *SEG*, 1104-1106.
- Todorov, V. (s.f.). *Integration of 3C-3D seismic data and well logs for rock property(M.Sc. thesis)*. Calgary,Al: University of Calgary.
- Van der Baan, M., & Juttenez, C. (2000). Neural networks in geophysical applications. *Geophysics*, 1032-1047.

- Velasquez, A. (2012). *3D Multicomponent seismic characterization of a clastic reservoir in the Middle Magdalena Valley Basin, Colombia (M. Sc thesis)*. the Board of Trustees of the Colorado School of Mines.
- Villamil, T. (1993). Mesozoic-Cenozoic Stratigraphy and Tectonic Evolution of the Caribbean Region - Northern South America: Implications for Eustasy from Exposed Sections of a Cretaceous-Eocene Passive Margin Setting. *Geol. Soc. Amer.*, paper C-8. Chap. Relative sea level, chronology, and a new sequence stratigraphy model for distal .
- Wang, P., Jion, F., & Taggart, I. (1995). A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability prediction. *Journal of Petroleum Geology*, 18, 1633-1446.
- Watrous, R. (1987). Learning algorithms for connectionist networks; Applied gradient of non linear optimization. *IEEE*, 619-627.
- Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *1969 IRE WESCON*, 65, 96-104.
- Wieland, A., & Leighton, R. (1987). Geometric analysis of neural network capabilities. *IEEE*, 385-392.

APÉNDICE A

TÉCNICAS DE OPTIMIZACIÓN SIN RESTRICCIÓN

Considere una *función de costo* $\xi(\mathbf{w})$ que es una función continuamente diferenciable de algún vector de pesos desconocido (parámetro) \mathbf{w} . La función $\xi(\mathbf{w})$ mapea los elementos de \mathbf{w} en números reales. Es una medida de cómo elegir el vector de peso (parámetro) \mathbf{w} de un algoritmo de aprendizaje para que se comporte de manera óptima. Queremos encontrar una solución óptima \mathbf{w}^* que satisfaga la condición :

$$\xi(\mathbf{w}^*) \leq \xi(\mathbf{w}) \dots A. 1$$

Esto es , se necesita resolver un *problema de optimización sin restricciones*, expresado de la siguiente manera:

Minimizar la función de costo $\xi(\mathbf{w})$ con respecto al vector de pesos \mathbf{w} A.3

La condición necesaria para la optimización es;

$$\nabla \xi(\mathbf{w}^*) = \mathbf{0} \dots A. 4$$

donde ∇ es el operador gradiente:

y $\nabla \xi(\mathbf{w})$ es el *vector gradiente* de la función de costo:

$$\nabla = \left[\frac{\partial}{\partial w_1}, \frac{\partial}{\partial w_2}, \dots, \frac{\partial}{\partial w_m} \right]^T \dots A. 5$$

$$\nabla \xi(\mathbf{w}) = \left[\frac{\partial \xi}{\partial w_1}, \frac{\partial \xi}{\partial w_2}, \dots, \frac{\partial \xi}{\partial w_m} \right]^T \dots A. 6$$

Una clase de algoritmos de optimización sin restricciones que es particularmente adecuada para el diseño de filtros adaptativos, pero también muy ampliamente aplicada en otros campos, se basa en la idea de descendencia iterativa local:

Comenzando con una conjetura inicial denotada por $\mathbf{w}(0)$, se genera una secuencia de vectores de ponderación $\mathbf{w}(1)$, $\mathbf{w}(2)$, . . . , tal que la función de costo $\xi(\mathbf{w})$ se reduce en cada iteración del algoritmo, como,

$$\xi(\mathbf{w}(n+1)) < \xi(\mathbf{w}(n)) \dots A. 7$$

donde $\mathbf{w}(n)$ es el valor anterior del vector de pesos y $\mathbf{w}(n+1)$ es su valor actualizado.

Se espera que el algoritmo finalmente converja en la solución óptima \mathbf{w}^* . Se dice se "espera" porque existe una clara posibilidad de que el algoritmo diverja (es decir, se vuelva inestable) a menos que se tomen precauciones especiales.

En esta sección se describen tres métodos de optimización sin restricciones que se basan en la idea del descenso iterativo de una forma u otra (**Bertsekas, 1995a**).

Método de descenso más pronunciado.

