

Detección en línea de onda-P en sismogramas 3-D utilizando redes neuronales

CARLOS ALZATE

GERMÁN CASTELLANOS

Grupo de Control y Procesamiento Digital de Señales.

Universidad Nacional de Colombia Sede Manizales, canalza@epm.net.co

RESUMEN

Se desarrolla un detector en línea del tiempo de arribo de ondas P sobre registros electrónicos de tres componentes de eventos tectónicos. Se lleva a cabo la detección en línea de la onda-p empleando redes neuronales (RN), utilizando dos diferentes formas: en la primera se sintetiza el detector de LRV, mientras en la segunda se desarrolla un clasificador estadístico con las respectivas clases (señal con onda-p presente y señal con solo ruido presente). La detección se realiza empleando redes neuronales de tipo perceptrón multi-capas, en las cuales se aprovechan las siguientes cualidades: su capacidad de mapeo no lineal de entrada-salida, buena generalización, bajo costo computacional, entre otras. Las anteriores cualidades hacen que las RN sean apropiadas para el proceso de señales en tiempo real. Las entradas a la RN corresponden a datos normalizados del registro sísmico y las características que miden el grado de polarización de la onda. Los registros procesados se obtuvieron de la base de datos del Instituto Iris y de Ingeominas, y su análisis da como resultado un mejor desempeño del sistema en la detección del tiempo de arribo, aunque no presenta una exactitud aceptable en la estimación del tiempo de arribo de la onda P .

PALABRAS CLAVE: REDES NEURONALES, PERCEPTRÓN MULTI-CAPA, DETECCIÓN ONDA P.

ABSTRACT

We have developed a P-wave arrival time detector of tectonic events over digital three-component records online. This process uses two different methods: the first one works with a LRV detector, and the second one uses a statistical classifier of classes (signal with P-wave and signal with seismic noise). Detection uses multi-layer perceptron neural network where is taken advantage of non-linear input-output, relation, well-generalization, low computational demands, etc. These appropriate qualities for real time signal processing. The input data for the system becomes a normalized vector of seismic signal and its polarization arguments. Testes were done with data from IRIS and INGEOMINAS database. Although the system has a good performance, its accuracy is not acceptable.

KEYWORDS: NEURAL NETWORK, PERCEPTION, MULTICAPA, DETECTING P-WAVE

INTRODUCCIÓN

La detección de la onda-p es referida a la tarea de detección de cambios abruptos, la cual consiste en identificar el instante del tiempo en el que las propiedades del proceso cambian intempestivamente, considerando que antes y después del momento del cambio, las propiedades del proceso son invariantes en el tiempo. Tal planteamiento permite que la derivación formal de algoritmos de detección de cambios abruptos se realice dentro de la clase de métodos estadísticos de análisis de procesos aleatorios estacionarios, para los cuales se consideran dos modelos de procesos

aleatorios: completamente independientes y los basados en modelos regresivos (AR). La prueba de hipótesis más empleada corresponde al logaritmo de relación de verosimilitud (LRV), de la cual se derivan los algoritmos analizados: la suma acumulativa (CUSUM), el LRV generalizado (GLRV), χ^2 CUSUM y CUSUM bilateral, entre otros. Sin embargo, en caso de tener ruido no gaussiano a la entrada del detector, la función LRV tiene un carácter no lineal con la entrada y, prácticamente, es muy difícil hallar una expresión explícita con el vector de entrada, lo que dificulta la síntesis práctica del detector. Así, el detector basado en el LRV para ruido no gaussiano debe considerarse como un dispositivo con mapeo no lineal. Por cuanto las redes neuronales pueden aproximar una gran clase de mapeo no lineal, su aplicación se ha considerado en la detección de señales (Lippmann *et al.*, 1989).

Manuscrito recibido para evaluación el 30 de octubre de 2002.

Artículo aceptado para publicación por el Comité Editorial el 30 de octubre de 2003.

En el presente artículo se lleva a cabo la detección en línea de la onda- p empleando RN, utilizando dos diferentes formas: en la primera se sintetiza el detector de LRV, mientras en la segunda se desarrolla un clasificador estadístico con las respectivas clases (señal con onda- p presente y señal con solo ruido presente). La detección en línea se realiza deslizando una ventana de N muestras a través de todo el registro sísmico de tres componentes. En cada instante de tiempo se obtienen las funciones de caracterización que corresponden a las entradas de la red neuronal, se realiza la propagación y se decide si en la ventana actual está el tiempo de arribo de la onda- p .

