

TESIS DOCTORAL

NUEVOS ESQUEMAS DE DETECCIÓN BASADOS EN EL USO DE PREDICCIÓN NO LINEAL: APLICACIÓN A LA DETECCIÓN DE INCENDIOS EN GRANDES ÁREAS

Autor: Pablo Andrés Bernabeu Soler

Director: Luís Vergara Domínguez

Departamento de Comunicaciones

Universidad Politécnica de Valencia

Índice

Capítulo I: Planteamiento de la Tesis	11
1. Introducción	13
2. Estructura de la tesis	18
Capítulo II: Esquema Básico de Predicción-Detección	21
1. Introducción	23
2. Esquema General	25
3. Predicción lineal del ruido infrarrojo	27
4. El detector	29
5. Diseño del sistema	31
6. Experimentos	35
7. Mejoras sobre el esquema planteado	42
Capítulo III: Técnicas de Predicción no Lineal	43
1. Objetivos	45
2. Introducción	45

3. Conceptos generales sobre predicción.....	46
4. Esquemas de predicción.....	48
4.1. Lineales.....	48
4.2. No lineales.....	50
4.2.1. Redes Neuronales.....	51
4.2.2. Funciones de Base Radial.....	54
4.2.3. Series de Volterra.....	59
4.2.4. Predictores Bilineales.....	62
5. Estructuras de Wiener y Hammerstein.....	64
Capítulo IV: Diseño del Predictor no Lineal: Estructura de Wiener.....	69
1. Introducción.....	71
2. Técnicas aplicadas al diseño de la no linealidad.....	73
2.1. Estimación directa.....	73
2.2. Estimación por aproximación polinomial.....	75
3. Evaluación de las técnicas de diseño.....	76
3.1. Evaluación técnicas en el bloque de datos de información A	78
3.1.1. Resultados obtenidos para el bloque A	78
3.1.2. Comentarios.....	83
3.2. Evaluación técnicas en el bloque de datos de información B	85
3.2.1. Resultados obtenidos para el bloque B	85
3.2.2. Comentarios.....	90
3.3. Evaluación técnicas en el bloque de datos de información C	92
3.3.1. Resultados obtenidos para el bloque C	92
3.3.2. Comentarios.....	97

4. Experimentos sobre el esquema completo.....	98
4.1. Análisis de la autocorrelación.....	99
4.2. Análisis de la función densidad de probabilidad	104
4.3. Probabilidades de falsa alarma	108
4.4. Probabilidad de detección.....	110
5. Estimación orden N del predictor lineal.....	113
6. Cota mínima del error de predicción lineal.....	115
7. Conclusiones	119
Capítulo IV: El detector	123
1. Introducción	125
2. Detector adaptado al subespacio de señal.....	127
2.1. Consideraciones teóricas	127
2.2. Ajuste de β y p	131
2.3. Validación del ajuste a los bloques A, B y C	139
3. Detector de crecimiento	148
3.1. Determinación de la distribución del estadístico	148
3.2. Experimentos	151
4. Análisis conjunto de ambos tipos de detectores	154
5. Conclusiones	158
Capítulo VI: Conclusiones y Líneas Futuras	159
1. Conclusiones	
1.1. Relativas a la predicción.....	161
1.2. Relativas al detector.....	163
2. Líneas futuras.....	164

ÍNDICE

Apéndice A: Obtención de la Fórmula para la Aproximación Polinómica de la Media Condicional.....	165
Apéndice B: Cálculo de la Descomposición de Wold	179
Referencias	187
Publicaciones y Congresos	197

Capítulo I:

Planteamiento de la Tesis

Capítulo I

PLANTEAMIENTO DE LA TESIS

1. Introducción:

La presente tesis se enmarca dentro de la problemática de la detección de eventos en sistemas de televigilancia para grandes áreas [CAR97]. Más en concreto, se considera la detección precoz de la presencia de un posible incendio en un entorno natural, en un área grande, mediante un sistema de infrarrojos. Aunque enfocada hacia la mencionada aplicación, se pretende obtener conclusiones generales, válidas para otros sistemas de detección en los que, en definitiva, el evento a detectar se manifieste mediante la presencia de una cierta señal en un fondo de ruido.

Asumimos de partida una estructura de datos 3-D: dos variables espaciales asociadas a la altura (o rango) y el acimut de la celda bajo análisis y una variable temporal asociada al barrido en que nos encontramos. Tal estructura es la que se genera en sistemas de teledetección basados en barrido de la zona vigilada, mediante un sensor cuyo diagrama de radiación recorre periódicamente dicha zona en altura (o rango) y en acimut, dirigido de forma mecánica, electrónica o mezcla de ambas (véase la figura 1).

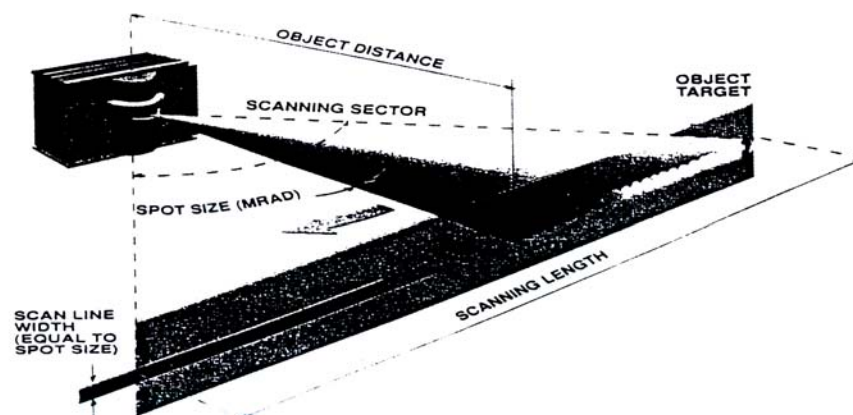


figura 1: Diagrama de barrido del sensor de infrarrojos: modelo "Thermoprofile 6 HT".

La mencionada estructura 3-D también puede generarse mediante una cámara de infrarrojos: sucesión de imágenes 2D.

En general una posible decisión se hará en base a un bloque de información, asociado a un cierto número de rangos, acimuts y barridos consecutivos.

Para facilitar el manejo de esta información, la estructuramos vectorialmente, para ello consideraremos un vector \mathbf{v} , de dimensión N , donde:

$$N = B \times A \times R$$

siendo B , A y R respectivamente el n° de barridos, n° de acimuts y n° de rangos incluidos en el mencionado bloque de información.

El problema a resolver es detectar la presencia de una señal de fuego no controlado en el vector anterior. Tendremos dos posibles hipótesis sobre el vector \mathbf{v} .

$$\begin{aligned} H_0: & \quad \mathbf{v} = \mathbf{r} \\ H_1: & \quad \mathbf{v} = \mathbf{r} + \mu \mathbf{s} \end{aligned}$$

Donde \mathbf{r} es ruido infrarrojo (señal captada por el sensor en ausencia de fuego) originada por el entorno natural bajo vigilancia y \mathbf{s} modela la forma en que la posible presencia de fuego afecta al vector \mathbf{v} .

El modelo que asumamos para \mathbf{s} dependerá de nuestro conocimiento previo de cómo se manifiesta un fuego, que puede también variar según la zona de la imagen infrarroja que consideremos, según el tipo de entorno, etc.

En primera instancia, si \mathbf{s} es determinista y conocida (sea $E_s = \mathbf{s}^T \mathbf{s}$ su energía), y \mathbf{r} es un vector $N(\mathbf{0}, \mathbf{I})$, es decir, sus elementos son variables aleatorias Gaussianas e incorreladas de media cero y varianza unidad, el detector óptimo es el de la figura 2 (filtro adaptado) [SCH91], donde el estadístico m es $N(\mu\sqrt{E_s}, 1)$, y podemos ajustar el umbral η para garantizar máxima probabilidad de detección (PD) para cierta probabilidad de falsa alarma.

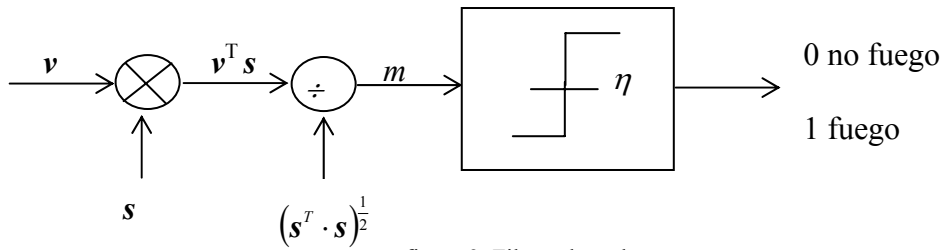


figura 2: Filtro adaptado

Sobre la anterior estructura existen diferentes extensiones que permiten considerar situaciones tales como:

1. El ruido es $N(\mathbf{0}, \mathbf{R})$, donde \mathbf{R} es la matriz de autocorrelación. En ese caso el filtro adaptado se aplica sobre los vectores [SCH91].

$$\mathbf{v}' = \mathbf{R}^{-1/2} \cdot \mathbf{v} \quad \mathbf{s}' = \mathbf{R}^{-1/2} \cdot \mathbf{s}$$

generándose el estadístico,

$$m' = \frac{\mathbf{v}'^T \cdot \mathbf{s}'}{(\mathbf{s}'^T \cdot \mathbf{s}')^{1/2}}$$

que tiene una distribución $N(\mu\sqrt{E_s'}, 1)$, donde $E_s' = \mathbf{s}'^T \mathbf{s}' = \mathbf{s}^T \mathbf{R}^{-1} \mathbf{s}$.

2. La señal no es conocida, aunque puede asumirse incluida en un subespacio cuya matriz de proyección es \mathbf{P} . La estructura del filtro adaptado se modifica, sustituyendo \mathbf{s} por $\mathbf{s}_e = \mathbf{P} \cdot \mathbf{v}$ (estimamos \mathbf{s} mediante proyección de \mathbf{v} en el subespacio de señal), el nuevo estadístico pasa a ser,

$$m_e = \mathbf{v}^T \cdot \mathbf{P} \cdot \mathbf{v}$$

que tiene la distribución $\chi_p^2(\mu^2 E_s)$, donde p es la dimensión del subespacio de señal. Es el denominado filtro adaptado al subespacio [SCH91][KAY98].

En relación a la última variante la referencia [GAR97] incluye un estudio exhaustivo sobre como diseñar detectores en función del grado de conocimiento asumido o disponible sobre la señal s a detectar. No obstante, en esta tesis nos centraremos en hipótesis realistas sobre nuestro posible conocimiento del fuego no controlado, para ello, como veremos, planteamos un filtro adaptado en subespacio, más otro detector (que denominaremos de crecimiento) que explote el hecho de que un fuego no controlado debe crecer en sucesivos barridos.

Otro aspecto a considerar es la posible no Gaussianidad de los datos del ruido, lo que, en general, complica bastante el diseño del detector. De hecho, primeros análisis de los datos de infrarrojos permiten comprobar que la hipótesis de Gaussianidad no es siempre correcta con las señales manejadas. En esta tesis, frente a la posibilidad de plantear esquemas de detección “ad hoc” para cada tipo de ruido, se han abordado los problemas de ruido mediante esquemas de predicción. Por una parte, la predicción permitirá reducir notablemente el nivel de ruido, dada la gran correlación que presentan los niveles de infrarrojos correspondientes al entorno bajo vigilancia. Adicionalmente, como veremos, la predicción “gaussianizará” el ruido, ajustándose por tanto a las condiciones asumidas por el filtro adaptado. Dada la posible no Gaussianidad, el predictor podrá incluir algún tipo de no linealidad que acerque las predicciones a la media condicionada: predicción óptima bajo el criterio de mínimo error cuadrático medio.

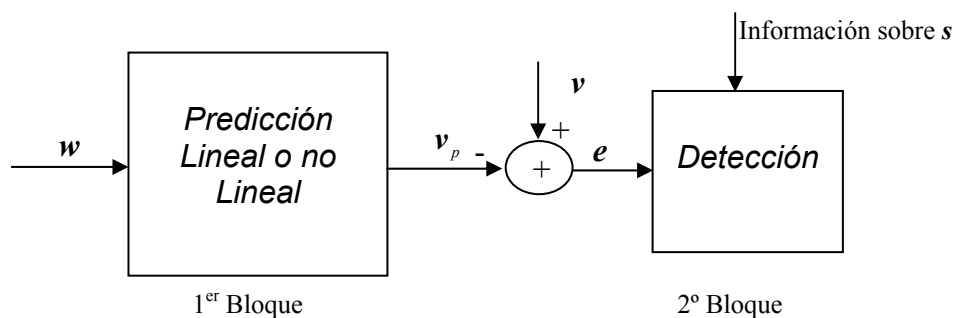


figura 3: Esquema propuesto de predicción/detección

En la figura 3 puede verse como a partir de un cierto vector w , cuyos elementos estén formados por valores correlados en mayor o menor medida con los elementos de v , por ejemplo correspondientes a las mismas celdas de v pero en barridos anteriores, y/o en acimuts próximos, y/o en rangos próximos (en general w puede tener distinta y normalmente superior dimensión a v), permita predecir v , de forma que el detector actúe sobre el vector de error e , en vez de sobre el vector original v . Si el vector e , al margen de reducir su nivel de ruido respecto al de v , es razonablemente Gaussiano, trasladamos gran parte de la complejidad del diseño del detector, al problema de diseño de un predictor en general no lineal.

El planteamiento del problema es, en principio, totalmente general: podemos ajustar las dimensiones y contenido de v y w según convenga. Así por ejemplo en la figura 4, el vector v pasa a ser un escalar (la muestra central de la retícula) y el vector w está formado por los valores de las celdas cercanas en acimut, rango y barrido, a la celda bajo análisis.

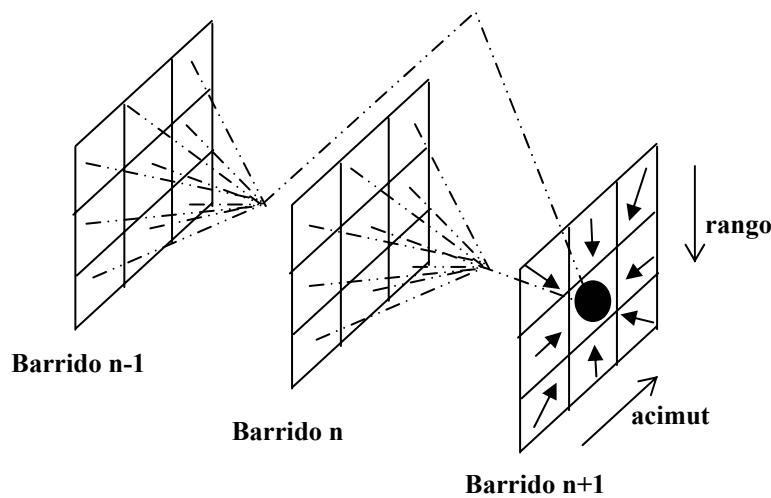


figura 4: Predicción de la celda central a partir de celdas próximas en rango, acimut y barrido

Podemos aumentar la dimensión de v involucrando un número de barridos mayor, lo que obliga a hacer predicciones a más largo plazo y condiciona asimismo el retardo que podemos permitir antes de tomar una decisión sobre existencia o no de alarma.

Igualmente en w puede no ser conveniente incluir celdas de distintos rangos, en especial en distancias alejadas, ya que el tamaño de la superficie de terreno asociado a dichas celdas varía (aumenta) considerablemente con la distancia lo que en general decorrelará las medidas asociadas.

2. Estructura de la tesis:

Tras este planteamiento general, en el siguiente Capítulo se establece una estructura concreta que incluye un predictor lineal y un filtro adaptado en subespacio. Dicha estructura se experimenta sobre un conjunto de señales de infrarrojo que serán sistemáticamente analizadas a lo largo de toda la tesis. Sobre la estructura presentada en dicho Capítulo, se plantean una serie de aspectos de mejora que serán estudiados en los Capítulos siguientes.

Así, una vez clarificado el interés de incluir no linealidades en el predictor, en el Capítulo III se hace una rápida revisión de posibles alternativas que llevan a la decisión de estudiar más detenidamente la denominada estructura de Wiener (predictor lineal seguido de una no linealidad sin memoria). Esta estructura se estudia en profundidad en el Capítulo IV, proporcionándose los algoritmos para el diseño de la no linealidad y experimentando sus prestaciones con los datos de infrarrojos anteriormente mencionados. Una conclusión importante de este Capítulo IV es la definición de criterios y procedimientos concretos para la elección de los diferentes parámetros que intervienen en el diseño tanto en la parte lineal como en la no lineal. Se incluyen asimismo un análisis de la cota de mínimo error cuadrático medio alcanzable por un predictor lineal.

En el Capítulo V, abordamos el estudio en detalle del detector, en particular en lo relativo al diseño del subespacio de señal. Asimismo se incluye en dicho Capítulo una propuesta de detector de crecimiento en la línea de lo anteriormente comentado sobre el comportamiento esperado de un fuego no controlado. Ambos detectores (el de subespacio y el de crecimiento) no sólo permiten una estimación de la señal de fuego s para contrastar su presencia frente al ruido de fondo, sino que ayudan a no generar falsas detecciones por posibles efectos ocasionales que pueden producir subidas

imprevistas de temperatura no debidas a fuegos descontrolados. Este último asunto de importancia práctica en nuestra aplicación, aparece ya con cierto detalle en el próximo Capítulo.

Capítulo II:

Esquema Básico de Predicción- Detección

Capítulo II

ESQUEMA BÁSICO DE PREDICCIÓN-DETECCIÓN

1. Introducción:

La televigilancia de grandes áreas mediante el barrido electrónico incluye un gran número de aplicaciones según diferentes tipos de energía: electromagnética, acústica, óptica, infrarroja, [IEE97] pero todas ellas tienen en común la detección automática, es decir la capacidad de resolver por sí mismo, un problema de test de hipótesis: ruido de fondo w presente, frente a señal, s , (alarma) más ruido presente.

$$H_0 : \mathbf{x} = \mathbf{w}$$

$$H_1 : \mathbf{x} = \mathbf{s} + \mathbf{w}$$

Para ser más específicos, vamos a considerar un sistema global para la televigilancia de fuegos en grandes áreas, en el que se dispondrá de un número de sensores infrarrojos localizados estratégicamente para cubrir la zona bajo vigilancia. La información captada por cada sensor será transmitida a una estación base, en la cual, se tomará la decisión de la presencia o no de fuego. Para mejorar las decisiones adoptadas, será conveniente, combinar la información proveniente de otros tipos de sensores, de los cuales se obtenga por ejemplo, información de las condiciones meteorológicas en la zona bajo vigilancia.

Con el fin de minimizar los recursos de comunicaciones entre los sensores y la estación base, parece obvio la necesidad de permitir una detección local de las alarmas, informando sólo a la estación base de las posibles alarmas detectadas.

Estamos interesados en la detección de “fuego descontrolado”, es decir, aquel que provoque un incremento continuo de la temperatura en un período de tiempo. Este tipo de fuego debería producir alarmas verdaderas. Por otro lado, nos podemos encontrar con “alarmas indeseadas”, provocadas por efectos ocasionales, como por ejemplo una

chimenea, la desaparición de una nube, un coche atravesando la zona bajo análisis, etc... Por último tenemos las “falsas alarmas”, causadas por el ruido de fondo infrarrojo debido al entorno. Nótese, por tanto la diferencia terminológica entre “alarma indeseada” y “falsa alarma”.

De acuerdo con lo anterior, estamos interesados en un sistema, que para un cierta probabilidad de falsa alarma (PFA), pueda maximizar la probabilidad de detección (PD) de fuego descontrolado, mientras minimice la probabilidad de detección de “alarmas indeseadas” (PDI).

En el planteamiento del problema, vamos a asumir que el área vigilada está dividida en diferentes celdas de resolución espacial. Las muestras de señal obtenida nos deben dar información de la temperatura en cada celda, de tal forma, que nos sea posible hacer un mapeado entre cada valor de la muestra y su celda. Para decidir automáticamente si hay una alarma en una celda debemos establecer un procedimiento. Una posibilidad simple podría ser establecer que una alarma corresponde a un muestra cuya amplitud sea mayor que un umbral predeterminado, de tal forma, que si disponemos de datos para caracterizar la distribución del ruido de fondo, el umbral puede ser ajustado para satisfacer una determinada PFA, y fijada ésta, entonces la PD obtenida (deseablemente la máxima) dependerá de la relación señal-ruido SNR .

Sin embargo, si consideramos una sola muestra, para realizar la detección, no habrá forma de distinguir fuegos descontrolados de los efectos ocasionales, esto es, la PD deberá ser parecida e incluso inferior a la PDI. Considerando que las características distintivas de un fuego real pueden incluir el incremento mantenido de la temperatura en el tiempo, nos lleva a pensar en la posibilidad de realizar la detección sobre varias medidas correspondientes a la misma celda en instantes de barridos consecutivos. Ordenamos las muestras de datos consecutivos en una forma vectorial, llevándolo así a un problema de detección vectorial (véase el Capítulo I). Para optimizar la utilidad de esta formulación vectorial, es necesario disponer de modelos para ambas señales de alarma (verdaderas y no deseadas). Para un funcionamiento robusto con aplicabilidad general, los modelos han de ser definidos en una forma general (debería ser

probablemente más apropiado hablar acerca de explotar propiedades generales que de modelos propiamente dichos). Por lo tanto, vamos a asumir simplemente que el modelo de vector de señal originado por un fuego descontrolado tiene la mayor parte de su energía dentro de un subespacio definido adecuadamente, mientras sólo un pequeño porcentaje de la energía de los vectores provocados por los efectos ocasionales se producen dentro de ese subespacio. Estas hipótesis se referirán principalmente a la separación de tipos de comportamiento paso bajo/paso alto. Dos parámetros simples permitirán la creación de un subespacio adaptado a las situaciones específicas.

2. Esquema General.

Siguiendo las ideas básicas proporcionadas en la introducción, vamos a proponer un esquema general, el cual constará de dos partes principales (figura 1): la primera corresponde a la etapa de predicción y la segunda a la etapa de detección.

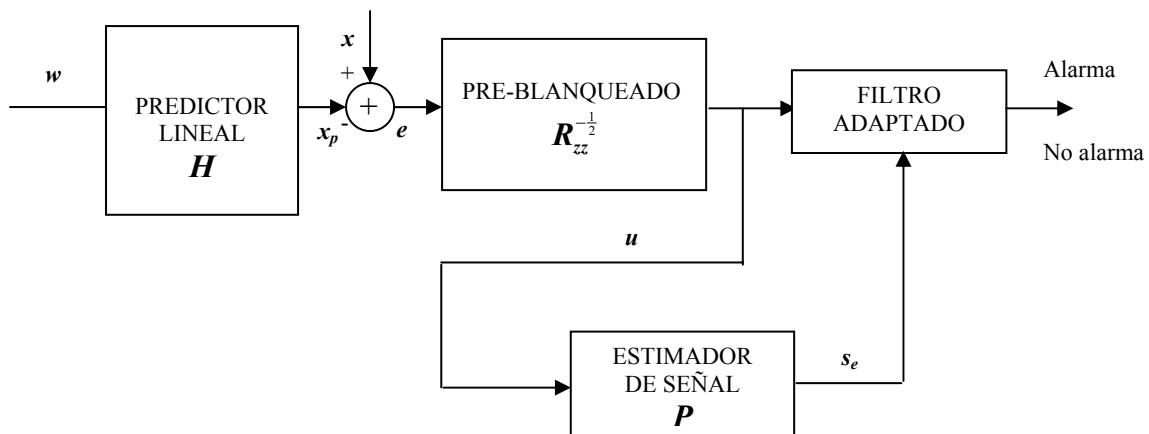


figura 1: Esquema de detección general

Asumimos que estamos realizando un barrido del área vigilada en acimut y rango. Cada celda corresponde a un par acimut-rango. Periódicamente, cada celda es iluminada por el haz del sensor. También podemos considerar el caso de vigilancia con cámara, si digitalizamos la imagen: cada píxel se corresponde con un par acimut-rango. Llamamos w al proceso estocástico de media cero compuesto por medidas de ruido de fondo correspondientes a una celda en sucesivos barridos, a los que asignamos instantes de tiempo n (discreto). En condiciones normales w debe ser un proceso fuertemente

correlado (predecible) ya que el entorno y su temperatura deben ser bastante estacionarios con el tiempo.

Las medidas serán descritas por el proceso estocástico \mathbf{x} , donde en general $\mathbf{x} = \mathbf{s} + \mathbf{w}$, siendo \mathbf{s} la señal determinista (pero desconocida) e independiente del ruido originada por la posible presencia de fuego descontrolado o efectos ocasionales.

El objetivo de la etapa de predicción es minimizar el ruido de fondo infrarrojo aprovechando la predictibilidad del mismo. Por esta razón, en esta etapa no es necesario distinguir entre fuego descontrolado y efectos ocasionales.

Refiriéndonos a la figura 1, hemos de definir el conjunto de vectores que intervienen.

El vector de ruido $\mathbf{w} = [w_n \ w_{n-1} \ \dots \ w_{n-N+1}]^T$, correspondiente a las muestras recogidas desde el $(n - N + 1)$ -ésimo barrido hasta el n -ésimo barrido en la misma celda, estamos asumiendo que no hay señal presente en esos barridos.

El vector $\mathbf{x} = [x_{n+1} \ x_{n+2} \ \dots \ x_{n+D}]^T$ de muestras tomadas desde el $(n + 1)$ -ésimo barrido al $(n + D)$ -ésimo barrido en la misma celda. Dicho vector siempre tendrá una componente de ruido, y cuando esté presente una alarma verdadera o indeseada una componente de señal, por ejemplo, si consideramos el vector $\mathbf{x} = \mathbf{s} + \mathbf{w}_D$, donde $\mathbf{s} = [s_{n+1} \ s_{n+2} \ \dots \ s_{n+D}]^T$ y $\mathbf{w}_D = [w_{n+1} \ w_{n+2} \ \dots \ w_{n+D}]^T$, estamos considerando la presencia de señal.

En principio, nosotros queremos decidir, si una alarma está presente sobre la base de la información contenida en el vector \mathbf{x} . Sin embargo, podemos intentar mejorar la *SNR* prediciendo el vector \mathbf{w}_D desde el vector \mathbf{w} , asumiendo que el ruido puede ser predicho hasta un cierto nivel. Por ello haremos una predicción \mathbf{w}_{pD} en principio lineal del vector \mathbf{w}_D , desde el vector \mathbf{w} , y se la restaremos al vector \mathbf{x} , entonces obtenemos un nuevo vector \mathbf{e} (error de predicción), el cual será la entrada al detector.

$$\mathbf{e} = \mathbf{x} - \mathbf{w}_{pD} = \mathbf{x} - \mathbf{H}\mathbf{w} = \mathbf{s} + \mathbf{z} \quad (1)$$

donde \mathbf{H} es una matriz $(D \times N)$ que implementa la transformación lineal de la predicción y \mathbf{z} es el vector de ruido residual después de la predicción.

Más adelante veremos que, si \mathbf{H} es óptima en el sentido de mínimo error cuadrático medio, la potencia media de ruido correspondiente a \mathbf{z} , será siempre menor que la potencia media original de ruido correspondiente a \mathbf{w}_D .

La segunda parte de este esquema general es el detector. Estamos interesados en alarmas que muestren (en la celda bajo análisis) una tendencia de incrementos suaves durante sucesivos barridos, asumiendo que esta conducta corresponde a fuego descontrolado. Este modelo de evolución del fuego puede ser explotado por medio de un subespacio adecuado de estimación de la señal, seguido de un detector convencional de filtro adaptado, preblanqueando el ruido residual \mathbf{z} para optimizar la detección.

3. Predicción lineal del ruido infrarrojo:

Debemos encontrar la transformación lineal óptima que minimice el error cuadrático medio $E\{\mathbf{e}^T \mathbf{e}\} = E\{\|\mathbf{e}\|^2\}$; éste es un caso particular de la ecuación de Wiener-Hopf [SCH91], con lo que:

$$\mathbf{H}_{opt} = \mathbf{R}_{xw} \cdot \mathbf{R}_{ww}^{-1} \quad (2)$$

donde $\mathbf{R}_{xw} = E\{\mathbf{x}\mathbf{w}^T\} (D \times N)$ y $\mathbf{R}_{ww} = E\{\mathbf{w}\mathbf{w}^T\} (N \times N)$. Si recordamos que estamos asumiendo que la posible señal presente es determinista y que el ruido es un proceso estacionario de media cero, resulta que, la matriz \mathbf{R}_{xw} tiene un elemento genérico (i, j) igual a

$$R_w(i+j-1) \quad 1 \leq i \leq D \quad 1 \leq j \leq N,$$

siendo $R_w(m)$ la función de autocorrelación del ruido. Por otro lado \mathbf{R}_{ww} tiene un elemento genérico:

$$R_w(i-j) \quad 1 \leq i \leq N \quad 1 \leq j \leq N.$$

Para analizar la reducción de potencia de ruido conseguida por la presencia del predictor, definiremos una figura de mérito, que denominamos “ganancia de relación señal ruido” (*GSNR*). Ésta evidentemente influirá en la *SNR* “que ve” el detector.

A partir de (1) y (2) obtenemos:

$$\mathbf{R}_{ee} = \mathbf{R}_{xx} - \mathbf{R}_{xw} \mathbf{R}_{ww}^{-1} \mathbf{R}_{xw}^T \quad (3)$$

A continuación, si tomamos el operador traza a ambos lados de la ecuación (3), y teniendo en cuenta que: $\mathbf{R}_{xx} = \mathbf{s}\mathbf{s}^T + E[\mathbf{w}_D \mathbf{w}_D^T]$, tenemos:

$$E\{\|\mathbf{e}\|^2\} = E_s + DR_w(0) - \text{Traza}(\mathbf{R}_{xw} \mathbf{R}_{ww}^{-1} \mathbf{R}_{xw}^T) = E_s + E\{\|\mathbf{z}\|^2\} \quad (4)$$

donde $E_s = \mathbf{s}^T \mathbf{s}$, es la energía de la señal. También además sabemos que:

$$E\{\|\mathbf{x}\|^2\} = E_s + DR_w(0) = E_s + E\{\|\mathbf{w}_D\|^2\} \quad (5)$$

A partir de lo anterior, definimos la *SNR* para el vector \mathbf{x} y para el vector \mathbf{e} .

$$\begin{aligned} SNR_x &= \frac{E_s}{E\{\|\mathbf{w}_D\|^2\}} = \frac{E_s}{E\{\|\mathbf{x}\|^2\} - E_s} \\ SNR_e &= \frac{E_s}{E\{\|\mathbf{z}\|^2\}} = \frac{E_s}{E\{\|\mathbf{e}\|^2\} - E_s} \end{aligned} \quad (6)$$

entonces la *GSNR* debida al hecho de realizar la predicción será.

$$GSNR = \frac{SNR_e}{SNR_x} = \frac{DR_w(0)}{DR_w(0) - \text{Traza}(\mathbf{R}_{xw} \mathbf{R}_{ww}^{-1} \mathbf{R}_{xw}^T)} \quad (7)$$

donde hemos hecho uso de las ecuaciones (4)-(6).

Denotar que $\text{Traza}(\mathbf{R}_{xw} \mathbf{R}_{ww}^{-1} \mathbf{R}_{xw}^T) > 0$ debido a la propiedad de definición positiva de \mathbf{R}_{ww}^{-1} , por eso, y en cualquier caso, $G\text{SNR} > 1$. En próximos apartados ilustraremos la influencia de $G\text{SNR}$ en la SNR a la entrada del detector.

4. El detector:

Una vez la SNR ha sido mejorada por la predicción del ruido, es necesario realizar una detección automática de la posible presencia de una alarma en el vector \mathbf{e} .

Si la señal (firma) debida a un fuego fuese conocida “a priori” podríamos usar un filtro adaptado precedido por un preblanqueado del ruido (ver figura 1, de nuevo). Este detector será óptimo en el caso Gaussiano. Sin embargo, el conocimiento de la firma del fuego no es una suposición práctica en nuestro caso, puesto que distintos fuegos firmarán de forma distinta, por ello tenemos que realizar una detección subóptima.

La peor situación posible es que no tengamos ningún conocimiento “a priori” de la señal a detectar. Bajo estas circunstancias, un simple detector de energía basado en la comparación de $\mathbf{u}^T \mathbf{u}$ (donde \mathbf{u} es la salida de la transformación de preblanqueado $\mathbf{R}_{zz}^{-1/2}$ aplicada a \mathbf{e}), véase la figura 1, con un umbral adecuado, es la solución más útil. Aparte de ser subóptimo, el detector de energía no permite ninguna distinción entre señales de alarmas verdaderas e indeseadas.

En alguna parte entre estos dos extremos, podemos concebir detectores que puedan tomar ventaja de algún conocimiento “a priori”, sobre el comportamiento esperado, de un fuego descontrolado. Es razonable considerar que un fuego descontrolado debería mostrar un incremento de temperatura en la celda correspondiente durante sucesivos barridos. Suponemos, además, que el tiempo entre barridos y las dimensiones de la celda son suficientes para que el comportamiento de los incrementos sea “suave” cuando lo comparamos con las variaciones entre barridos de las alarmas indeseadas.

Asumiremos que el comportamiento suave del fuego descontrolado equivale a considerar que su firma (después del preblanqueado) puede estar incluida en un

subespacio (paso bajo) de dimensión $p < D$. Mientras que las firmas de las alarmas indeseadas se extiende al espacio completo de dimensión D (al igual que lo hacen las firmas del ruido de fondo infrarrojo). Entonces, una estimación de la señal [GAR94], puede obtenerse por medio de la proyección del vector \mathbf{u} en el correspondiente subespacio de señal. Esta estimación puede ser utilizada como la señal necesaria en la implementación de un filtro convencional adaptado (ver figura 1). Es el denominado filtro adaptado al subespacio [SCH91].

Si llamamos \mathbf{P} a la matriz de proyección, el estadístico del detector a comparar con un umbral será.

$$\mathbf{r} = \mathbf{u}^T \mathbf{P} \mathbf{u}$$

Asumimos que \mathbf{P} es capaz de reducir significativamente la energía de una señal debida a efectos ocasionales, reduciendo así la PDI.

Nos queda una cuestión importante, relativa al diseño de la matriz \mathbf{P} . Para ello, tenemos que tener en cuenta las diferencias antes mencionadas entre los modelos de alarmas verdaderas e indeseadas. Así, nosotros proponemos el uso de las secuencias dominantes de Slepian [MUL91][THO82], las cuales son las respuestas impulsionales de filtros FIR paso-bajo que tienen propiedades selectivas óptimas. Incluimos una breve explicación al respecto.

Sea el vector \mathbf{h} (dimensión D) la respuesta impulsional del filtro FIR y $H(\omega)$ la correspondiente respuesta en frecuencia. Nosotros queremos $H(\omega)$ para concentrar la máxima cantidad de energía en el intervalo $-\pi\beta < \omega < \pi\beta$ bajo la restricción $\mathbf{h}^T \mathbf{h} = 1$. Las soluciones a este problema son los autovectores dominantes de la matriz $\mathbf{M}(D \times D)$, definida por

$$\mathbf{M}(m, n) = \beta \operatorname{sinc}[\beta\pi(m - n)] \tag{9}$$

Estos autovectores son las secuencias dominantes de Slepian. El primer autovector dominante es aquel que posee el correspondiente máximo autovalor y que concentra el

máximo de energía en el intervalo $-\pi\beta < \omega < \pi\beta$. Los autovectores dominantes sucesivos concentrarán cada vez menos energía en el mismo intervalo. Nosotros podemos elegir los primeros p autovectores dominantes para definir a un subespacio paso-bajo, donde el vector \mathbf{u} será proyectado. La matriz de proyección \mathbf{P} vendrá dada por,

$$\mathbf{P} = \mathbf{A}\mathbf{A}^T \quad (10)$$

donde \mathbf{A} es una matriz $D \times p$ cuyas columnas son los p autovectores dominantes de \mathbf{M} .

Debe apreciarse que la capacidad de rechazar alarmas indeseadas, o aceptar alarmas verdaderas, dependerá solamente de la cantidad de energía del correspondiente vector de señal incluido en el subespacio considerado. Indicar que por medio de β y p nosotros podemos seleccionar el subespacio más adecuado para un escenario dado.

5. Diseño del sistema:

Consideramos interesante incluir este apartado para centrar en lo posible el enfoque aplicado de los algoritmos propuestos. Los Capítulos IV y V de esta tesis volverán sobre los aspectos prácticos relativos a criterios de diseño para fijar los diferentes parámetros del esquema propuesto, incluyendo en el mismo las diferentes mejoras que se estudian a lo largo de la tesis.

Calibración: Para empezar, hemos de realizar una estimación de la función de autocorrelación del ruido presente $R_w(m)$. En general el escenario no será estacionario: diferentes zonas y diferentes barridos temporales requerirán una estimación de la autocorrelación diferente. Por ello, proponemos dividir el escenario en ventanas bidimensionales (rango y acimut) de tal forma que la estacionaridad pueda ser asumida dentro de cada ventana, por ejemplo, la correlación del ruido barrido a barrido es la misma para todas las celdas dentro de una ventana dada. Después de un cierto intervalo de tiempo, podemos actualizar las autocorrelaciones estimadas para considerar las variaciones en las condiciones generales del escenario (día o noche, variaciones climáticas, etc.)

Predicción: Tenemos que fijar D y N . Hemos de recordar que la decisión de alarma/no alarma se hace a partir del vector \mathbf{x} , el cual está constituido por las muestras recogidas durante D barridos sucesivos. En principio, deberíamos intentar usar el valor más alto posible de D , con el objetivo de optimizar la explotación de la información sobre los modelos implicados. Sin embargo, tenemos un límite superior sobre D , debido al máximo retardo permitido cuando una situación de alarma es registrada. Por otro lado, N depende de la función de autocorrelación: no tiene sentido hacer predicciones a partir de valores que no estén correlados con los valores a predecir ($N+D$ no debería superar la longitud de la autocorrelación). Una vez fijados D y N , diseñaremos el predictor óptimo \mathbf{H} por medio de la ecuación (2).

Detector: Tenemos que fijar β y p para definir la matriz \mathbf{P} , y debemos seleccionar el umbral T , con el que compararemos el estadístico r . Los parámetros β y p dependerán del tipo de alarmas indeseadas que tenemos que evitar. Preferiblemente, ambos parámetros deberían ser tan pequeños como fuera posible, con el objetivo de reducir severamente la energía producida por los efectos ocasionales dentro del subespacio delimitado por \mathbf{P} , pero tenemos que evitar la cancelación de señales producidas por fuegos descontrolados. El valor de T será fijado para obtener una determinada PFA. Esto es sencillo si asumimos que el ruido en el vector \mathbf{u} es Gaussiano multivariable, porque bajo estas circunstancias, el estadístico r tiene una distribución conocida: chi-cuadrado con p grados de libertad [SCH91][KAY98].

A continuación vamos a considerar brevemente las probabilidades de detección PDI y PD. Naturalmente, dependerán de la SNR . La SNR correspondiente a r es el parámetro no central de la distribución chi-cuadrado [SCH91][KAY98], que viene dada por la ecuación (11).

$$SNR = \mathbf{s}^T \mathbf{R}_{zz}^{-1/2} \mathbf{P} \mathbf{R}_{zz}^{-1/2} \mathbf{s} \quad (11)$$

donde el vector \mathbf{s} corresponde a fuego descontrolado o a efectos ocasionales. Vemos que, para una matriz dada \mathbf{P} , la SNR depende de cada \mathbf{s} y \mathbf{R}_{zz} particulares. Sin embargo, la expresión (11) indica claramente que la SNR equivale a la energía de la señal

(después del preblanqueado) que existe en el interior del subespacio delimitado por \mathbf{P} . Entonces \mathbf{P} debería seleccionarse adecuadamente para que los modelos de señal producidos por fuegos descontrolados quedasen dentro de ese subespacio. O dicho de otra forma, un efecto ocasional debería producir señales donde la mayor parte de su energía quedase fuera del subespacio delimitado por \mathbf{P} . Esto es conveniente para separar el comportamiento paso bajo esperado de los modelos de alarmas verdaderas de los comportamientos de banda ancha esperados en las alarmas indeseadas.