En el método de descenso más pronunciado, los ajustes sucesivos aplicados al vector de peso w están en la dirección de descenso más pronunciado, es decir, en una dirección opuesta al vector de gradiente $\nabla \xi(w)$. Para la conveniencia de la presentación se escribe,

$$\mathbf{g} = \nabla \xi(\mathbf{w}) \dots \dots \mathbf{A. 8}$$

Por lo tanto, el algoritmo de descenso más pronunciado esta formalmente descrito por:

$$\mathbf{w}(\mathbf{n} + 1) = \mathbf{w}(\mathbf{n}) - \eta \mathbf{g}(\mathbf{n}) \dots \dots \mathbf{A. 9}$$

donde η es una constante positiva llamada *parámetro de rapidez de aprendizaje*, y $\mathbf{g}(\mathbf{n})$ es el vector gradiente evaluado en el punto $\mathbf{w}(\mathbf{n})$. En el avance de la iteración n a $n+1$ el algoritmo aplica la corrección;

$$\begin{aligned} \Delta \mathbf{w}(\mathbf{n}) &= \mathbf{w}(\mathbf{n} + 1) - \mathbf{w}(\mathbf{n}) \\ &= -\eta \mathbf{g}(\mathbf{n}) \dots \dots \mathbf{A. 10} \end{aligned}$$

Para demostrar que la formulación del algoritmo de descenso más pronunciado satisface la **condición A.7** para un proceso iterativo, usamos la serie de expansión de Taylor de primer orden alrededor de $\mathbf{w}(\mathbf{n})$ para aproximar $\nabla \xi(\mathbf{w}(\mathbf{n} + 1))$ como:

$$\xi(\mathbf{w}(\mathbf{n} + 1)) \approx \xi(\mathbf{w}(\mathbf{n})) + \mathbf{g}^T(\mathbf{n}) \Delta \mathbf{w}(\mathbf{n})$$

El uso de la cual es justificado para valores pequeños de η . Sustituyendo la **A.10** en esta aproximación, obtenemos;

$$\begin{aligned} \xi(\mathbf{w}(\mathbf{n} + 1)) &\approx \xi(\mathbf{w}(\mathbf{n})) - \eta \mathbf{g}^T(\mathbf{n}) \mathbf{g}(\mathbf{n}) \\ &= \xi(\mathbf{w}(\mathbf{n})) - \eta \|\mathbf{g}(\mathbf{n})\|^2 \end{aligned}$$

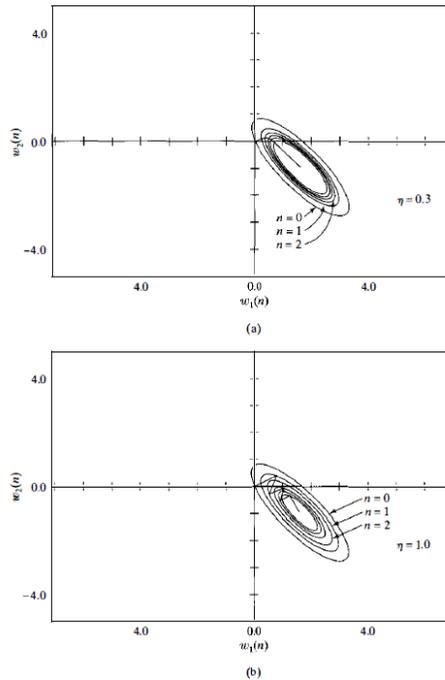
que muestra que, para un parámetro de rapidez de aprendizaje positiva, η , la función de costo disminuye a medida que el algoritmo progresa de una iteración a la siguiente. El razonamiento presentado aquí es aproximado en el sentido de que este resultado final solo es cierto para tasas de rapidez de aprendizaje lo suficientemente pequeñas.

El método de descenso más pronunciado converge a la solución óptima w^* lentamente. Además, el parámetro rapidez de aprendizaje tiene una profunda influencia en su comportamiento de convergencia:

- Cuando η es pequeño, la respuesta transitoria del algoritmo está sobre amortiguada, ya que la trayectoria trazada por $w(\mathbf{n})$ sigue una trayectoria suave en el plano w , como se ilustra en la **Fig. 67a**.

- Cuando η es grande, la respuesta transitoria del algoritmo está sub-amortiguada, ya que la trayectoria de $w(n)$ sigue una trayectoria en zigzag (oscilante), como se ilustra en la **Fig. 67b**.
- Cuando η excede un cierto valor crítico, el algoritmo se vuelve inestable (es decir, diverge).

Figura 67. Trayectoria del método de descenso más pronunciado en un espacio 2D para dos diferentes valores de η . a) $\eta=0.3$ y b) $\eta=1$. Las coordenadas w_1 y w_2 son los elementos del vector w .



Modificado de Haykin, S. (1999). *óp. cit.*, p. 123.