DETECCIÓN EN LÍNEA DE ONDA- P UTILIZANDO REDES NEURONALES

Detector de LRV. La detección de señales implica decidir si una señal objetivo está presente o no, a partir de datos observados. Existen dos criterios comunes para el diseño e implementación de un detector de señales: el criterio de Bayes y el criterio de Neyman-Pearson. Ambos criterios dan como resultado el detector de LRV (Luo *et al.*, 1997), el cual está compuesto de dos partes: la primera determina la LRV de la función densidad de probabilidad (fdp) en presencia de la señal objetivo y de la fdp en ausencia en la señal objetivo (ruido). La segunda parte determina la detección (salida del detector igual a uno) si la razón de verosimilitud es mayor que un umbral λ \square

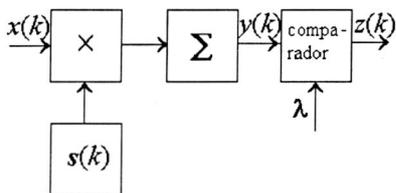


Figura 1

La detección de un cambio abrupto de la onda- p , se puede simplificar al caso de análisis cuando se tiene un solo patrón dado $s(k)$, $1kn$, que representa el cambio abrupto más ruido blanco gaussiano (k) , $1 \leq k \leq n$, con media $E[(\eta)] = 0$ y varianza $E[\eta^2(k)] = \sigma^2$. El algoritmo de detección puede ser simplificado hasta el correlador que calcule el producto escalar de y con s [4]:

$$Z(n) = (y, s) = \sum_{k=1}^n s(k)x(k) > \lambda_0,$$

Acordes con la estructura del correlador de la figura 1, se sugiere una red de tres capas con n neuronas en la capa de entrada, m neuronas en la capa oculta y función de transferencia $f_1(\cdot)$ y una neurona como capa de salida con función de transferencia $f_2(\cdot)$. En principio, la estructura del detector secuencial descrito puede ser

implementado con una red neuronal del tipo retropropagación (figura 2). Los pesos $w_{ij}^{(1)}$, $1 \leq i \leq n$ denotan la conexión entre la neurona i en la capa de entrada y la neurona j en la capa oculta; mientras $y_j(k)$ y $f_2(k)$, $1 \leq j \leq m$ corresponden a su salida y función de activación, respectivamente; $w_{jl}^{(2)}$ denota la conexión entre la neurona j de la capa oculta y la neurona en la capa de salida. Se define respectivamente $z(k)$, $f_3(k)$ y γ como la salida, la función de activación, y el umbral para la neurona en la capa de salida. Sin embargo, en caso de tener ruido no gaussiano a la entrada del detector, la función de verosimilitud tiene un carácter no lineal con la entrada y, prácticamente, es muy difícil hallar una expresión explícita con el vector de entrada, lo que dificulta la síntesis del detector. El detector basado en la relación de verosimilitud para ruido no gaussiano debe considerarse por tanto, como un dispositivo con mapeo no lineal. Por cuanto las redes neuronales del tipo MLP pueden aproximar una gran clase de mapeo no lineal, su aplicación se ha considerado en la detección de señales. En Luo *et al.*, (1997) se sugiere la realización de la función de verosimilitud para señales de entrada con diferente fdp a la gaussiana, entrenando adecuadamente los pesos de la RN para obtener el nivel de no linealidad requerido por el detector.

Detector basado en la clasificación simple de eventos: En este caso, se desarrolla un clasificador estadístico con las respectivas clases (señal con onda- p presente y señal con solo ruido presente). La RN identifica la clase a la cual el patrón pertenece debido a su capacidad de extracción de características importantes a partir del conjunto de entrenamiento. La clasificación es una transformación que mapea los puntos intermedios obtenidos de la extracción de características a puntos correspondientes a cada clase en un espacio de decisión n -dimensional, donde n es el número de clases por reconocer. En este esquema los perceptrones multicapa realizan reconocimiento de patrones debido a que poseen un sistema no supervisado para extracción de características (capas ocultas) y un sistema supervisado para clasificación (capa de salida).