Ahora volvamos a la utilidad de la predicción para mejorar la SNR . Asumiendo que el vector de señal (después del preblanqueado) está dentro del subespacio \mathbf{P} , entonces:

$$SNR = \mathbf{s}^T \mathbf{R}_{zz}^{-1} \mathbf{s} \tag{12}$$

Consideramos la situación más favorable en la cual la señal es enteramente paso bajo, una componente continua,

$$\mathbf{s} = (1/\sqrt{D})[1 \ 1 \ 1 \dots \dots 1]^T \text{ (hemos normalizado } \mathbf{s} \text{ para tener } E_s = 1 \text{, sin pérdida de generalidad)}$$

y el ruido residual es enteramente blanco ($\mathbf{R}_{zz} = R_{zz}(0)\mathbf{I}$). Entonces la ecuación (12) queda:

$$SNR = \frac{1}{R_{zz}(0)} = \frac{D}{E\{\|\mathbf{z}\|^2\}} = \frac{D}{DR_w(0) - \text{Traza}(\mathbf{R}_{xw} \mathbf{R}_{ww}^{-1} \mathbf{R}_{xw}^T)} \tag{13}$$

donde hemos considerado la ecuación (4). Ahora , comparando la ecuación (13) con la ecuación (7) y teniendo en cuenta que, para una señal normalizada en energía, tenemos

que, $SNR_x = \frac{1}{DR_w(0)}$, obtenemos la expresión:

$$SNR = \frac{GSNR}{R_w(0)} = GSNR \cdot SNR_x \cdot D \tag{14}$$

la ecuación (14) nos da una clara indicación de la importancia de incluir un paso de predicción para mejorar la relación señal ruido original SNR_x , así como de la proporcionalidad entre SNR y el valor de D .

Por otro lado, teniendo en cuenta que la hipótesis de banda ancha del ruido residual es una hipótesis utilizada para deducir la ecuación (14), consideramos también de interés medir cómo el predictor es capaz de decorrelar el ruido. En otras palabras, cuánto la matriz R_{zz} está más cerca de la matriz identidad que $R_{w_D w_D}$. Una posibilidad obvia para determinar esta proximidad es comparar las distancias de las matrices $R_{w_D w_D}$ y R_{zz} con la matriz identidad.

Consideremos una medida estándar de matrices, a saber la *norma de Frobenius* [GOL89], la cual es la suma de los cuadrados de todos los elementos de la correspondiente matriz, e igual a la traza del producto ortogonal de la matriz. Nosotros podemos comparar la *norma de Frobenius* de las diferencias de la matrices (adecuadamente normalizadas) $R_{w_D w_D}$ y R_{zz} con la matriz I .

Para ello proponemos el parámetro “ganancia de decorrelación”(GDECOR) que lo definimos como:

$$GDECOR = \frac{\text{Traza}((R_{w_D w_D} - I)^T (R_{w_D w_D} - I))}{\text{Traza}((R_{zz} - I)^T (R_{zz} - I))} \quad (15)$$

donde la normalización de la matriz está definida de la forma,

$$R_{vvn} = A^{-1/2} R_{vv} A^{-1/2} \quad (16)$$

siendo A una matriz diagonal con la misma diagonal principal que R_{vv} . Nótese que $R_{w_D w_D n} = (1/R_{ww}(0))R_{w_D w_D}$, es la potencia media normalizada de la matriz de autocorrelación. Este no es el caso de R_{zz} , porque en general el error de predicción del ruido se incrementará con el intervalo de predicción, y la matriz R_{zz} tendrá valores diferentes (incrementándose) a lo largo de su diagonal principal.

Los parámetros $GSNR$ y $GDECOR$ no son parámetros de diseño. Sin embargo, podrían ayudar en la determinación del orden N del predictor: Incrementamos N hasta no obtener mejoras significativas en $GSNR$ y $GDECOR$. En cualquier caso, suministran indicaciones que consideran el interés de incluir el paso de predicción.

$GSNR$ y $GDECOR$ serán evaluadas en esta aplicación en los experimentos de la sección 6, con datos reales, cuando consideremos las estimaciones de la autocorrelación del ruido infrarrojo de fondo.

6. Experimentos:

El esquema de la figura 1 ha sido aplicado a datos infrarrojos reales recogidos por un radar pasivo infrarrojo, localizado en una zona montañosa de Alcoi (Alicante). El tipo de sensor usado es el Thermoprofile 6 HT, (Sistemas Infrarrojos AGEMA). Dicho sensor es usado con barrido electrónico en rango y mecánico en acimut. Los datos grabados dependen de tres variables: tiempo (o número de barrido), acimut y rango (o número de muestra).

Para este experimento hemos seleccionado tres bloques, la selección de dichos bloques no ha sido arbitraria.

El primero de ellos **bloque A** (figura 2a) presenta una elevada correlación en el dominio de predicción (barrido n) lo que supondrá una elevada ganancia en SNR al aplicar el predictor. Sin embargo su estadística es poco Gaussiana, siendo en un principio un bloque candidato a la inclusión de no linealidades en la mejora de la predicción.

Por el contrario, el segundo **bloque B** (figura 2b), presenta una baja correlación en n pero una función densidad de probabilidades bastante Gaussiana.

Consideraremos asimismo un bloque mayor **bloque C**, que incluye al **bloque A** y parte del **bloque B**, (figura 2c) ya que la definición de lo que es un bloque estacionario no deja de ser algo subjetivo y por tanto interesa analizar la robustez del sistema cuando se consideran bloques más grandes, pero claramente no homogéneos. Debe notarse que

cuanto más pequeños son los bloques de calibración, mayor será la complejidad del sistema.

Estos tres bloques por separado, serán usados de forma sistemática a lo largo de la tesis.

Cada uno de los bloques, está formado por 84 registros, los cuales aparecen superpuestos en las figuras 2a, 2b y 2c, respectivamente. En un registro dado, cada muestra corresponde a una celda particular, por eso, en las figuras 2a, 2b y 2c, cada registro individual nos muestra las variaciones de amplitud cuando nos movemos en rango, para los correspondientes acimuts y barridos. El valor de 84, 21×4 registros, corresponde a 21 barridos consecutivos (el período de barrido es 1 minuto, desde las 17:00 a las 17:20), tomados en cuatro acimuts consecutivos (empezando a 7.3° con respecto a la normal al sensor y avanzando en saltos de 0.23°).

Cada registro de los **bloques A** y **B** tiene 51 muestras, mientras que el **bloque C** tiene 100 muestras, cada muestra corresponde a un rango de distancia.

El primer bloque empieza en la muestra 250, que corresponde a aproximadamente 10 Km. de distancia y finaliza en la muestra 300, la cual corresponde a 5 Km. de distancia. De manera similar en el segundo bloque los rangos van desde 4.5 Km. a 800 metros de la posición del sensor. El **bloque C**, corresponde a 100 rangos consecutivos tomados a partir de la muestra 250 y finalizando en la muestra 349, de tal forma que incluye al **bloques A** y parte del **B**

La técnica propuesta ha sido aplicada a cada uno de los bloques, asumiendo un comportamiento estacionario dentro de cada uno de ellos, (la presencia de no estacionaridad en los datos, en general procede de los diferentes segmentos de muestras en rango, debido al incremento del tamaño de las celdas cuando nos vamos alejando del sensor).

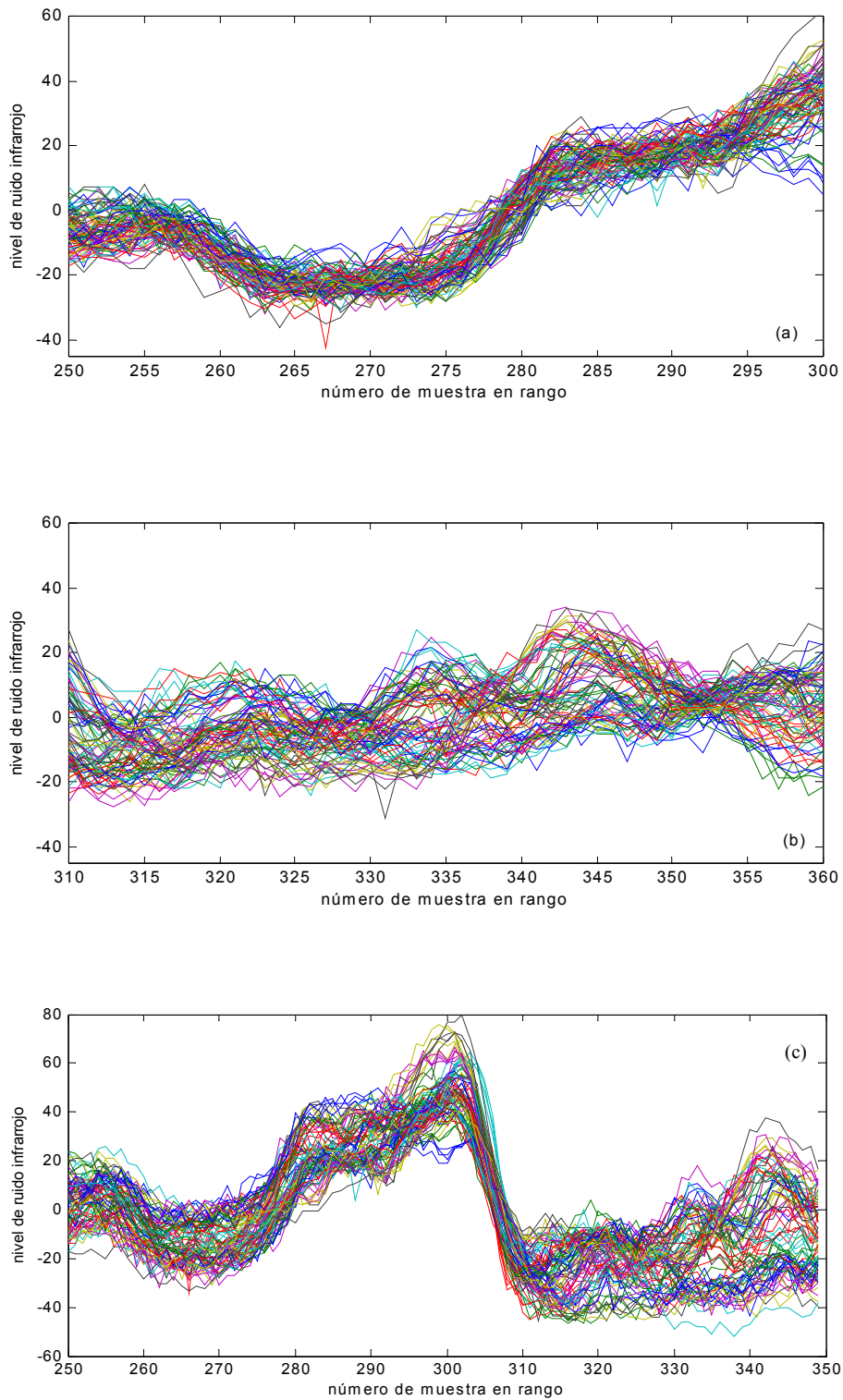


figura 2: Bloques de datos infrarrojos: (a) Bloque A, (b) Bloque B, (c) Bloque C

Hemos elegido en estos primeros experimentos los valores de $N = 2$ y $D = 4$ en cada celda, permitiéndonos predecir los valores correspondientes a los instantes $n+1$ y $n+4$, a partir de los valores medidos en los instantes de tiempo n y $n-1$.

Las estimaciones de la autocorrelación requeridas han sido calculadas, para el diseño del predictor óptimo H (4×2) para cada uno de los bloques. Las figuras 3a, 4a y 5a nos muestran la autocorrelación estimada normalizada (línea sólida), respectivamente para cada bloque. Estas estimaciones corresponden a la autocorrelación en tiempo o en la dimensión del número de barrido, pero no en la dimensión de rango, por eso, la mayor autocorrelación estimada en el primer bloque está de acuerdo con la menor dispersión o varianza que se observa en la figura 2a cuando la comparamos con la figura 2b. Además se observa como la autocorrelación estimada del bloque conjunto, figura 5a, es intermedia a la de los “**bloques A y B**”, debido a que presenta mayor dispersión que el bloque de la figura 2a, pero menor que el bloque de la figura 2b. Para ilustrar los efectos de la decorrelación del predictor, mostramos también la secuencia de autocorrelación (línea de puntos) en las figuras 3a, 4a y 5a correspondientes a z , el vector de ruido residual tras el predictor. Lo que estamos representando aquí es la primera fila de la matriz de autocorrelación R_{zz} . Los efectos de la decorrelación conseguidos se manifiestan en el valor mayor que 1 obtenido por *GDECOR*, indicado en las mismas figuras.

En las figuras 3b, 4b y 5b, se muestran los histogramas de los datos originales en línea sólida, junto con los histogramas de los datos del error de predicción (ruido residual) en línea de puntos, obtenidos por medio de aplicar a los datos originales el predictor óptimo diseñado con sólo presencia de ruido.

Observamos el comportamiento no Gaussiano del primer bloque de datos originales, esto nos sugiere que el uso de un predictor no lineal mejoraría la *GSNR*. Sin embargo, el histograma de los datos de error de predicción es claramente Gaussiano en todos los casos, por tanto el detector diseñado se beneficiará de la usual hipótesis de Gaussianidad. En particular, podremos fijar fácilmente una PFA deseada. Los valores calculados para *GSNR*, aparecen también indicados en la figura. Como esperábamos

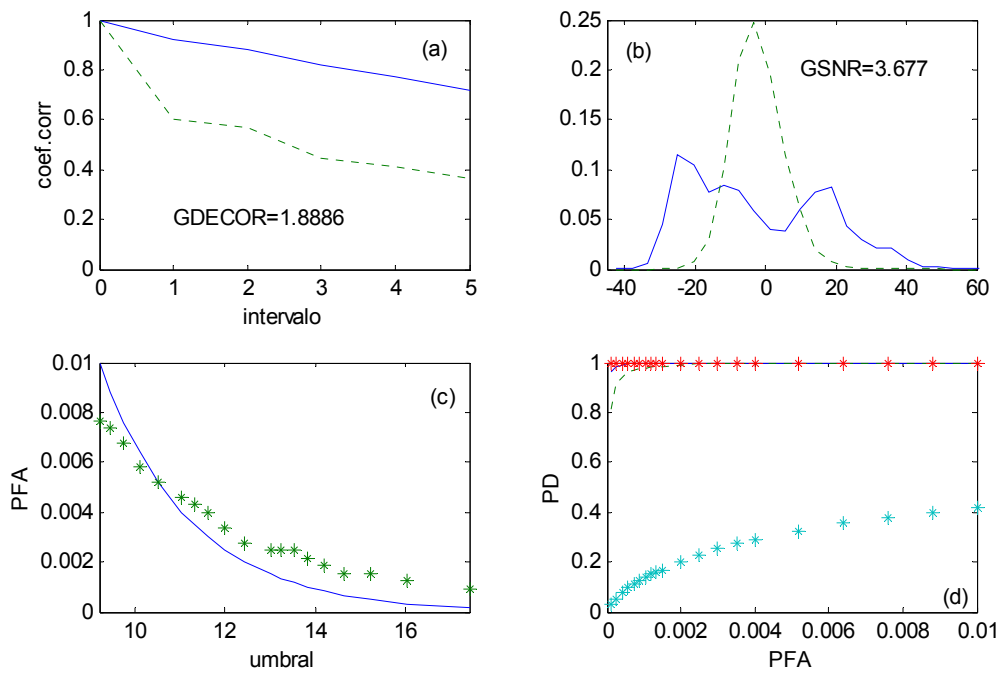
hemos obtenido una mayor $GSNR$ en el primer bloque debido a la claramente mayor correlación de ruido de los datos originales en la dimensión número de barrido.

El resto de las figuras están relacionadas con el detector. Primeramente, las figuras 3c, 4c y 5c nos muestran la PFA teórica (línea sólida), asumiendo la Gaussianidad del error de predicción y la comparan con la PFA estimada. Ésta ha sido medida posteriormente usando todos los posibles vectores \mathbf{z} de cuatro elementos permitidos en cada bloque después de predecir los datos originales (3264 vectores, en los bloques primero y segundo y 6400 para el bloque conjunto), para los valores seleccionados de $N = 2$, $D = 4$. Estas figuras nos dan un idea de la capacidad para fijar una PFA deseada. La matriz \mathbf{P} (4x4), será calculada a partir de (10), donde \mathbf{A} está constituida por los dos autovectores dominantes de \mathbf{M} (4x4) para $\beta = 0.1$.

Finalmente, las figuras 3d, 4d y 5d son las más importante para mostrar el interés de la técnica propuesta. En ellas, representamos la PD y PDI estimadas frente a la PFA en dos diferentes situaciones: la primera, la detección de un posible fuego, éste ha sido simulado sumando una tendencia lineal en los cuatro elementos de los vectores \mathbf{z} disponibles. El valor inicial de la tendencia lineal es igual a la desviación estándar del ruido, y el valor final (en este caso cuatro minutos más tarde) es cuatro veces el valor inicial. En las gráficas, la línea continua corresponde al uso del filtro adaptado al subespacio y la línea de puntos al detector de energía, para la figura 3d, en ambos casos es siempre igual a 1. La mejora conseguida debido a la inclusión de la matriz \mathbf{P} , que es muy significativa en el segundo bloque (figura 4d), puede claramente notarse. Por otro lado, la curva de asteriscos se corresponde a la PDI estimada de alarmas indeseadas, las cuales las hemos simulado añadiendo un valor, solamente al tercer elemento de cada uno de los vectores \mathbf{z} . Para hacer una correcta comparación, el valor sumado es igual a la raíz cuadrada de la tendencia lineal para la simulación de un fuego creciente. Observamos una curva de asteriscos en la parte superior, que corresponde a un detector de energía, la PDI estimada es igual a 1 en todo los casos, esto significa que el estadístico r estará siempre por encima del correspondiente umbral, cuando los efectos ocasionales simulados estaban sumados a los 3264 vectores disponibles, de los bloques

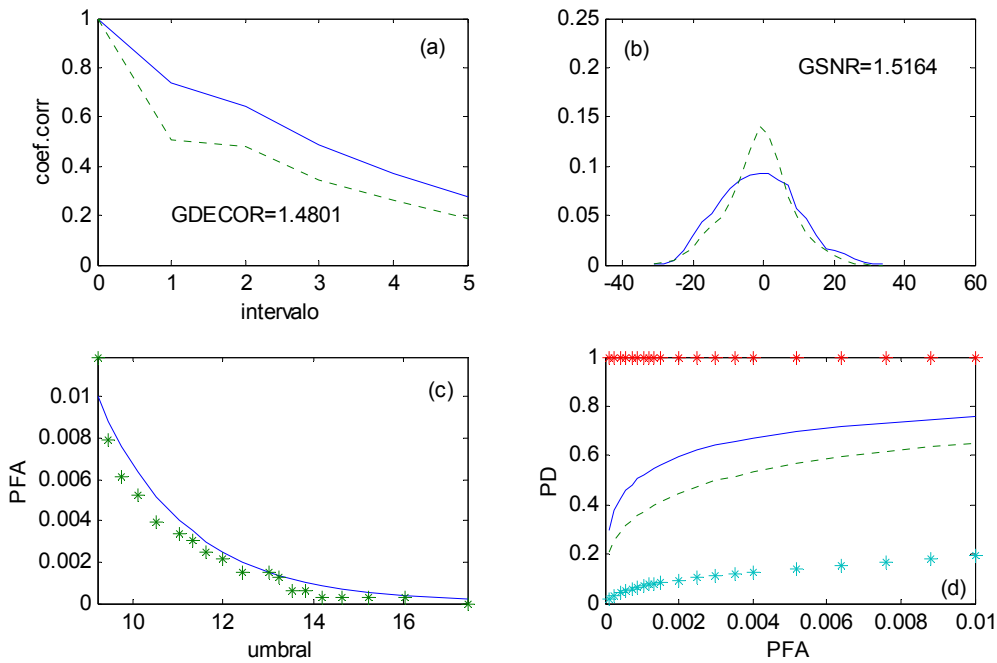
primero y segundo, y a los 6400 del bloque conjunto. Sin embargo, al incluir P (curva de asteriscos inferior), la PDI queda drásticamente reducida.

Como conclusión, podemos añadir, como en los tres bloques seleccionados, hemos conseguido con la presencia del detector adaptado al subespacio, dos efectos: en primer lugar mejorar la PD de fuego descontrolado con respecto al detector de energía, observable en las figuras 4d y 5d; y en segundo lugar la drástica reducción de los efectos ocasionales.



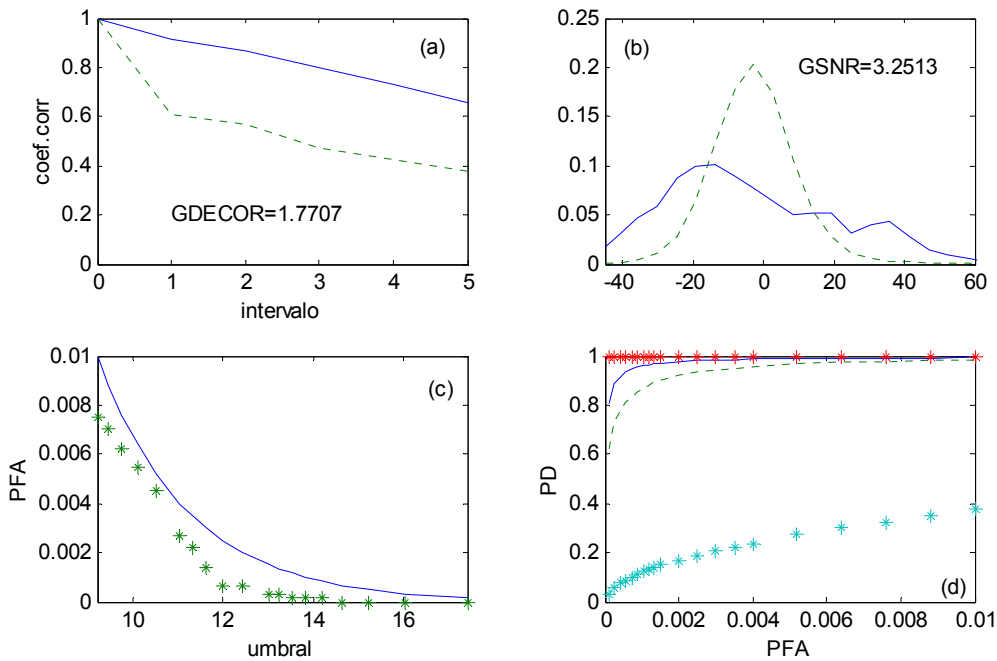
Bloque A

figura 3: (a) Autocorrelación estimada de los datos originales (línea sólida) y del error de predicción (línea puntos)
 (b) Histograma de los datos originales (línea sólida) y del error de predicción (línea de puntos)
 (c) PFA teórica (línea sólida) y PFA estimada (línea de asteriscos).
 (d) PD frente PFA para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector energía (línea puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior).



Bloque B

figura 4: (a) Autocorrelación estimada de los datos originales (línea sólida) y del error de predicción (línea puntos)
 (b) Histograma de los datos originales (línea sólida) y del error de predicción (línea de puntos)
 (c) PFA teórica (línea sólida) y PFA estimada (línea de asteriscos)
 (d) PD frente PFA para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector energía (línea puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior).



Bloque C

figura 5: (a) Autocorrelación estimada de los datos originales (línea sólida) y del error de predicción (línea puntos)
 (b) Histograma de los datos originales (línea sólida) y del error de predicción (línea de puntos)
 (c) PFA teórica (línea sólida) y PFA estimada (línea de asteriscos)
 (d) PD frente PFA para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector energía (línea puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior).

7. Mejoras sobre el esquema planteado.

Básicamente, pueden considerarse dos conjuntos de mejoras. El primer conjunto relativo al predictor y el segundo al detector.

En relación con el predictor y dada la (en general) no Gaussianidad de los datos, puede plantearse el diseño de predictores no-lineales [PRI88]. Este es el aspecto que se desarrolla en el Capítulo IV de esta tesis, tras una revisión de técnicas de predicción no lineal en el Capítulo III, cuyo enfoque como veremos, no pierde nunca de vista la importancia de proponer esquemas suficientemente simples y realizables en la práctica, pero que puedan mejorar sustancialmente la *SNR* a la entrada del detector, así como “Gaussianizar” el ruido de forma que el filtro adaptado sea óptimo.

En relación con el detector, las mejoras obvias van en la línea de incluir cuanta información sea posible sobre la señal de fuego descontrolado. Ciertamente, mayor información suele ir unida a menor generalidad: restringimos más el conjunto de señales detectables. No obstante resulta inmediato incluir lo que denominamos un “detector de crecimiento” que tenga en cuenta el carácter creciente en tiempo de un fuego descontrolado. Todo ello se desarrolla en el Capítulo V de la tesis.

Por otra parte, tanto en el Capítulo IV (predictor) como en el V(detector), se profundiza en los criterios de ajuste de los diferentes elementos del sistema: longitud del predictor (N), inclusión o no de no linealidad, número de barridos en los que basar la predicción (D), dimensión del subespacio de señal (p) y frecuencia de corte asociada a dicho subespacio (β).

Capítulo III:

Técnicas de Predicción no Lineal

Capítulo III**TÉCNICAS DE PREDICCIÓN NO LINEAL****1. Objetivos:**

Presentamos en este Capítulo una revisión de la problemática de la predicción no lineal, así como de las familias de métodos existentes para realizar predictores no lineales. Pretendemos con ello proporcionar cierto nivel de autocontenido a la tesis y presentar una perspectiva de la predicción no lineal que permita al lector interesado explorar otras vías de las aquí elegidas. Porque, finalmente, vamos a optar por una estructura simple, la denominada estructura o sistema de Wiener, tal como indicamos en el último apartado de este Capítulo. En el siguiente se presentarán dos técnicas de diseño originales, de dicha estructura, y se aplicarán exhaustivamente a los datos reales de infrarrojo anteriormente presentados. No obstante consideramos interesante incluir aquí esta revisión fruto del esfuerzo de búsqueda y clasificación de técnicas que ha permitido llegar a la selección de una opción concreta, realista para la aplicación bajo estudio.

2. Introducción:

El procesado de señales o procesos estocásticos abarca un área donde la física y la matemática se encuentran e interactúan para resolver un amplio rango de problemas. En particular los métodos clásicos para el procesado de procesos estocásticos se fundamentaban en tres suposiciones: linealidad, estacionaridad y gaussianidad.

Sin embargo, la mayoría de las señales físicas que forman parte de las aplicaciones reales son generadas por procesos dinámicos que son simultáneamente, no lineales, no estacionarios y no gaussianos.

Muchos sistemas no lineales pueden ser descritos satisfactoriamente usando modelos que constan de combinaciones lineales junto con combinaciones no lineales.

Estos modelos nos permiten caracterizar numerosas estructuras no lineales como por ejemplo: saturación de sensores, control de válvulas, ciertos sistemas biológicos, limitadores paso banda en receptores de radio, etc.

El problema principal en la identificación de sistemas es encontrar una buena estructura que nos permita una adecuada modelización de la misma [SUY98][LJU87].

Si vamos evolucionando desde estructuras de identificación lineales a no lineales, el problema de identificación aumenta como consecuencia del mayor tipo de modelos no lineales frente a modelos lineales.

El problema de seleccionar el modelo general puede describirse como un compromiso entre la contribución en varianza y en sesgo. Uno quiere ofrecer mucha flexibilidad (lo que aumenta la varianza) para que el modelo pueda adaptarse prácticamente a cualquier función no lineal, mientras al mismo tiempo impedir una sobreadaptación, por lo que uno no quiere usar modelos excesivamente flexibles (lo que aumenta el sesgo). El modo tradicional para hacerlo es experimentar y testear muchos modelos de estructuras diferentes ofreciendo flexibilidad en diferentes formas y cantidades. Al comparar los modelos uno entonces elige aquél en el cual obtiene el mejor requerimiento, el compromiso sesgo-varianza.

3. Conceptos generales sobre predicción:

El objetivo de las técnicas de predicción se centra en la capacidad de realizar predicciones futuras de estados de un sistema, a partir de cualquier tipo de conocimiento previo sobre la dinámica del mismo.

El hecho de predecir los valores futuros de las muestras de una señal, a partir del conocimiento de sus valores anteriores nos puede permitir:

- Tomar decisiones anticipadas: Predicción bursátil, predicción de fenómenos naturales, seguimiento de móviles.

- Supresión de ruido/interferencias: Predecir los valores del ruido o interferencias y restarlos de los valores medidos; con lo que podemos mejorar la SNR en detección radar o suprimir la interferencia entre símbolos en transmisión digital (igualadores), (este es el interés en nuestra tesis).
- Codificación/compresión: Codificar sólo el error de predicción y/o los parámetros de un modelo predictivo (codificación de voz, audio, imagen).
- Clasificar las señales según lo predecibles que sean.

En primer lugar deberíamos plantearnos “¿Por qué es posible predecir señales?”. Para responder a esta cuestión analizaremos el comportamiento de las señales deterministas y las señales o procesos estocásticos, sobre estas últimas es sobre las que adquiere una mayor significación, la capacidad para determinar en mayor medida estructuras capaces de predecir estados futuros de las mismas.

a) Señales deterministas.

Siguen una evolución definida por una cierta ecuación, que nos permite conocer el valor de sus muestras para cualquier instante n , a partir de valores de las muestras en instantes anteriores.

$$x(n) = f[x(n-1), x(n-2), \dots]$$

por lo que serán perfectamente predecibles si conocemos la ecuación que define su dinámica y además si la evolución es muy “evidente” podemos tratar de estimar dicha ecuación a partir de las medidas.

Si estudiamos su dinámica mediante la evolución de vectores formados por N muestras consecutivas comprobaremos que dichos vectores se mueven en un subespacio “atractor”. Si la dimensión de dicho espacio es fraccionaria la señal es caótica: su evolución es muy sensible a las condiciones iniciales y la ecuación de evolución es poco evidente, de todo ello, surge un área de investigación que trata de

caracterizar cómo se comportan los predictores con señales caóticas [CAS89] [SUG90][GUE97][IEE98].

b) Señales (procesos) estocásticos.

Si en un proceso estocástico la probabilidad de determinar el valor en un instante dado está condicionada por los valores anteriores, tendremos un cierto grado de predictibilidad,

$$P[x(n) / x(n-1), x(n-2), \dots] \neq P[x(n)]$$

por lo tanto podemos predecir el valor futuro si somos capaces de medir dicho grado de predictibilidad.

Por ejemplo, una forma sencilla de medir el grado de predictibilidad lineal es mediante la autocorrelación (o su transformada de Fourier, el espectro).

A continuación revisaremos diferentes esquemas de predicción tanto lineales como no lineales.

4. Esquemas de predicción:

4.1. Lineales:

La posibilidad más obvia es la de predecir el valor futuro con un valor igual al más reciente disponible, o realizar la predicción promediando ponderadamente los últimos valores anteriores disponibles, este planteamiento nos lleva de forma natural a los **predictores lineales**, en que el promedio se diseña tratando de alcanzar una cierta *optimización* en la predicción. Bajo esta perspectiva establecemos dos esquemas de predictores lineales.

a) Predictores lineales de duración finita.

El esquema de los predictores lineales de duración finita, se corresponde con el representado en la figura 1, donde podemos observar que la predicción de la

muestra futura $\hat{x}(n)$, se obtiene mediante la combinación lineal de los instantes anteriores multiplicados por unos coeficientes de ponderación a_k , valores que hacen referencia al peso de la contribución en la predicción de la muestra $x(n-k)$.

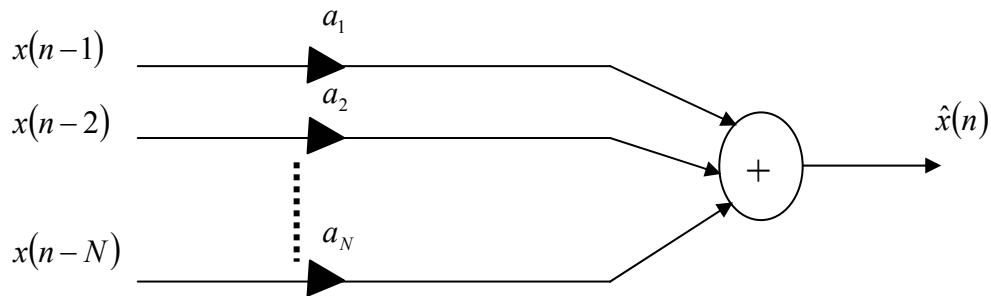


figura 1: Esquema de predictor lineal de duración finita

Los coeficientes a_k se diseñan tratando de minimizar (según cierto criterio) el error de predicción, $e(n) = x(n) - \hat{x}(n)$.

$$\hat{x}(n) = \sum_{k=1}^N a_k x(n-k)$$

b) Predictores lineales de duración infinita.

En la figura 2, representamos el esquema correspondiente a predictores lineales de duración infinita, donde se observa que se incluyen los errores de predicción anteriores, lo que equivale a predecir con todas las muestras pasadas.

De tal forma que la predicción correspondiente a la muestra futura $\hat{x}(n)$, se obtiene a partir de la combinación lineal

$$\hat{x}(n) = \sum_{k=1}^N a_k x(n-k) + \sum_{i=1}^L b_i e(n-k)$$

donde los coeficientes b_i nos indican el grado de influencia del error de predicción según el instante.

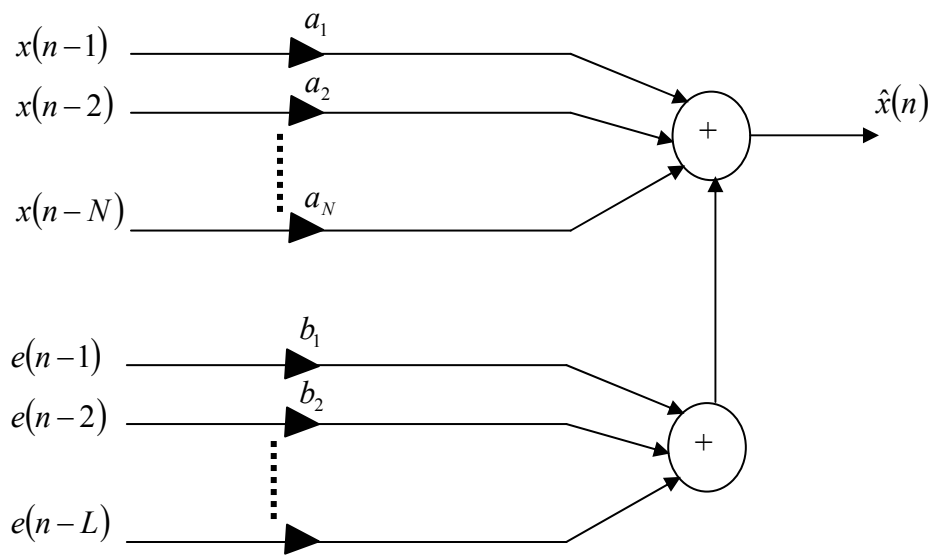


figura 2: Esquema de predictor lineal de duración infinita

Como conclusión, faltaría preguntarnos bajo qué circunstancias estos esquemas de predicción, basados en estructuras lineales, son suficientes, o por el contrario cuando hemos de evolucionar hacia esquemas de predicción no lineales con el fin de mejorar el grado de predictibilidad y en consecuencia mejorar más la *SNR* a la salida del bloque predictor.

4.2. No lineales:

Si elegimos como criterio de optimización la minimización del error cuadrático medio (ecm), la predicción óptima es la media condicionada del valor a predecir a los valores anteriores. [PRI88]

$$\hat{x}_{ecm}(n) = E[x(n) / x(n-1), x(n-2), \dots]$$

esta expresión es en general una función no lineal, por lo que el uso de predictores no lineales permitirá una mejor aproximación de la media condicionada. [KUB98]

Sólo si el proceso estocástico es Gaussiano, la media anterior es una función lineal, por lo que el predictor óptimo es lineal.

En esta sección vamos a analizar cuatro diferentes esquemas de predictores no lineales, los cuales difieren fundamentalmente en dónde colocan la no linealidad y que engloban la casi totalidad de estructuras de predicción no lineal que aparecen en la literatura.

- Redes Neuronales.
- Funciones de base radial.
- Series de Volterra.
- Bilineales.

4.2.1. Redes neuronales:

Las redes neuronales pueden tomar estructuras muy complicadas y variadas, pero todas ellas se basan en combinadores lineales seguidos por linealidades sin memoria [PRO96] [PRO97] [HAY96] [HAY95] [ATY96] [LIU97] [CHE90] [SUY96]. Es decir, en cada neurona, tendríamos un sistema constituido por dos bloques: el primero lineal y el segundo no lineal, tal como se puede observar en la figura 3.

La salida del primer bloque corresponde a la combinación lineal de las N muestras anteriores al instante de predicción.

$$x_{CL}(n) = \sum_{k=1}^N w_k x(n-k)$$

siendo $x_{CL}(n)$, la salida del combinador lineal, la cual, es ahora la entrada al bloque no lineal. Dicho bloque está caracterizado por una función no lineal que nos determinará el valor de la salida para la entrada proveniente del bloque lineal.

En consecuencia una red neuronal, estará constituido por un conjunto de neuronas, distribuidas en capas. Esta distribución de neuronas proporciona estructuras de redes neuronales más o menos complejas.

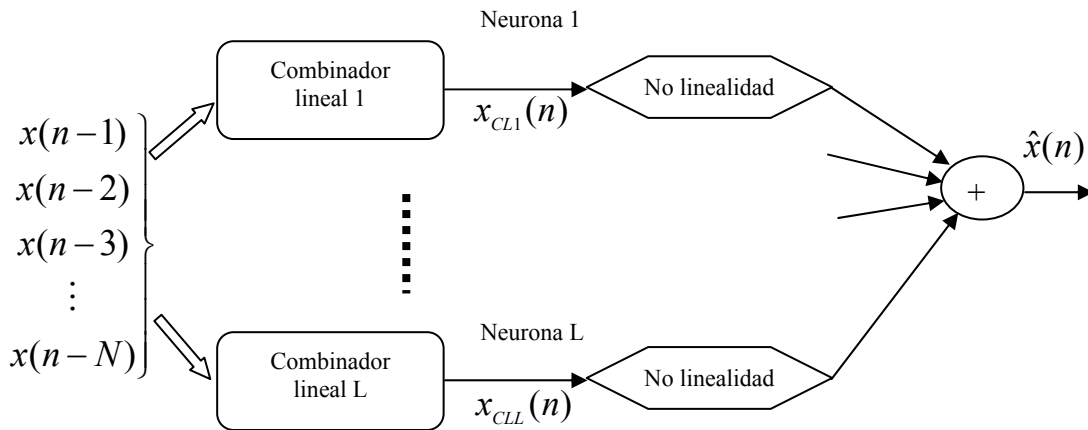


figura 3: Esquema correspondiente a una red neuronal.

Las redes neuronales tiene un importante número de propiedades que hacen apropiado su uso en aplicaciones de procesamiento de señal [HAY96].

El modelo de la neurona y la arquitectura de la red neuronal describe como una red transforma su entrada en su correspondiente salida. El modelo de la neurona y la arquitectura en cada caso determinan el grado de complejidad computacional de la misma.

En una neurona con una sola entrada sin sesgo (figura 4) el escalar de entrada p se multiplica por un factor de ponderación w , para formar el producto $w * p$. Este escalar será ahora el argumento de la función de transferencia no lineal F , dando lugar al escalar de salida a . Si además la neurona presenta un sesgo distinto de cero, b , el argumento de la función de transferencia no lineal será ahora $w * p + b$.

El hecho de que cada unidad de procesamiento (neurona) de una red neuronal presenta una función en su interior que es no lineal, permite que las redes neuronales se caractericen por tener una habilidad inherente para modelar no

linealidades subyacentes contenidas en los mecanismos que caracterizan unos determinados datos de entrada.

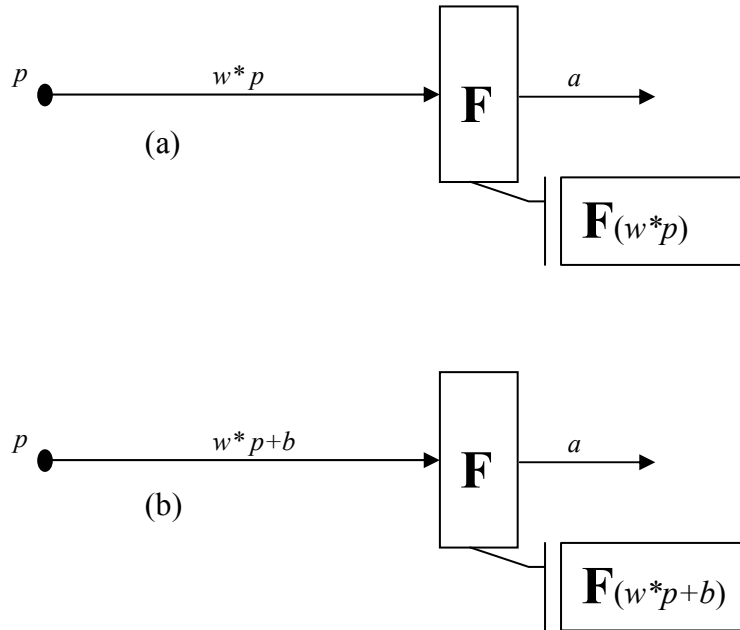


figura 4: Esquema correspondiente a una neurona con entrada escalar.
(a) Con sesgo $b=0$, (b) Con sesgo $b \neq 0$.