Método de Newton

La idea básica del método de Newton es minimizar la aproximación cuadrática de la función de costo $\varepsilon(w)$ alrededor del punto actual $w(n)$; esta minimización se realiza en cada iteración del algoritmo. Específicamente, usando una expansión de la serie Taylor de segundo orden de la función de costo alrededor del punto $w(n)$, podemos escribir:

$$\begin{aligned} \Delta\xi(w(n)) &= \xi(w(n+1)) - \xi(w(n)) \\ &= g^T(n)\Delta w(n) + \frac{1}{2}\Delta w^T(n)H(n)\Delta w(n) \dots A.11 \end{aligned}$$

Como antes, $g(n)$ es el vector gradiente m-por-1 de la función de costo $\varepsilon(w)$ evaluada en el punto $w(n)$. La matriz $H(n)$ es la matriz de Hessian de m-por-m, también evaluada en $w(n)$. El Hessiano de $\varepsilon(W)$ se define por:

$$H = \nabla^2\xi(w)$$

$$= \begin{bmatrix} \frac{\partial^2 \xi}{\partial w_1^2} & \frac{\partial^2 \xi}{\partial w_1 \partial w_2} & \dots & \frac{\partial^2 \xi}{\partial w_1 \partial w_m} \\ \frac{\partial^2 \xi}{\partial w_2 \partial w_1} & \frac{\partial^2 \xi}{\partial w_2^2} & \dots & \frac{\partial^2 \xi}{\partial w_2 \partial w_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \xi}{\partial w_m \partial w_1} & \frac{\partial^2 \xi}{\partial w_m \partial w_2} & \dots & \frac{\partial^2 \xi}{\partial w_m^2} \end{bmatrix} \dots \text{A. 12}$$

La **Ec. A.12** requiere que la función de costo $\xi(w)$ sea dos veces continuamente diferenciable con respecto a los elementos de w . Diferenciando la **Ec. A.11** con respecto Δw , el cambio $\Delta \xi(w)$ se minimiza cuando;

$$\mathbf{g}(n) + \mathbf{H}(n)\Delta \mathbf{w}(n) = \mathbf{0}$$

Resolviendo esta ecuación para $\Delta w(n)$, se obtiene;

$$\Delta \mathbf{w}(n) = -\mathbf{H}^{-1}(n)\mathbf{g}(n)$$

Esto es,

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) + \Delta \mathbf{w}(n) \\ &= \mathbf{w}(n) - \mathbf{H}^{-1}(n)\mathbf{g}(n) \dots \text{A. 13} \end{aligned}$$

donde $\mathbf{H}^{-1}(n)$ es la matriz inversa del Hessiano de $\xi(w)$.

En términos generales, el método de Newton converge rápidamente de manera asintótica y no muestra el comportamiento en zigzag que a veces caracteriza el método de descenso más pronunciado. Sin embargo, para que el método de Newton funcione, el $\mathbf{H}(n)$ Hessiano tiene que ser una matriz definida positiva para todo n . Desafortunadamente, en general no hay garantía de que $\mathbf{H}(n)$ sea positivo definido en cada iteración del algoritmo. Si el Hessiano $\mathbf{H}(n)$ no es positivo, la modificación del método de Newton es necesaria (**Powell, 1987; Bertsekas, 1995a**).

Método de Gauss-Newton

El método de Gauss-Newton es aplicable a una función de costo que se expresa como la suma de cuadrados de error. Es decir,

$$\xi(n) = \frac{1}{2} \sum_{i=1}^n e^2(i) \dots \text{A. 14}$$

donde se incluye el factor de escala 1/2 para simplificar las cuestiones de análisis posteriores. Todos los términos de error en esta fórmula se calculan sobre la base de un vector de peso w que se fija durante todo el intervalo de observación $1 \leq i \leq n$.

La señal de error $e(i)$ es una función del vector de peso ajustable w . Dado un punto de operación $w(n)$, se linealiza la dependencia de $e(i)$ en w escribiendo;

$$e'(i, w) = e(i) + \left[\frac{\partial e(i)}{\partial w} \right]_{w=w(n)}^T (w - w(n)), i = 1, 2, \dots, n \dots A. 15$$

Equivalentemente, usando una notación matricial, se tiene,

$$e'(i, w) = e(n) + J(n)(w - w(n)) \dots A. 16$$

donde $e(n)$ es el vector error.

$$e(n) = [e(1), e(2), \dots, e(n)]^T$$

y $J(n)$ es la matriz Jacobina de n por m de $e(n)$;

$$J(n) = \begin{bmatrix} \frac{\partial e(1)}{\partial w_1} & \frac{\partial e(1)}{\partial w_2} & \dots & \frac{\partial e(1)}{\partial w_m} \\ \frac{\partial e(2)}{\partial w_1} & \frac{\partial e(2)}{\partial w_2} & \dots & \frac{\partial e(2)}{\partial w_m} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial e(n)}{\partial w_1} & \frac{\partial e(n)}{\partial w_2} & \dots & \frac{\partial e(n)}{\partial w_m} \end{bmatrix} \dots A. 17$$

La matriz Jacobina $J(n)$ es la transpuesta de la matriz gradiente m por n , $\nabla e(n)$, donde,

$$\nabla e(n) = [\nabla e(1), \nabla e(2), \dots, \nabla e(n)]^T$$

El valor actualizado del vector w ($n+1$), entonces estará definido por:

$$w(n+1) = \arg \min_w \left\{ \frac{1}{2} \|e'(n, w)\|^2 \right\} \dots A. 18$$

Usando la **Ec. A.15** para evaluar la norma cuadrada Euclidiana de $e'(n, w)$, obtenemos:

$$\frac{1}{2} \|e'(n, w)\|^2 = \frac{1}{2} \|e(n)\|^2 + e^T(n)J(n)(w - w(n)) + \frac{1}{2} (w - w(n))^T J^T J(n) (w - w(n))$$