ENTRENAMIENTO DEL SISTEMA

Funciones características

Con el objeto de aumentar la efectividad del estimador durante la extracción de parámetros, se emplean las funciones características. Entre las más empleadas están las siguientes:

Razón de energía. En las señales sísmicas de tres componentes, se puede construir una función para detectar movimiento sísmico coherente en un plano bidimensional. La razón entre la energía en el plano horizontal y la energía total en el instante de tiempo i está dado:

$$ER = \frac{\sum_{n=1}^{i+N-1} (x_1(n)^2 + x_2(n)^2)}{\sum_{n=1}^{i+N-1} (x_1(n)^2 + x_2(n)^2 + x_3(n)^2)} \quad (1)$$

donde x_1, x_2 y x_3 son las tres componentes de la señal sísmica, N es el tamaño de la ventana. ER tiende a 1 si la mayoría de la energía sísmica está concentrada en el plano xy , y tiende a 0 si está concentrada en la componente vertical.

Short Term Average/Long Term Average (STA/LTA). Se calcula la razón entre el valor medio de la amplitud absoluta de dos ventanas de tiempo consecutivas, una corta (STA) y una larga (LTA), y se compara con un umbral predeterminado que declara la presencia del evento sísmico. La determinación del tamaño óptimo de las ventanas depende del contenido frecuencial del sismo (Oonincx, 1998). Se considera una señal x con N muestras, con la que se construye una función característica dada por:

$$k(i) = x(i)^2 + C_u(i)(x(i) - x(i-1))^2 \quad (2)$$

para $i \geq 1$, con $C_x(i)$ como parámetro dependiente del tiempo. Se define a $\square(0) = x(0)^2$, y en forma recursiva los promedios sobre las ventanas cortas y largas se establecen como:

$$STA(i) = C_1 \square(i) + (1 - C_1)STA(i-1) \quad (3)$$

$$LTA(i) = C_2 \square(i) + (1 - C_2)LTA(i-1) \quad (4)$$

para $i \geq 1$, y valores iniciales $STA(0) = LTA(0) = \square(0)$. Los valores de las constantes C_1 y C_2 están entre 0 y 1, valores que se deben sintonizar en la práctica según la aplicación. La determinación del cambio es declarada como:

$$I_p = \inf \{i \in N \mid STA(i) / LTA(i) > C_3\} \quad (5)$$

siendo C_3 el umbral de decisión.

Función rectilínea F . Ésta mide el grado de polarización lineal de una señal sísmica de tres componentes, la cual aumenta si la energía sísmica se concentra en una dirección principal. La onda- p está generalmente polarizada de forma lineal, de esta forma se espera el máximo valor de F en el tiempo de arribo de la onda- p (Anant *et al.*, 1995).

$$F = 1 - \frac{\lambda_2}{\lambda_1} \quad (6)$$

donde λ_1 es el mayor valor propio y λ_2 es el segundo mayor valor propio de la matriz de covarianza:

$$M = \begin{bmatrix} Var(x_1) & Cov(x_1, x_2) & Cov(x_1, x_3) \\ Cov(x_1, x_2) & Var(x_2) & Cov(x_2, x_3) \\ Cov(x_1, x_3) & Cov(x_2, x_3) & Var(x_3) \end{bmatrix} \quad (7)$$

donde x_1, x_2 y x_3 , y son series de tiempo correspondientes a los tres componentes de una señal sísmica. $Var(x_1)$ es la varianza de x_1 y $Cov(x_1, x_2)$ es la covarianza entre x_1 y x_2 :

$$Cov(x_1, x_2) = \frac{1}{N_t} \sum_{i=1}^{N_t} (x_1(i) - \mu_1)(x_2(i) - \mu_2) \quad (8)$$

donde μ_1 y μ_2 son los valores medios de x_1 y x_2 respectivamente y N_t es el número de muestras. La función rectilínea F es igual a 1 cuando una onda está totalmente polarizada de forma lineal (arribo de la onda- p) y a 0 cuando no existe polarización.

Función grado de polarización DOP . Mide el grado de polarización de una señal de tres componentes:

$$DOP = \frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1 + \lambda_2 + \lambda_3)^2} \quad (9)$$

donde $\lambda_1, \lambda_2, \lambda_3$ son los valores propios de la matriz de covarianza M . Esta función característica puede ser usada para detectar ondas- p y ondas- s debido a que éstas poseen alto grado de polarización (Oonincx, 1998).