Por ejemplo, un perceptrón multinivel, [HAY94a][LIP87][HUS93] que corresponde a una estructura habitual de representación de una red neuronal, consta de un elevado número de neuronas dispuestas en forma de niveles, donde cada neurona en un particular nivel está conectada a un elevado número de fuentes nodos o neuronas del nivel previo. Esta forma global de interconectividad presenta el potencial de ser tolerante a fallos.

Las redes neuronales aprenden por medio de una sesión de entrenamiento durante la cual sus parámetros son ajustados de una forma sistemática con el fin de minimizar una función coste. Típicamente la función coste corresponde al criterio de minimizar el error cuadrático medio, en el que la señal de error es definida como la diferencia entre la respuesta deseada y la salida de la red producida como respuesta a la correspondiente señal de entrada.

4.2.2. Funciones de Base Radial:

Podemos ver en la figura 5 una estructura de predicción basada en las funciones de base radial (FBR). [CHN96][NIS96][MUL96]

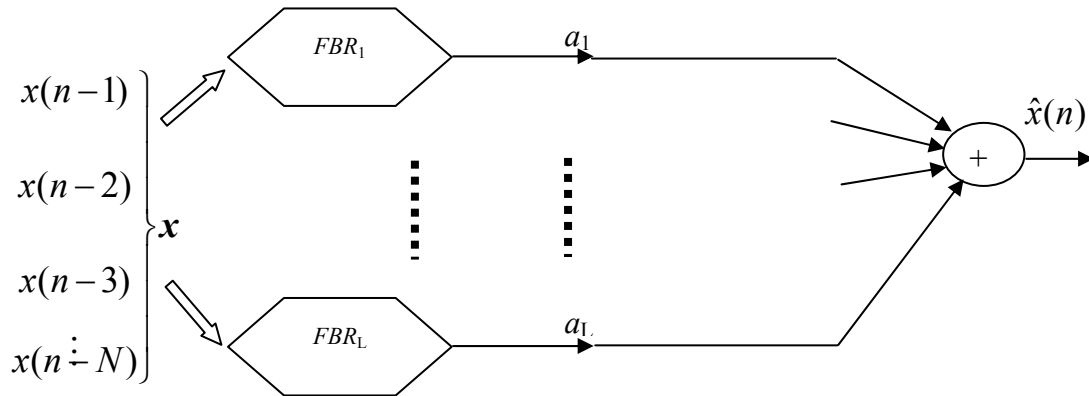


figura 5: Esquema correspondiente a una red de funciones de base radial

Para predecir el valor de señal $\hat{x}(n)$, el vector de entrada de la red FBR

$$\mathbf{x} = [x(n-1), x(n-2), \dots, x(n-N)]^T$$

es un vector N dimensional que contiene las N muestras anteriores al instante de predicción. Una elección típica es la función Gaussiana,

$$FBR_j = \exp(-\alpha \|\mathbf{x} - \mathbf{c}_j\|^2)$$

como la no linealidad. De tal forma, que la respuesta correspondiente al j -ésimo nodo para el vector de entrada \mathbf{x} viene dada por,

$$\theta_j = \exp(-\alpha \|\mathbf{x} - \mathbf{c}_j\|^2)$$

donde \mathbf{c}_j es un vector “center” N dimensional y α es una constante positiva que determina el ancho de la respuesta simétrica de los nodos. Investigaciones teóricas al respecto han mostrado que la elección de la no linealidad para los

nodos no es crucial y un ancho uniforme para cada nodo es suficiente para la aproximación universal. La salida de la red es definida como,

$$\hat{x}(n) = \sum_{j=1}^L \theta_j a_j$$

donde a_j son los pesos de conexión de la red y L es el número de nodos, que podemos diseñar con técnicas similares a las del predictor lineal. Los “centers”, c_j , $1 \leq j \leq L$, los podemos seleccionar entrenando con \mathbf{x}_k , $1 \leq k \leq M$. Donde el valor de los pesos pueden ajustarse usando el método de minimizar el error cuadrático medio.

Las redes de funciones de base radial FBR han tenido un considerable éxito en aplicaciones de predicción de series temporales no lineales. Pero en la mayoría de los casos su éxito se basaba en la predicción de señales estacionarias. El funcionamiento del predictor FBR para señales no estacionarias basadas en el esquema de la figura 4 es menos satisfactorio. Este es el motivo de que la red FBR, así como otros modelos basados en redes neuronales, no caractericen bien las variaciones temporales. Debido a que las señales de trabajo reales no sólo son no Gaussianas sino que además son no estacionarias, es necesario analizar sistemas basados en FBR que exhiben ambas características [CHN96] entre ellas podemos citar la red GFBR, que como en el caso de la FBR, es una red neuronal realimentada con un solo nivel oculto. En la red GFBR sin embargo, el vector de entrada a la red es generando mediante diferencias de los datos originales. El orden de la diferenciación determina el orden de la red GRBF. Por ejemplo, el vector de entrada de una red GFBR de primer orden en el instante n viene dado por,

$$\mathbf{x}'_n = \mathbf{x}_n - \mathbf{x}_{n-1} = [x(n-1) - x(n-2), x(n-2) - x(n-3), \dots, x(n-N) - x(n-N-1)]^T$$

donde \mathbf{x}_n y \mathbf{x}_{n-1} , son los vectores de entrada originales de la red FBR en los instantes n e $n-1$, respectivamente. Los elementos de \mathbf{x}'_n muestran el índice de

cambio en la trayectoria de la serie temporal conforme evolucionan las N muestras.

En la figura 6, vemos representada la estructura de la red. En ella podemos observar que aunque la función Gaussiana es todavía usada como función no lineal que compara la similitud del vector de entrada con los “centers” de los datos ocultos, ahora la respuesta de la función Gaussiana es multiplicada por un término adicional $(x(n-1) + \delta_j)$, de tal forma que la respuesta del j -ésimo nodo oculto de la red GFBR de primer orden a la entrada del vector x_i' viene dada por la expresión,

$$\theta_j = \exp\left(-\alpha\|x_n' - c_j'\|^2\right) \times (x(n-1) + \delta_j) \tag{1}$$

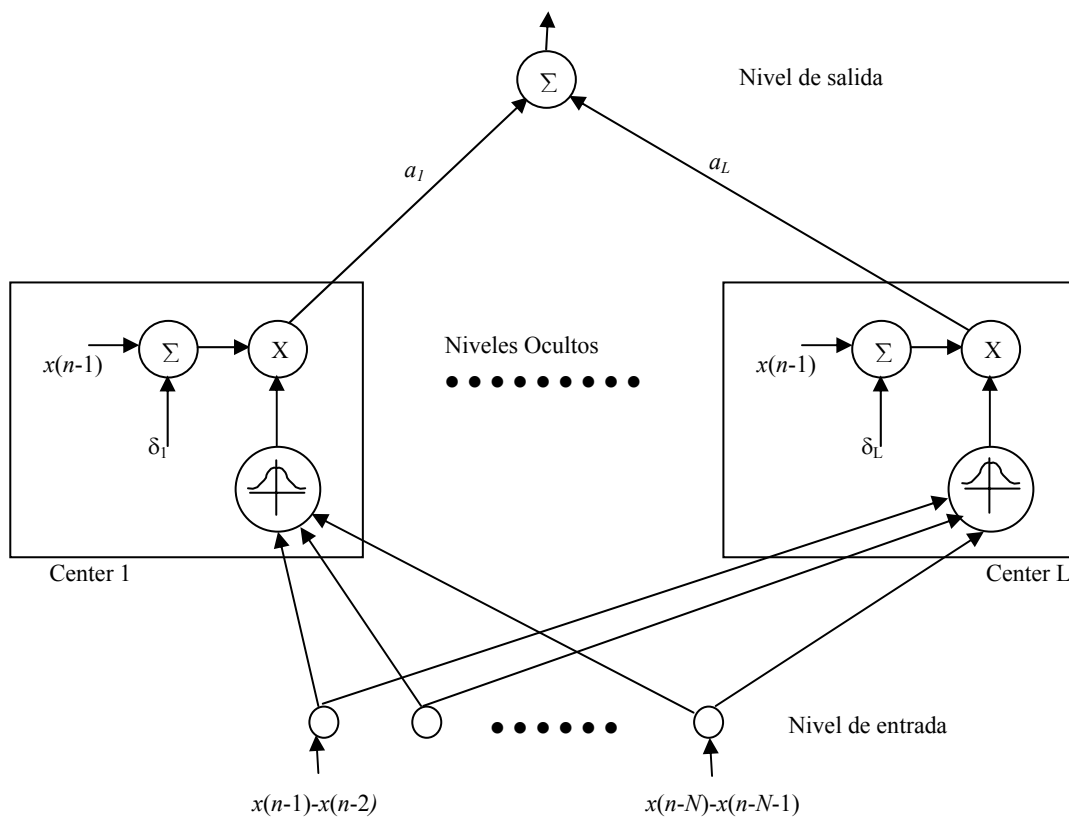


figura 6: Topología de la red GFBR de primer orden.

donde \mathbf{c}_j' es un vector “center” N -dimensional y δ_j es un valor constante asociado al correspondiente “center”. El término $(x(n-1) + \delta_j)$ puede ser interpretado como un simple paso de predicción local de $x(n-1)$ por medio del nodo oculto j -ésimo. Desde la expresión (1), si el vector de entrada es muy parecido al “center” j -ésimo, el valor de la función Gaussiana estará próximo a 1.0 y el predictor $(x(n-1) + \delta_j)$ se hace totalmente activo. Como en el caso de la red FBR, el nivel de salida es un combinador lineal con pesos a_j , con $1 \leq j \leq L$.

Los “centers” \mathbf{c}_j' y los escalares δ_j , $1 \leq j \leq L$, los podemos elegir durante el entrenamiento de los datos para entrenar \mathbf{x}_k' , $1 \leq k \leq M$, como podemos observar en la figura 7. Para cada vector de entrada de entrenamiento, \mathbf{x}_k' , definimos $d_k = x(k) - x(k-1)$.

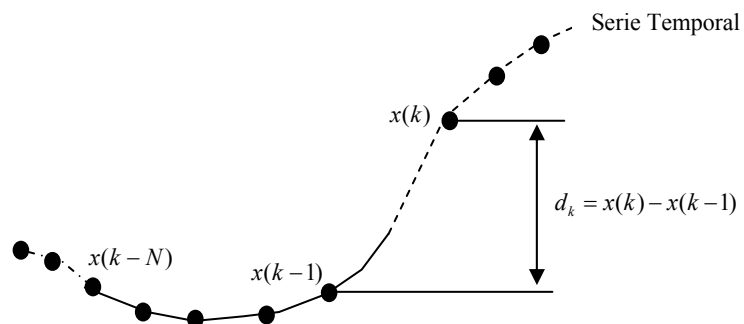


figura 7: Modo de entrenamiento del j -ésimo nodo oculto. Si el k -ésimo vector \mathbf{x}_k' de entrenamiento lo elegimos como el “center” \mathbf{c}_j' , δ_j es colocada a d_k .

Si \mathbf{x}_k' lo seleccionamos como el j -ésimo “center” \mathbf{c}_j' , colocamos δ_j igual a d_k para asegurar que el j -ésimo nodo oculto es el predictor perfecto de $x(k)$. De esta forma, el problema de construir una red es equivalente a la tarea de seleccionar una subclase de modelo de L términos $\{\mathbf{c}_j', \delta_j\}_{j=1}^L$ desde el modelo total de M términos $\{\mathbf{x}_k', d_k\}_{k=1}^M$.

En base a lo visto anteriormente el modelo de la red GFBR llega a ser obvio cuando ésta realice operaciones de predicción. Cada nodo oculto compara el vector de entrada a la red x_n' con su center c_j' . La respuesta Gaussiana de cada nodo oculto nos indica el grado de igualdad entre x_n' y c_j' . Así los nodos ocultos son sensibles al gradiente de la serie antes que a la propia serie como sucede en el modelo de la red FBR. El término $(x(n-1) + \delta_j)$ tiene también un claro significado geométrico, si nos referimos a la figura 8, podemos observar que si el j -ésimo “center” es igual al gradiente x_n' de la serie, $(x(n-1) + \delta_j)$ es el valor apropiado para ser una buena aproximación de $x(n)$.

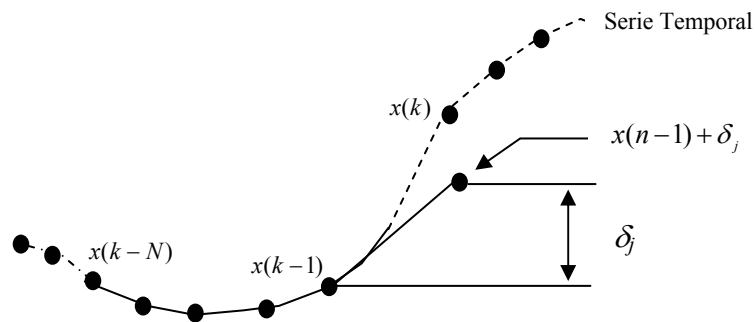


figura 8: Modo de predicción del j -ésimo nodo oculto. Si el “center” c_j' es igual a x_k' , $(x(n-1) + \delta_j)$ es una buena aproximación de $x(n)$.

Hemos analizado con cierto detalle la estructura de red GFBR como una técnica de predicción no lineal para series temporales no lineales y no estacionarias. Como vemos los nodos de esta red se diseñan para responder al gradiente de la serie temporal antes que a su trayectoria. Esto puede normalmente mejorar la exactitud de la predicción, particularmente para serie temporales homogéneas no estacionarias. Nuestra intención con esta presentación es poner de manifiesto como, incluso en estructuras no lineales relativamente sencillas de ajustar, la complejidad del diseño puede resultar muy limitadora para su aplicación práctica.

4.2.3. Series de Volterra:

La estructura de las series de Volterra es idéntica a la estructura correspondiente a las Funciones de Base Radial, las cuales presentan una arquitectura común, [NIS96][PIT90][OZD96][CHA98][MAT96] una expansión no lineal seguida de un combinador lineal, como podemos observar en la figura 9.

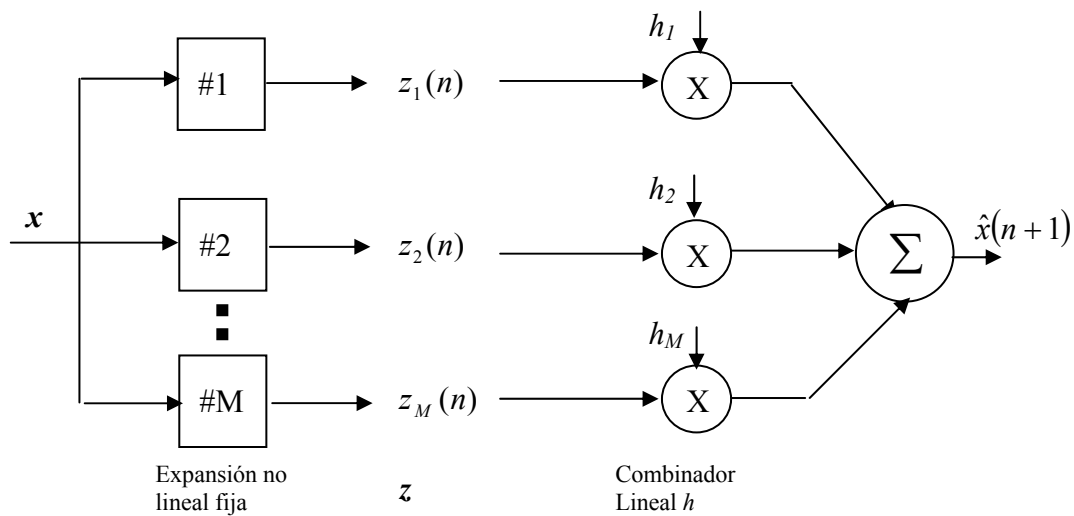


figura 9: Estructura de un sistema predictor no lineal para N muestras del vector x y M coeficientes del vector h

La entrada del predictor será como de costumbre,

$$x = [x(n)x(n-1)\dots x(n-N+1)]^T$$

a partir del cual obtenemos una estima $\hat{x}(n+1)$ del próximo dato $x(n+1)$, o en general de $x(n+D)$.

El vector x es la entrada a una expansión fija de no linealidades para formar el vector z . En este caso la predicción puede expresarse como el producto escalar del vector intermedio z con el vector h , que contiene los M coeficientes o pesos, $\hat{x}(n+1) = h^T z$. Donde la calidad de la predicción es por lo general medida por el error cuadrático medio de predicción.

La estructura básica de las series de Volterra corresponde a una serie de términos polinómicos. Estos términos se construyen a partir de los N valores conocidos del vector \mathbf{x} . De tal forma que una estima de $\hat{x}(n+1)$ la podemos obtener mediante la serie.

$$\hat{x}(n+1) = a + \sum_{i_1=0}^{N-1} a_{i_1} x(n-i_1) + \sum_{i_1 \leq i_2} a_{i_1 i_2} x(n-i_1)x(n-i_2) + \sum_{i_1 \leq i_2 \leq \dots \leq i_p} a_{i_1 i_2 \dots i_p} x(n-i_1)x(n-i_2)\dots x(n-i_p)$$

esta expresión correspondería a la serie de Volterra de p términos. Esta expresión es similar al desarrollo truncado de la serie de Taylor multivariable.

Analizando la estructura de la red, es evidente, que el número de términos implicados llegará a ser prácticamente inmanejable a medida que los parámetros p y N sean más altos . Así por ejemplo, para $p=3$, el número de términos M crece como un polinomio cúbico, por ejemplo,

$$M = \frac{1}{6} (6 + 11N + 6N^2 + N^3)$$

en definitiva, cada serie será subóptima en función del truncado realizado.

Por ejemplo, en la serie de Volterra con $N=p=2$, tendremos que el vector \mathbf{x} será de la forma,

$$\mathbf{x} = [x(n)x(n-1)]^T$$

y la predicción la representaremos por el producto interior $\hat{x}(n+1) = \mathbf{h}^T \mathbf{z}$, donde $\mathbf{h}^T = [a \ a_0 \ a_1 \ a_{00} \ a_{01} \ a_{11}]$ y $\mathbf{z}^T = [1 \ x(n) \ x(n-1) \ x^2(n) \ x(n)x(n-1) \ x^2(n-1)]$.

Para ajustar los coeficientes suponiendo el criterio de minimización de mínimo cuadrados, los pesos del vector \mathbf{h} pueden expresarse como la solución de un sistema de M ecuaciones lineales:

$$\boldsymbol{\varphi}_{zz} \mathbf{h} = \boldsymbol{\varphi}_{zx}$$

donde $\boldsymbol{\varphi}_{zz} = \sum_{k=N}^{K-1} \mathbf{z}_k \mathbf{z}_k^T$ y $\boldsymbol{\varphi}_{zx} = \sum_{k=N}^{K-1} \mathbf{z}_k x(k+1)$, siendo K es el número de elementos disponibles de la serie temporal $x[n]$, $\{\mathbf{x}_k : 1 \leq k \leq K\}$. Habitualmente, tendremos más ecuaciones que incógnitas y por tanto buscaremos un vector solución \mathbf{h} , cuya norma sea mínima (\mathbf{h}_{min}), usando cualquiera de las múltiples técnicas existentes [NIS96].

Mientras el diseño de un predictor lineal convencional requiere el conocimiento de una estima del segundo orden de la señal, en este tipo de estructura de predicción basadas en series de Volterra, se explotan los momentos más altos de la señal. En un caso particularmente fácil donde $p=3$ y $N=2$, \mathbf{z} contiene términos del tipo $x(n)x(n-1)$ y $x^2(n)$, evaluar $\boldsymbol{\varphi}_{zz}$ implica calcular términos como,

$$\sum_k x^3(k)x(k-1)$$

que se corresponde a la estima del momento de cuarto orden. En este ejemplo particular tenemos que estimar hasta el momento de orden sexto para construir el predictor. Si lo comparamos con una estructura basada en Funciones de Base Radial, los momentos de todos los órdenes son necesarios para construir el predictor, donde la no linealidad fija, puede ser expandida como una serie de Taylor, por ejemplo cuando $N=2$.

$$s(\|\mathbf{x}_k - \mathbf{c}_j\|) = \sum_{i_1+i_2=0}^{\infty} \beta_{i_1 i_2} x^{i_1}(k) x^{i_2}(k-1)$$

Por tanto las series de Volterra nos permiten caracterizar sistemas no lineales, y presentan al igual que las estructuras basadas en FBR una complejidad con un número de estados elevado. Se basa en definitiva en la

aproximación de no linealidades mediante aproximaciones polinómicas de grado $p \times N$.

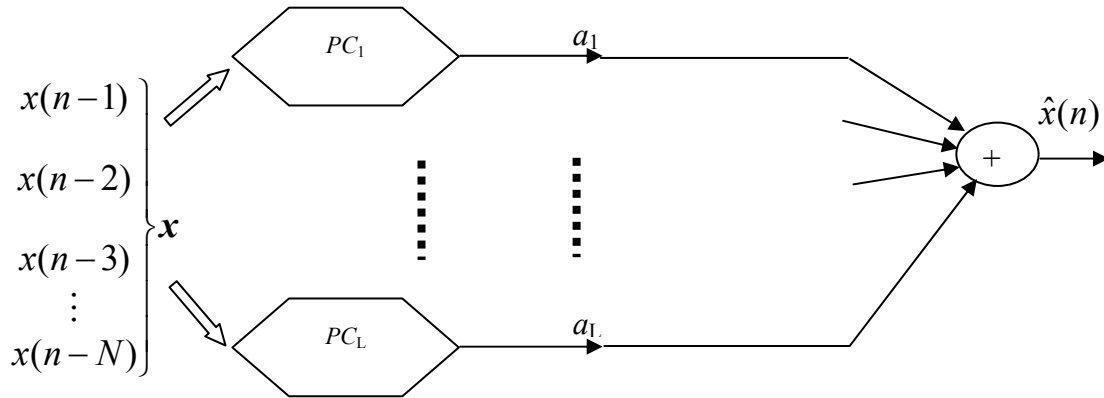


figura 10: Esquema correspondiente a una red basada en Series de Volterra

Como conclusión, la estructura de un predictor no lineal basado en Series de Volterra podemos verla en la figura 10. Donde PC_i corresponden a los diferentes productos cruzados que intervienen, cada uno de ellos corresponde a las no linealidades expresadas en la figura 9 (compárese con la figura 5).

4.2.4. Predictores bilineales:

Los predictores bilineales, presentan una estructura idéntica a los filtros lineales de orden infinito, pero incluyendo además términos cruzados del tipo $x(\cdot)e(\cdot)$, lo que le convierte en una estructura no lineal, tal como podemos observar en la figura 11.

Los predictores bilineales también equivalen a las series de Volterra de orden infinito.

El valor de la predicción de una serie temporal bilineal satisface la siguiente ecuación en diferencias, [NEU97][LEE94][PRI88].

$$\hat{x}(n) = \sum_{i=1}^N a_i x(n-i) + \sum_{i=1}^L b_i e(n-i) + \sum_{i=1}^K \sum_{j=1}^J c_{ij} x(n-i)e(n-j)$$

siendo $e(n) = x(n) - \hat{x}(n)$ el error de predicción.

En particular podemos observar que si los coeficientes c_{ij} son cero, el modelo de predicción equivale a una estructura lineal de duración infinita (ver figura 2).

A pesar de la simplicidad de los modelos de sistemas bilineales, hay un gran tipo de sistemas no lineales que podemos modelar adecuadamente mediante sistemas o estructuras bilineales.

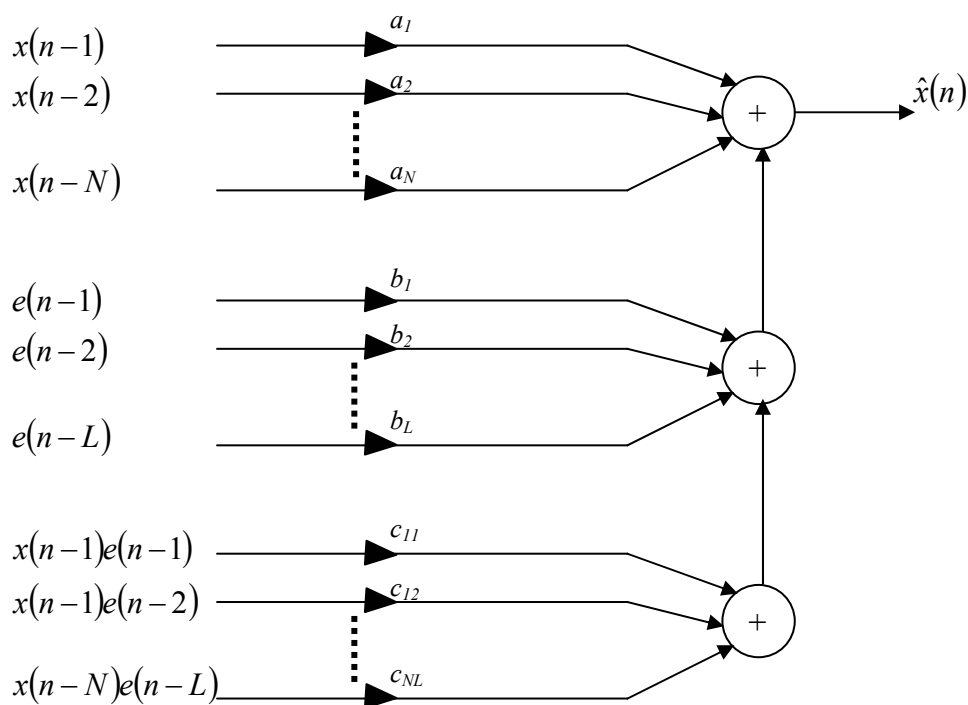


figura 11: Esquema correspondiente a un predictor bilineal

Una importante cuestión asociada con los modelos de sistemas bilineales es su estabilidad [LEE94]. Debido a la estructura no lineal, el problema de estabilidad es mucho más complicado que el asociado con sistemas recursivos lineales. Por ejemplo, es posible encontrar señales de entrada limitada que pueden causar que la salida de casi todos los sistemas bilineales sea ilimitada.

De esta manera se ha de evaluar las condiciones de estabilidad de las estructuras bilineales, así las estructuras bilineales adaptativas RLS extendidas,

[LEE93] han sido enseñadas para ser estables en el sentido que el error cuadrático medio este limitado siempre que la señal deseada esté limitada en el mismo sentido.

5. Estructuras de Wiener y Hammerstein:

En la sección anterior hemos presentado diferentes tipos de estructuras no lineales. En definitiva, se trata de sistemas con los que podemos describir con mayor o menor éxito comportamientos no lineales, en consecuencia mejorar nuestra relación SNR a la salida del predictor. Pero, dichas estructuras presentan en general una elevada complejidad en el diseño de las mismas debido al gran número de coeficientes necesarios, y a lo complicado de los algoritmos de entrenamiento, que al final repercute también en un elevado tiempo de computación, al margen de problemas específicos como la inestabilidad de los predictores bilineales.

En la presente sección vamos a describir estructuras muy simples con las que caracterizar predictores no lineales, con el fin de proporcionar sistemas capaces de mejorar la predicción en entornos no-Gaussianos. De las estructuras descritas, justificaremos aquélla que mejor se adapte a las necesidades de la aplicación.

En general puede comprobarse como las diferentes arquitecturas revisadas en los apartados anteriores, se caracterizan por combinar partes lineales con partes no lineales, pudiendo estas últimas aparecer antes o a continuación de las lineales. Partiendo de esa idea, planteamos el estudio de estructuras muy simples que permitan su aplicabilidad al problema que nos ocupa. Si la mejora obtenida con estas estructuras simples es cercana a la mejor que podríamos conseguir, no tiene sentido recurrir a arquitecturas no lineales complejas de diseño complicado y coste computacional alto. Por otra parte, siempre será posible mejorar el comportamiento (si el esfuerzo lo merece) a base de incluir más estructuras simples actuando conjuntamente, lo que nos lleva a las estructuras más complejas anteriormente presentadas. Por todo ello, optamos en principio por la opción de las denominadas estructuras de Wiener, de Hammerstein y de Wiener-Hammerstein, que pasamos a presentar, cuyos análisis teóricos se realizaron entre los años 60 y 70, y que están suscitando mucho interés en los últimos tiempos, [CEL01] [NOR01]

[NOR98] [BER01] [RAL97] [SHI94] [CHE97] [MAK98] [DIJ97] [HUA98] [MAH98] [KAN98] [BOU93] [HAY94b]

- **Sistema de Wiener:** [WIE58] [HAL69] [BUS74] [FRE76] [BIL77] [RUD79] [SCH81] [KOR86] [HAB87]

Consta de un bloque lineal seguido por una ganancia no lineal, (sistema no lineal sin memoria) como podemos ver en la figura 12.

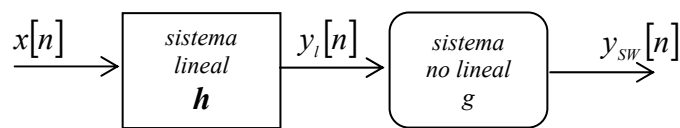


figura 12: Modelo de Wiener

- **Sistema de Hammerstein:** [CHA71] [GAL76] [BIL79] [CHU88] [ESK91] [PAW91] [RAL95]

El orden de los bloques es el inverso al modelo anterior, figura 13.

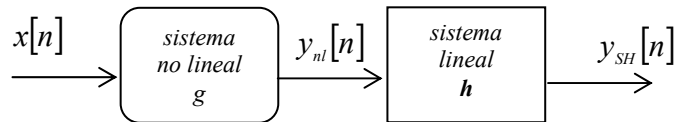


figura 13: Modelo de Hammerstein.

- **Sistema de Wiener-Hammerstein:** [BOU95]

Sistema constituido por dos bloques (estructuras) lineales y un bloque no lineal, tal como puede observarse en la figura 14.

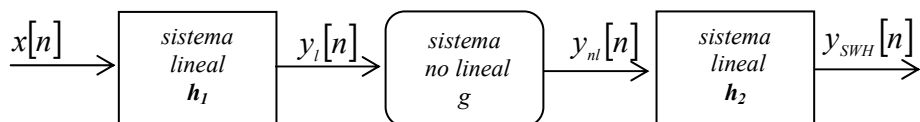


figura 14: Modelo de Wiener-Hammerstein.

Debe hacerse notar que, estructuras más o menos complejas pueden construirse a partir de bloques simples conectados en paralelo o serie. Al respecto señalar que Korenberg [KOR91] demostró que cualquier sistema que se caracterice por medio de una serie infinita de Volterra, puede ser exactamente representado por una suma finita de sistemas Wiener en paralelo.

El sistema seleccionado en la tesis corresponde al de Wiener, el motivo fundamental se centra en que tras analizar las diferentes estructuras, el número de bloques tanto lineales como no lineales necesarios en nuestra aplicación es considerablemente menor en este sistema.

Como ejemplo aclaratorio, para ciertos valores de N y D (recuérdese que N es el número de barridos con los que predecir y D el número de barridos a predecir), en la estructura de Wiener sería necesario disponer de D bloques lineales y de D bloques no lineales, en cambio para caracterizar la estructura de Hammerstein serían necesarios $N \times D$ bloques no lineales y D bloques lineales, tal como podemos observar en las figuras 15 y 16.

Las estructura de Wiener puede considerarse como una red neuronal formada por una sólo neurona. También, si la no linealidad sin memoria es de tipo polinómico, conduce a una estructura de Volterra o incluso a una Función de Base Radial (FBR), si elegimos adecuadamente dicha no linealidad. Por tanto, dentro de su simplicidad aglutina las diferentes aproximaciones para la implementación de predictores no lineales revisados en la sección 4.

En el siguiente Capítulo procederemos a un estudio detallado de esta estructura de Wiener, muy orientada como se verá, a la aplicación concreta que nos ocupa.

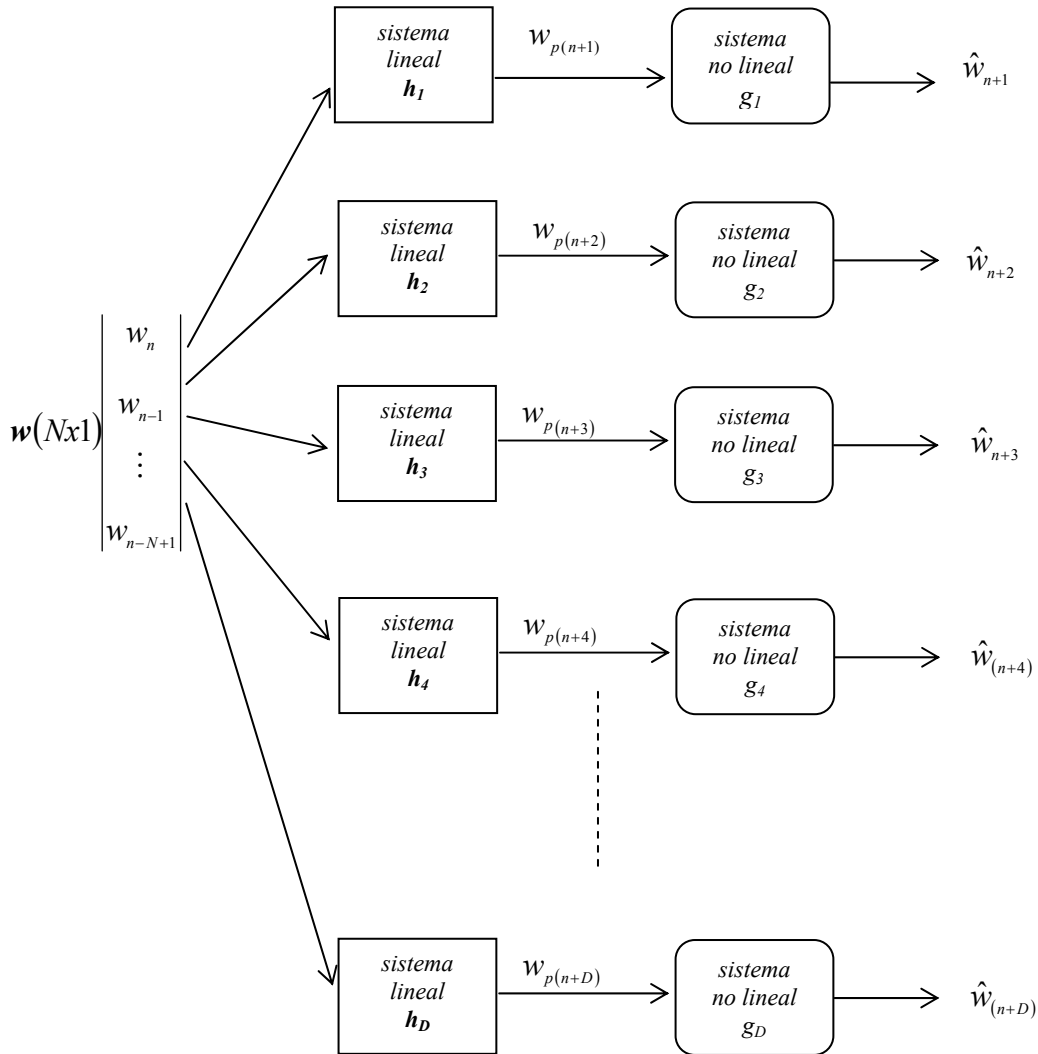


figura 15: Estructura de Wiener para: Longitud predictor N y D barridos en los que basar la predicción.

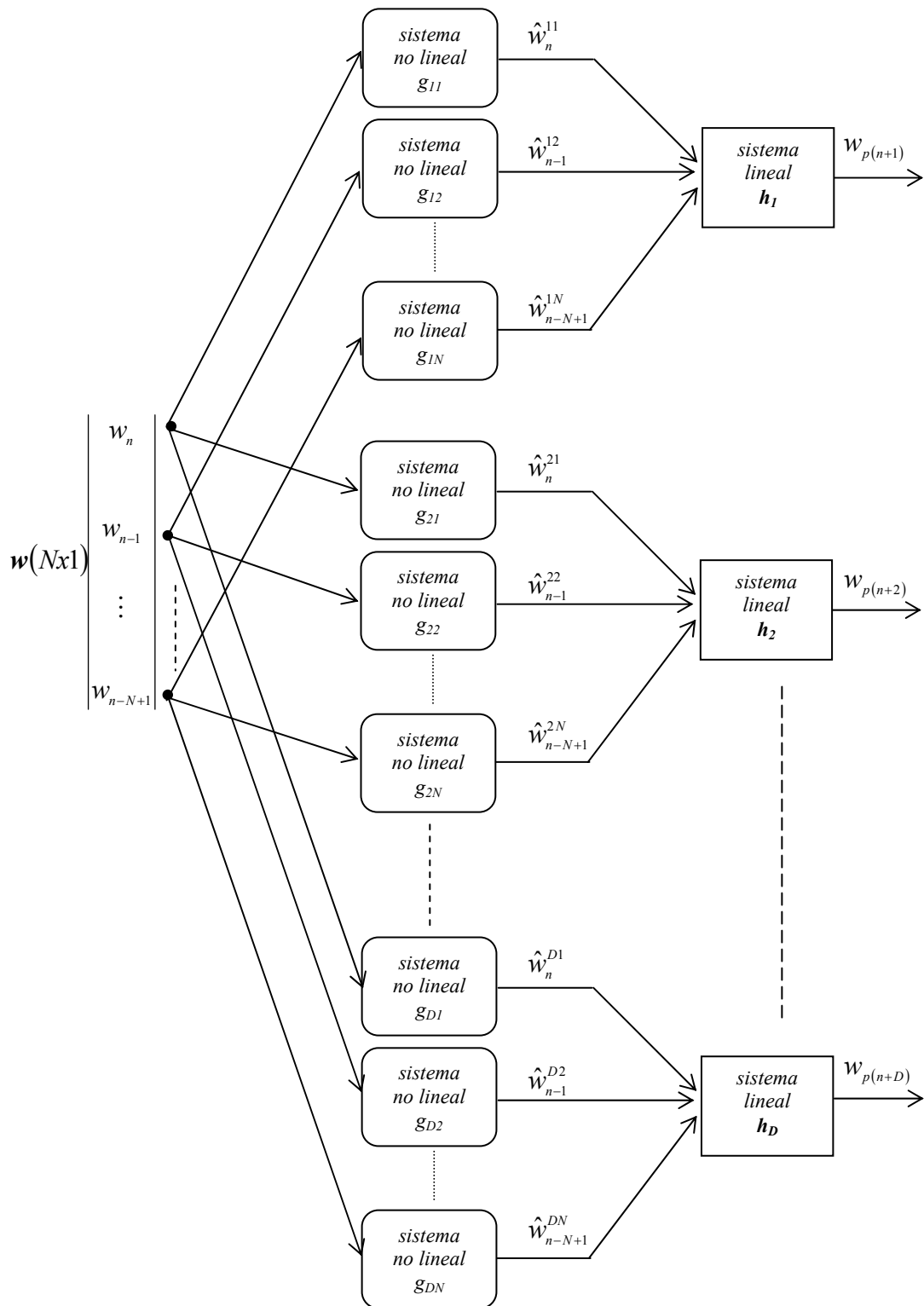


figura 16: Estructura de Hammerstein para: Longitud predictor N y D barridos en los que basar la predicción.

Capítulo IV:

Diseño del Predictor no Lineal: Estructura de Wiener

Capítulo IV

DISEÑO DEL PREDICTOR NO LINEAL: ESTRUCTURA DE WIENER

1. Introducción:

En la figura 1, podemos observar la no linealidad en la función de distribución $f(x, x_p)$ obtenida al representar la nube de puntos de las predicciones x_p a la salida del predictor lineal frente a los datos de entrada procedentes del sensor de infrarrojos, en un caso particular, en concreto, para el **bloque C** (acimuts primero y segundo), con longitud del predictor ($N=6$) y distancia de predicción ($D=6$). Si tenemos en cuenta que el mínimo del error cuadrático medio de la predicción es la media condicional de la variable aleatoria correspondiente a la muestra predicha dada las muestras anteriores obtenidas por el predictor, [PRI88] ésta tendrá una distribución no lineal.