Por lo tanto, derivando esta expresión con respecto a w e igualando la expresión a cero, obtenemos;

$$J^T(n)e(n) + J^T(n)J(n)(w - w(n)) = 0$$

Resolviendo esta ecuación para w , podemos escribir con base a la **A.18**

$$w(n+1) = w(n) - \left(J^T(n)J(n)\right)^{-1} J^T(n)e(n) \dots \text{A.19}$$

que describe la forma pura del **método de Gauss-Newton**.

A diferencia del método de Newton que requiere el conocimiento de la matriz de Hess de la función de costo $\varepsilon(n)$, el método de Gauss-Newton solo requiere la matriz Jacobiana del vector de error $e(n)$. Sin embargo, para que la iteración de Gauss-Newton sea calculable, el producto de matriz $J^T(n)J(n)$ debe ser no singular.

Con respecto al último punto, reconocemos que $J^T(n)J(n)$ siempre es un no negativo definido. Para garantizar que no sea singular, la matriz Jacobina $J(n)$ debe tener un rango de fila n ; es decir, las n filas de $J(n)$ en la **Ec. A.17** deben ser linealmente independientes. Lamentablemente, no hay garantía de que esta condición siempre se mantenga. Para evitar la posibilidad de que $J(n)$ sea de rango deficiente, la práctica habitual es agregar la matriz diagonal δI a la matriz $J^T(n)J(n)$. El parámetro δ es una pequeña constante positiva elegida para asegurar que:

$$J^T J(n) + \delta I: \text{positivo definido para toda } n$$

Con base a esto, el método de Gauss-Newton es implementado en la forma ligeramente modificada:

$$w(n+1) = w(n) - \left(J^T(n)J(n) + \delta I\right)^{-1} J^T(n)e(n) \dots \text{A.20}$$

El efecto de esta modificación se reduce progresivamente a medida que el número de iteraciones, n , aumenta. Note también que la ecuación recursiva **A.18** es la solución de la función de costo modificada:

$$\xi(w) = \frac{1}{2} \left\{ \delta \|w - w(0)\|^2 + \sum_{i=1}^n e^2(i) \right\} \dots \text{A.21}$$

Donde $w(0)$ es el valor inicial del vector de peso $w(i)$.

Ahora, se está equipado con las herramientas de optimización que necesitamos para abordar los problemas específicos involucrados en el diseño de redes multicapa de flujo hacia adelante.

APÉNDICE B

ATRIBUTOS SÍSMICOS

Los *atributos físicos* se definen como aquellos atributos que están directamente relacionados con la propagación de la onda, la litología y otros parámetros. Se tomaron las definiciones de atributos sísmicos dadas en **Taner et al. (1994)**, **Taner (1997)**, **Taner (2001)**, **Chopra y Marfurt (2007)**. Se tomó en un inicio una colección de 19 atributos sísmicos, los cuales se sabía de antemano que respondían a rasgos estratigráficos.

Estos atributos se basan en la definición de la traza compleja dada por:

$$C(t) = s(t) + ih(t).. B. 1$$

donde t es el tiempo, C(t) es la traza compleja, s(t) es la traza sísmica, h(t) es la transformada de Hilbert de la traza. Reescribiendo la **Ec. B.1** en su forma polar, se obtiene:

$$C(t) = E(t) * e^{i\phi(t)} \dots B. 2$$

donde

$$E(t) = \sqrt{(s(t))^2 + h(t)^2} \dots B.3$$

es la amplitud de la envolvente, representa la energía total instantánea y su magnitud es del mismo orden que la de la traza sísmica. La envolvente se relaciona con los contrastes de impedancia acústica y la fase instantánea esta dado por:

$$\phi(t) = \tan^{-1}\left(\frac{h(t)}{s(t)}\right) \dots B.4$$

el cual es un buen indicador de la continuidad lateral. Así mismo, puede usarse para el cálculo de la velocidad de la fase y también incluir su primera y segunda derivada.

La **Ec. B.2** también puede reescribirse como:

$$C(t) = E(t) * (\cos\phi(t) + i\sin\phi(t)) \dots B. 5$$

Otro atributo instantáneo puede ser calculado diferenciando con respecto al tiempo la fase instantánea, como se muestra a continuación:

$$\omega(t) = \frac{d\phi(t)}{dt} = \frac{s(t)h'(t) - s'(t)h(t)}{s^2(t) + h^2(t)} \dots B.6$$

Este atributo es conocido como frecuencia instantánea, y se relaciona con la propagación de la onda y los ambientes de depositación. Puede ser usado como indicador de zonas de fractura, y también para la estimación de la atenuación sísmica.