Función valor absoluto. Definida de la siguiente forma:

$$F_{abs} = abs \{x(i)\}, \forall i = 1, 2, \dots, N, \quad (10)$$

donde N es el número de muestras y x es una señal sísmica. Esta función tiene alta fidelidad y velocidad de proceso (Dai *et al.*, 1997).

Función envolvente analítica compleja. Usada ampliamente en sistemas de detección con el algoritmo de STA/LTA y está definida como (Kuang *et al.*, 1998):

$$E(t) = \sqrt{x^2(t) + \tilde{x}^2(t)} \quad (11)$$

donde $\tilde{x}(t)$ es la transformada Hilbert de la señal sísmica $x(t)$.

Función de planaridad. Registra el grado de polarización de una señal de tres componentes en un plano, y está definida por (Oonincx, 1998):

$$P = \frac{\lambda_1 + \lambda_2 - 2\lambda_3}{\lambda_1 + \lambda_2} \quad (12)$$

donde $\lambda_1, \lambda_2, \lambda_3$ son los valores propios de la matriz de covarianza M . La onda- s está polarizada de tal forma que el movimiento de sus partículas está confinado a un plano; por tanto se espera el máximo valor de P en el tiempo de arribo de esta onda.

Topología de la RN

La función valor absoluto tiene alta fidelidad y velocidad de proceso (Dai *et al.*, 1996), mientras la función rectilínea (6) y el grado de polarización (DOP) (9) son medidas discriminantes debido a que miden el grado de polarización de una serie de tiempo, y la onda- p es una onda compresional que generalmente está polarizada de forma lineal. De esta forma, el espacio de entrada lo conforman la función valor absoluto, la función rectilínea y la función DOP .

Número de neuronas de entrada. Está determinado por el número de muestras de la ventana correspondiente a la función valor absoluto, la función rectilínea y la función DOP . El número

de muestras de la ventana es determinado a través de la función rectilínea y la norma *varimax* (Anant *et. al.*, 1995):

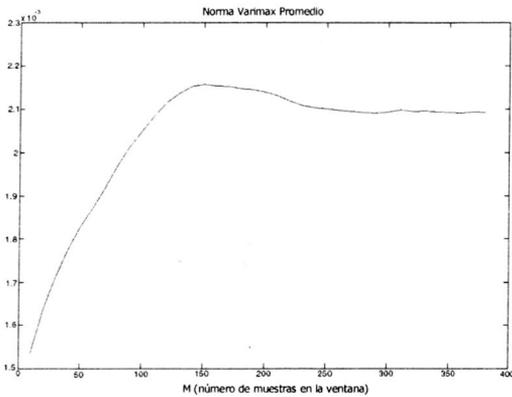


Figura 3. Norma varimax promedio.

$$V_c = \frac{\sum_i x(i)^4}{\left(\sum_i x(i)^2\right)^2} \quad (13)$$

donde x es la serie de tiempo correspondiente a la función recti-

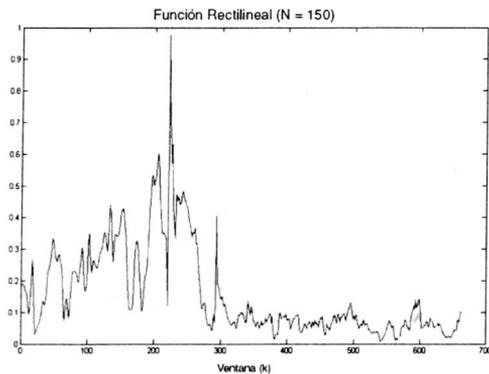


Figura 4. Ejemplo de función rectilínea.

lineal para ventanas deslizantes de M muestras, $M = 2, 3, 4, \dots, M_{MAX}$. La idea es obtener el valor óptimo de M para que la función rectilínea tenga sólo un pico prominente que corresponde a la ventana donde está la onda- p , es decir, minimizar la entropía de la función rectilínea. La norma *varimax* mide la entropía de una serie de tiempo; mientras más alta es, menor número de picos tiene la señal. Se determinó que el valor óptimo promedio de M es de 150 muestras en cada ventana como se muestra en la figura 3; por tanto, el número de neuronas de entradas es 150 para la función valor absoluto, una para la función rectilínea y otra para la función *DOP*. En total, el número de neuronas de entrada es 152.