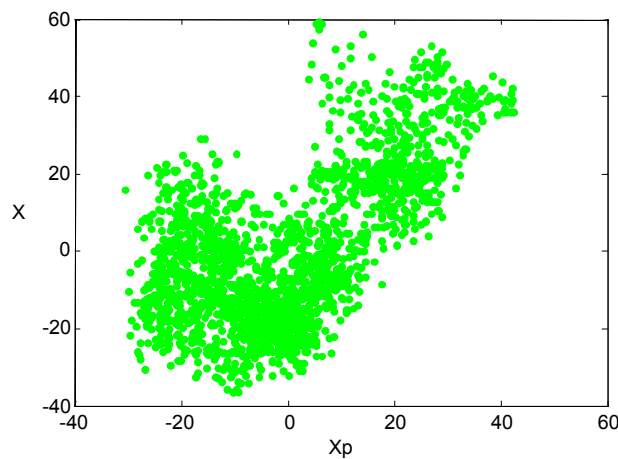


figura 1: Ejemplo de función de distribución de los datos procedentes del sensor

El objetivo de este Capítulo se centra en diseñar la estructura de Wiener general de la figura 2. La parte lineal de la estructura puede concebirse siguiendo cualquiera de los métodos estándares, en nuestro caso aplicamos lo analizado en el Capítulo II, sección 3 de la presente tesis. Para el diseño de la no linealidad, proponemos dos técnicas. El objetivo de cada una es lograr la media condicional de la predicción no lineal dada la

lineal, para de esta forma reducir el nivel de ruido de fondo y en consecuencia mejorar la relación *SNR* en la aplicación de detección automática de fuegos descontrolados. Con este planteamiento en el diseño, pretendemos mantener la simplicidad del predictor no lineal mientras aprovechamos la solución no lineal para implementar la media condicional del valor predicho, dada la salida del predictor lineal.

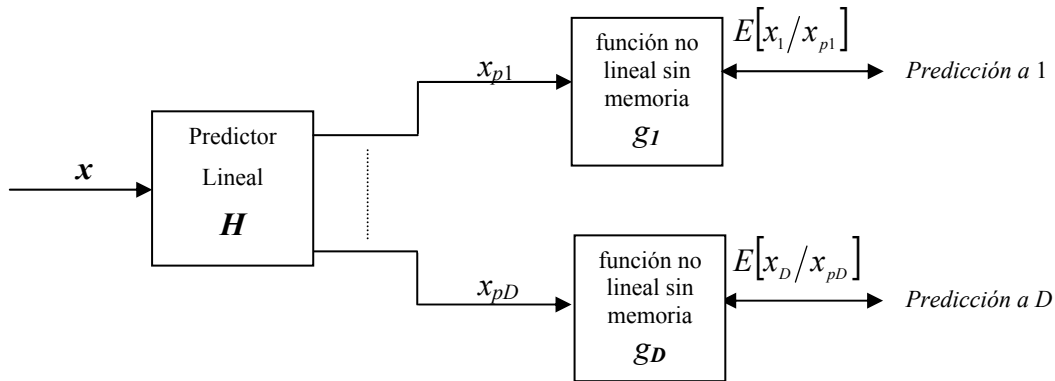


figura 2: Estructura de Wiener propuesto para la predicción no lineal

Para ilustrar el diseño de la estructura trabajamos con los mismos bloques considerados en el Capítulo II. Seleccionando el primer y segundo acimut para entrenar la estructura, y, el tercero y cuarto para testarla. En consecuencia, los **bloques A y B**, están constituidos por 21 registros, para cada uno de los 21 barridos consecutivos tomados cada minuto, en cada registro hay 102 muestras en rango (*Ra*). El valor de 102, 51x2, corresponde a 51 rangos seguidos tomados en dos acimuts consecutivos (primero y segundo para entrenar, tercero y cuarto para testear la estructura). Para el **bloque C**, tendremos 21 registros de 200 muestras cada uno (100x2).

Las estimación de cada una de las *l* distintas funciones no lineales sin memoria, necesarias para implementar las *D* estructuras de Wiener (ver Capítulo III, sección 5) se realiza a partir de todas las *V* posibles muestras predichas x_{pl} , a distancia de predicción *l* (*l* variando de 1 a *D*), con $V = Ra \times (21 - N - D + 1)$, obtenidas al pasar los datos de entrenamiento por el predictor lineal (de orden *N* y *D* barridos en los que basar la predicción).

De tal forma que para el diseño de la función l -ésima no lineal sin memoria, tenemos que evaluar la media condicional del proceso estocástico x_{pl} (las V muestras predichas a distancia l) con sus V muestras correspondientes a los datos originales a distancia l , x_l .

En la sección 2, presentamos las dos técnicas propuestas para estimar la media condicional, haciendo hincapié en la segunda de las técnicas propuestas, ya que nos permite encontrar una formulación cerrada (polinómica) con la que estimar a partir de los datos de entrenamiento la media condicional. En la sección 3, evaluamos cada una de las técnicas propuestas en cada uno de los diferentes bloques seleccionados a priori, cuyas características han sido expuestas en el Capítulo II, sección 6, definiendo el factor potencia de error normalizado (NPEMP), factor que nos permitirá determinar cuándo conviene incluir la no lineal en el predictor. De esta forma presentamos un procedimiento para decidir si resulta necesario incluir o no la no linealidad, basado directamente en la estimación del error de predicción, que consideramos mucho más práctico que otras alternativas presentadas en [KUB98][DZW99][BERN98][PON98]. Si además tenemos en cuenta que el predictor óptimo para un proceso estacionario es lineal, el factor NPEMP podría servir también como indicativo del grado de gaussianidad del proceso. En la sección 4, se verifica la validez de la estructura descrita sobre los experimentos realizados en el Capítulo II, sección 3, con el fin de determinar las curvas de PD y PDI, al considerar la no linealidad. Por último en la sección 5, se establece un procedimiento con el fin de evaluar el valor adecuado del orden del predictor N , así como una técnica para valorar si podríamos mejorar el predictor lineal sustituyendo las estructuras FIR implícitas en las filas de la matriz de predicción H por estructuras IIR.

2. Técnicas aplicadas al diseño de la no linealidad:

2.1. Estimación directa:

Necesitamos estimar la función no lineal $g_l(\cdot)$ correspondiente a predicciones a distancia l . Suponemos disponibles para dicho diseño un conjunto de V parejas de

valores $\{(x_{pl}, x_l)\}$, donde x_{pl} es la predicción lineal óptima (para un cierto N) del valor a predecir x_l .

A continuación desplazamos una ventana de anchura Δ muestras a lo largo del registro doble formado por el conjunto $\{x_{pl}\}$ ordenado ascendentemente y sus correspondientes parejas. En cada posición de la ventana podemos calcular una estima de la función (media condicionada).

$$g(x_{pl}) = E[x_l/x_{pl}] \tag{1}$$

tal y como indicamos a continuación.

Sea $x_{pl}^{(i)}$, con $i=1, \dots, V$ el conjunto de predicciones ordenado de menor a mayor y $x_l^{(i)}$, con $i=1, \dots, V$ el conjunto de los valores a predecir asociados a cada uno de los valores $x_{pl}^{(i)}$ (nótese que $x_l^{(i)}$ no necesariamente guardará un orden ascendente), de esta forma construimos el estimador como,

$$\hat{g}(x_{pl}^{(i)}) = \frac{1}{\Delta} \sum_{i=1}^{i+\Delta-1} x_l^{(i)} \tag{2}$$

Esto nos lleva a una estimación de $g(\cdot)$ en los puntos $x_{pl}^{(i)}$, con $i=1, \dots, V$. Por interpolación podemos calcular $g(\cdot)$ en cualquier otro valor de su argumento x_{pl} .

El ancho y la forma de la ventana controla el grado de suavidad de las estimas de $g(\cdot)$.

Si disponemos de un valor V de muestras elevado, la técnica aplicada será la mejor aproximación de la media condicional, y en consecuencia será la referencia para determinar la calidad en cada caso del segundo método propuesto.

2.2. Estimación por aproximación polinomial:

Pretendemos encontrar una forma cerrada de solución para el estimador no lineal. En [ERT97][ERT00] se presenta una fórmula general para la media condicional (multidimensional), la cual puede ser utilizada para realizar una aproximación polinomial de la no linealidad.

A partir de esta expresión multidimensional nosotros particularizamos la fórmula propuesta para nuestro caso unidimensional, llegando a, (ver apéndice A).

$$g(x_{pl}) = \sum_{m=1}^{\infty} \frac{1}{m!} C_m(x_l, x_{pl}) H_m(x_{pl}) \tag{3}$$

En la expresión (3), denominamos $C_m(x_l, x_{pl})$ al cumulante cruzado definido de la forma,

$$C_m(x_l, x_{pl}) = cum \left(x_l, \overbrace{x_{pl}, \dots, x_{pl}}^{m \text{ veces}} \right) \tag{4}$$

y H_m es el polinomio de Hermite de orden m .

En la ecuación (3), tenemos que asumir que x_{pl} es una variable aleatoria Gaussiana de media cero y varianza unidad, así como estacionaridad para poder estimar el cumulante cruzado. El poder asumir Gaussianidad es una suposición en principio realista ya que, en todo caso, a la salida del predictor lineal los datos presentarán siempre una distribución más Gaussiana que a la entrada, al ser la salida del bloque lineal la combinación de N muestras anteriores de la secuencia original.

Si tenemos en cuenta este condicionante, las limitaciones respecto a la capacidad de aproximar la expresión teórica de la media condicional con la que se obtendría directamente se centrará entorno a:

- El grado de gaussianidad de los datos predichos a la salida del predictor lineal.
- El truncamiento realizado en la ecuación (3), en nuestra aplicación hemos considerado $m=3$.
- La estacionaridad asumida para estimar $C_m(x_l, x_{pl})$.

Si desarrollamos la ecuación (3), tal como podemos ver en el apéndice A, y realizamos por ejemplo una aproximación de tercer orden ($m=3$), obtenemos la expresión polinómica con la que caracterizamos la media condicional teórica.

$$g(x_{pl}) = \frac{1}{6}(a_3 - 3a_1)x_{pl}^3 + \frac{1}{2}a_2x_{pl}^2 + \left(\frac{5}{2}a_1 - \frac{1}{2}a_3\right)x_{pl} - \frac{1}{2}a_2 \quad (5)$$

donde $a_i = E[x_l x_{pl}^i]$ son los momentos cruzados que deben ser estimados a partir de los datos de entrenamiento.

Nótese que sustituyendo x_{pl} por su combinación lineal de los datos (predictor lineal) llegamos a una expresión que constituye un caso particular de serie de Volterra.

3. Evaluación de las técnicas de diseño:

A partir de los datos recogidos por el sensor infrarrojo pasivo en los 21 barridos considerados, determinaremos las mejoras en *SNR* al aplicar el esquema de la figura 2 en la predicción de los datos originales.

Para evaluar las técnicas propuestas, en cada uno de los tres bloques de información, en el diseño de la función no lineal sin memoria y las mejoras que produce este esquema no lineal frente al lineal definido en el Capítulo II, procedemos de la siguiente forma. A partir de los datos de entrenamiento estimamos por un lado el predictor lineal $H(NxD)$ (ecuación (2), Capítulo II) y por otro las D funciones no lineales sin memoria, por los dos métodos propuestos en el anterior apartado. Utilizaremos para el diseño en cada

bloque, los subbloques obtenidos a partir de los dos primeros acimuts. Y para medir las prestaciones (potencia media del error de predicción) los subbloques de acimuts tercero y cuarto. Adicionalmente, nos interesará determinar cuándo es o no conveniente introducir la no linealidad en la predicción, para ello deberíamos considerar la mejora en *SNR* al realizar la predicción no lineal, para cada una de las técnicas propuestas, frente a la predicción lineal.

Para establecer una adecuada comparación en diferentes instantes de predicción D , definimos la potencia del error cuadrático medio de predicción normalizado (NPEMP), ecuación (6), obtenido al pasar el bloque de testeo (correspondientes al tercer y cuarto acimut) sobre las estimaciones correspondientes sólo al predictor lineal o al predictor lineal más la no linealidad sin memoria obtenida mediante la aproximación polinómica de la media condicional, ecuación (5), o al predictor lineal más la no linealidad sin memoria obtenida por estimación directa de la media condicional.

$$NPEMP_1 = \frac{E[\|e_l\|^2]}{E[\|x_l\|^2]} = \frac{\text{Potencia error de predicción}}{\text{Potencia media de señal}} \quad (6)$$

donde e_l es el error de predicción correspondiente a cada caso.

$$e_l = x_l - x_{pl} \quad (\text{predictor lineal}) \quad (7)$$

$$e_l = x_l - g_1(x_{pl}) \quad (\text{predictor no lineal, estimación directa}) \quad (8)$$

$$e_l = x_l - g_2(x_{pl}) \quad (\text{predictor no lineal, aproximación polinómica}) \quad (9)$$

Con estas consideraciones, representamos para cada uno de los bloques seleccionados y caracterizados en el Capítulo II, por un lado la función de distribución definida en la figura 1, para longitudes del predictor: $N=1$, $N=2$, $N=4$, $N=6$ y $N=8$ y distancias de predicción l , variando de 1 a 6; a la que superponemos las curvas de estimación de la media condicional al aplicar las técnicas propuestas en la sección 2 de este Capítulo. Y por otro lado el valor normalizado de la potencia media del error predicción. El ancho de la ventana considerada en el método directo es 100.

3.1. Evaluación técnicas en el bloque de datos de información A:

3.1.1. Resultados obtenidos para el bloque A:

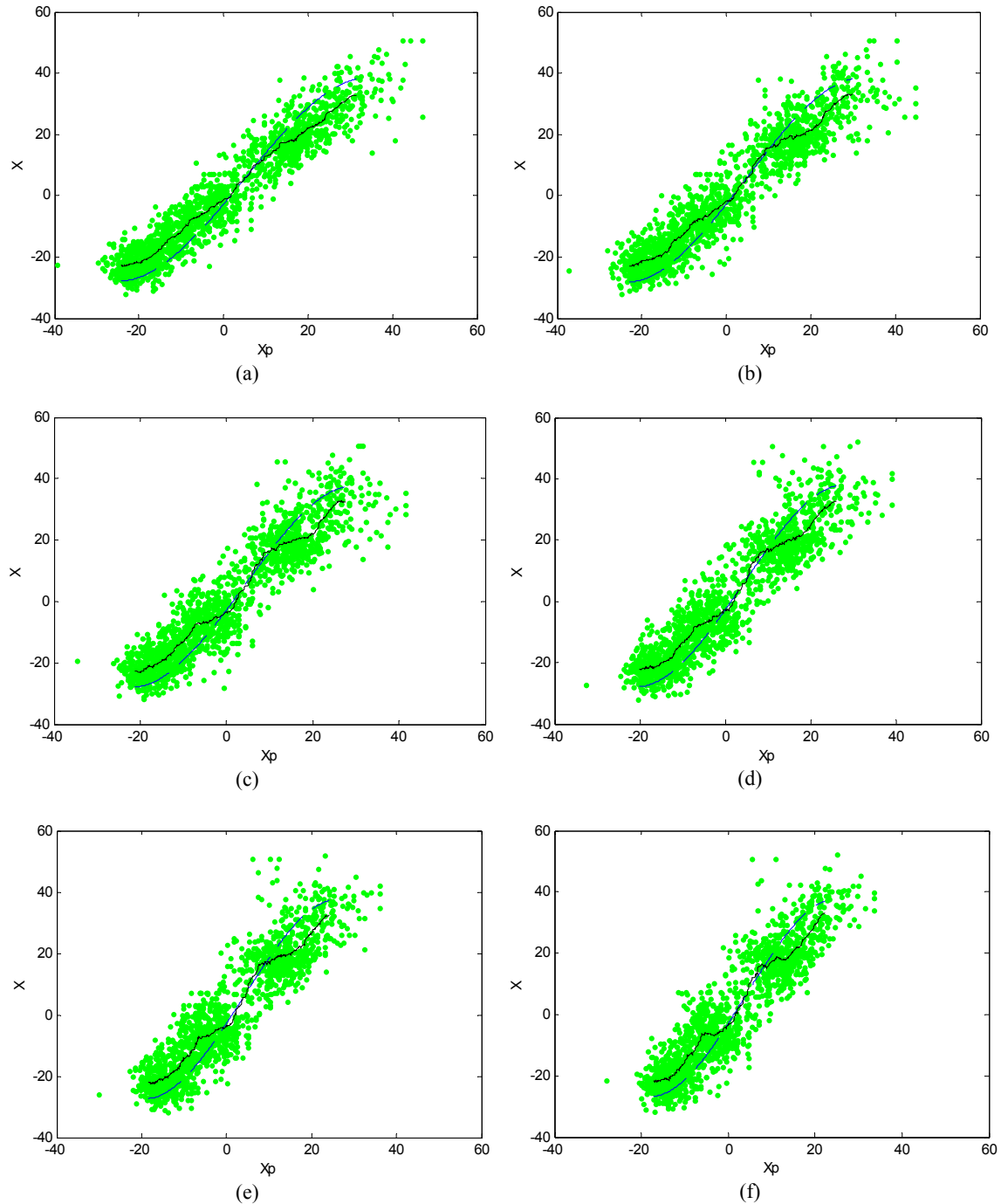


figura 3: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=1$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque A.**

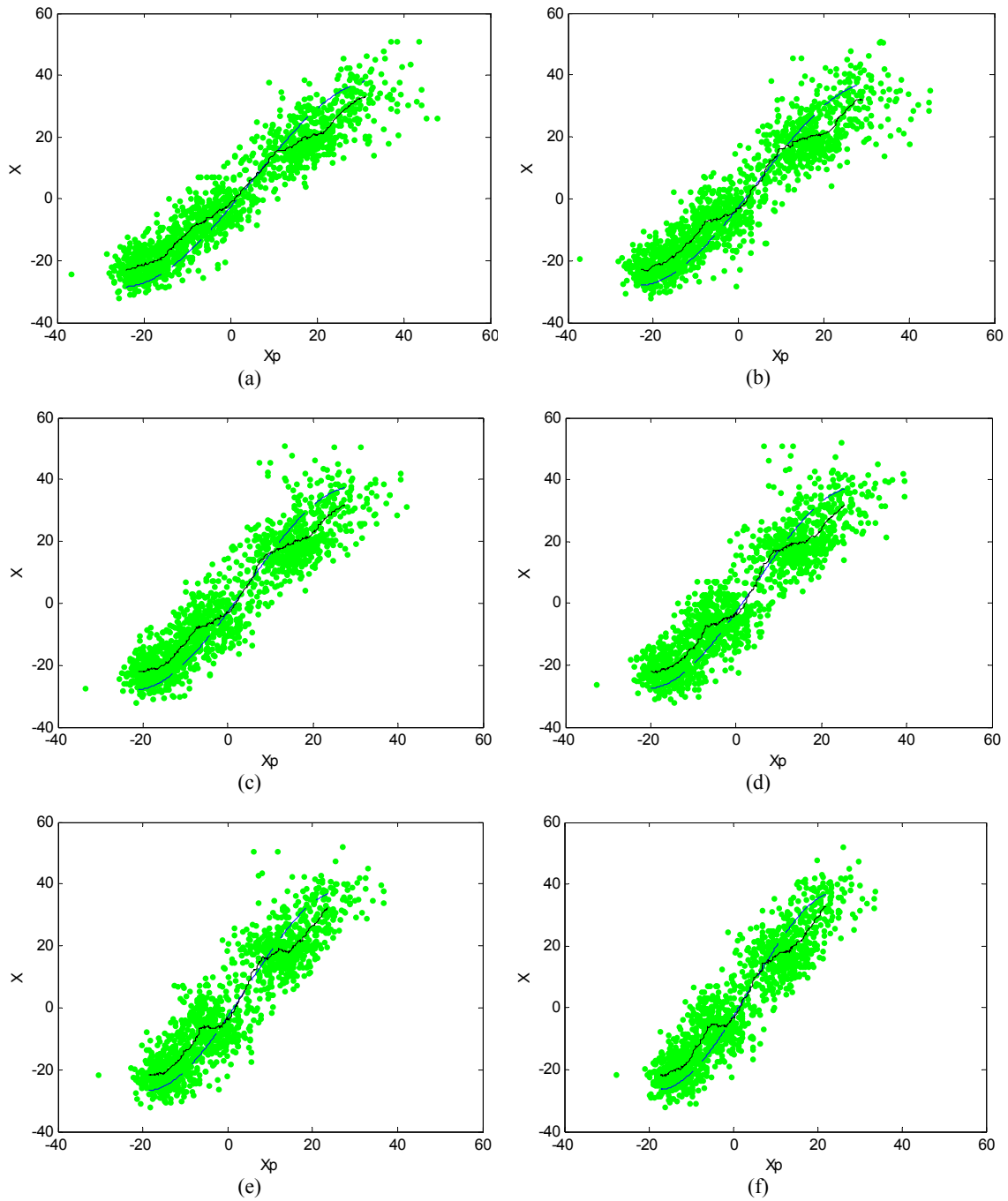


figura 4: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=2$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque A.**

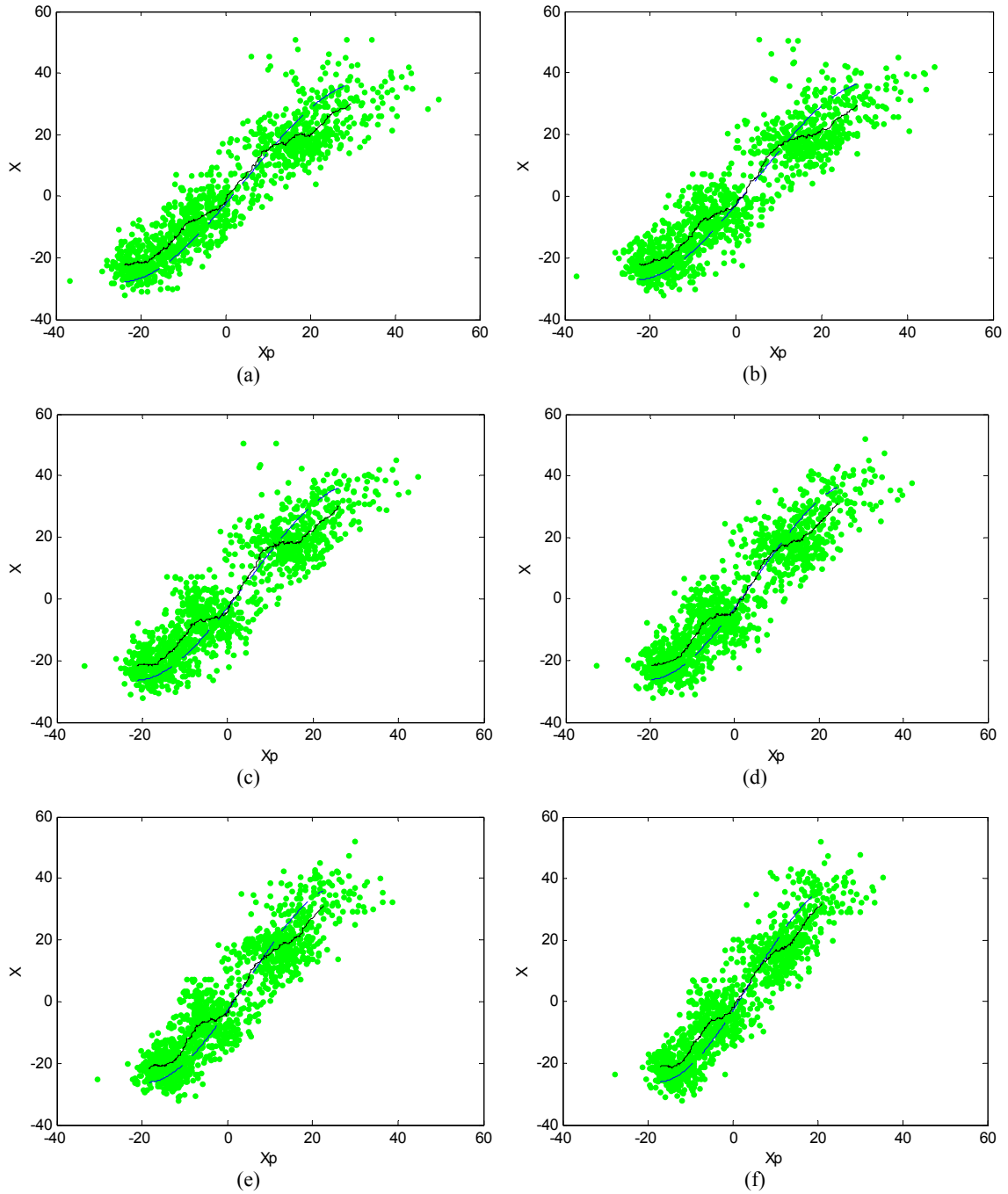


figura 5: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=4$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque A.**

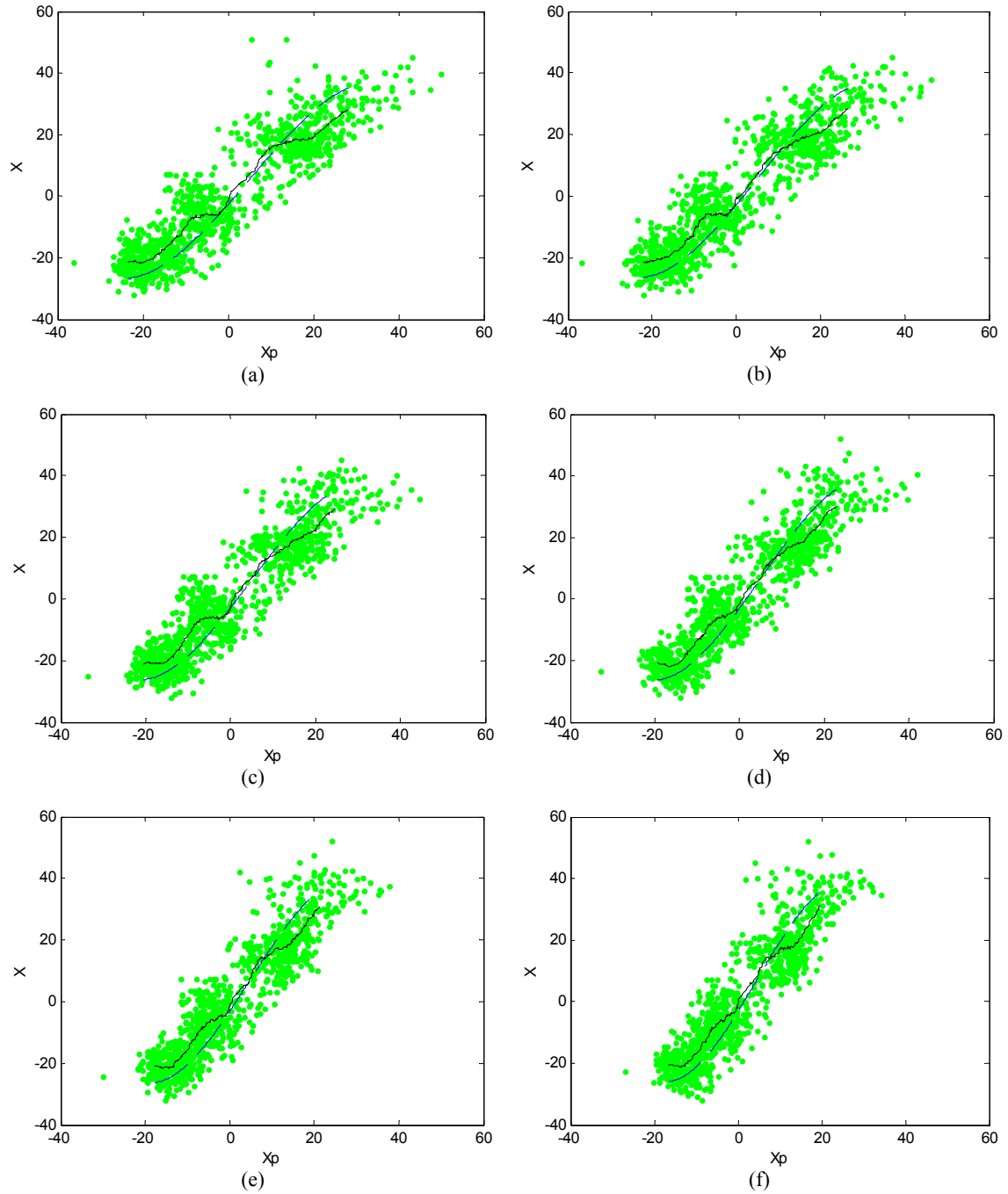


figura 6: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=6$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque A.**

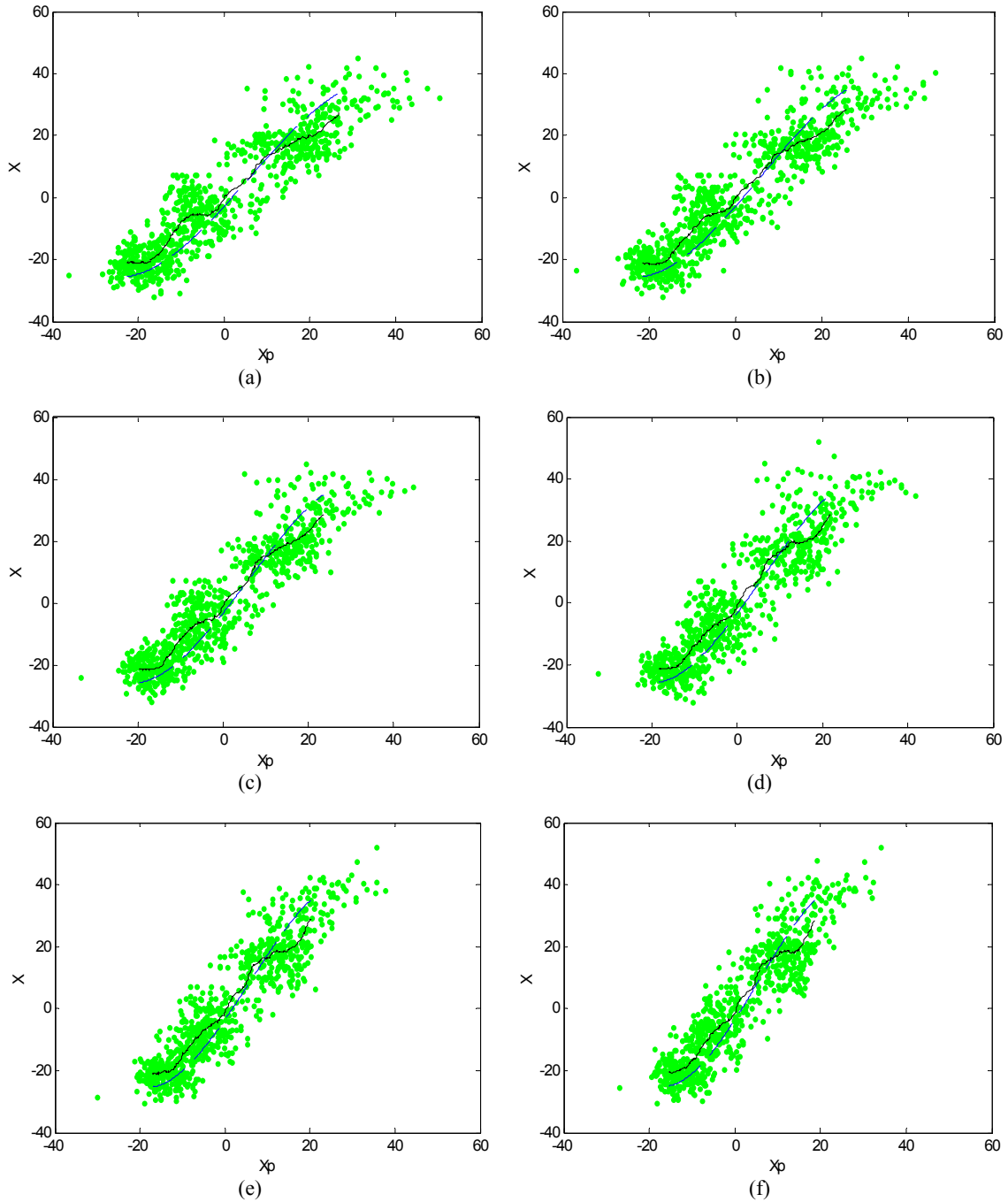


figura 7: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=8$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque A.**

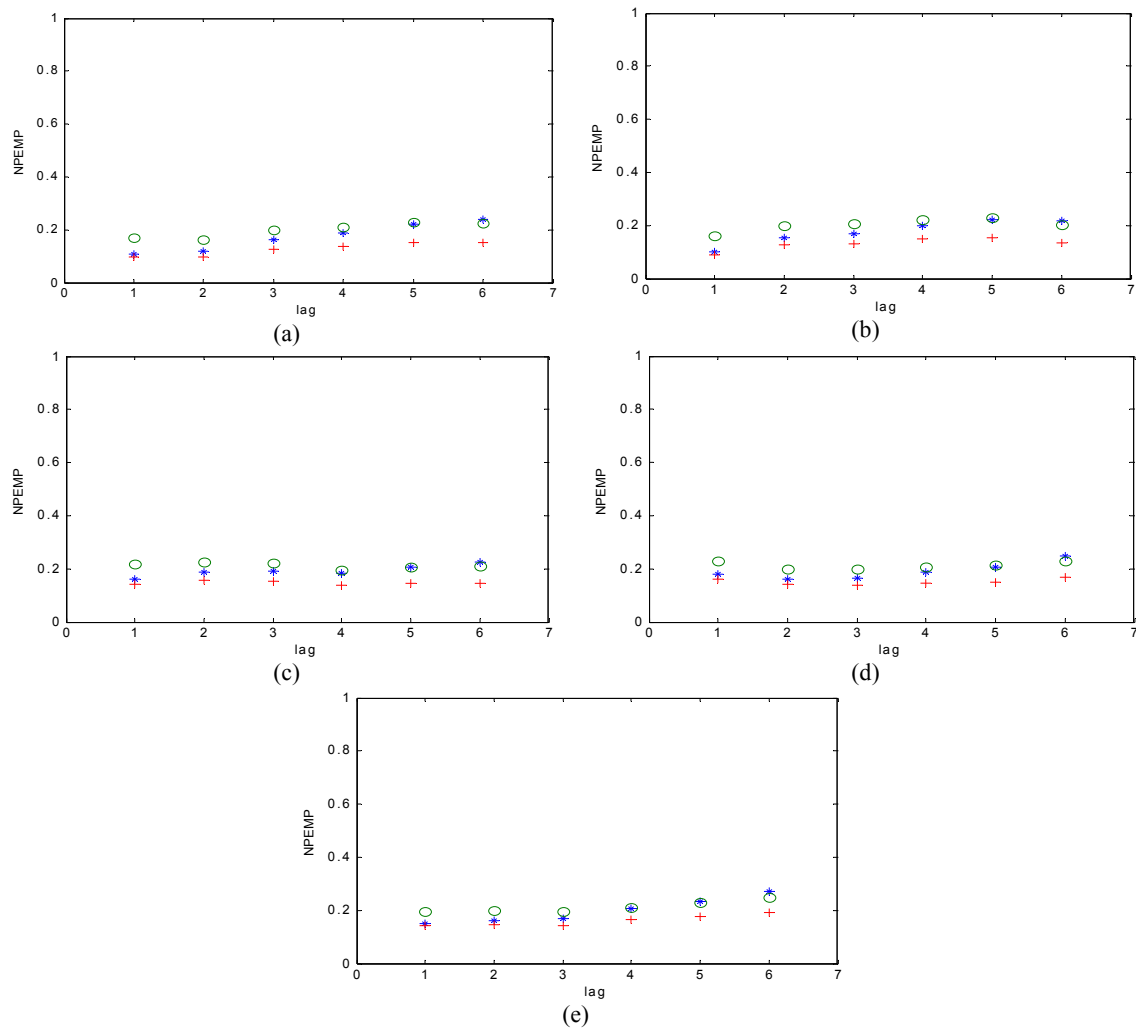


figura 8: Potencia de Error de Predicción Normalizado (**bloque A**) para diferentes valores del intervalo de predicción l (lag) y diferentes valores del orden del predictor (a) $N=1$; (b) $N=2$, (c) $N=4$, (d) $N=6$, (e) $N=8$.
 * Sólo el predictor lineal (NPEMP*)
 o Predictor lineal más la no linealidad sin memoria, aproximación por el polinomio de tercer orden de la media condicional (NPEMP^o).
 +Predictor lineal más la no linealidad sin memoria, estimación directa de la media condicional (NPEMP⁺).

3.1.2. Comentarios:

Como podemos observar en las figuras 3 a 7, al tener una distribución de puntos $f(x, x_p)$ bastante lineal, la inclusión de la no linealidad no nos aporta importantes ventajas (figura 8), a pesar de ser un bloque con una función densidad de probabilidad no Gaussiana. En realidad, como puede observarse en la figura 2a del Capítulo II, así como en la figura 3a del mismo Capítulo, este bloque presenta un comportamiento bimodal; para cada modo, el comportamiento es “localmente” Gaussiano y por ello pueden distinguirse con

claridad dos núcleos o “clusters” de puntos en las figuras 3 a 7 anteriores. En cada núcleo el ajuste es bastante lineal y con una pendiente parecida.

Realmente, si la correlación en barrido (la utilizada para el diseño del predictor lineal) es la misma salvo por un factor, de nivel, en ambos núcleos, el predictor lineal diseñado es el óptimo y similar al que obtendríamos haciendo un diseño separado para cada núcleo. Como éstos son “localmente” Gaussianos, no es de extrañar que el predictor lineal sea comparable al no lineal.

En la figura 8, observamos que de todos modos, la estimación directa (mejor aproximación de la media condicional) para los distintos valores de N y D es la mejor estimación posible de $f(x, x_p)$, (menor valor de NPEMP^+ frente al caso lineal, NPEMP^*). No obstante los valores de NPEMP^+ y NPEMP^* son muy parecidos, con lo que concluimos que la inclusión de la no linealidad no aporta ventajas sustanciales.

Para la aproximación polinómica, de orden $m=3$, ecuación (5), el valor de NPEMP^0 es en la mayoría de los casos incluso mayor que el obtenido al considerar únicamente la predicción lineal. Este hecho es debido a la no estricta Gaussianidad de las predicciones obtenidas al aplicar el predictor no lineal, con lo que se comete un error en la aproximación polinómica, ya que para la expresión cerrada suponíamos Gaussianidad en la predicción, ver apéndice A, que se aprecia en las diferentes gráficas de la figura 8. Por otra parte para conseguir una mejor aproximación (límite estimación directa) deberíamos aumentar el orden m de la predicción.

Por otro lado se observa que por el hecho de variar el orden del predictor N , el valor NPEMP , al considerar la no linealidad, apenas varía.

Finalmente comentar, que con independencia de la mayor o menor “Gaussianidad”, este es un bloque muy correlado, por lo que el predictor lineal consigue una fuerte reducción del ruido infrarrojo de fondo (factor 0.2 de reducción, véase la figura 8), que resulta difícil de mejorar con la no linealidad.