El atributo de varianza esta dado por;

$$\sigma_t^2 = \frac{\sum_{j=t-L/2}^{j=t+L/2} w_{j-t} \sum_{i=1}^l (x_{ij} - \bar{x}_j)^2}{\sum_{j=t-L/2}^{j=t+L/2} w_{j-t} \sum_{i=1}^l (x_{ij})^2} \dots \dots B. 7$$

donde x_{ij} es el valor de la muestra en una posición horizontal i y j es una muestra en tiempo vertical. w_{ij} es un factor de suavizado vertical sobre una ventana de longitud L.

La raíz cuadrada de la amplitud promedio (por su acrónimo en inglés, RMS) está dada por:

$$A_{RMS} = \sqrt{\frac{1}{N} \sum_{k=N/2}^{N/2} (s(t+k))^2} \dots \dots B. 8$$

donde N representa en número de muestras de tiempo por traza.

La banda de ancho instantánea esta dado por:

$$\omega_B(t) = \left| \frac{\frac{dE(t)}{dt}}{E(t)} \right| \dots \dots B. 9$$

El suavizado (sweetness) está definido de la siguiente manera:

$$Suavizado(t) = \frac{E(t)}{\sqrt{f_a(t)}} \dots \dots B. 10$$

donde $f_a(t)$ es la frecuencia promedio y E(t) la envolvente.

Los atributos; local flatness, echado iluminado (dip illumination), dip deviation, Chaos, Amplitud, Edge, AVO, y first derivative = (ds/dt) , fueron obtenidos usando el software de Petrel.

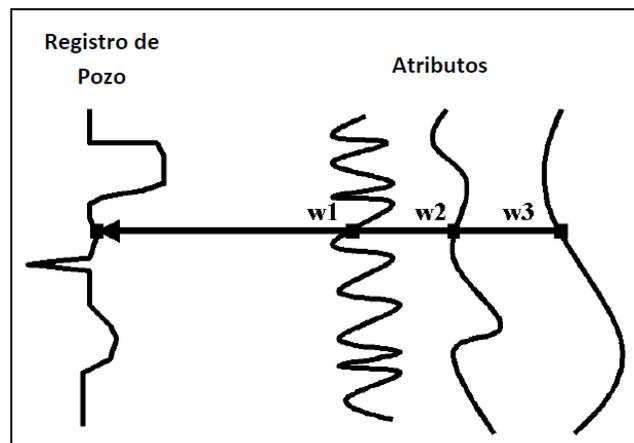
Todos estos atributos fueron calculados sobre la traza 532, la cual es la que pasa más cerca del pozo X-2

APÉNDICE C

REGRESION LINEAL MULTIVARIABLE CONVOLUCIONAL

En el caso de un estudio de más de un atributo es necesario utilizar regresión lineal multivariada. De cada registro de pozo, de la propiedad petrofísica objetivo, se toma una muestra que será modelada como una combinación lineal de atributos sísmicos en el mismo momento de tiempo, como se muestra en la (Fig. 68)

Figura 68. Asumiendo un caso para tres atributos sísmicos. Cada pozo muestra es modelo de una combinación lineal de los atributos en el mismo tiempo



Modificado de Hampson et al., 2001.

En cada muestra de tiempo el registro de pozo es modelado mediante una ecuación lineal:

$$L(t) = w_0 + w_1A_1(t) + w_2A_2(t) + w_3A_3(t) \dots C. 1$$

Los pesos pueden ser derivados minimizando el error cuadrático medio de predicción, la cual es una extensión de la ecuación de minimización del error para la regresión lineal simple (Ec. C.2), la cual se muestra a continuación;

$$E^2 = \frac{1}{N} \sum_{i=1}^N (y_i - a - bx_i)^2 \dots C. 2$$

$$E^2 = \sum_{i=1}^N (L_i - w_0 + w_1A_{1i} + w_2A_{2i} - w_3A_{3i})^2 \dots C.3$$

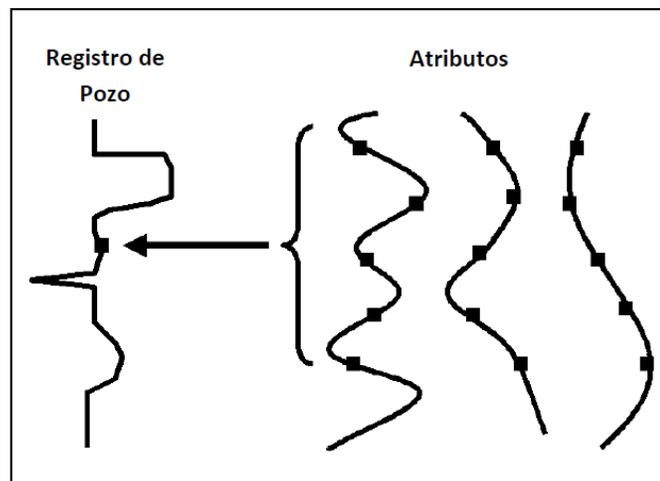
La solución para los cuatro pesos produce la ecuación normal estándar:

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} N & \sum A_{1i} & \sum A_{2i} & \sum A_{3i} \\ \sum A_{1i} & \sum A_{1i}^2 & \sum A_{1i}A_{2i} & \sum A_{1i}A_{3i} \\ \sum A_{2i} & \sum A_{1i}A_{2i} & \sum A_{2i}^2 & \sum A_{2i}A_{3i} \\ \sum A_{3i} & \sum A_{1i}A_{3i} & \sum A_{2i}A_{3i} & \sum A_{3i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum L_i \\ \sum A_{1i}L_i \\ \sum A_{2i}L_i \\ \sum A_{3i}L_i \end{bmatrix} \dots\dots\dots C.4$$

Al igual que en el caso de atributo único, el error cuadrático medio, **Ec. C.3**, calcula utilizando los pesos derivados, que constituye una medida de bondad de ajuste para la transformación. (**Hampson et al., 2001**).

Para resolver las diferencias de frecuencia entre el dato del registro de pozo y los datos sísmicos (atributos) se realiza un “operador convolucional” (**Fig. 69**).

Figura 69. Usando un operador convolucional de 5 puntos para relacionar los atributos sísmicos al pozo objetivo.



Modificado de Hampson et al., 2001.

El operador convolucional consiste en utilizar varios valores vecinos de cada atributo y no solo la muestra que le corresponde en un tiempo determinado (**Hampson et al., 2001**).

La extensión de la **Ec. C.1** con operador convolucional *, se expresa de la siguiente manera:

$$L(t) = w_0 + w_1 * A_1(t) + w_2 * A_2(t) + w_3 * A_3(t) \dots C.5$$

En este trabajo se tomó un operador convolucional de 7.

APÉNDICE D

PRUEBA GAMMA

La prueba Gamma es una herramienta de modelado y análisis no lineal que nos permite examinar la naturaleza de una relación hipotética de entrada / salida en un conjunto de datos numéricos. **La estadística Gamma Γ** es el parámetro fundamental y esencial calculado con **la prueba Gamma**. Primero reportado en **Stefánsson et al. (1997)** con la conjetura de que un algoritmo muy simple (la prueba Gamma) podría usarse para estimar directamente a partir de un conjunto dado de datos de entrada / salida la medida en que los datos fueron derivados de un modelo suave subyacente, aunque este modelo fuese desconocido.

La prueba Gamma es una técnica totalmente no paramétrica y sus resultados se aplican independientemente de los métodos particulares utilizados posteriormente para construir el modelo. La prueba Gamma reúne dos ideas importantes: **Es un estimador de error cuadrático medio, que se aplica a todos los modelos no lineales suavizados**. En los estudios de regresión lineal, el coeficiente de correlación y los estimadores del Error Medio Cuadrado (MSE) similares se usan para calcular la precisión de ajuste del modelo. Si el MSE es cero, los datos se encuentran a lo largo de la línea de regresión, por lo que claramente existe una fuerte relación entre la entrada y la salida. Si las estimaciones de error son altas, la línea de regresión no tiene sentido. Para el modelado no lineal, existen numerosas técnicas para el ajuste de curvas, incluidas las redes neuronales, ver **Bishop (1996)**, regresión lineal local y máquinas de vectores de soporte

Procedimiento para calcular la estadística Gamma Γ

Supongamos que recibimos un conjunto de datos de entrada-salida

$$\{\mathbf{x}_1(i), \mathbf{x}_2(i), \dots, \mathbf{x}_m(i), \mathbf{y}_i\} = \{(\mathbf{x}_i, \mathbf{y}_i) | 1 \leq i \leq M\} \dots D.1$$

donde m es la dimensión del espacio de entrada y M es el número de puntos de datos. Pensamos en el vector $\mathbf{x} = (x_1, x_2, \dots, x_m) \in \mathbb{R}^n$ como la entrada, confinada a un conjunto limitado cerrado $C \subset \mathbb{R}^n$, y (sin pérdida de generalidad) el correspondiente escalar $y \in \mathbb{R}^n$, como la salida, y la gran pregunta es: ¿en qué medida la salida está determinada por la entrada? Supongamos que la relación subyacente entre la entrada y la salida se puede descomponer en un componente suave y un componente ruidoso. Escribimos esto como:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \mathbf{r} \dots D.2$$

donde \mathbf{x} es un vector de entrada, \mathbf{f} es una función suave con primera y segunda derivadas, esta es la única suposición requerida. En particular, se observa que la prueba Gamma no es directamente aplicable a problemas que involucran datos categóricos. Uno necesita tener datos de precisión con 2 o más dígitos del rango de medición efectiva tanto en entradas como en salidas. La relación entre la respuesta \mathbf{y} y el vector explicativo \mathbf{x} tiene la forma dada en la **Ec. D.2** donde $\mathbf{f}(\mathbf{x}) = \boldsymbol{\varepsilon}(\mathbf{y}|\mathbf{x})$ es la verdadera función de regresión y \mathbf{r} es el residuo o ruido.