En las figuras 4 y 5 se muestra la función rectilínea y la función *DOP*, respectivamente, de la señal de tres componentes de la figura 6 para una ventana de 150 muestras con 5 de traslape. El máximo de las funciones corresponde a la ventana que contiene el tiempo de arribo de la onda- p .

Número de capas ocultas. Cualquier espacio de decisión arbitrario puede ser aproximado por una red neuronal de perceptrones multicapa con una capa oculta de neuronas no lineales (Haykin, 1999). De esta forma es necesaria sólo una capa oculta para separar el espacio de entrada que corresponde al conjunto de entrenamiento, pero esto no implica que sea el valor óptimo en el sentido de generalización.

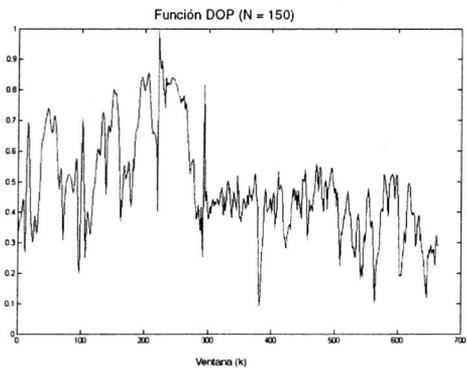


Figura 5. Ejemplo de función DOP.

Número de neuronas ocultas. Este número determina la cantidad de hiperplanos que van a formar el espacio de decisión. En general, la cantidad de neuronas ocultas debe ser la mínima posible para evitar sobreentrenamiento y mala generalización. Se realizaron pruebas empezando con dos neuronas ocultas y entrenando con el algoritmo de *Levenberg-Marquardt* y el criterio de

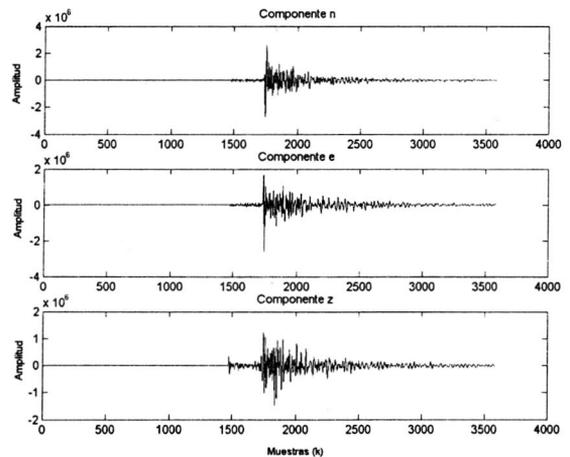


Figura 6. Señal de prueba de tres componentes.

Early Stopping [14]. Los resultados mostraron que el número mínimo de neuronas para obtener una buena generalización fue de 5 neuronas en la capa oculta.

Conjunto de entrenamiento. Inicialmente se utilizaron conjuntos pequeños de entrenamiento correspondientes a 10 datos de onda-p y 10 datos de ruido. El desempeño del sistema fue pobre, con una elevada tasa de falsas alarmas y falso rechazo. Posteriormente, se aumentó de forma significativa el número de patrones de ruido. Finalmente, el sistema fue entrenado con 465 patrones de ruido y 13 patrones de onda-p.

Algoritmo de entrenamiento. Éste corresponde al de *Levenberg-Marquardt*, el cual es una modificación del método de minimización de *Gauss-Jordan*. Este algoritmo tiene menor tiempo de convergencia que la mayoría de los métodos clásicos como la *Regla Delta Generalizada*. Se utilizó el criterio de *Early Stopping* para detener el entrenamiento cuando el error de validación alcanzó el mínimo. Este método garantiza que el sistema entrenado tiene buena generalización para un conjunto predeterminado de patrones de validación. El error de entrenamiento fue del orden de 10^{-5} . EL esquema del sistema neuronal se ilustra en la figura 7:

En el caso del detector neuronal de razón de verosimilitud, la síntesis se realizó con los siguientes parámetros:

Tamaño de la ventana (N). Debe ser determinado a partir del tipo y del contenido frecuencial de la señal sísmica (Oonincx, 1998). Se determinó el valor de *N* a partir del contenido frecuencial de las señales sísmicas de tipo regional y local, lo que resulta en un valor de *N*=80 muestras, que corresponde a 4s para señales muestreadas a 20 Hz.

Número de neuronas ocultas. Se determinó empezando con dos neuronas y probando el desempeño de la red, para un resultado final de 25 neuronas en la capa oculta.