3.2. Evaluación técnicas en el bloque de datos de información **B**:

3.2.1. Resultados obtenidos para el **bloque B**:

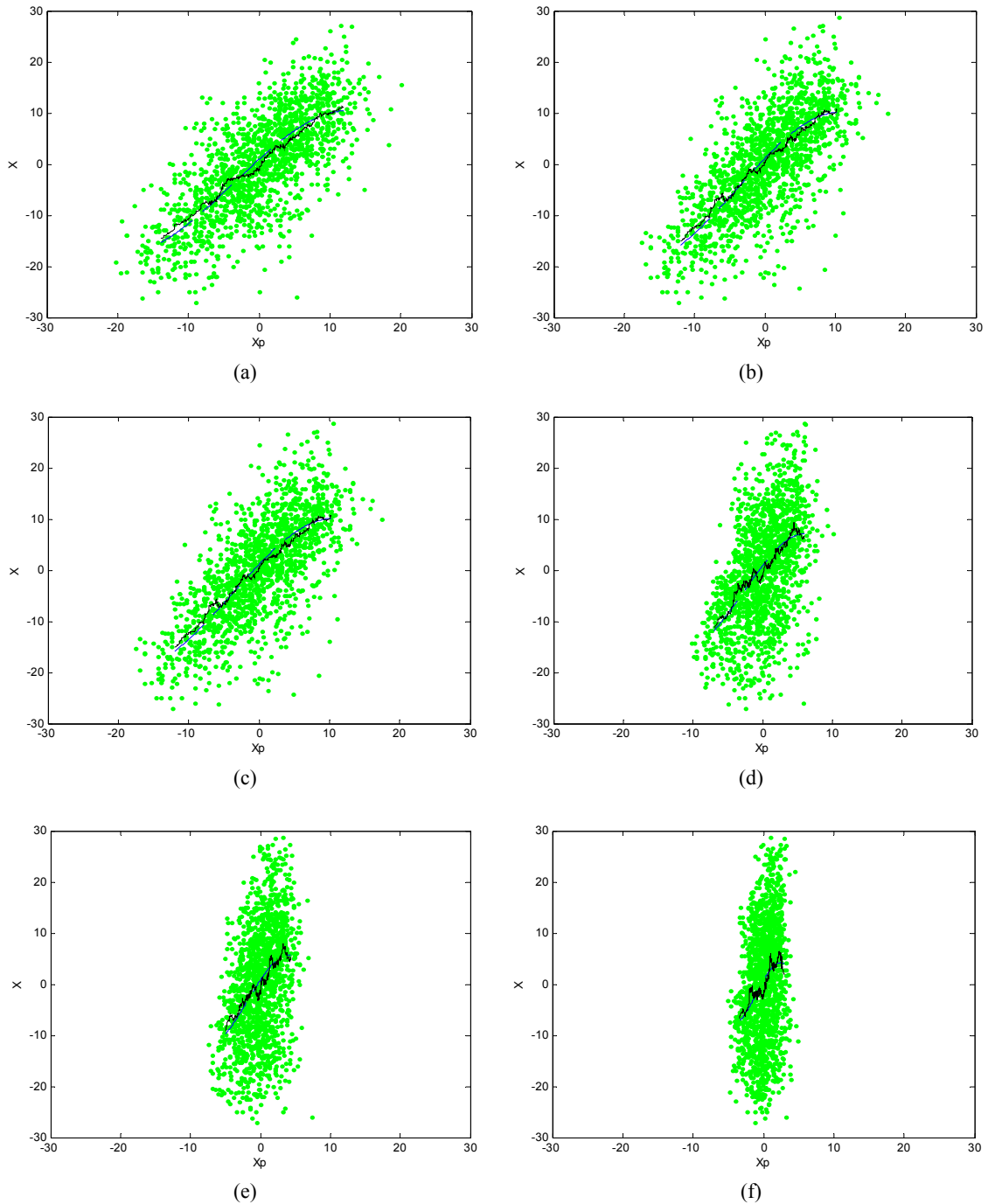


figura 9: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=1$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque B.**

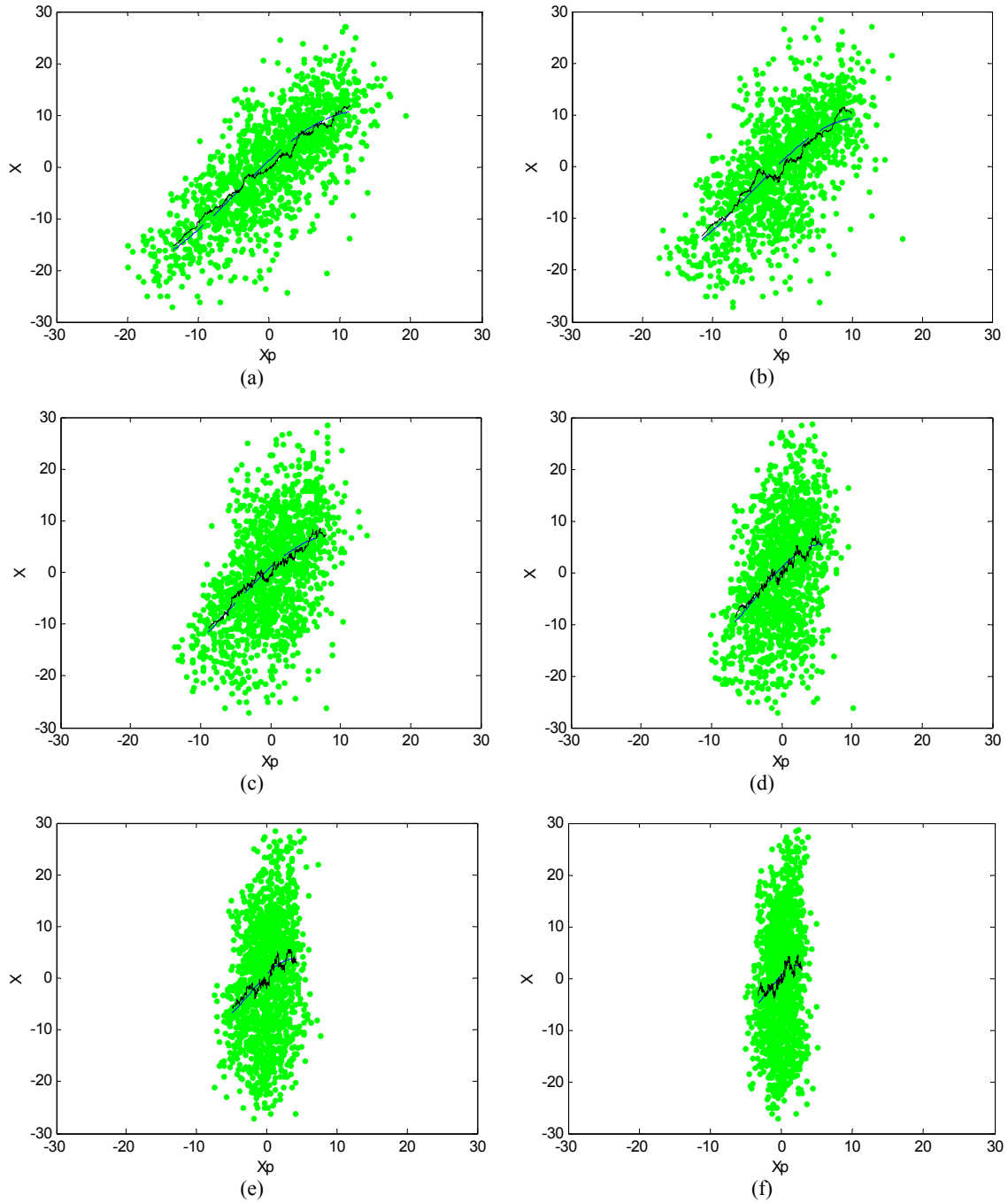


figura 10: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=2$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque B.**

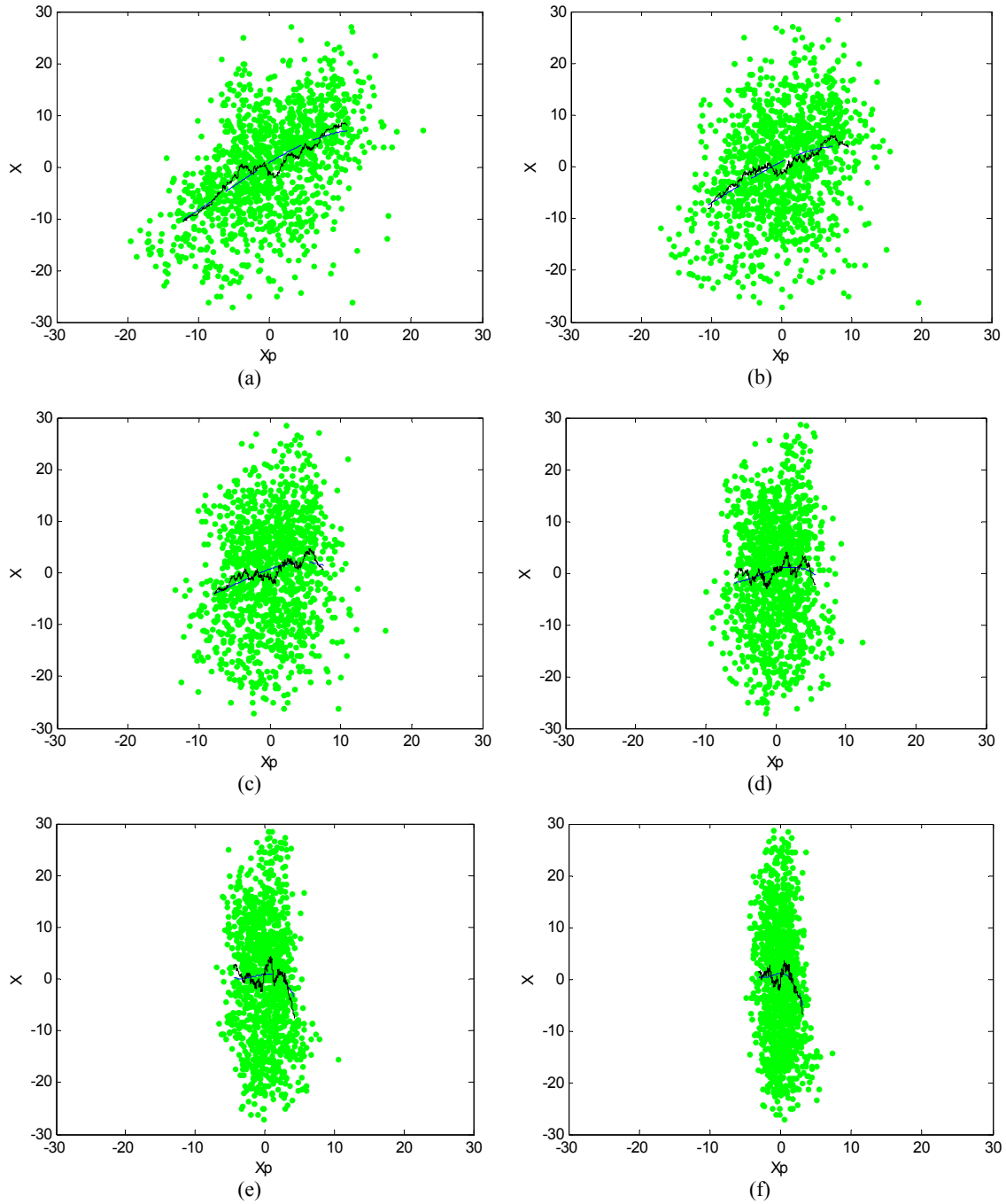


figura 11: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=4$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque B.**

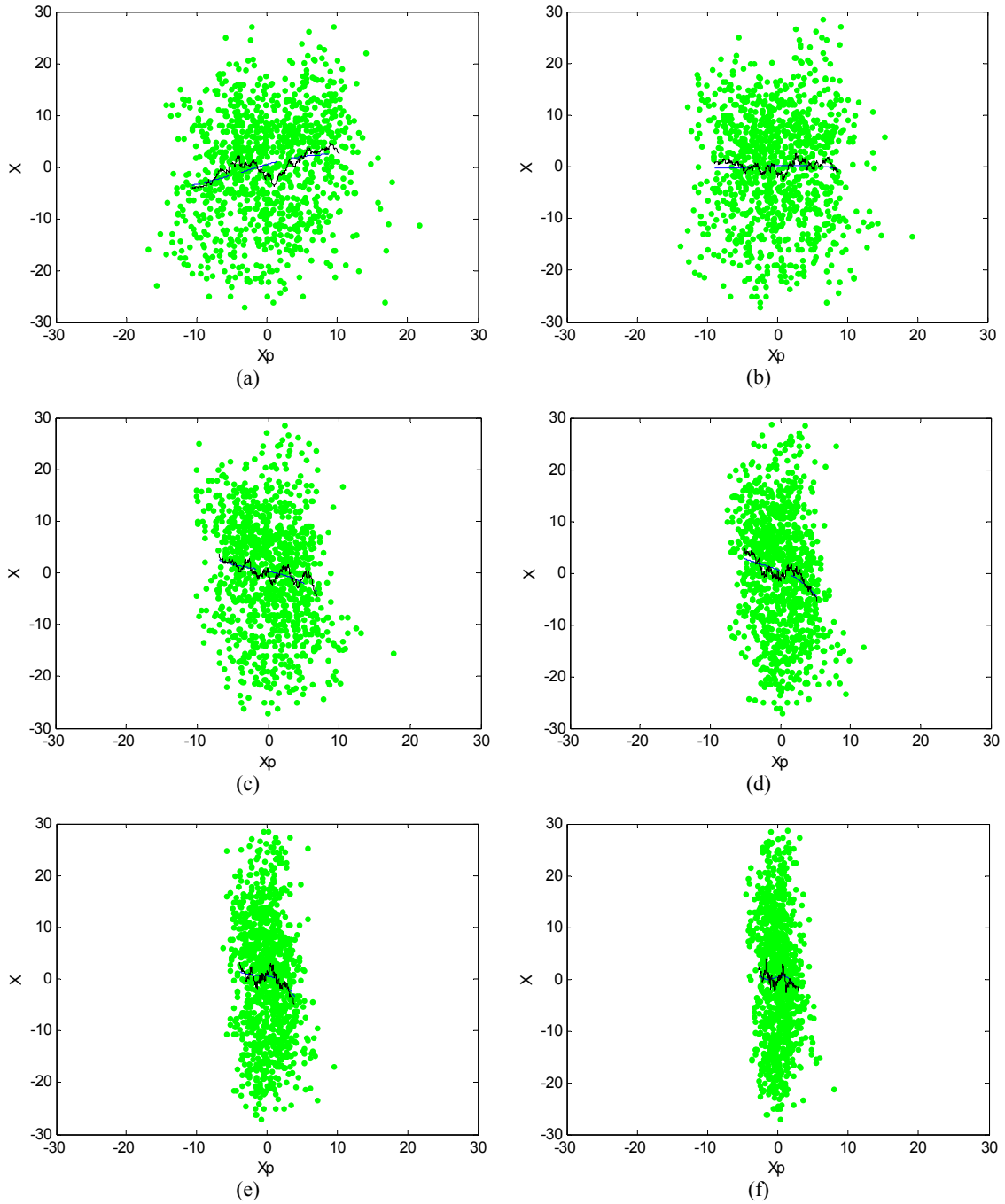


figura 12: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=6$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque B.**

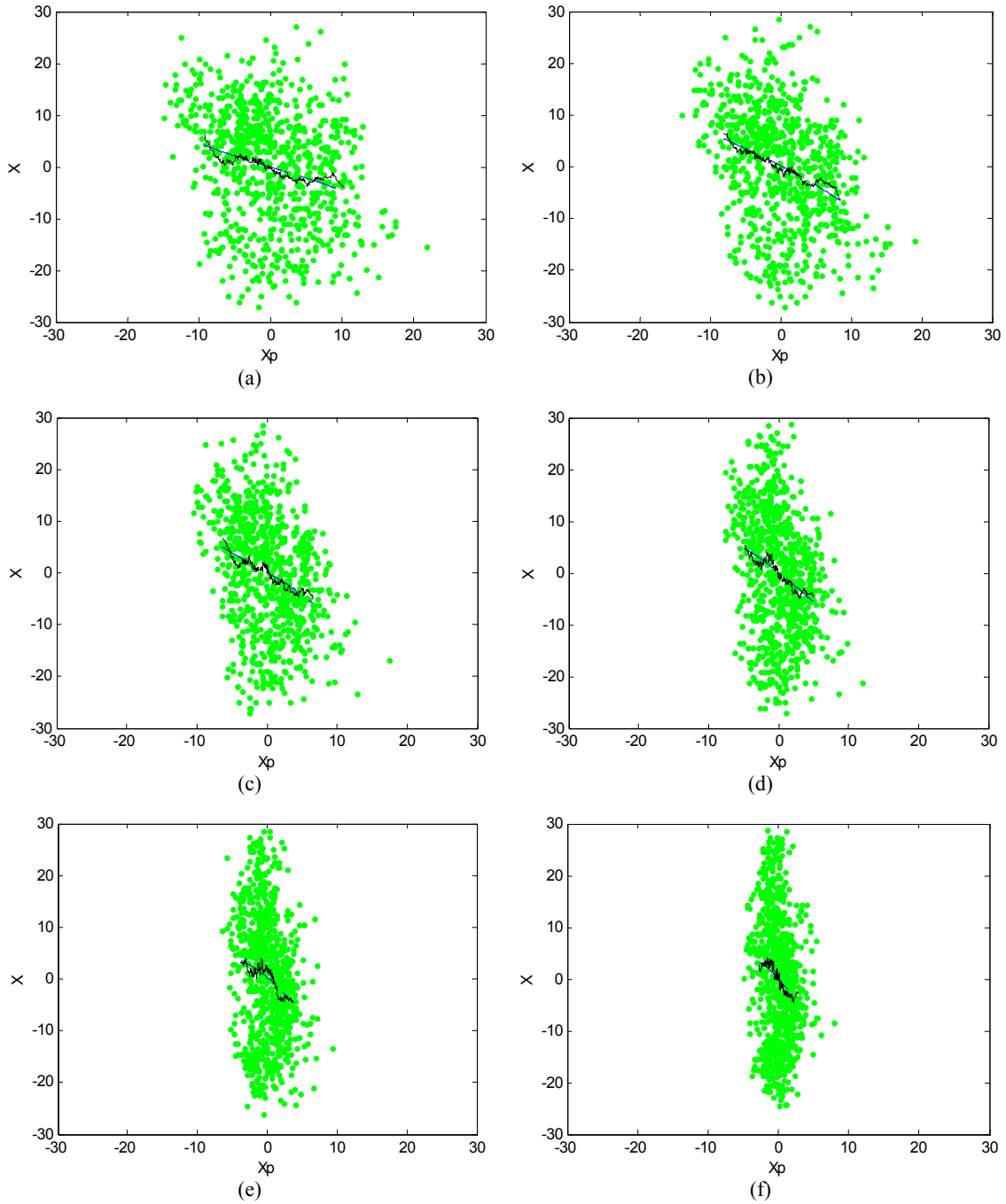


figura 13: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=8$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$.

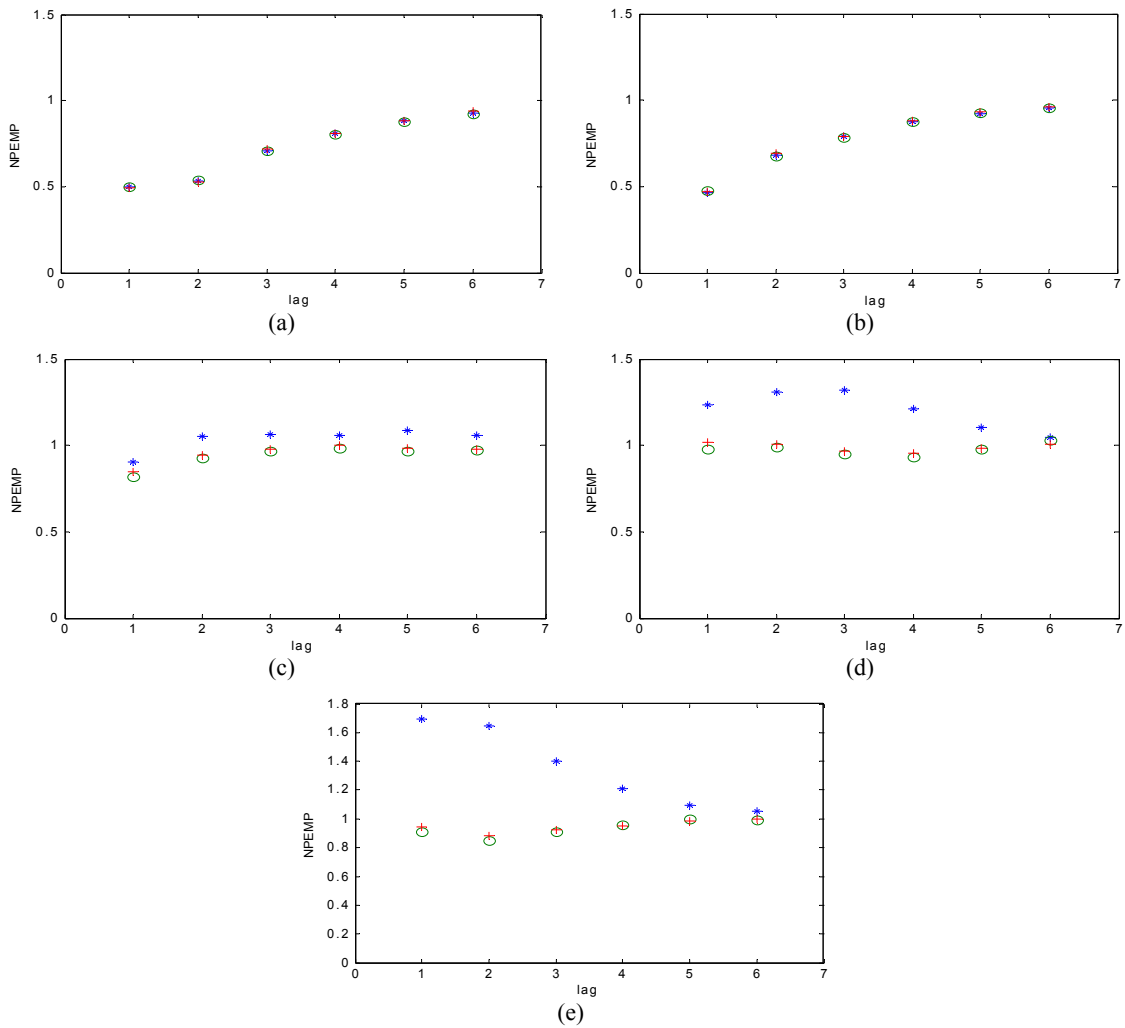


figura 14: Potencia de Error de Predicción Normalizado (**bloque B**) para diferentes valores del intervalo de predicción l (lag) y diferentes valores del orden del predictor (a) $N=1$; (b) $N=2$, (c) $N=4$, (d) $N=6$, (e) $N=8$.
 * Sólo el predictor lineal (NPEMP*)
 o Predictor lineal más la no linealidad sin memoria, aproximación por el polinomio de tercer orden de la media condicional (NPEMP^o).
 +Predictor lineal más la no linealidad sin memoria, estimación directa de la media condicional (NPEMP⁺).

3.2.2. Comentarios:

Este es un bloque con comportamiento Gaussiano por lo que en principio la inclusión de la no linealidad no debería introducir mejoras con respecto al predictor lineal. No obstante al ser un bloque con baja correlación en n , se producen ciertos fenómenos que pueden hacer interesante incluir la no linealidad.

Así observamos en las figuras 9 a 13 como la relación lineal entre x y x_p es evidente, pero cuando l sube de valor, la baja correlación existente tiende a hacer muy pequeño el valor x_p , comprimiendo la nube de puntos verticalmente. Además para $N > 3$ se observa que con l pequeños la correlación se difumina y las nubes de puntos se dispersan. Todo ello lleva a las medidas de error de predicción de la figura 14, donde sólo para $N=1$ ó $N=2$ y $l < 5$ se obtienen reducciones del ruido infrarrojo significativas (factores de reducción de 0.5 a 0.7) iguales para los tres tipos de predictores como cabía esperar.

Si N aumenta, el error de predicción se dispara y para $l < 4$, se produce incluso una amplificación del mismo si usamos sólo el predictor lineal. Con los predictores no lineales, sin embargo “contenemos” esta amplificación. Dado que el orden N puede eventualmente elegirse a partir de otros criterios, la inclusión de las no linealidades nos garantiza una mayor robustez del sistema.

Por último indicar que el comportamiento del predictor no lineal basado en la aproximación polinómica es prácticamente idéntico al basado en el método directo. Nótese que en este caso x_p es Gaussiano al ser el bloque Gaussiano y la hipótesis de Gaussianidad necesaria para la estricta validez de la ecuación (5) se cumple. La truncación en $m=3$ no supone en este caso pérdida de calidad.

3.3. Evaluación técnicas en el bloque de datos de información C:

3.3.1. Resultados obtenidos para el **bloque C**:

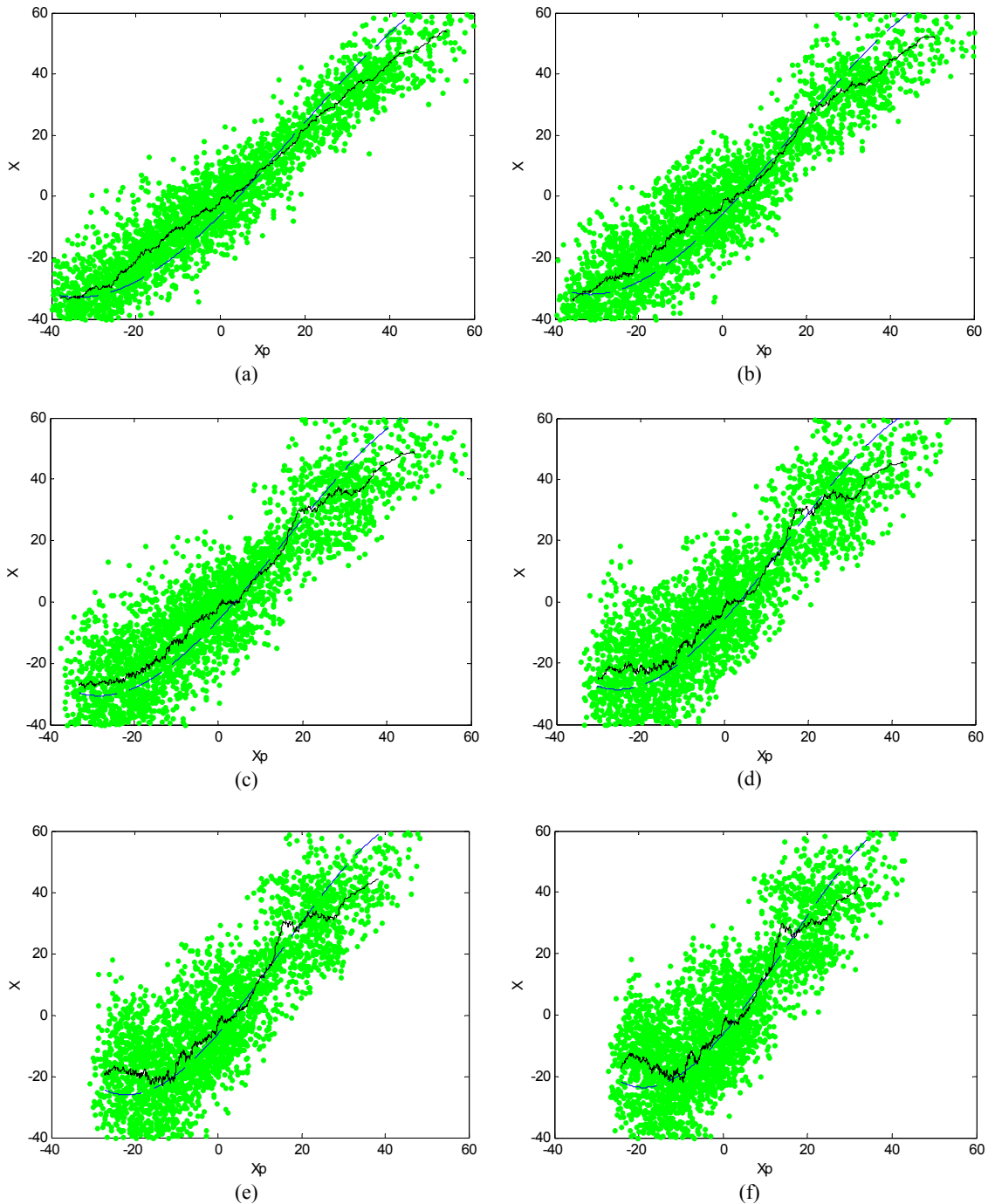


figura 15: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=1$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque C.**

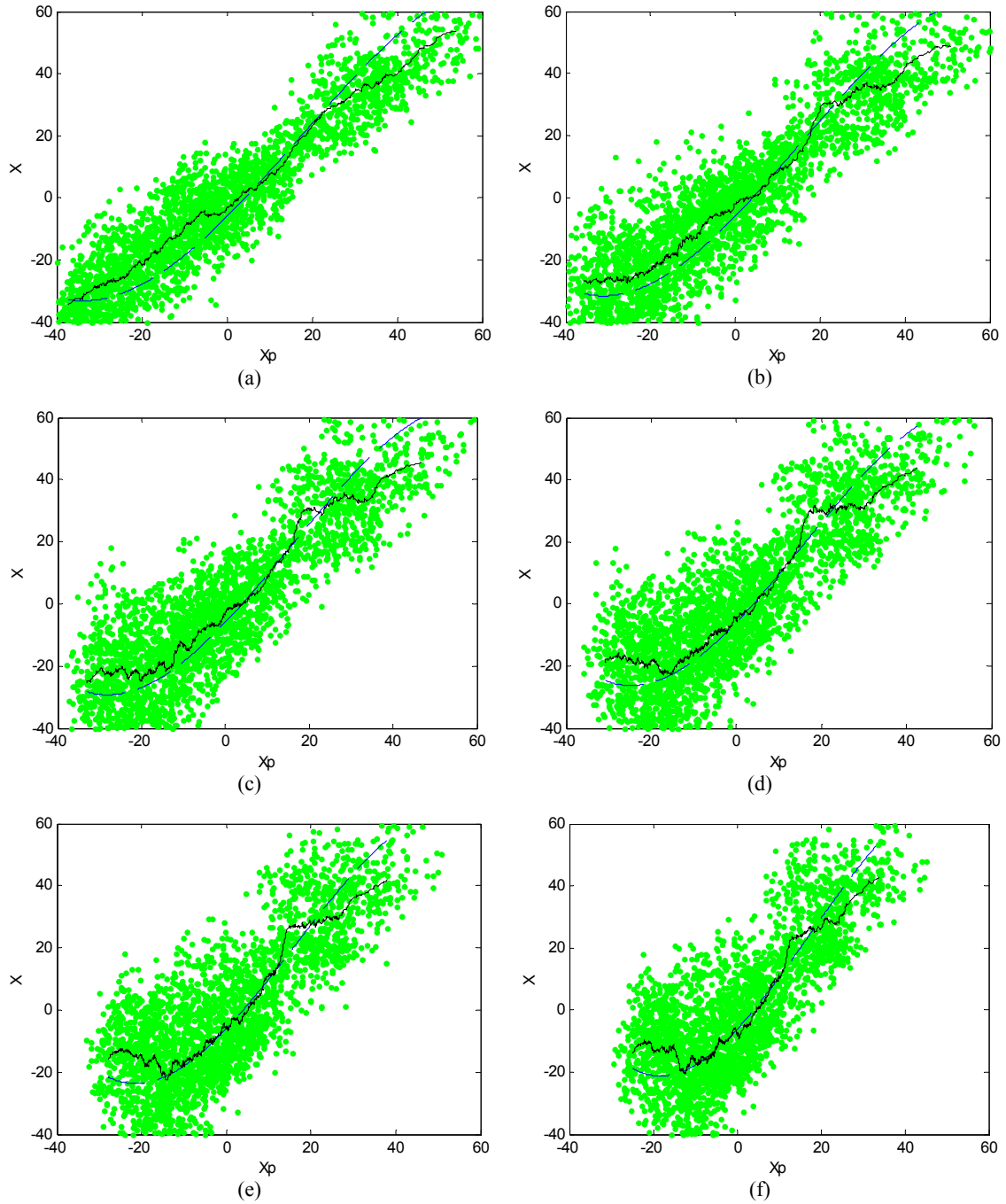


figura 16: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=2$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque C.**

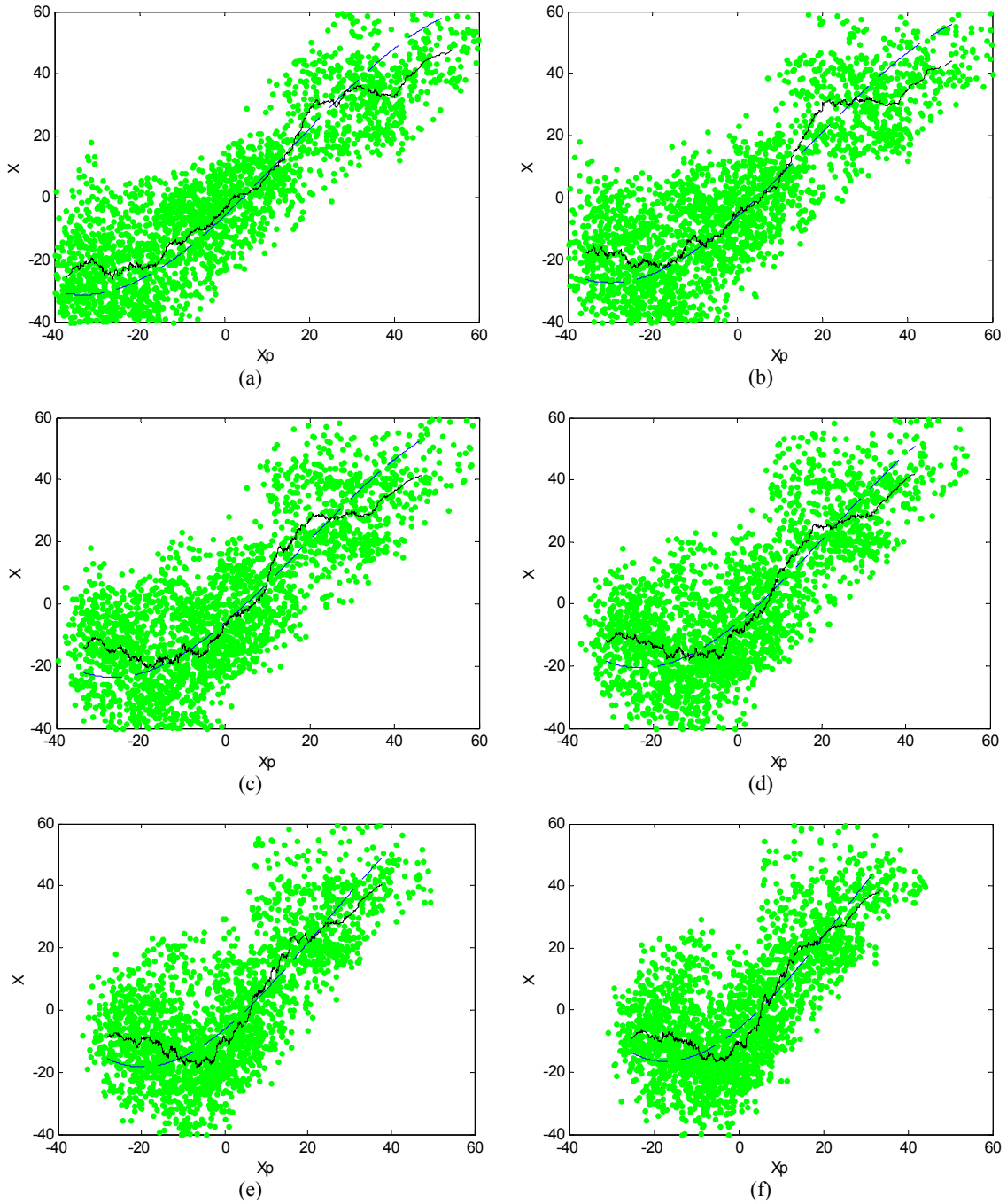


figura 17: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=4$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque C.**

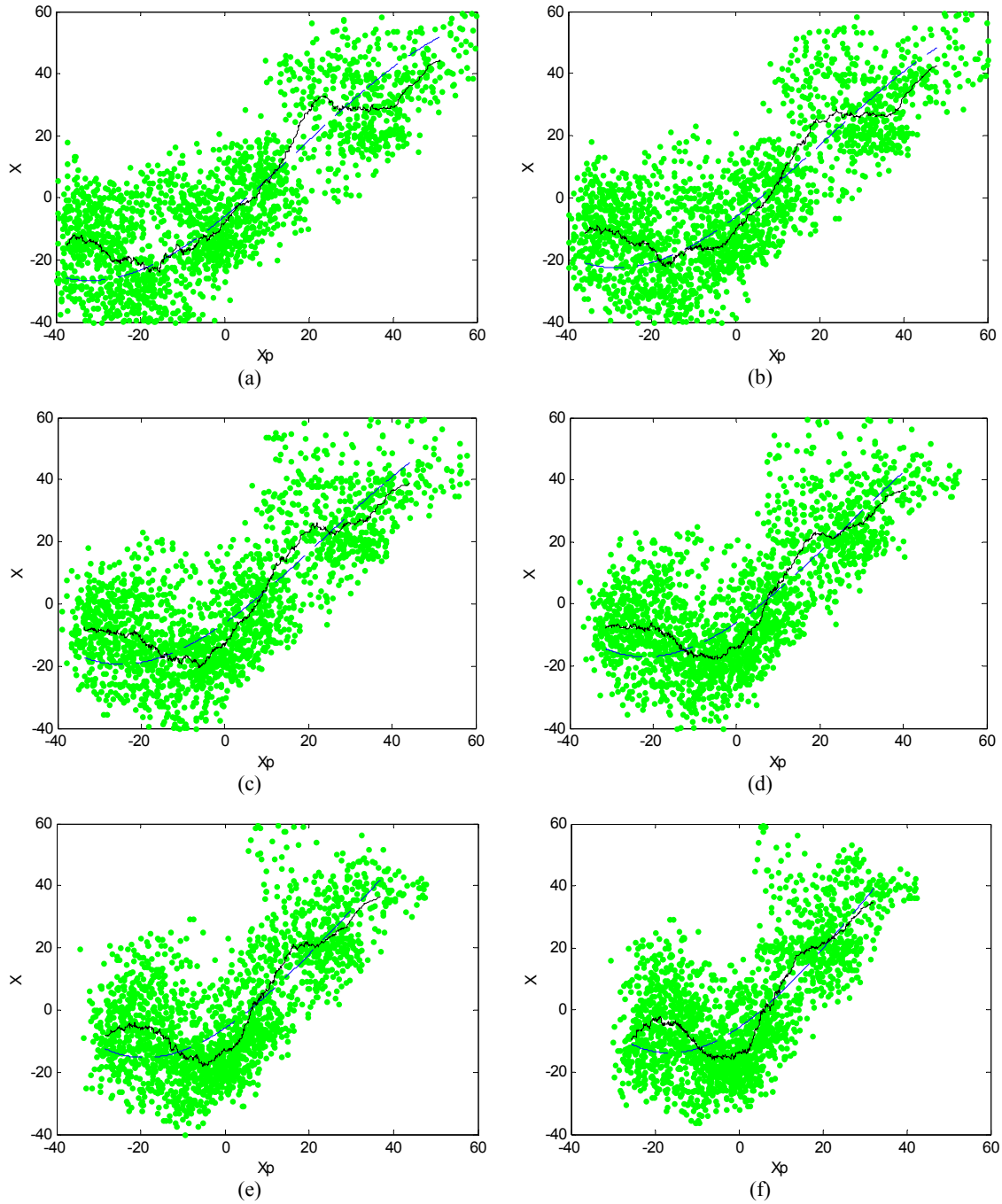


figura 18: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=6$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque C.**

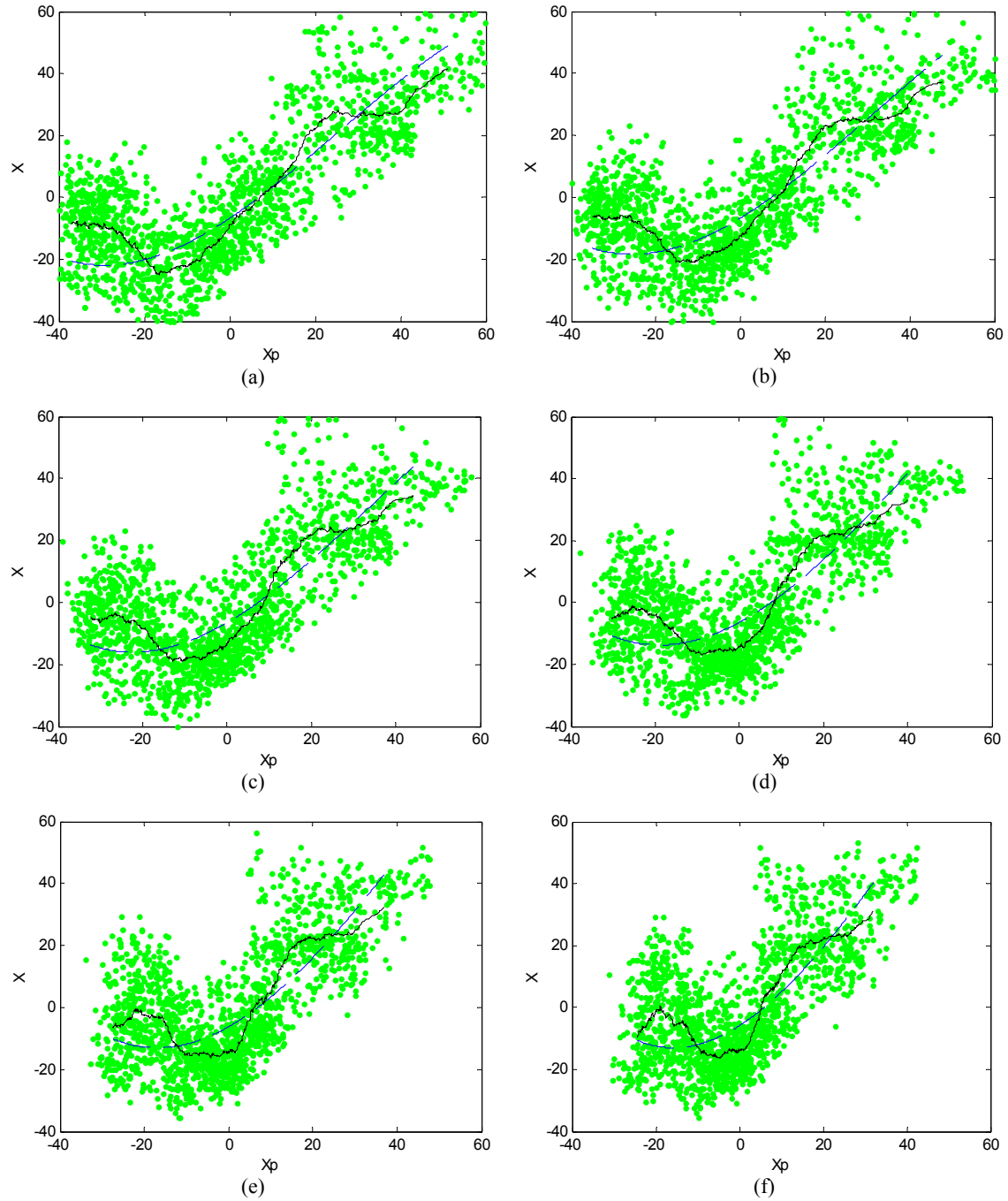


figura 19: Nube de puntos $f(x, x_p)$ para los datos de entrenamiento. Superpuestas, la función no lineal sin memoria: Método directo (línea sólida negra), método teórico (línea de trazos azul). Para $N=8$ y diferentes valores de: (a) $l=1$, (b) $l=2$, (c) $l=3$, (d) $l=4$, (e) $l=5$, (f) $l=6$. **Bloque C.**