En cierto sentido, uno puede pensar esto como la expectativa de ver y dada la observación x como en el enfoque Bayesiano. La variable estocástica r representa el ruido, con media cero y varianza finita. Más precisamente, r representa la parte de la salida que no puede ser contabilizada por ningún modelo de datos sin problemas. Podemos suponer que la media de r es igual a cero ya que cualquier sesgo constante puede ser absorbido en la función desconocida f . La prueba Gamma proporciona una estimación de la varianza $\text{Var}(r)$ de r . Una vez que sabemos qué parte de la variación de datos se determina y qué parte se debe al ruido, ajustar una curva a los datos es considerablemente más fácil y confiable.

Considérese que x y x' representan dos puntos en el espacio de entrada C . Tomando como base la **Ec. D. 2** $y = f(x) + r$, y similarmente $y' = f(x') + r'$. Por lo tanto,

$$\frac{1}{2}(y' - y)^2 = \frac{1}{2}(r' - r)^2 + (f(x') - f(x))^2 \dots \mathbf{D.3}$$

La continuidad de f implica que:

$$|f(x') - f(x)| \rightarrow 0 \text{ as } |x' - x| \rightarrow 0 \dots \mathbf{D.4}$$

Por la **Ec. D.2**, se obtiene,

$$\frac{1}{2}(y' - y)^2 \rightarrow \frac{1}{2}(r' - r)^2 \text{ as } |x' - x| \rightarrow 0 \dots \mathbf{D.5}$$

Por lo tanto, tomando la **expectación** en ambos lados de la **Ec. D.5**, se obtiene que,

$$\varepsilon\left(\frac{1}{2}(y' - y)^2\right) \rightarrow \text{Var}(r) \text{ as } |x' - x| \rightarrow 0 \dots \mathbf{D.6}$$

De hecho, si los puntos x y x' son idénticos, entonces $\varepsilon\left(\frac{1}{2}(y' - y)^2\right) = \text{Var}(r)$. Sin embargo, dado cualquier conjunto de datos finitos, no podemos hacer la distancia $|x' - x|$ entre los dos puntos x y x' arbitrariamente pequeños y por lo tanto, no podemos evaluar el límite dado en **Ec. D.6** directamente. Por lo tanto, por **Ec. D.6** vemos que el error involucrado en la estimación de $\text{Var}(r)$, por la expectativa $\varepsilon\left(\frac{1}{2}(y' - y)^2\right)$, esencialmente depende sólo de la distancia entre los puntos x y x' . Este error se minimizará si se considera pares de valores de salida y y y' para los cuales las entradas correspondientes x y x' son los vecinos más cercanos en el espacio de entrada C (es decir, los pares de puntos x y x' para los cuales la distancia $|x' - x|$ es minimizada). La base de la idea es que, si x y x' son vecinos cercanos en el espacio de entrada y f es continuo, entonces y y y' deberían estar cerca en el espacio de salida.

Considere que x'_i denote el vecino más cercano de x_i entre los puntos de entrada $x = (x_1, x_2, \dots, x_m)$ definido como un punto que está más cerca de x_i que cualquier otro. Las relaciones vecinas más cercanas para un conjunto de M puntos se pueden encontrar en $O(M \ln(M))$ tiempo utilizando la técnica **árbol kd** desarrollada por **Bentley (1975)**. Esta búsqueda se puede realizar de manera eficiente mediante el uso de las propiedades de árbol para eliminar rápidamente grandes porciones del espacio de búsqueda. **Un árbol kd es un**

árbol binario en el que cada nodo es un punto k-dimensional. Vamos a definir \mathbf{p} como el número de vecinos más cercanos para calcular la estadística Gamma Γ . La pregunta que surge es **¿cuál es el valor óptimo de \mathbf{p} en una situación particular?** Cuando tenemos muchos datos disponibles, se puede uno dar el lujo de tener una cantidad proporcionalmente mayor. Sin embargo, es un hecho notable que en muchos miles de experimentos (con conjuntos de datos no tan grandes) informados en los documentos teóricos y aplicados relacionados con la prueba Gamma (véase **Durrant, 2001; Evans y Jones, 2002; Jones, 2004**), demuestran que tomando $p \approx 10$ generalmente se obtienen resultados bastante buenos, y que el tiempo empleado en buscar optimizar \mathbf{p} a menudo no vale las ganancias normalmente marginales. Por lo tanto, se usa $\mathbf{p} = 10$ como punto de partida para llevar a cabo los experimentos en este trabajo.