En cuanto al conjunto de entrenamiento, éste fue determinado de la misma forma que el clasificador antes analizado, empezando

con conjuntos pequeños y luego incrementando dependiendo del desempeño del sistema. Al final de las pruebas el conjunto de entrenamiento estuvo compuesto de 2300 patrones de ruido y 48 patrones de onda-p. El algoritmo de entrenamiento fue una modificación de la *Regla Delta Generalizada* descrito en Luo et al. (1977).

Resultados y conclusiones

Los sistemas fueron probados utilizando 77 señales sísmicas de tres componentes muestradas a 20 Hz, que corresponden a sismos regionales y locales de la zona de California y Japón. Adicionalmente se agregó ruido blanco gaussiano a las señales para determinar el desempeño de los algoritmos frente a condiciones de ruido.

La relación señal a ruido se calculó hallando la varianza del ruido antes de la onda-p, y hallando la varianza de la señal+ruido después de la onda-p en una ventana estacionaria:

$$SNR = 10 \log_{10} \left(\frac{\sigma_{S+N}^2}{\sigma_N^2} - 1 \right)$$

Los resultados del tiempo de proceso se representan en la tabla 1, los cuales se midieron para los algoritmos en la estación SUN Enterprise 3500, como también en Matlab y en Fortran 95 en un PC sobre Pentium 4.

Tabla 1

Detector	Tiempo proceso (SUN)	Tiempo proceso (PC F95)	Tiempo proceso (PC Matlab)
Clasificador	212 ms	950 ms	4 s
LRV	59 ms	400 ms	1 s

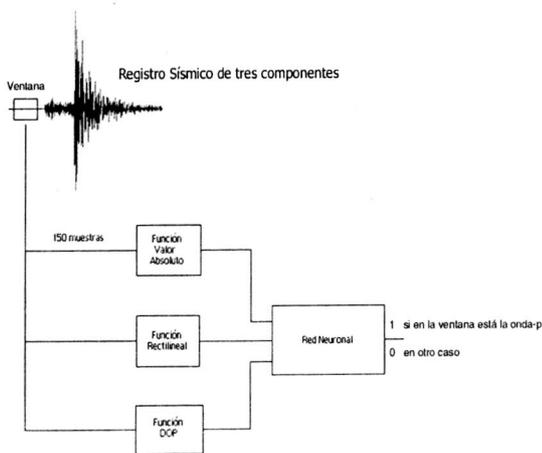


Figura 7. Diagrama del detector basado en el clasificador.

En el caso del detector neuronal de razón de verosimilitud, el desempeño dio como resultado un error promedio de 9,1% para una relación señal a ruido mínima promedio de 15dB. Mientras, para el clasificador, el desempeño del detector resultó en un error promedio de 1,3 % y la relación señal a ruido mínima promedio fue de 9,67dB. Como se observa el trabajo del clasificador es superior al detector basado en la LRV.

Se presentan ejemplos de detección con el clasificador neuronal sobre señales sísmicas de tres componentes con alta SNR (Figura 8) y baja SNR (Figura 9). El recuadro indica la ventana donde se encuentra el tiempo de arribo de la onda-p.

BIBLIOGRAFÍA

KUANG, C. and MENDEL, J. (1994). First break refraction event picking using fuzzy logic systems, IEEE Transactions on Fuzzy Systems, Vol 2, No. 4. November 1994.

- LUO, F. and UNBEHAUEN, R. (1997). *Applied neural networks for signal processing*, University of Erlangen-Nuremberg, 1997.
- DAI, H. and MACBETH, C. (1997). The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings, *British Geological Survey*, Edinburgh, Scotland.
- ANANT, K. S. and DOWLA, F. A. (1995). *Wavelet transform methods for phase identification in three-component seismograms*, University of California, Davis and Lawrence Livermore National Laboratory.
- OONINCX, P. J. (1998). *Automatic Phase Detection in Seismic Data using the Discrete Wavelet Transform*, Centrum voor Wiskunde en Informatica.
- HAYKIN, S. (1999). *Neural Networks: A comprehensive foundation*, Prentice Hall.
- LIPPMANN, R. and BEKMAN P. 1989. "Adaptive Neural Ney Preprocessing for signal detection in non gaussian noise", *Adv. Neural Information Processing System*, Vol 1, pp. 124-132.