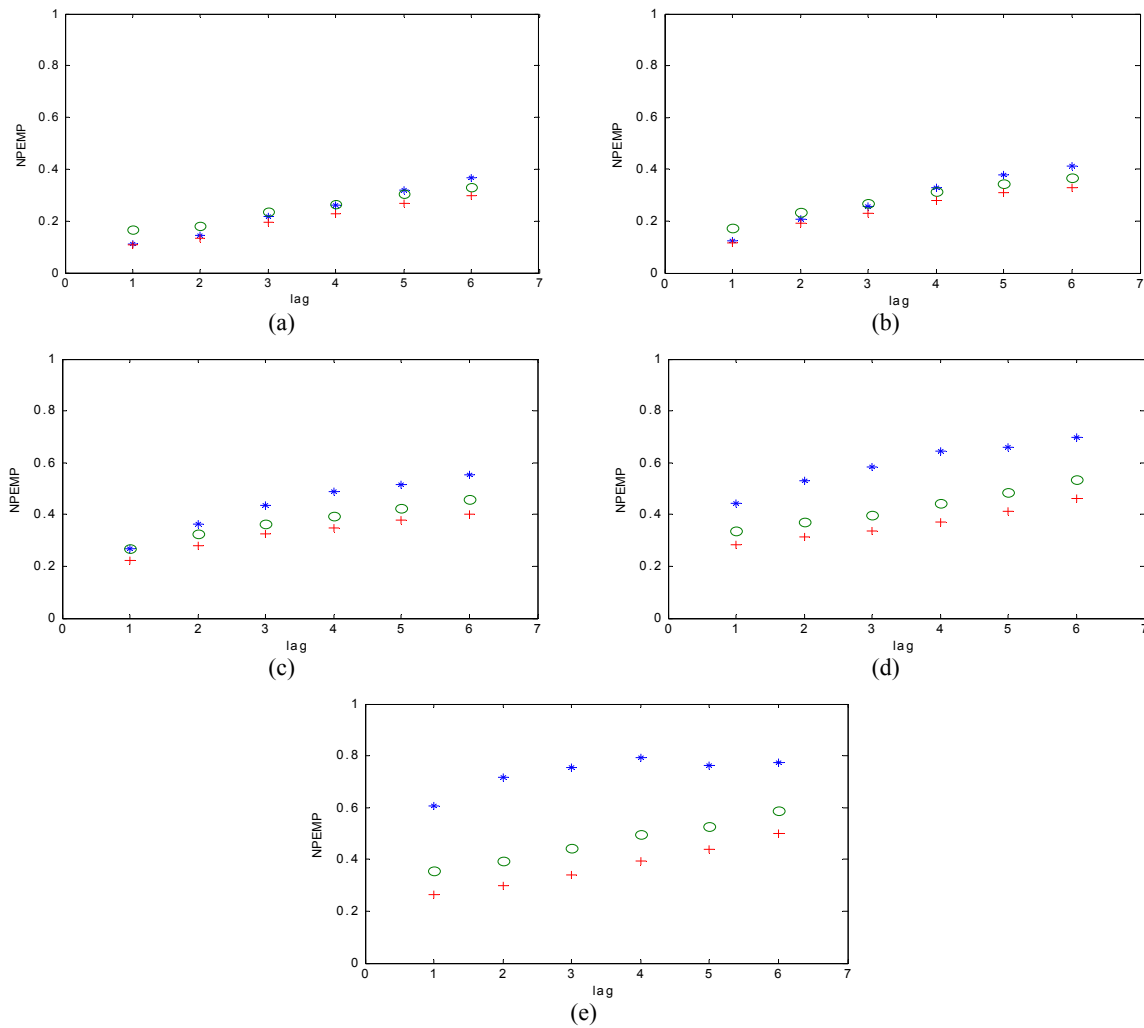


figura 20: Potencia de Error de Predicción Normalizado (**bloque C**) para diferentes valores del intervalo de predicción l (lag) y diferentes valores del orden del predictor (a) $N=1$; (b) $N=2$, (c) $N=4$, (d) $N=6$, (e) $N=8$, * Sólo el predictor lineal (NPEMP*) o Predictor lineal más la no linealidad sin memoria, aproximación por el polinomio de tercer orden de la media condicional (NPEMP^o) +Predictor lineal más la no linealidad sin memoria, estimación directa de la media condicional (NPEMP⁺).

3.3.2. Comentarios:

Se trata de un bloque que nos determina a priori la viabilidad de entrenar con bloques de información mayores, de esta forma minimizamos las zonas de entrenamiento a considerar para caracterizar las diferentes áreas bajo vigilancia.

Las relaciones de carácter no lineal son lo bastante evidentes en las figuras 15 a 19, en especial a medida que N y/o l aumentan. En relación a la figura 20 se

observa que en todos los casos la estimación de la función no lineal de memoria cero obtenida al aplicar la ecuación (5), toma valores de NPEMP⁰ intermedios entre el caso lineal (NPEMP^{*}) y la estimación de las no linealidades de forma directa (NPEMP⁺).

Como podemos observar en las diferentes gráficas de la figura 20, al aumentar N , la presencia de la no linealidad provoca nuevamente una “contención” del error de predicción con respecto al caso del predictor lineal.

En definitiva la presencia de la no linealidad hace al sistema más robusto.

Nótese finalmente que los valores NPEMP en este bloque son en general intermedios entre los obtenidos en los **bloques A** y **B** (figuras 8, 19 y 20).

4. Experimentos sobre el esquema completo:

En este apartado vamos a considerar la inclusión de la no linealidad sobre el esquema completo presentado en el Capítulo II.

Hasta ahora hemos verificado la mayor o menor reducción en el nivel de ruido infrarrojo conseguida mediante la predicción no lineal pero el interés final nos lo dará el comportamiento del detector, por lo que resulta muy importante considerar el esquema completo.

Analizamos sucesivamente las funciones representadas en la figuras 3, 4, y 5 del Capítulo II, para analizar la influencia del nuevo esquema en los diferentes aspectos de interés:

- Grado de decorrelación del error de predicción.
- Funciones densidad de probabilidad tanto de las predicciones x_p como del error de predicción.
- Capacidad de ajuste de la PFA.

- Probabilidad de detección de fuegos descontrolados y alarmas indeseadas con y sin inclusión del filtro de subespacio P .

El esquema de la figura 21 ha sido aplicado a los datos infrarrojos obtenidos por el sensor, para cada uno de los tres diferentes bloques de información bajo análisis.

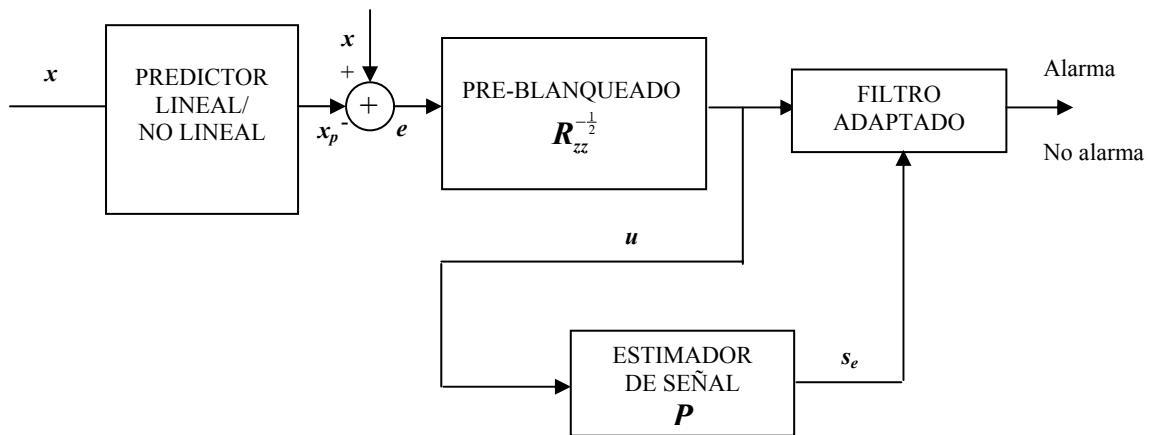


figura 21: Esquema de detección usando predicción no lineal

Asumiendo comportamiento estacionario dentro de cada uno de los bloques considerados, vamos a determinar las mejoras en la detección por el hecho de introducir no linealidades en el paso de predicción y comparar los resultados con el planteamiento lineal tratado en el Capítulo II.

Hemos elegido para estos experimentos los valores de longitud del predictor lineal $N=1, 2, 4, 6$ y distancia máxima de predicción $D = 6$.

4.1. Análisis de la autocorrelación:

En las figuras 22, 23, 24, se muestran los efectos de decorrelación del predictor sobre para cada uno de los datos de entrenamiento para los **bloques A, B y C** respectivamente. Lo que estamos aquí representado es la primera fila de la matriz de autocorrelación R_{zz} estimada (siendo z el ruido residual $e=s+z$, en ausencia de señal $e=z$) estimada a partir del error de predicción obtenido al aplicar sólo el predictor lineal, o al aplicar el predictor lineal más la no linealidad aproximada, ecuación (5),

o al aplicar la linealidad más la no linealidad evaluada por el método directo, con $e = [e_1, e_2, \dots, e_D]^T$, siendo e_l la variable aleatoria error de predicción a distancia l obtenido a partir de las ecuaciones (7 a 9).

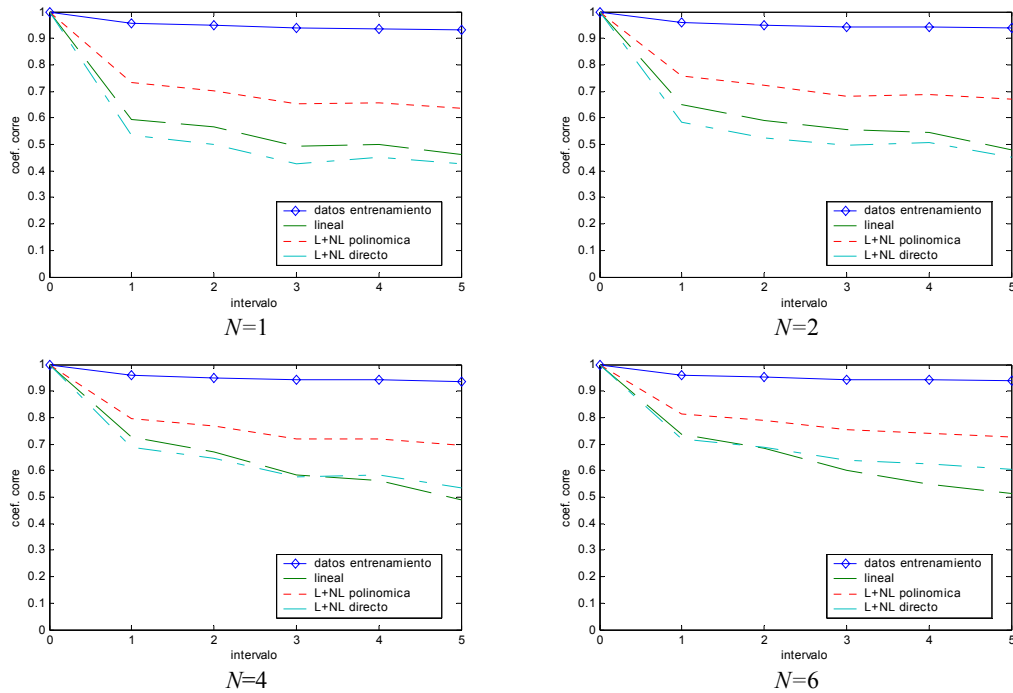


figura 22: Estimación de la matriz de autocorrelación para el bloque A

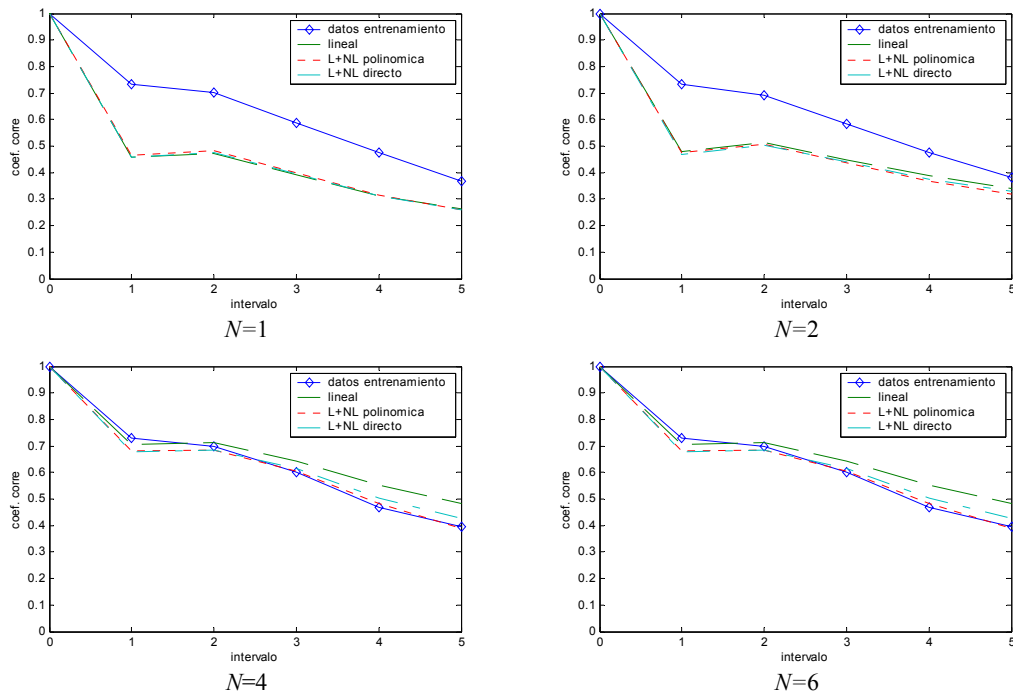


figura 23: Estimación de la matriz de autocorrelación para el bloque B

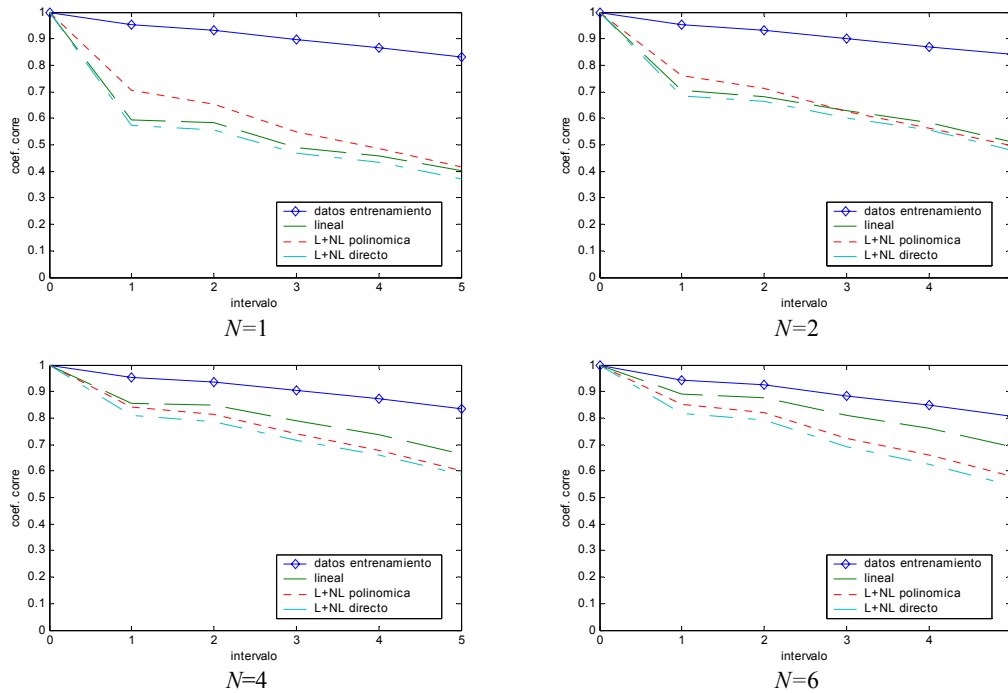


figura 24: Estimación de la matriz de autocorrelación para el **bloque C**

Si disponemos de V vectores e para la calibración, podemos estimar la matriz de autocorrelación,

$$R_{zz} = \frac{1}{V} \sum_{i=1}^V e_i e_i^T \tag{10}$$

La estimación obtenida tendrá mayor o menor varianza según el número de vectores (V) de D elementos obtenidos al realizar la predicción, en todo caso será insesgada.

En las figuras 22, 23 y 24 se puede observar la mayor decorrelación del error de predicción con respecto a los datos originales usados para entrenar, al representar la autocorrelación de los datos originales. Para estimar la matriz de autocorrelación de los datos originales hemos procedido de idéntica forma definida a la definida en la ecuación (10).

$$R_{xx} = \frac{1}{V} \sum_{i=1}^V x_i x_i^T \tag{11}$$

donde x_i son los V vectores de entrada al predictor disponibles para la calibración.

En la figura 22, **bloque A**, observamos la elevada correlación de los datos de entrenamiento de este bloque, también podemos ver que para la aproximación polinómica la decorrelación obtenida es menor, probablemente debido a que no se dan las condiciones de Gaussianidad necesarias para la validez de la ecuación (5).

En la figura 23, **bloque B**, observamos el mayor grado de incorrelación de los datos de entrenamiento, así como, al ser un bloque de información con función densidad de probabilidades Gaussiana, la estima de la autocorrelación para el predictor no lineal obtenida la no linealidad a partir de la ecuación (5), es prácticamente idéntica a la obtenida por el método directo. Para los casos $N=4$ y $N=6$ apenas se modifica la correlación de los datos originales, ya que las predicciones son de valor muy pequeño por la fuerte decorrelación existente y las secuencias de error de predicción coinciden prácticamente con los propios datos a predecir.

En la figura 24, **bloque C**, podemos una vez más observar, una situación intermedia con respecto a los bloques anteriores.

Los efectos de la decorrelación conseguidos se manifiestan en el valor mayor que 1 obtenido por *GDECOR* (parámetro definido en el Capítulo II, sección 5) tal como podemos observar en las Tablas I a III. En ellas se especifica para cada bloque el valor del parámetro “ganancia de decorrelación” para la predicción lineal y para las predicciones no lineales tras aplicar cada una de las técnicas propuestas, para distintos valores del orden del predictor y $D=6$.

Para el **bloque A**, (tabla I) se consigue mayor *GDECOR* en el sistema de Wiener calculada la no linealidad mediante el método directo, siendo esta diferencia apreciable, a medida que N disminuye, lo que determina el interés en este caso de incluir no linealidades en la predicción. Por otro lado debido a la no gaussianidad de los datos de entrada se consiguen *GDECOR* para la aproximación polinómica de la

no linealidad menores que para el caso lineal, lo que implicaría tener que aumentar a priori el orden de la aproximación.

BLOQUE A ($D=6$)	$GDECOR_{LINEAL}$	$GDECOR_{NL+POL}$	$GDECOR_{NL+DIR}$
	$N=1$	2.08	1.67
$N=2$	2.01	1.62	2.81
$N=4$	1.87	1.52	2.02
$N=6$	1.83	1.47	1.91

Tabla I: $GDECOR$ correspondiente al **bloque A**.

BLOQUE B ($D=6$)	$GDECOR_{LINEAL}$	$GDECOR_{NL+POL}$	$GDECOR_{NL+DIR}$
	$N=1$	1.38	1.41
$N=2$	1.27	1.28	1.29
$N=4$	0.92	1.03	1.02
$N=6$	0.89	1.04	1.01

Tabla II: $GDECOR$ correspondiente al **Bloque B**

Para el **bloque B**, (tabla II) apenas mejoramos la $GDECOR$ por el hecho de incluir la no linealidad. Por otro lado, y debido a la baja correlación como hemos indicado con anterioridad el predictor lineal apenas es capaz de predecir adecuadamente el valor de las muestras a cierta distancia l , hecho que es apreciable a medida que D y N aumentan. También este parámetro nos indica cómo la presencia de la no linealidad en este tipo de bloques de información provoca $GDECOR$ siempre mayores que 1, así como del valor prácticamente idéntico para la estructura no lineal al considerar cualquiera de los dos métodos de diseño propuesto, debido a la función de distribución Gaussiana de los datos de entrada. En

consecuencia y tal como nos indicaba el parámetro NPEMP la presencia de la no linealidad corrige el error de predicción obtenido al aplicar sólo el predictor lineal.

Para el **bloque C**, se observa (tabla III) las mejoras obtenidas al aplicar el predictor no lineal para, así como la situación intermedia obtenida al implementar la no linealidad a partir de la ecuación (5).

BLOQUE C (D=6)	$GDECOR_{LINEAL}$	$GDECOR_{NL+POL}$	$GDECOR_{NL+DIR}$
	N=1	1.81	1.72
N=2	1.53	1.54	1.66
N=4	1.23	1.34	1.43
N=6	1.14	1.33	1.46

Tabla III: $GDECOR$ correspondiente al **Bloque C**.

Nótese como valores $N=1$ o $N=2$ resultan los más adecuados en estos bloques desde el punto de vista de la parámetro $GDECOR$.

4.2. Análisis de la función densidad de probabilidad:

Siguiendo con los experimentos realizados, estamos interesados en comprobar el grado de gaussianidad del error de predicción obtenido al aplicar la predicción no lineal, para de esta forma facilitar el diseño del detector, tal como hemos planteado en el Capítulo I. Para ello y una vez calibrado el sistema y diseñado el detector lineal óptimo y la estructura de Wiener mediante las técnicas presentadas en la sección 2, con los datos de entrenamiento, introducimos a dicho esquema los datos para testear la estructura (segundo y tercer acimut). En las figuras 25, 26, y 27, aparecen por un lado representados los diferentes histogramas de cada uno de los datos de testeo originales para cada uno de los diferentes bloques considerados, junto con los histogramas errores de predicción (ruido residual), obtenidos al aplicar

a los datos originales (de testeo) los distintos esquemas de predicción planteados, (diseñados con los datos de entrenamiento).

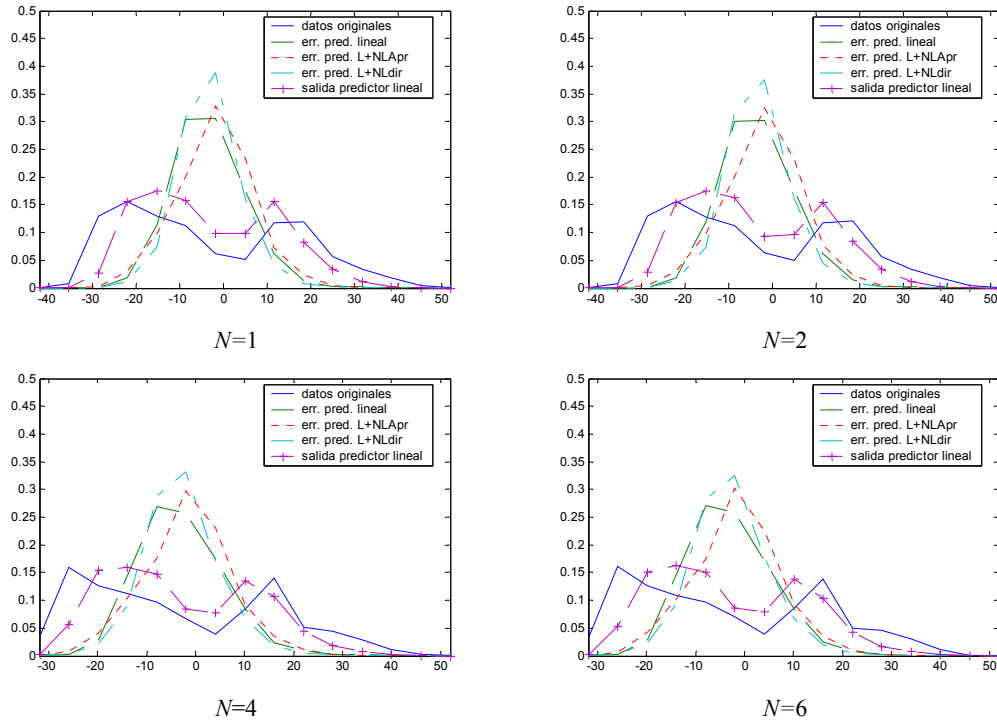


figura 25: Histogramas de los datos originales, sus predicciones y los diferentes errores de predicción. **Bloque A**

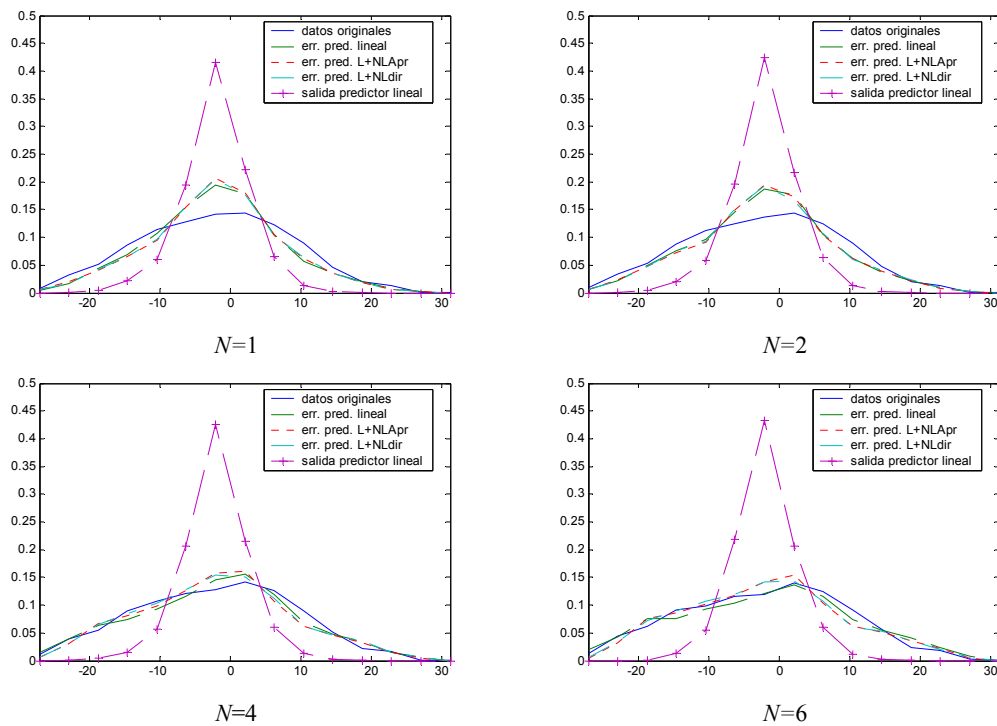


figura 26: Histogramas de los datos originales, sus predicciones y los diferentes errores de predicción. **Bloque B**

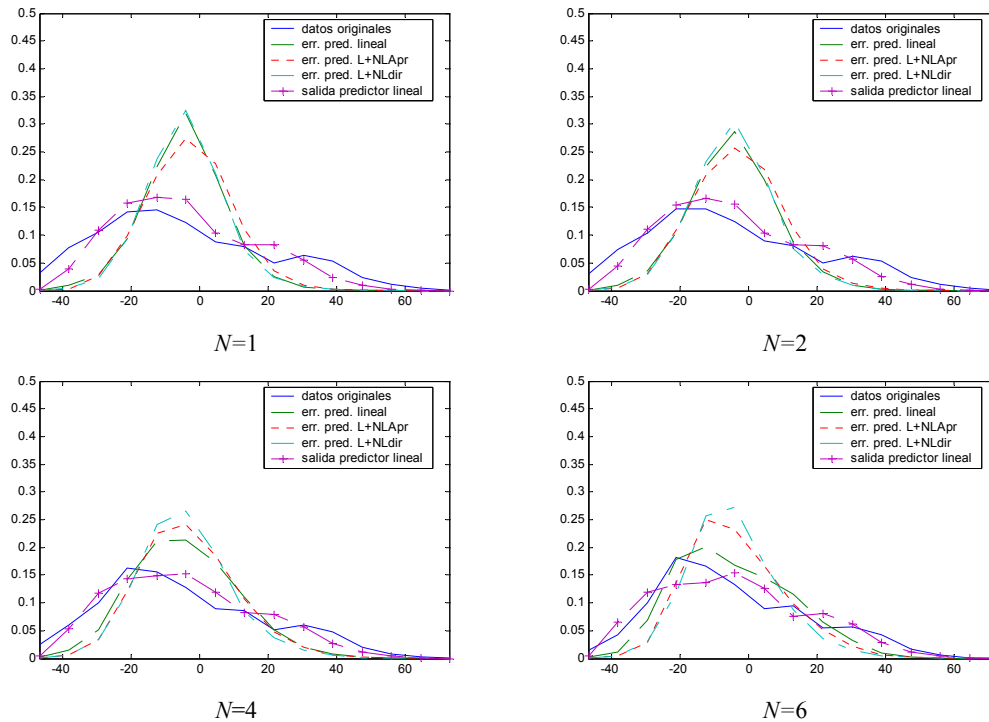


figura 27: Histogramas de los datos originales, sus predicciones y los diferentes errores de predicción. **Bloque C**

Por otro lado, hemos representado en las mismas gráficas el histograma correspondiente a las predicciones obtenidas al aplicar sólo el predictor lineal, para de esta forma determinar la validez estricta de la ecuación (5). En este caso cuando las predicciones tengan una función densidad de probabilidad más Gaussiana la fórmula propuesta para evaluar la media condicional será una mejor aproximación.

De esta forma la mayor gaussianidad de las predicciones obtenidas en el **bloque B**, determina la validez de la técnica propuesta en este caso, siendo peor a medida que el grado de gaussianidad empeora (**bloque A**). En cambio, en el **bloque C**, al encontrarse en una situación intermedia la distribución obtenida es razonablemente Gaussiana y en consecuencia la fórmula propuesta en la ecuación (5) es una buena aproximación.

En todos los casos se observa, que los histogramas de los datos de error de predicción son claramente Gaussianos, por tanto el detector diseñado se beneficiará de la usual hipótesis de Gaussianidad. De esta forma y al igual que vimos en el Capítulo II, podremos fijar fácilmente una PFA deseada.

A continuación representamos en las tablas IV, V y VI la *GSNR*, (ganancia en *SNR*), tal como definimos en el Capítulo II.

$$GSNR = \frac{SNR_e}{SNR_x} = \frac{1}{NPEMP} \tag{12}$$

Como esperábamos hemos obtenido una mayor *GSNR* en el primer bloque (tabla 1) debido a la mayor correlación de ruido de los datos originales en la dimensión número de barrido.

BLOQUE A (<i>D</i> =6)	<i>GSNR</i> _{LINEAL}	<i>GSNR</i> _{NL+POL}	<i>GSNR</i> _{NL+DIR}
	<i>N</i> =1	6.07	5.30
<i>N</i> =2	5.92	5.09	8.21
<i>N</i> =4	5.40	4.74	7.06
<i>N</i> =6	5.18	4.64	6.77

Tabla IV: *GSNR* correspondiente al **bloque A**.

BLOQUE B (<i>D</i> =6)	<i>GSNR</i> _{LINEAL}	<i>GSNR</i> _{NL+POL}	<i>GSNR</i> _{NL+DIR}
	<i>N</i> =1	1.37	1.40
<i>N</i> =2	1.22	1.24	1.24
<i>N</i> =4	0.98	1.07	1.08
<i>N</i> =6	0.85	1.02	1.03

Tabla V: *GSNR* correspondiente al **Bloque B**.

BLOQUE C ($D=6$)	$GSNR_{LINEAL}$	$GSNR_{NL+POL}$	$GSNR_{NL+DIR}$
	$N=1$	4.39	4.14
$N=2$	3.66	3.59	4.43
$N=4$	2.33	2.68	3.20
$N=6$	1.70	2.33	2.89

Tabla VI: $GSNR$ correspondiente al **Bloque C**.

En el **bloque C**, se observa la mayor $GSNR$ obtenida al aplicar la no linealidad excepto en la aproximación polinómica para $N \leq 2$. Lo que implicaría en estos casos tener que aumentar el orden m en la aproximación.

En general $N=1$ es la mejor elección desde el punto de vista del parámetro $GSNR$.

4.3. Probabilidades de falsa alarma:

El resto de las figuras están relacionadas con el detector. Las figuras 28, 29 y 30 nos muestran para cada uno de los bloques y para la longitudes del predictor N consideradas, la PFA teórica (línea sólida) asumiendo la Gaussianidad del error de predicción y la comparamos con la PFA estimada para cada uno de los sistemas propuestos. Esta ha sido medida usando todos los posibles vectores e de 6 elementos permitidos en cada bloque después de predecir los datos originales de testeo (tercer y cuarto acimut) sobre la estructura diseñada a partir de los datos de entrenamiento (primer y segundo acimut). Estas figuras nos indican la capacidad para fijar una PFA. La matriz P , es calculada a partir de las expresiones determinadas en el Capítulo II, con parámetros $\beta = 0.1$ y $p=2$ (dos primeros autovectores dominantes de Slepian).

El número de vectores de error disponibles en cada caso, nos determina la capacidad para estimar valores pequeños de PFA (al disponer de mayor o menor número de muestras V). Para estimar la PFA hemos procedido de la misma forma

que en los histogramas: una vez calibrado el predictor a partir de los datos de entrenamiento estimamos la PFA en el detector para los datos de testeo.

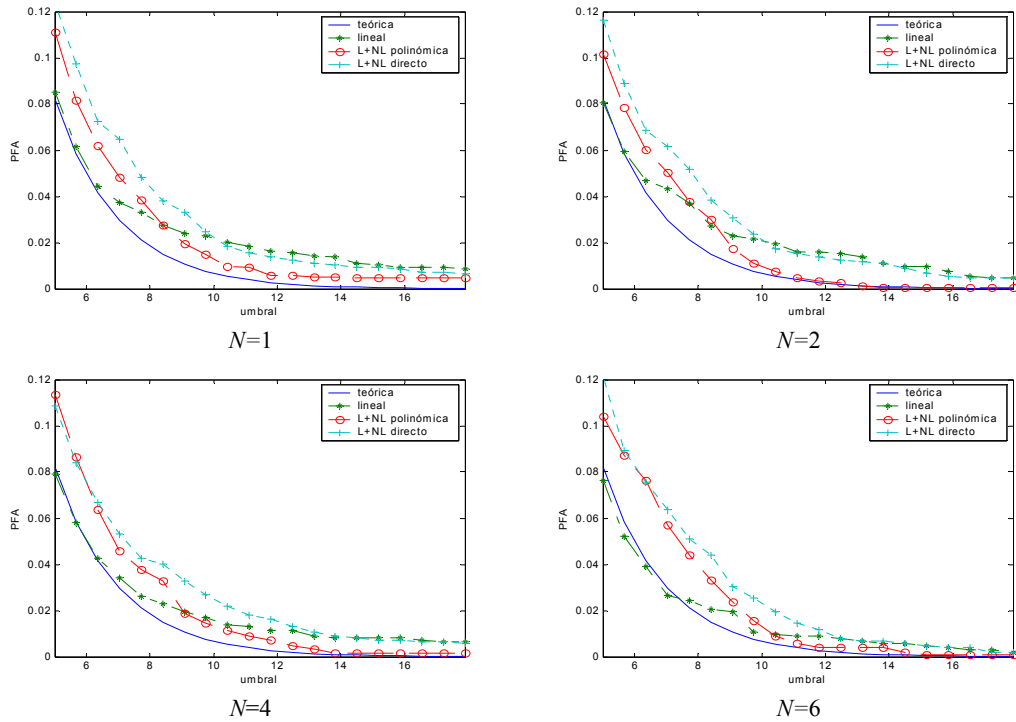


figura 28: PFA teórica frente a PFA estimada para cada una de los distintos sistema. **Bloque A**

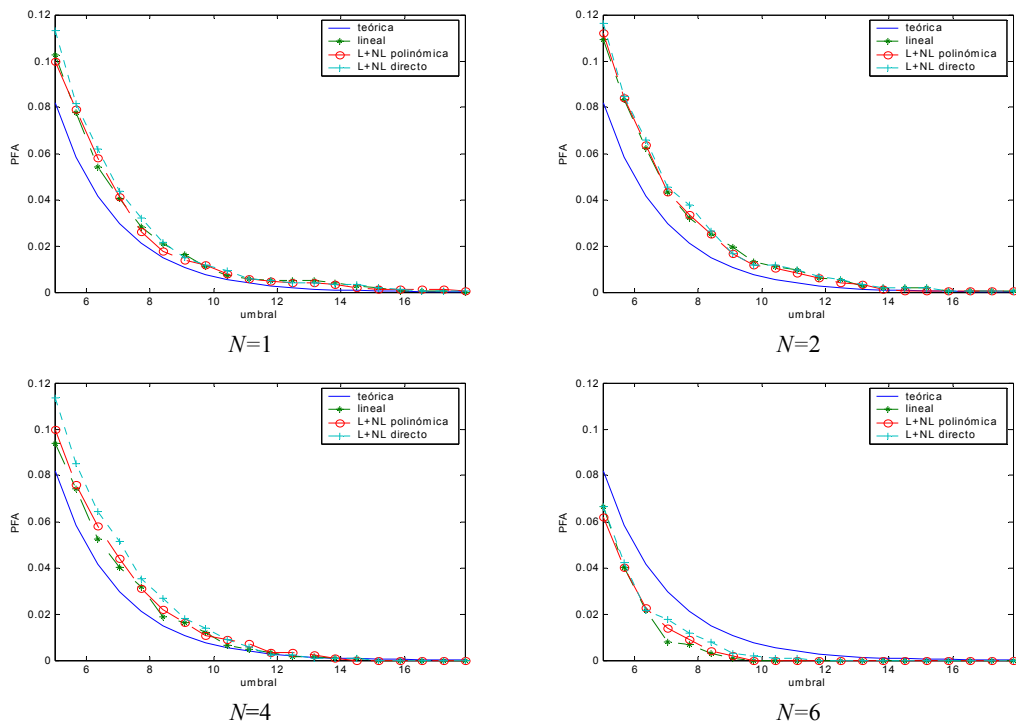


figura 29: PFA teórica frente a PFA estimada para cada una de los distintos sistema. **Bloque B**

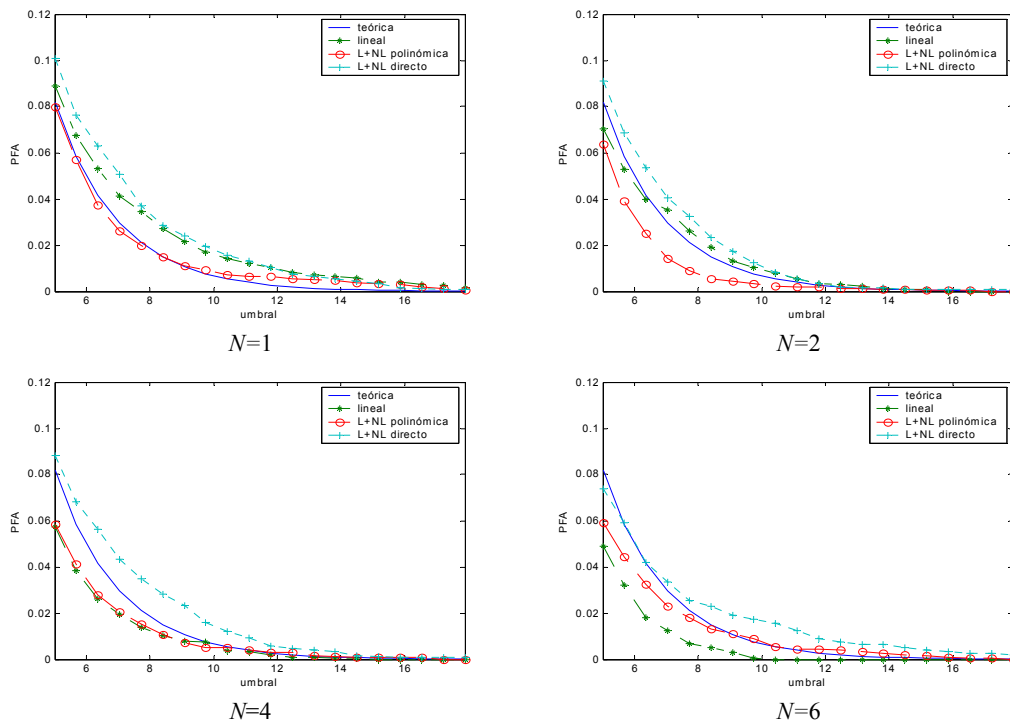


figura 30: PFA teórica frente a PFA estimada para cada una de los distintos sistema. **Bloque C**

Las curvas anteriores sugieren una razonable capacidad de controlar el diseño de manera que se garantice una cierta PFA.