La prueba Gamma estima $\mathbf{Var}(\mathbf{r})$ construyendo primero un árbol kd usando los vectores de entrada \mathbf{x}_i ($1 \leq i \leq M$) y luego usando el árbol kd para construir una lista de los k -ésimos ($1 \leq k \leq \mathbf{p}$) vecinos más cercanos a $\mathbf{x}_{N(i,k)}$ ($1 \leq i \leq M$) de \mathbf{x}_i , donde se define $\mathbf{N}[i, k]$ como una lista de índices correspondientes al conjunto de k -ésimos vecinos más cercanos equidistantes de \mathbf{x}_i . El algoritmo entonces calcula la muestra promedio de las variables aleatorias promedio $\gamma_M(\mathbf{k})$ y $\delta_M(\mathbf{k})$ de la manera mostrada a continuación:

$$\gamma_M(\mathbf{k}) = \frac{1}{2M} \sum_{i=1}^M |\mathbf{y}_{N[i,k]} - \mathbf{y}_i|^2 \dots D. 7$$

donde $|\cdot|$ denota la distancia Euclidiana. Para \mathbf{k} in en rango $1 \leq \mathbf{k} \leq \mathbf{p}$. La **Ec. D.8** puede considerarse como una secuencia de estimados para $\varepsilon \left(\frac{1}{2} (\mathbf{y}' - \mathbf{y})^2 \right)$, y en cada caso, se obtiene una indicación del error por calcular la distancia media cuadrada de los **k -ésimo** vecinos más cercanos, definido por:

$$\delta_M(\mathbf{k}) = \frac{1}{M} \sum_{i=1}^M |\mathbf{x}_{N[i,k]} - \mathbf{x}_i|^2 \dots D. 8$$

donde $\mathbf{y}_{N(i,k)}$ es el valor de salida asociado con $\mathbf{x}_{N(i,k)}$ (nótese que $\mathbf{y}_{N(i,k)}$ no es necesariamente el k -ésimo vecino más cercano de \mathbf{y}_i en el espacio de salida). Intuitivamente, se puede observa que la **Ec. D.6** en cierta forma es equivalente a

$$\gamma_M(\mathbf{k}) \rightarrow \mathbf{Var}(\mathbf{r}) \text{ as } \delta_M(\mathbf{k}) \rightarrow \mathbf{0} \dots D. 9$$

Sin embargo, dado un conjunto finito de datos, no se pueden hacer las distancias entre los vecinos más cercanos arbitrariamente pequeñas como en el caso de la **Ec. D.6**, así que tampoco se puede evaluar este límite de manera directa. La prueba Gamma se basa en la afirmación de que la relación entre $\gamma_M(\mathbf{k})$ y $\delta_M(\mathbf{k})$ es aproximadamente lineal cuando $\delta_M(\mathbf{k}) = \mathbf{0}$, es decir existe una constante \mathbf{A} de tal forma que;

$$\gamma_M(\mathbf{k}) = \mathbf{Var}(\mathbf{r}) + \mathbf{A} \delta_M(\mathbf{k}) + \vartheta(\delta_M(\mathbf{k})) \text{ as } \delta_M(\mathbf{k}) \rightarrow \mathbf{0} \dots D. 10$$

Como resultado (asumiendo que el espacio de entrada no tiene puntos aislados se puede esperar que la distancia promedio entre los vecinos más cercanos converja a cero (en probabilidad) a medida que el número de puntos de datos M aumenta, esto es:

$$\delta_M \rightarrow 0 \quad \text{as } M \rightarrow \infty \dots D. 11$$

y se ve que se puede establecer una relación lineal entre $\gamma_M(\mathbf{k})$ y $\delta_M(\mathbf{k})$ sujeto a la condición de que tengamos suficientes puntos de datos. Habiendo establecido la linealidad entre $\gamma_M(\mathbf{k})$ y $\delta_M(\mathbf{k})$ se estima el límite de la **Ec. D. 9** de $\gamma_M(\mathbf{k})$, cuando $\delta_M(\mathbf{k}) \rightarrow 0$ realizando regresión lineal en los pares $\{(\delta_M(\mathbf{k}), \gamma_M(\mathbf{k})) : \mathbf{i} \leq \mathbf{p} \leq \mathbf{b}\}$ de la manera siguiente;

$$\gamma_M(\mathbf{k}) = \tau + A\delta_M(\mathbf{k}) \dots D. 12$$

donde la intersección vertical Γ devuelve la estimación para **Var (r)**. El parámetro de la pendiente **A** también se obtiene, ya que normalmente contiene información sobre la complejidad de la superficie desconocida $\mathbf{y}=\mathbf{f}(\mathbf{x})$. El principal resultado de **Evans y Jones (2002)** es que si **C** es un cuerpo convexo compacto en \mathbb{R}^m y muestras de datos $\mathbf{x} \in \mathbf{C}$ seleccionadas con una densidad de muestreo positiva suave, φ , entonces el número Γ devuelto por el algoritmo converge en probabilidad a **Var (r)** cuando $M \rightarrow \infty$. Una discusión rigurosa completa matemática de estos problemas teóricos que conforman la base de la prueba Gamma se puede encontrar en **Evans (2002)**.