4.4. Probabilidad de detección:

Por último, en las figuras 31, 32, 33, donde se ha seleccionado $N=2$ y $D=6$, hemos representado la PD (probabilidad de detección de fuego descontrolado) y la PDI (probabilidad de detección de alarmas indeseadas) frente a la PFA para cada una de las estructuras de predicción analizadas considerando por un lado un detector de energía, y por otro lado el detector basado en el filtro adaptado en subespacio (ver figura 21), con $\beta=0.1$ y $p=2$.

Hemos sumado una tendencia lineal a los 6 elementos ($D=6$) del vector error de predicción (e), obtenido al pasar los datos de testeo por el esquema de energía 30, para simular un fuego descontrolado. De esta forma se ha estimado la PD para el detector de energía (línea de puntos) y el detector basado en el filtro adaptado al subespacio (línea sólida). En todas las gráficas se observa la mejora obtenida en la

PD al incluir el subespacio adaptado \mathbf{P} , pudiendo observar también como a su vez la PD mejora al incluir la no linealidad (máxima PD en el caso de estimación directa).

De esta forma concluimos que el detector se aprovecha de la inclusión de la no linealidad debido a la disminución de la potencia de error de predicción en las seis distancias de predicción.

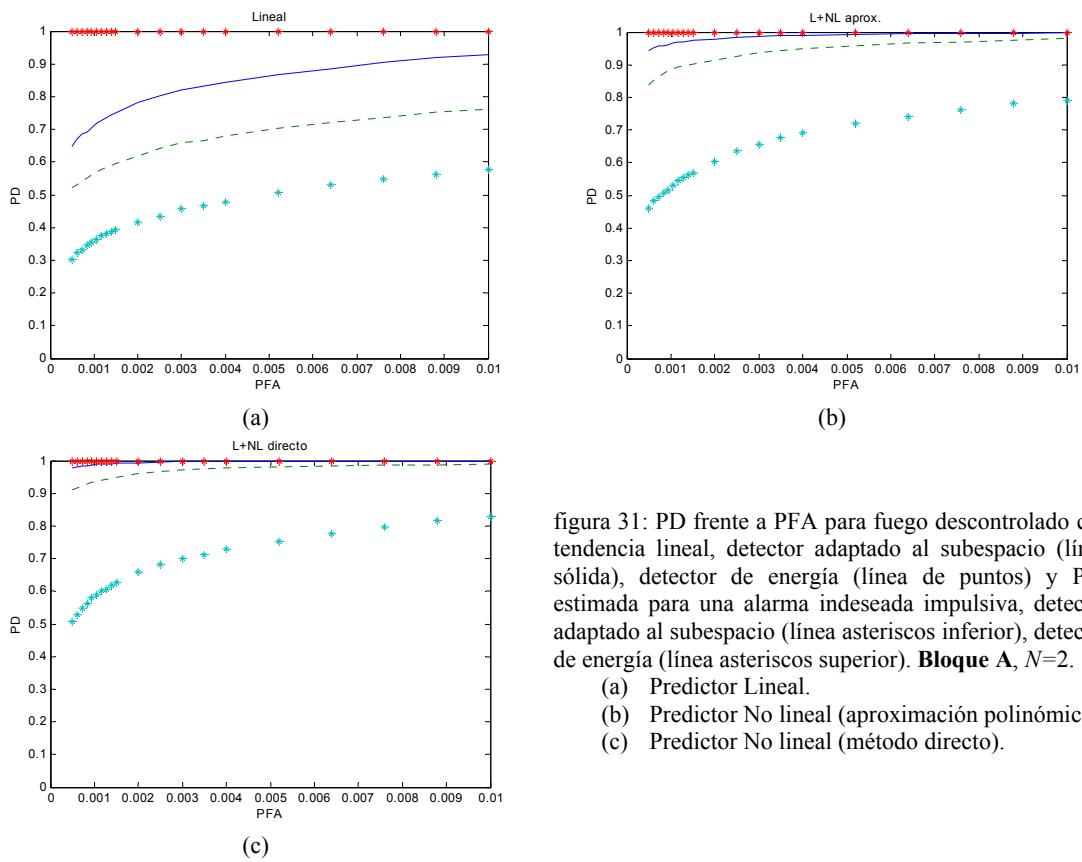


figura 31: PD frente a PFA para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). **Bloque A**, $N=2$.
 (a) Predictor Lineal.
 (b) Predictor No lineal (aproximación polinómica).
 (c) Predictor No lineal (método directo).

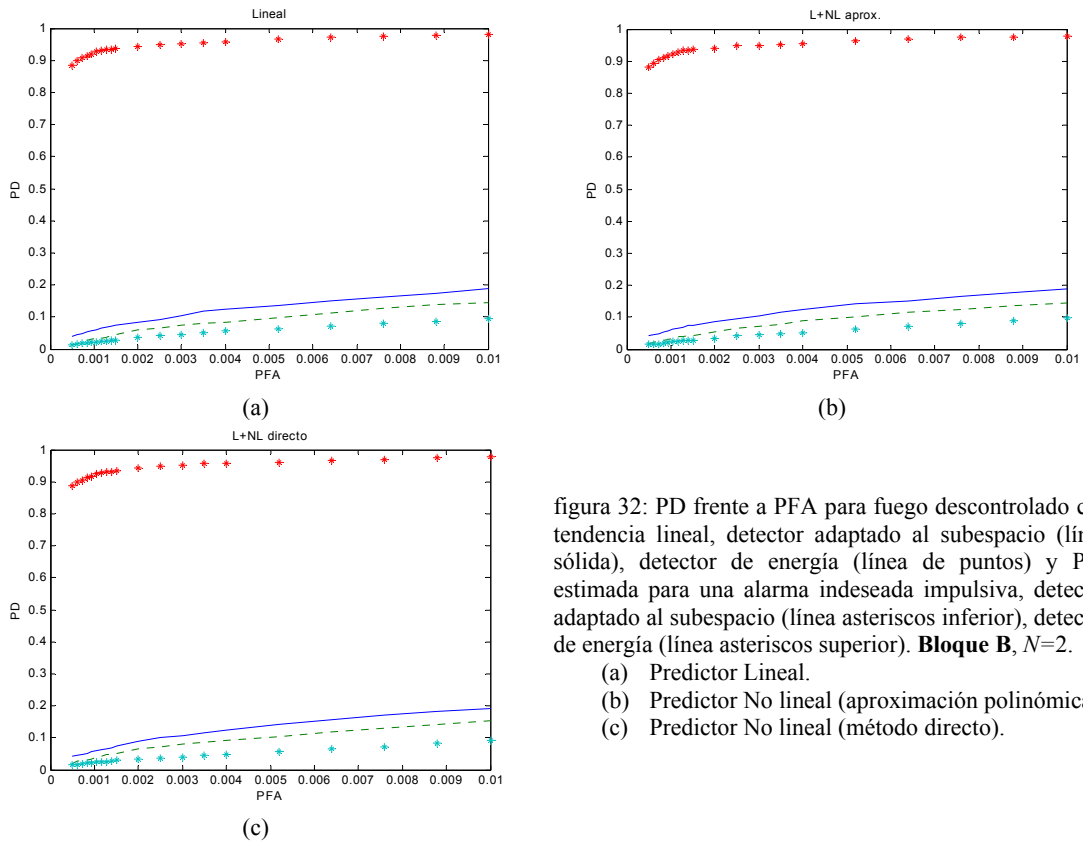


figura 32: PD frente a PFA para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). **Bloque B**, $N=2$.

- (a) Predictor Lineal.
- (b) Predictor No lineal (aproximación polinómica).
- (c) Predictor No lineal (método directo).

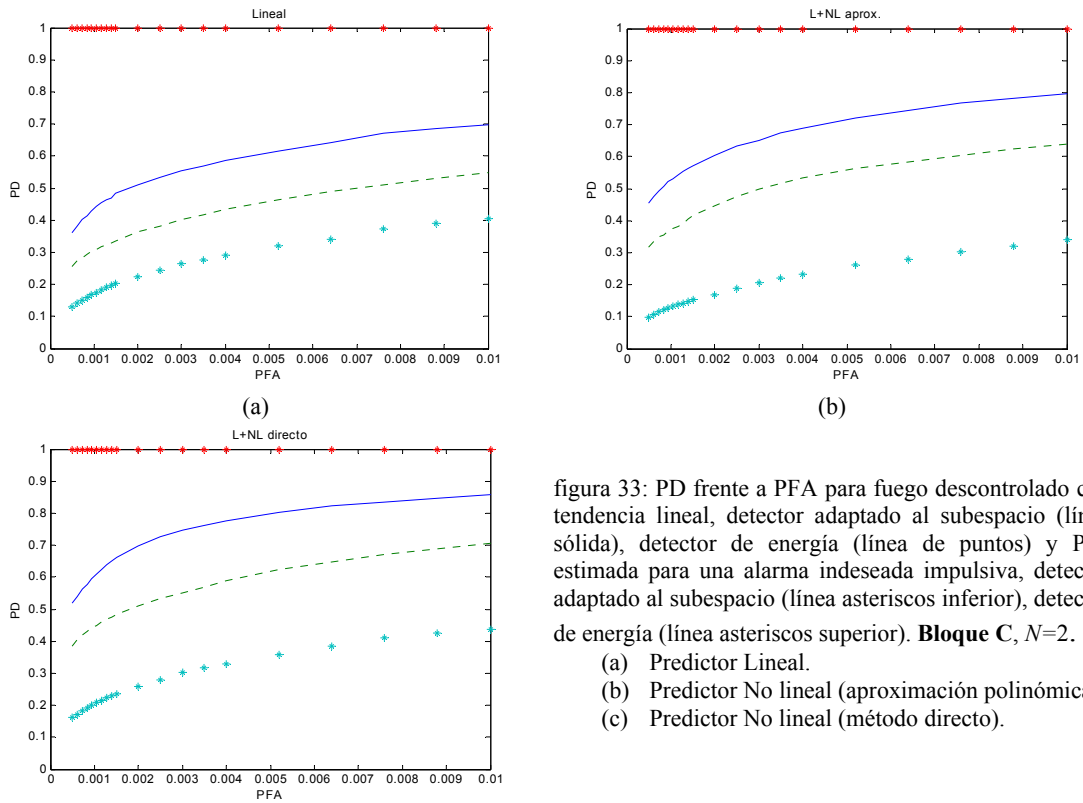


figura 33: PD frente a PFA para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). **Bloque C**, $N=2$.

- (a) Predictor Lineal.
- (b) Predictor No lineal (aproximación polinómica).
- (c) Predictor No lineal (método directo).

Las diferentes curvas de asteriscos corresponden a la probabilidad de detección de alarmas indeseadas, éstas han sido simuladas sumando al quinto elemento (distancia 5) de cada uno de los vectores de error e una muestra de amplitud $\sqrt{30}$, para de esta forma realizar una adecuada comparación (misma energía de señal) con respecto a la simulación de la señal de fuego creciente (tendencia lineal de energía 30). En las figuras 31, 32 y 33 se observa la drástica reducción de la PDI al incluir el estimador de señal P (curva de asteriscos inferior) para cada una de las diferentes estructuras propuestas. Por lo general, se observa que la PDI aumenta ligeramente al incluir la no linealidad.

En el capítulo V analizaremos en mayor profundidad el detector adaptado al subespacio, así como analizaremos las ventajas de considerar un nuevo esquema de detección que denominamos “detector de crecimiento”.

5. Estimación orden N del predictor lineal:

Para el diseño del predictor lineal óptimo H , (figura 2) hemos de fijar los valores N y D .

En un principio debido a que la decisión de alarma/no alarma se hace a partir de vectores de D elementos (D barridos sucesivos en los que basar la predicción), el parámetro D debería ser lo más alto posible, con el fin de optimizar la explotación de la información sobre los modelos implicados, pero tenemos un límite en el máximo retardo permisible ante una alarma de fuego. Dado un cierto retardo admisible podemos aumentar D a base de aumentar la velocidad de barrido (disminuir el tiempo entre barridos) pero entonces el límite lo impondrán los recursos computacionales para realizar todo el procesado necesario en cada barrido.

Por otra parte debemos fijar el orden del predictor N . Este es un problema clásico en modelado paramétrico y análisis espectral para el que se han propuesto numerosas alternativas [MAR87][KAY88][KAY81]. No obstante todas ellas asumen un comportamiento estacionario, poco realista en nuestro caso, especialmente se en la calibración consideramos bloques de datos que abarcan una zona especial extensa (por

ejemplo el **bloque C** es prácticamente la unión de dos bloques el **A** y el **B**, uno más lejano y otro más cercano al sensor).

Por ello proponemos aquí una técnica más realista que siempre tiene presente la conveniencia de ser “conservador” en el orden N , es decir N debe ser un valor no muy alto en general. A partir de la ecuación (3) del Capítulo II podemos estimar la matriz de autocorrelación del error de predicción de los datos de entrenamiento. A lo largo de su diagonal principal tenemos las potencias medias de los errores de predicción para distancias $l=1$ a $l=D$, denominémoslas P_{el} .

Definimos el cociente,

$$c_l^N = \frac{P_{el}^N}{P_x} \tag{13}$$

como la potencia media de error de predicción a distancia l normalizada por la potencia media de los datos a predecir, para un predictor de longitud N . Así en la figura 34 hemos representado dicho cociente para $N=2$ y $N=15$ (entre N pequeño y N grande) y para l variando de 1 a 6. Se observa que apenas hay mejora por incrementar N de 2 a 15 de manera que elegiremos $N=2$, dada la previsible no estacionaridad de los datos. Si pasar a $N=15$ supusiese una considerable mejora en las curvas anteriores deberíamos comprobar que dicha mejora es real con los datos de testeo, es decir deberíamos ver si con,

$$\hat{c}_l^N = \frac{\hat{P}_{el}^N}{\hat{P}_x} \tag{14}$$

donde \hat{P}_{el}^N es el parámetro NPEMP representado en las figuras 8, 14 y 20 para cada bloque, se sigue manteniendo o no la mejora (en nuestro caso obviamente no).

6. Cota mínima del error de predicción lineal:

Una cuestión de cierto interés es saber hasta qué punto nuestro predictor lineal queda lejos o cerca de la cota mínima de potencia de error alcanzable por un predictor lineal de orden suficientemente elevado. Esto es especialmente interesante en nuestro caso dado que en la práctica, la no estacionaridad obliga a disponer de órdenes muy bajos para el predictor lineal. Si la conclusión es que estamos muy lejos de la cota mínima, deducimos que resultaría más adecuada la inclusión de un predictor IIR (predictor lineal de respuesta al impulso infinita). No obstante, el diseño de filtros IIR como predictores no será considerado en esta tesis y se dejará como una línea futura de trabajo.

Para encontrar la cota mínima utilizaremos la representación Wold [PRO92] [PRI88]. Si $\{x(n)\}$ es un proceso estocástico estacionario con espectro $S_x(\omega)$ continuo y $\log S_x(\omega)$ integrable, puede obtenerse como la salida de un filtro de respuesta al impulso $h(n)$ causal, de fase mínima, ante una entrada $\{e(n)\}$, ruido blanco de media cero y varianza unidad.

$$x(n) = \sum_{m=0}^{\infty} h(m)e(n-m) \tag{15}$$

Puede demostrarse [PRI88] que

$$\sigma_m^2 = \sum_{n=0}^{m-1} h^2(n) \tag{16}$$

es la cota mínima del error de predicción lineal a distancia m .

En el Apéndice B indicamos dos métodos para calcular $h(n)$, a partir de la cual resolveremos la ecuación (16) para encontrar el valor de la cota mínima a distancia de predicción m y compararla con el valor obtenido para el N seleccionado (ver figura 34), donde podemos comparar la cota con la potencia de error para $N=2$ y $N=15$. Para una adecuada comparación normalizamos la cota mínima por la potencia media de los datos

a predecir. De la observación de la figura 34 deducimos que con un predictor lineal de orden $N=2$ estamos todavía algo lejos de la cota mínima. Por tanto, dado que por la longitud de los registros, así como por su no estacionaridad no podemos aumentar mucho N , podríamos plantear la alternativa de un predictor IIR.

$$cota_m = \frac{\sigma_m^2}{P_x} \tag{17}$$

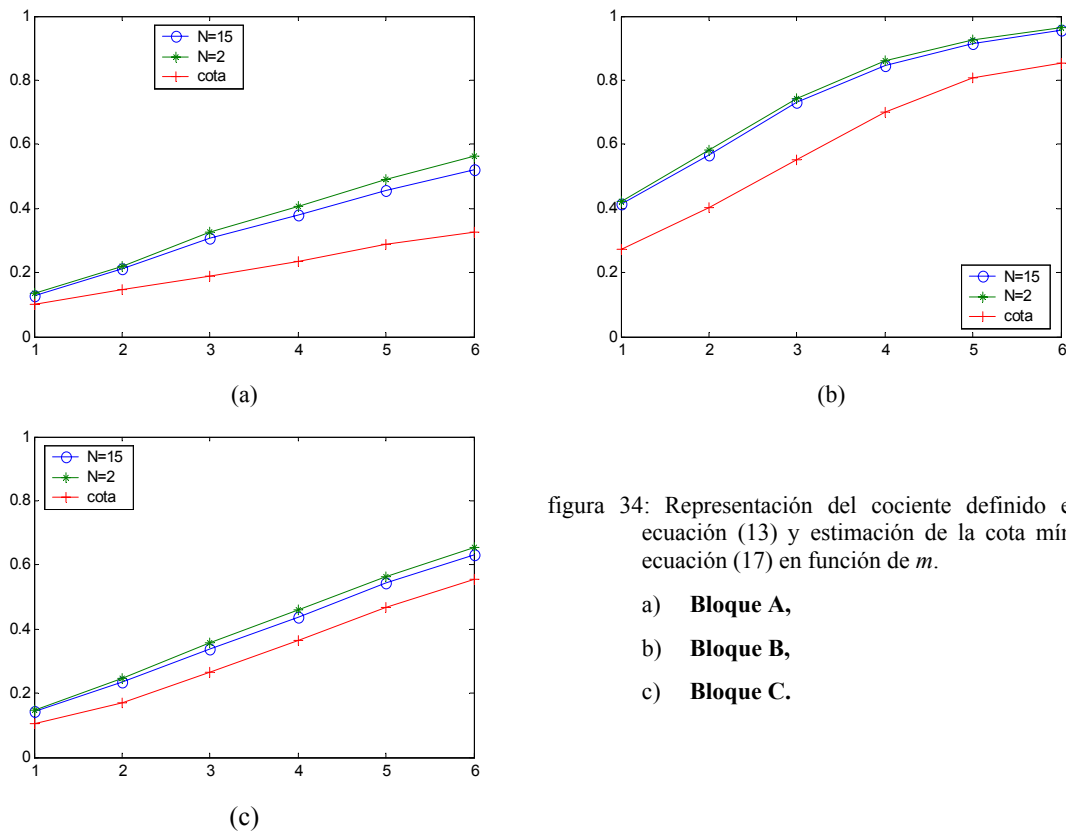


figura 34: Representación del cociente definido en la ecuación (13) y estimación de la cota mínima, ecuación (17) en función de m .

- a) **Bloque A,**
- b) **Bloque B,**
- c) **Bloque C.**

Para la obtención de $h(n)$ por cualquiera de los dos métodos presentados en el apéndice B, necesitamos previamente estimar el espectro de los datos $x[n]$ (en nuestra aplicación los diferentes datos de entrenamiento (primer y segundo acimut) correspondientes a los **bloques A, B, C**). El método usado ha sido el de Blackman-Tukey.

Este método consiste [HAY91] en realizar la transformada de Fourier, para un determinado enrejillado (M), del vector r de tamaño M , definido en la ecuación (18)

$$\mathbf{r} = [0, \dots, 0, R(-L), R(-L+1), \dots, R(-1), R(0), R(1), \dots, R(L-1), R(L), 0, \dots, 0]^T \quad (18)$$

con $R[m]$ la función de autocorrelación estimada a partir de los datos de entrenamiento, donde debemos fijar el número de correlaciones necesarias $(L+1)$ (ancho de la ventana) para que la ecuación (17) tome el valor mínimo.

Para ello, proponemos una técnica para determinar el número de correlaciones necesarias. Dicha técnica consiste en determinar la inversa del ancho de banda equivalente (τ_L) del espectro estimado $S_x(\omega)$ al realizar la transformada de Fourier a la ecuación (18) para cada uno de los posibles vectores \mathbf{r} construidos a partir de considerar en el mismo 1, 2 hasta el máximo de correlaciones posibles de los datos a predecir, (21 barridos temporales disponibles en nuestro caso), de tal forma que el número $L+1$ de correlaciones necesarias en el vector \mathbf{r} coincide en el primer punto donde se produce un cambio significativo de pendiente de la función obtenida al representar τ_L , (ver figura 35).

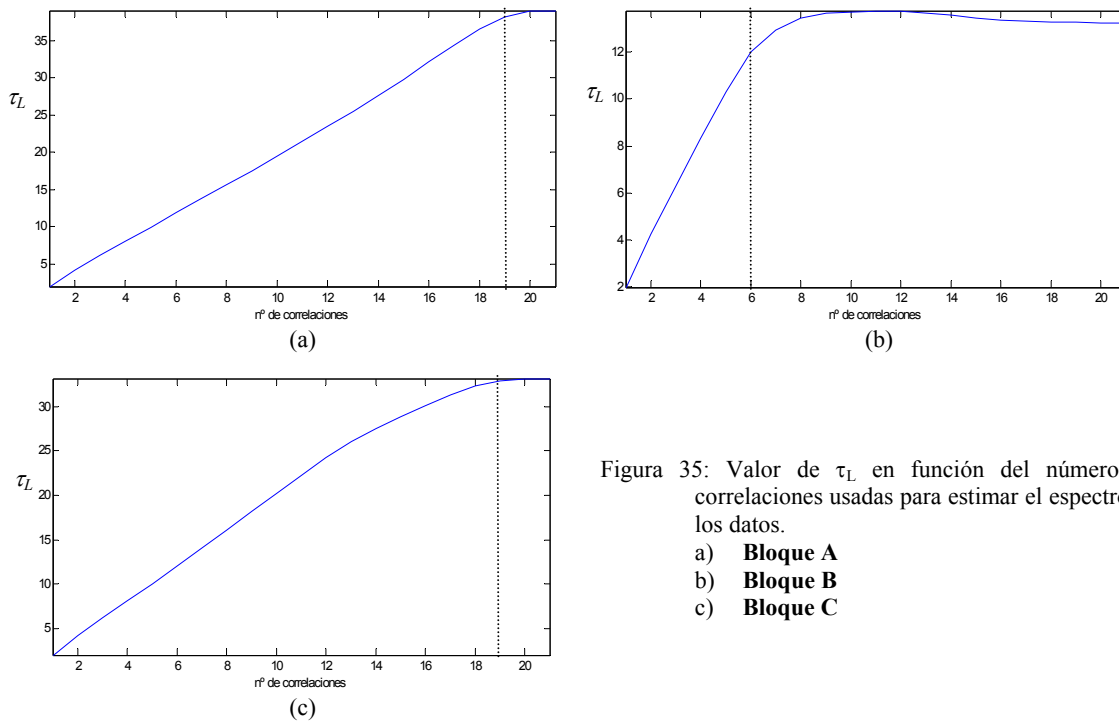


Figura 35: Valor de τ_L en función del número de correlaciones usadas para estimar el espectro de los datos.

- a) **Bloque A**
- b) **Bloque B**
- c) **Bloque C**

El valor τ_L puede considerarse como una medida indirecta de la duración efectiva de la autocorrelación de los datos bajo análisis. A partir de la figura 35 obtenemos que el número de correlaciones necesarias para el **bloque A** es de 19, para el **bloque B** de 6 y para el **bloque C** de 19 (valores que hemos comprobado posteriormente y que determinan el valor mínimo de la cota). Una vez determinado el número de correlaciones necesarias en el método de Blackman-Tukey para estimar el espectro y a partir de lo visto en el Apéndice B, determinamos la cota mínima para decidir o no la inclusión de filtros IIR en la parte lineal del esquema de predicción propuesto.

El objetivo se centra ahora en determinar el número de correlaciones necesarias a partir de la secuencia discreta τ_L de manera automática.

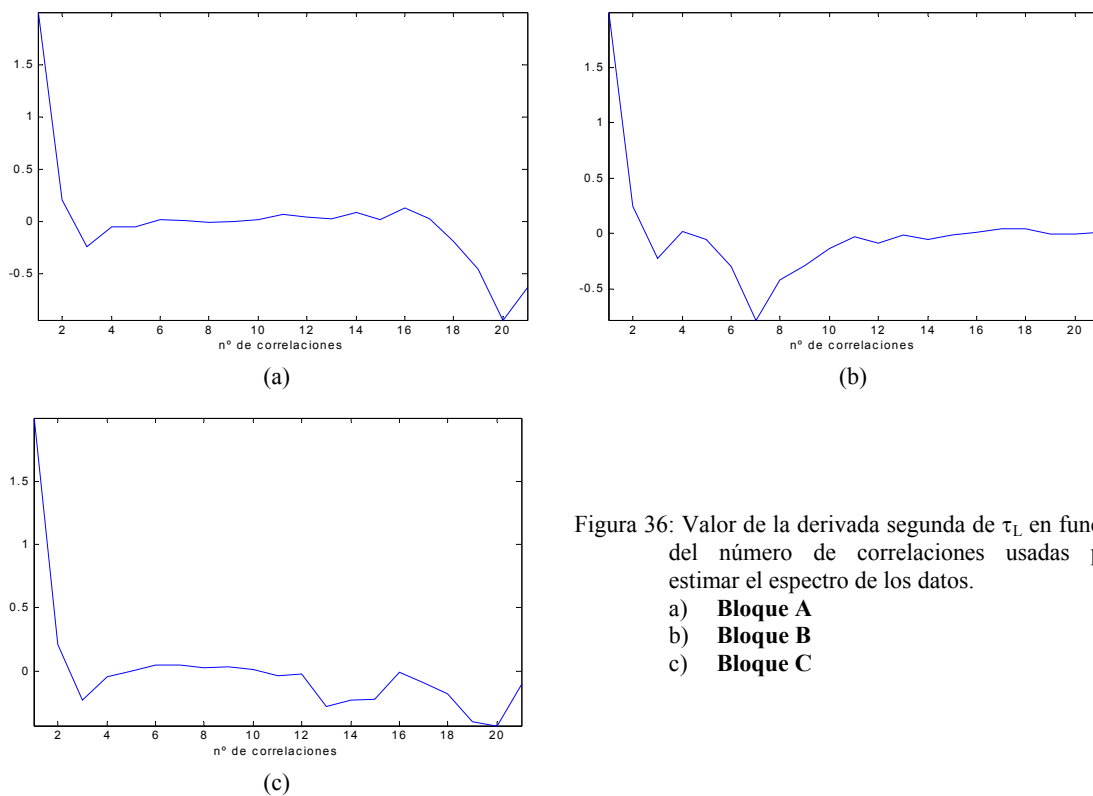


Figura 36: Valor de la derivada segunda de τ_L en función del número de correlaciones usadas para estimar el espectro de los datos.

- a) **Bloque A**
- b) **Bloque B**
- c) **Bloque C**

Para ello proponemos realizar la derivada segunda del parámetro τ_L (figura 36), ya que estamos interesados en determinar la mayor variación de la pendiente (esta variación será de signo negativo al pasar de una pendiente creciente a decrecer hasta pendiente cero) de las gráficas representadas en la figura 35. Con este planteamiento

deberemos buscar el valor de número de correlaciones correspondiente al mínimo de la derivada segunda de τ_L , si además tenemos en cuenta (figura 35) que el valor es aquél donde se produce la variación, el valor del número de correlaciones necesarias para aplicar al método de Blackman-Tukey corresponde al anterior a aquél donde se obtiene el mínimo de la derivada segunda (figura 36).

De tal forma en la figura 36 los mínimos obtenidos en cada gráfica corresponden a los instantes 20, 7 y 20 respectivamente, en consecuencia el número de correlaciones necesarias será de 19 (**bloque A**), 6 (**bloque B**) y 19 (**bloque C**).

7. Conclusiones:

Se han propuesto dos técnicas nuevas de diseño de la parte no lineal de la estructura de Wiener, una, denominada método directo, estima de forma directa la media condicionada de los valores a predecir, la otra utiliza una aproximación polinómica cuyos coeficientes deben ser calculados a partir de estimas de momentos cruzados de orden superior. La parte lineal del predictor se diseña de forma convencional mediante la ecuación de Wiener-Hopf.

Las técnicas propuestas se han aplicado a los tres bloques seleccionados, obteniéndose algunas conclusiones que permiten cierto grado de generalización a saber:

- Situaciones de no-Gaussianidad, pero de comportamiento bi o multimodal en el que para cada modo hay Gaussianidad, pueden llevar a reducciones aparentemente muy poco significativas en la potencia media del error de predicción al incluir una no linealidad, si el coeficiente de correlación en la dirección de predicción es el mismo en cada modo. Esto es especialmente cierto en situaciones de alta correlación, en las que el predictor lineal consigue ya una elevada reducción de error de predicción (caso del **bloque A**).
- Situaciones de Gaussianidad no precisan en principio de la inclusión de una no linealidad, sin embargo si se combinan con la presencia de baja

correlación pueden darse amplificaciones del error de predicción lineal si el orden del predictor lineal y/o la distancia a predecir toman valores altos. Al incorporar la no linealidad “contenemos” el error de predicción, dotando al sistema de mayor robustez (caso **bloque B**).

- La presencia de la no linealidad resulta conveniente también en situaciones altamente no estacionarias, producidas por el entrenamiento a partir de bloques de datos que abarcan zonas espaciales muy amplias, tanto desde el punto de vista de la reducción del error de predicción, como de la robustez frente a amplificaciones de error de predicción (caso **bloque C**).
- El método directo de ajuste de la no linealidad es siempre el que da mejores resultados a cambio de su mayor complejidad de diseño e implementación. La aproximación polinómica de orden 3 da prácticamente los mismos resultados cuando la salida del predictor lineal puede considerarse Gaussiana (**bloque B**), dado que entonces se satisfacen las hipótesis asumidas en la derivación de la ecuación correspondiente.
- Las probabilidades de falsas alarmas admiten un razonable ajuste dado que en todos los casos el error de predicción (ruido de entrada al detector) pueden considerarse Gaussiano.
- La probabilidad de detección mejora al incluir la no linealidad junto al predictor lineal, siempre que la distribución de los datos sea no Gaussiana. Esto ocurre así incluso en casos como el del **bloque A** en que la reducción del error de predicción es aparentemente poco significativa. Debe tenerse en cuenta que la detección se hace sobre las D muestras, y que pequeñas mejoras en la predicción de cada elemento, pueden llevar a una mejora global significativa en la detección.

Se ha propuesto una técnica concreta de ajuste del orden del predictor lineal, que tiene en cuenta la conveniencia de no colocar órdenes altos dada la no estacionaridad de los datos.

Se ha señalado un procedimiento para el cálculo de la cota mínima de potencia mínima del error de predicción que puede resultar útil para evaluar cómo de cerca estamos de dicha cota con el predictor lineal diseñado y, caso de estar lejos, optar por un predictor IIR, previo a la no linealidad.

Capítulo V:

El Detector

Capítulo V

EL DETECTOR

1. Introducción:

En este Capítulo pretendemos analizar con detalle el detector. Estamos interesados en caracterizar esquemas de detección que nos permitan incrementar la probabilidad de detección de fuegos reales (PD) y minimizar la probabilidad de alarmas indeseadas (PDI) tomando como referencia el detector de energía. A lo largo de la presente tesis, se ha propuesto para el esquema de detección el denominado “detector adaptado al subespacio de señal” (figura 1) sin establecer criterios para justificar el valor tomado en los parámetros de diseño del mismo (β y p , ver Capítulo II, sección 4). Así pues, en la sección 2 de este Capítulo, analizamos con detalle la influencia de los parámetros de diseño para mejorar la PD y minimizar la PDI.

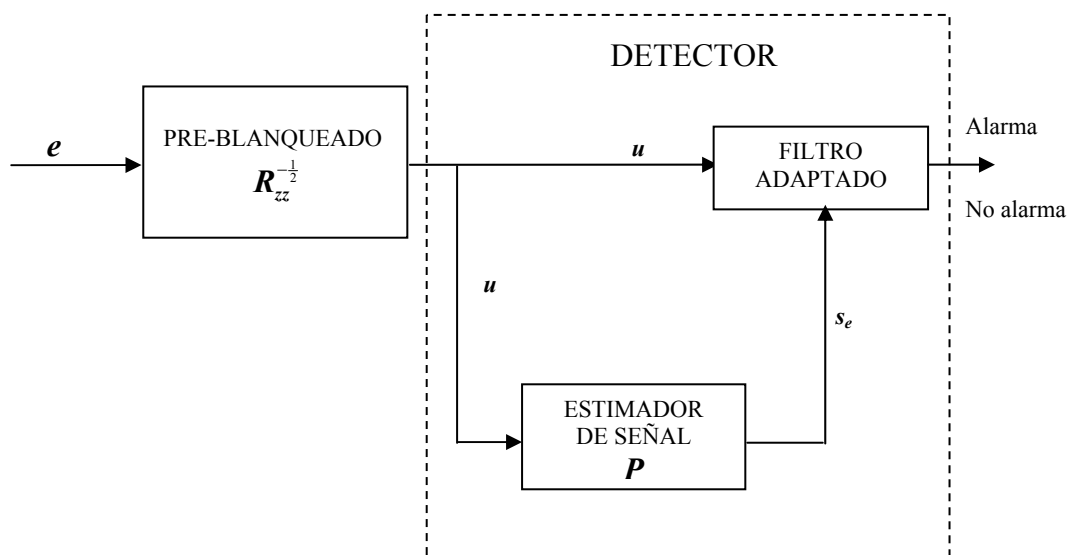


figura 1: Detector adaptado al subespacio de señal

Por otro lado, en la sección 3, proponemos un nueva estructura para el bloque de detección, que denominamos “detector de crecimiento” (figura 2). El esquema permite

determinar (al tener en cuenta que un fuego evolucionará incrementado la temperatura media de la celda bajo análisis en cada barrido a partir del instante en que se produzca), tendencias crecientes (lineales o no lineales) en los vectores de D elementos en los cuales el detector debe decidir entre alarma y no-alarma.

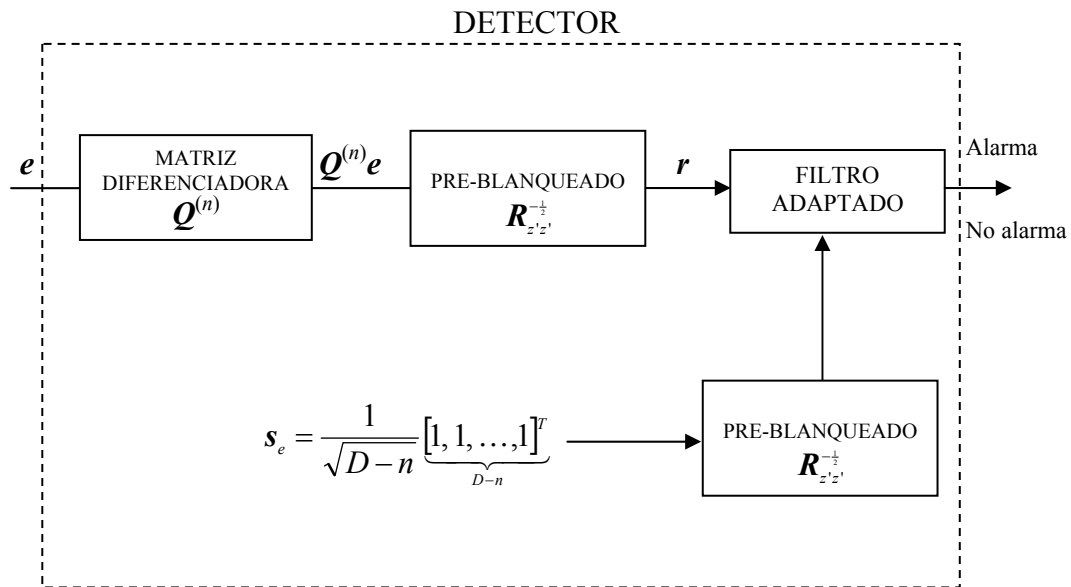


figura 2: Detector de crecimiento

En ambos detectores (el de subespacio y el de crecimiento) caracterizaremos en primer lugar teóricamente el esquema propuesto para diferentes señales correspondientes a modelos de fuegos y a modelos de alarmas indeseadas, para posteriormente y una vez fijado teóricamente sus parámetros de diseño, validar los resultados obtenidos para cada uno de los bloques información considerados a lo largo de la presente tesis.

En la sección 4 propondremos y analizaremos para los **bloques A, B y C**, la mejora obtenida en la detección al considerar como esquema de detección ambos detectores (filtro adaptado al subespacio y el de crecimiento) simultáneamente para decidir entre alarma y no-alarma.

2. Detector adaptado al subespacio de señal:

En esta sección se analizan con detalle los parámetros de diseño del subespacio adaptado de señal, $(\beta$ y p), para los siguientes modelos de señal con los que caracterizar fuegos descontrolados y alarmas indeseadas.

-Modelos de fuegos descontrolados.

- Mantenimiento de la temperatura a lo largo de los D barridos en los que basar la predicción.
- Crecimiento lineal con una cierta pendiente en los D barridos.

-Modelo de alarmas indeseadas.

- Pico de señal (delta) situada en el barrido l -ésimo.

Para una adecuada comparación entre los modelos de fuegos y de falsas alarmas considerados la energía del correspondiente vector debe ser la misma, tal como lo hemos aplicado a los experimentos de los Capítulos II y IV.

2.1. Consideraciones teóricas:

Si observamos el esquema de la figura 1, obtenemos una estimación de la señal de fuego a partir de la proyección del vector \mathbf{u} (salida de la transformación de preblanqueado aplicada al vector \mathbf{e}) en el correspondiente subespacio de señal determinado por la matriz de proyección \mathbf{P} .

De esta forma el estadístico del detector a comparar con un umbral es,

$$r = \mathbf{u}^T \mathbf{P} \mathbf{u} \quad (1)$$

Si suponemos que todos los posibles vectores \mathbf{u} , para un bloque de información considerado a la salida de la transformación de preblanqueado, corresponden a muestras de ruido de fondo de media cero y varianza unidad. La media del estadístico r vendrá dada por,

$$E[r] = E[\mathbf{u}^T \mathbf{P} \mathbf{u}] = E\left[\sum_{i=1}^D \sum_{j=1}^D \mathbf{P}_{ij} u_i u_j^T\right] = \sum_{i=1}^D \sum_{j=1}^D \mathbf{P}_{ij} E[u_i u_j^T] \quad (2)$$

donde $E[u_i u_j^T] = \delta_{ij}$ debido al preblanqueado y ser de varianza 1, con lo que la expresión (2) nos queda

$$E[r] = E[\mathbf{u}^T \mathbf{P} \mathbf{u}] = \sum_{i=1}^D \mathbf{P}_{ii} = \text{Traza}(\mathbf{P}) \quad (3)$$

expresión que nos indica que la media del estadístico r es igual a la *traza* de la matriz de proyección $\mathbf{P}(D \times D)$. Por otra parte, si tenemos en cuenta que,

$$\mathbf{P} = \underbrace{\mathbf{A}(\mathbf{A}^T \mathbf{A})}_{\mathbf{I}} \mathbf{A}^T = \mathbf{A} \mathbf{A}^T \quad (4)$$

donde \mathbf{A} tal como indicamos en el Capítulo II, sección 4, es una matriz $D \times p$ cuyas columnas son los p autovectores dominantes de la matriz $\mathbf{M}(D \times D)$ [THO82]

$$\mathbf{M}_{m,n} = \sin c[\beta(m-n)] \quad (5)$$

donde de (5) observamos que \mathbf{M} una matriz simétrica. Los autovectores de \mathbf{M} [KAY88] [HAY91] se obtienen al resolver el sistema de ecuaciones.

$$\mathbf{M} \mathbf{a} = \lambda \mathbf{a} \quad (6)$$

con \mathbf{a} el autovector y λ su correspondiente autovalor asociado. Se cumple [HAY91] que

- $\mathbf{a}_i^T \mathbf{a}_j = \delta_{ij}$, y λ_i real (7)

- $\mathbf{M} = \sum_{i=1}^D \lambda_i \mathbf{a}_i \mathbf{a}_i^T$ (8)

donde definimos \mathbf{A} ($D \times p$) como la matriz constituida por los p ($p \leq D$) autovectores dominantes de \mathbf{M} (los p autovectores de mayor autovalor).

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p] \quad (p \text{ grados de libertad}) \quad (9)$$

Entonces la *traza*(\mathbf{P}) a partir de la expresión (4) y teniendo en cuenta (7) es igual al número de autovectores considerados p .

$$\text{Traza}(\mathbf{P}) = \sum_{i=1}^p \text{Traza}(\mathbf{a}_i \mathbf{a}_i^T) = \sum_{i=1}^p \|\mathbf{a}_i\|^2 = p \quad (10)$$

con lo que la media del estadístico r es igual al número de autovectores dominantes de Slepian considerados.

Si no incluimos el subespacio adaptado de señal \mathbf{P} , (caso detector de energía) la media del estadístico r es igual a D , en este caso la demostración es sencilla, ya que la ausencia del estimador de señal equivale a considerar $\mathbf{P}=\mathbf{I}$ (matriz identidad).

$$\text{Traza}(\mathbf{I}) = D \quad (11)$$

Consideremos ahora únicamente la presencia de señal, caracterizada en el vector \mathbf{s} , de esta forma el r estará definido por la expresión,

$$\mathbf{s}^T \mathbf{P} \mathbf{s} = \mathbf{s}^T \mathbf{A} \mathbf{A}^T \mathbf{s} = \sum_{i=1}^p (\mathbf{s}^T \mathbf{a}_i)^2 \quad (12)$$

De tal forma que podemos definir una ganancia en SNR (G_{SNR_p}) por el hecho de incluir la matriz de proyección \mathbf{P} para p grados de libertad.

- Sin \mathbf{P}
$$SNR = \frac{\mathbf{s}^T \mathbf{s}}{D} \quad (13)$$

- Con \mathbf{P}
$$SNR_p = \frac{\mathbf{s}^T \mathbf{P} \mathbf{s}}{p} = \frac{\sum_{i=1}^p (\mathbf{s}^T \mathbf{a}_i)^2}{p} \quad (14)$$

Por tanto al incluir la matriz de proyección \mathbf{P} .

$$GSNR_p = \frac{SNR_p}{SNR} = \frac{D \mathbf{s}^T \mathbf{P} \mathbf{s}}{p \mathbf{s}^T \mathbf{s}} \quad (15)$$

Si ahora consideramos como modelo de fuego estable la señal definida en el vector \mathbf{s} ($E_s=1$).

$$\mathbf{s} = \frac{1}{\sqrt{D}} [1, 1, \dots, 1]^T \quad (16)$$

y para la alarma indeseada el modelo caracterizado en el vector de \mathbf{s} ($E_s=1$),

$$\mathbf{s} = [0, \dots, 1, \dots, 0]^T \quad (17)$$

que toma el valor 1 en la posición j -ésima, a partir de la ecuación (15) obtenemos las expresiones (18) y (19) (para fuego estable y alarma indeseada respectivamente).

$$GSNR_p = \frac{D}{p} \sum_{i=1}^p \frac{1}{D} \left(\sum_{j=1}^D \mathbf{a}_i(j) \right)^2 = \frac{1}{p} \sum_{i=1}^p \left(\sum_{j=1}^D \mathbf{a}_i(j) \right)^2 \quad (18)$$

$$GSNR_p = \frac{D}{p} \sum_{i=1}^p \mathbf{a}_i^2(j) = \frac{\mathbf{P}_{jj} \cdot D}{p} \quad (19)$$

De la expresión (19) se deduce que la $GSNR_p$ para el modelo de señal correspondientes a las falsas alarmas depende de elemento j -ésimo de la diagonal principal de la matriz de proyección \mathbf{P} , de tal forma que el caso peor corresponderá con el vector de la forma (17), cuyo valor 1 se sitúa en la posición j -ésima para la cual \mathbf{P}_{jj} es máximo.

En la tabla I, podemos analizar las expresiones (18) y (19) para los diferentes valores posibles de $p=1$ a 6 y $\beta=0.1$, al evaluar el modelo de señal de fuego definido en (16) y el de alarma indeseada (17) con $j=3$.

$\beta = 0.1$	Fuego Estable $GSNR_p$	Alarma Indeseada $GSNR_p$
$p = 1$	5.99	1.08
$p = 2$	2.99	0.59
$p = 3$	2.00	0.75
$p = 4$	1.50	0.70
$p = 5$	1.20	0.73
$p = 6$	1.00	1.00

Tabla I: Valores de $GSNR_p$, para $D=6$

Si observamos el valor de $p=6$, que corresponde al caso del detector de energía, tanto el modelo de fuego estable como el modelo de alarma indeseada presenta una ganancia unidad, sin embargo, si analizamos los distintos grados de libertad, el parámetro $GSNR_p$ para fuego estable es siempre mucho mayor que al considerar la señal de tipo alarma indeseada.

A partir del análisis teórico presentado, vamos a analizar para los modelos de señal considerados cómo influyen los parámetros β y p en la probabilidad de detección tanto de fuegos descontrolados como de falsas alarmas.

2.2. Ajuste de β y p :

El objetivo del detector adaptado al subespacio consiste en ajustar β y p con el fin de reducir severamente la energía producida por los efectos ocasionales que conducirían a alarmas indeseadas dentro del subespacio delimitado por \mathbf{P} mientras tenemos que evitar la cancelación de señales producidas por fuegos descontrolados (tabla I).

A partir del análisis anterior y debido a la complejidad matemática para encontrar expresiones que relacionen las ecuaciones (18) y (19) con los parámetros de diseño β y p , (en [THO82] se presentan fórmulas para valores de p altos y β pequeños, que no es nuestro caso) vamos a evaluar mediante distintas curvas cómo influyen los parámetros citados en la probabilidad de detección de modelos de señal que tipifiquen fuegos descontrolados frente a modelos de señal correspondientes a

alarmas indeseadas, de esta forma dispondremos de criterios objetivos con los que elegir los parámetros de diseño del filtro adaptado al subespacio β y p .

Para las distintas simulaciones, consideramos que el ruido de fondo presente es Gaussiano multivariante de media cero y varianza unidad, de esta forma, el estadístico r , ecuación (1), presenta una distribución chi-cuadrado con p grados de libertad [SCH91][KAY98]. Con esta consideración y fijando una determinada probabilidad de falsa alarma (PFA) determinamos el umbral (T) necesario para decidir la presencia o no de alarmas. Una vez obtenidos los umbrales a diferentes probabilidades de falsa alarma, representaremos gráficamente la probabilidad de detección versus la probabilidad de falsa alarma para distintos valores de los parámetros β y p , al considerar únicamente modelos de señales s (de energía E_s) en el vector u , para ello [SCH91] sabemos que el estadístico r , ecuación (12), presenta una distribución chi-cuadrado no central cuyo parámetro de no centralidad es $s^T P s$.

Para la obtención de las diferentes graficas definimos el parámetro normalizado “betan” como,

$$\text{betan} = \frac{\beta}{1/2D} = 2\beta \cdot D \quad (20)$$

con D distancia máxima en la que basar la predicción.

En todos los casos la energía de la señal (E_s) es 24, lo que implica una $SNR=24$. En el modelo de señal de alarma indeseada ecuación (17), $j=3$.

En las figuras 3 a 9, se analiza como influye el parámetro “betan” fijado el parámetro p , para los modelos de fuegos descontrolados y de alarma indeseada definidos en la sección 2 del presente Capítulo. Si observamos las figuras 3 y 4 (donde $p=1$) concluimos que a medida que “betan” aumenta la probabilidad de detección de modelos de fuegos disminuye, mientras que por el contrario la probabilidad de detección de alarmas indeseadas aumenta, efecto que también se manifiesta si analizamos las parejas de figuras 5-6, 7-8 y la figura 9, en esta última

se observa como para $p=D=6$, los parámetros β y p no influyen ya que en este caso la matriz de proyección \mathbf{P} es la matriz identidad, es decir, no existe el filtro adaptado al subespacio de señal tal como hemos analizado teóricamente y en consecuencia todos los modelos de señal tanto de fuego como de alarma indeseada presentan la misma PD para una determinada SNR . Como conclusión podemos determinar que para conseguir los objetivos de minimizar la PDI y maximizar la PD frente a la que se obtendría al considerar un detector de energía conviene usar valores β pequeños.

En las figuras 3 a 9 podemos ver como a medida que p aumenta, la PD para modelos de fuegos disminuye, mientras aumenta la PD de alarmas indeseadas, esto nos lleva a pensar que hemos de considerar valores de p entorno a 1 ó a 2.

En la figura 10, se observa como a partir de $\beta < 1$ dicho parámetro apenas influye en la PD, en consecuencia podemos fijar una β del orden de 1.

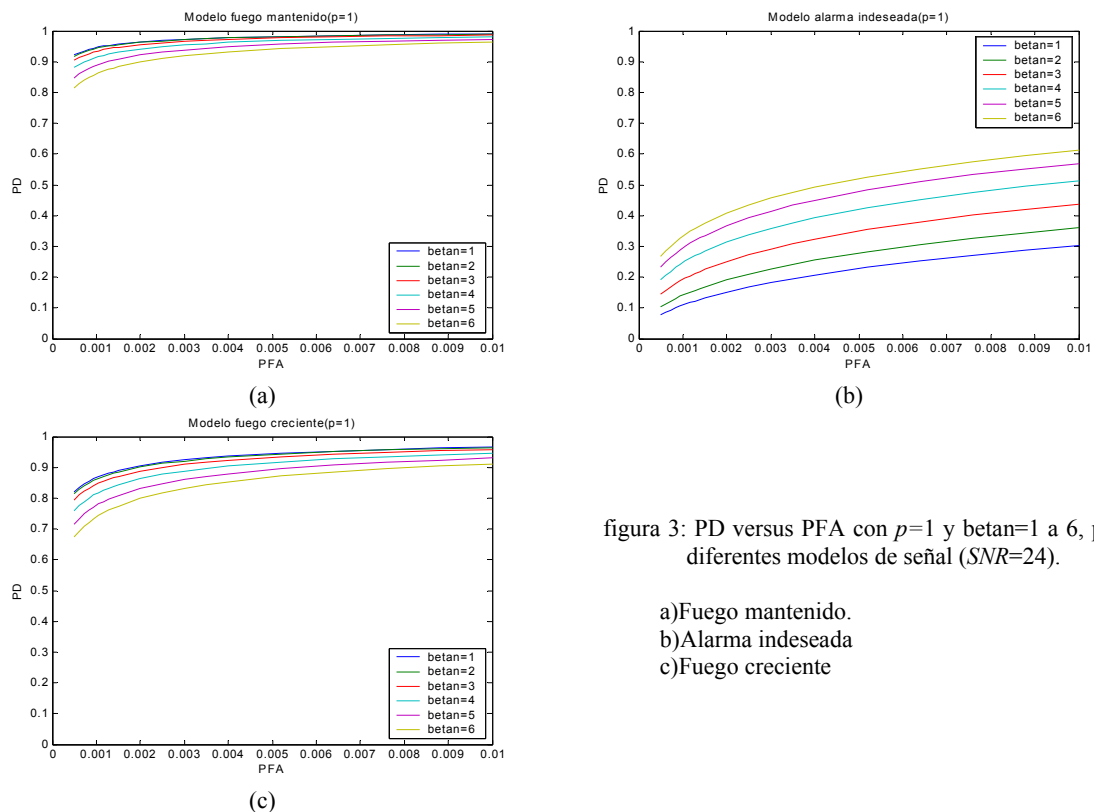
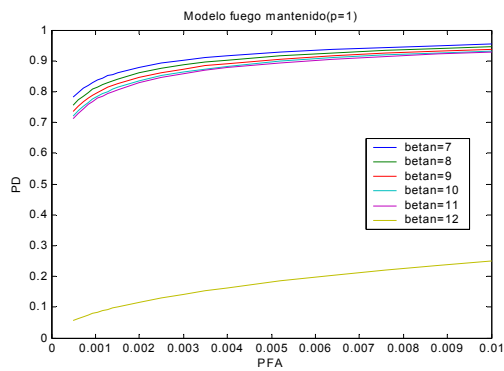
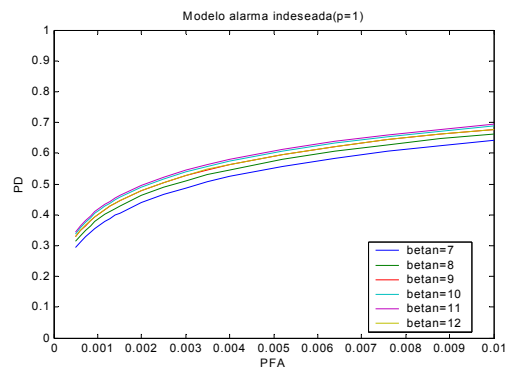


figura 3: PD versus PFA con $p=1$ y $\beta=1$ a 6, para diferentes modelos de señal ($SNR=24$).

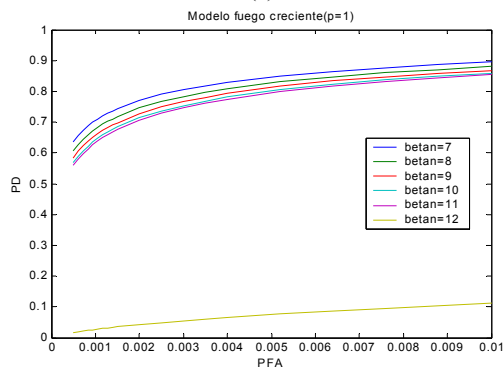
- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente



(a)



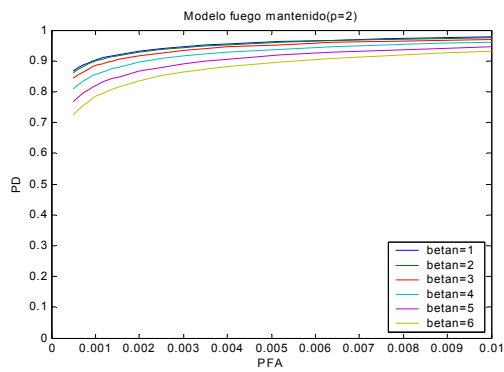
(b)



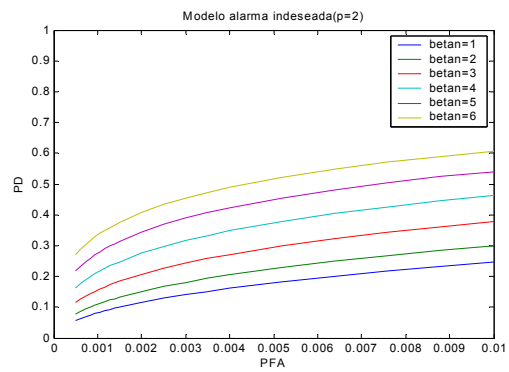
(c)

figura 4: PD versus PFA con $p=1$ y $\text{betan}=7$ a 12 , para diferentes modelos de señal ($\text{SNR}=24$).

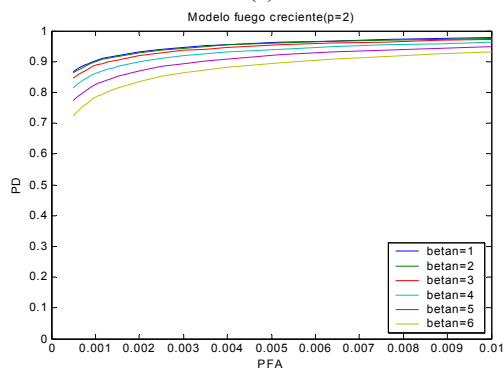
- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente



(a)



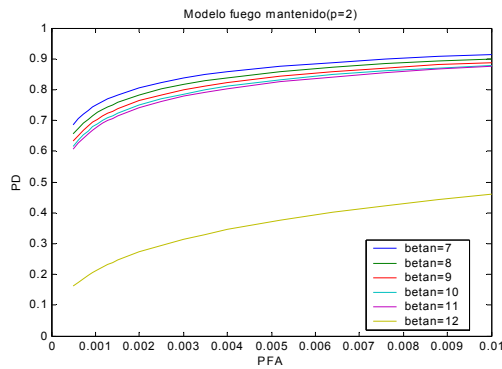
(b)



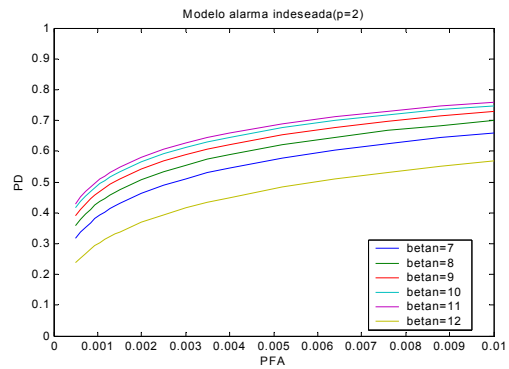
(c)

figura 5: PD versus PFA con $p=2$ y $\text{betan}=1$ a 6 , para diferentes modelos de señal ($\text{SNR}=24$).

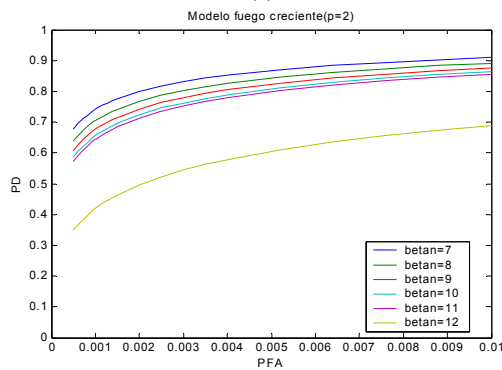
- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente



(a)



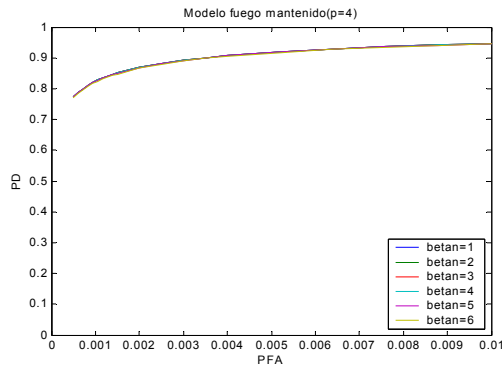
(b)



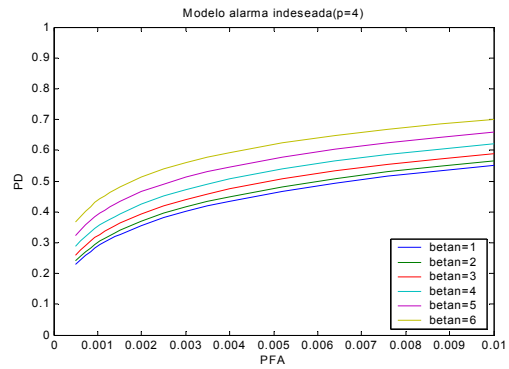
(c)

figura 6: PD versus PFA con $p=2$ y $\text{betan}=7$ a 12 , para diferentes modelos de señal ($\text{SNR}=24$).

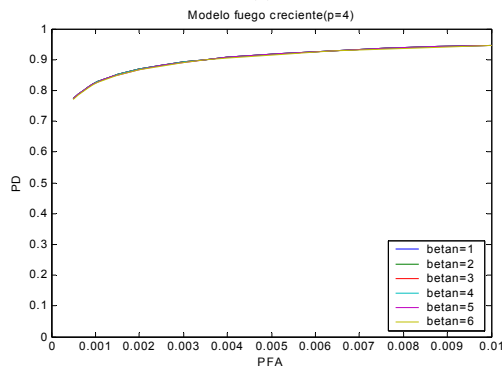
- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente



(a)



(b)



(c)

figura 7: PD versus PFA con $p=4$ y $\text{betan}=1$ a 6 , para diferentes modelos de señal ($\text{SNR}=24$).

- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

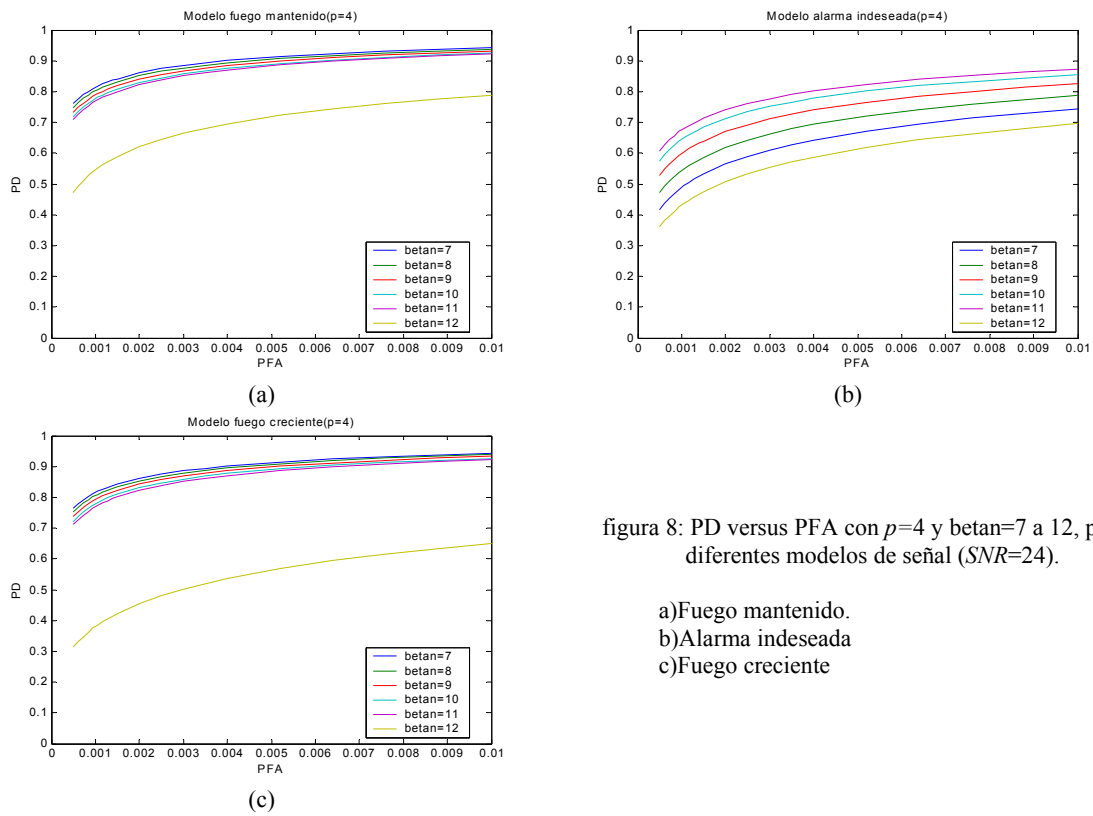


figura 8: PD versus PFA con $p=4$ y $\beta=7$ a 12 , para diferentes modelos de señal ($SNR=24$).

- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

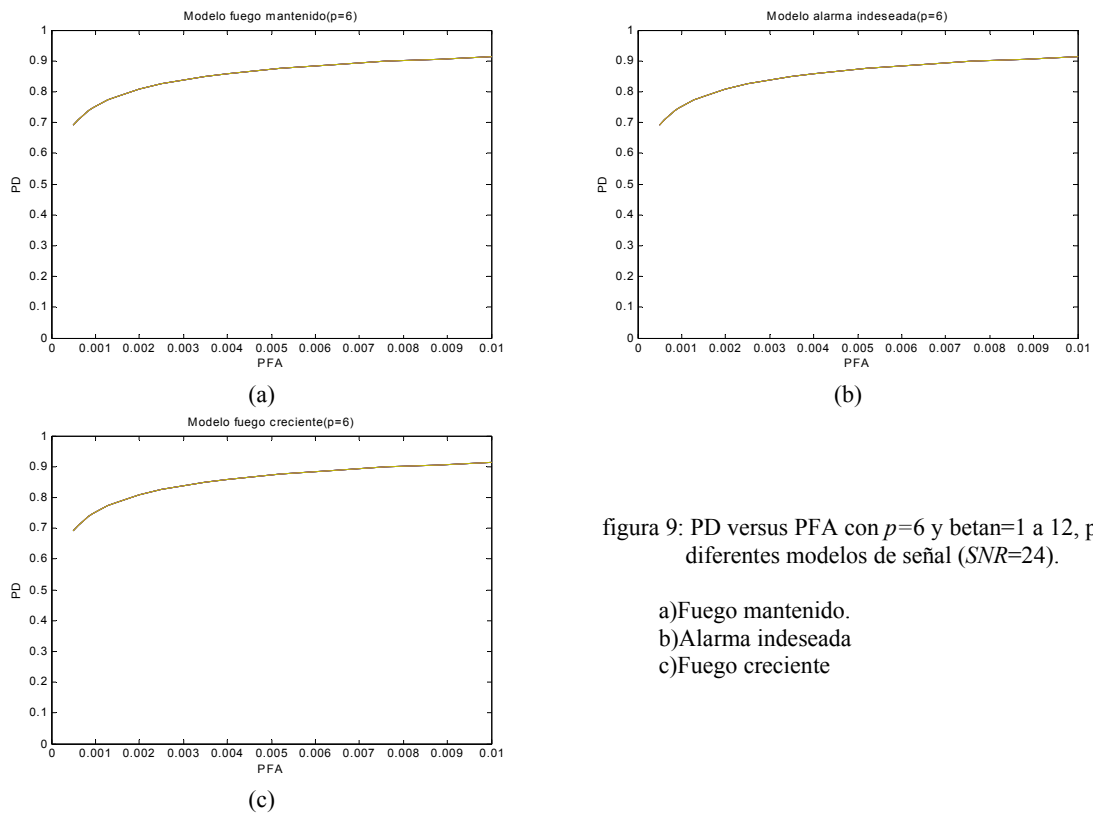


figura 9: PD versus PFA con $p=6$ y $\beta=1$ a 12 , para diferentes modelos de señal ($SNR=24$).

- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

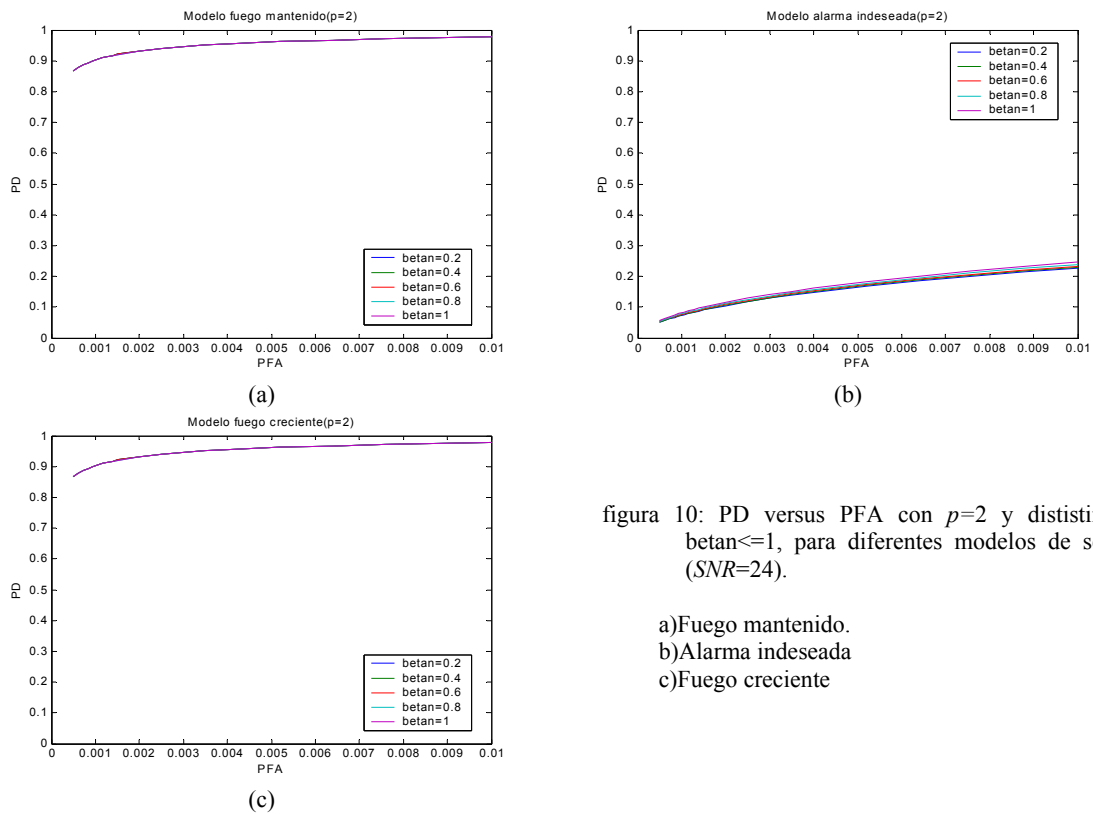


figura 10: PD versus PFA con $p=2$ y distintos $\beta \leq 1$, para diferentes modelos de señal ($SNR=24$).

- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

Las figuras 11 a 13 nos permiten determinar el valor adecuado del parámetro p (número de autovectores dominantes de Slepian considerados), si analizamos con detalle las figuras 11b, 12b y 13b, observamos que la PDI se minimiza para $p=2$, al igual que para este mismo valor conseguimos la mayor PD para el modelo de fuego creciente (figuras 11c, 12c y 13c), mientras que el valor óptimo para el fuego estable (figuras 11a, 11b y 11c) corresponde al valor $p=1$. Esto nos lleva a determinar cómo un valor adecuado $p=2$.

Si comparamos las diferentes gráficas de las figuras 11 a 13, se confirma el interés de ajustar β con valores del orden de

$$\beta \cong \frac{1}{2D} \tag{21}$$

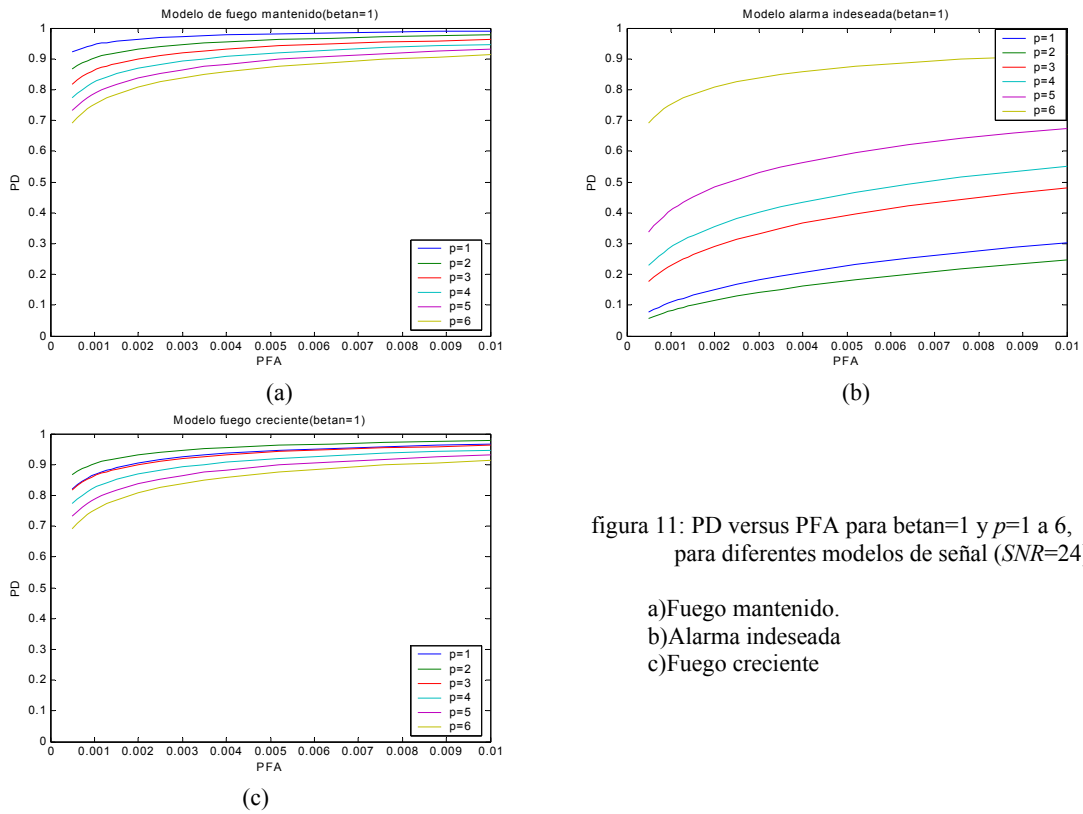


figura 11: PD versus PFA para $\beta=1$ y $p=1$ a 6, para diferentes modelos de señal ($SNR=24$).

- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

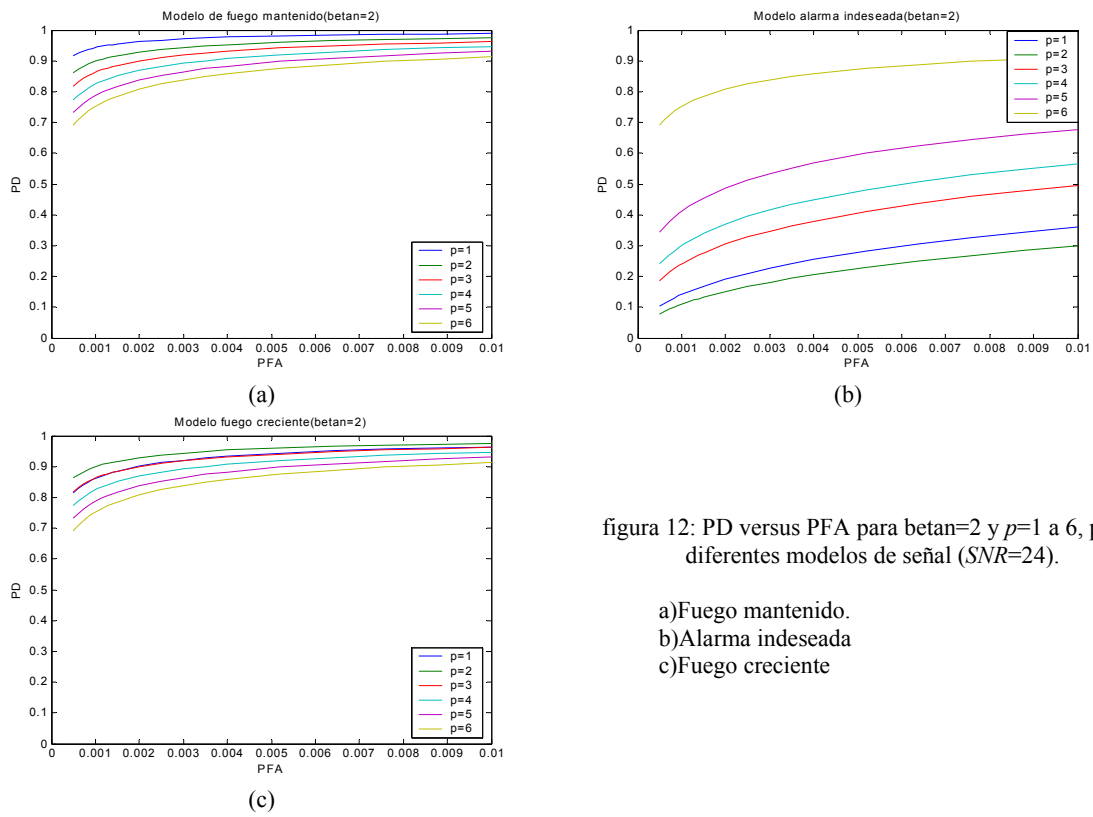


figura 12: PD versus PFA para $\beta=2$ y $p=1$ a 6, para diferentes modelos de señal ($SNR=24$).

- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

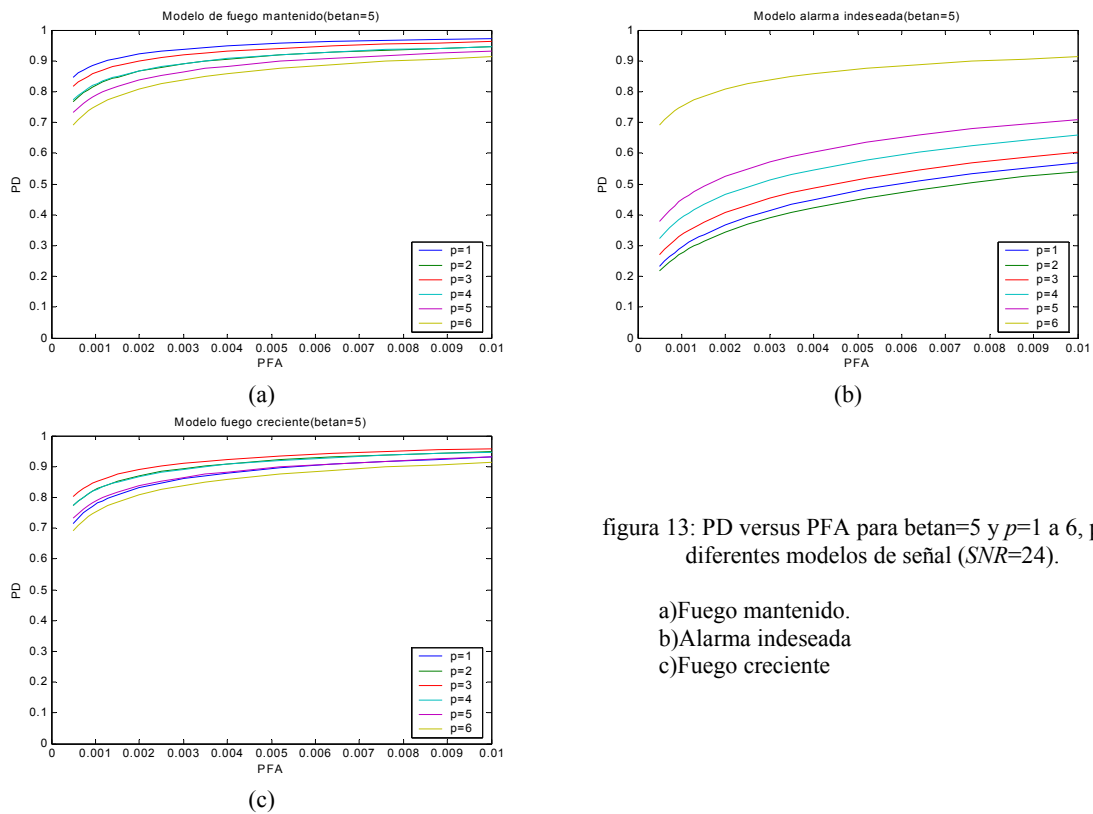


figura 13: PD versus PFA para $\beta=5$ y $p=1$ a 6 , para diferentes modelos de señal ($SNR=24$).

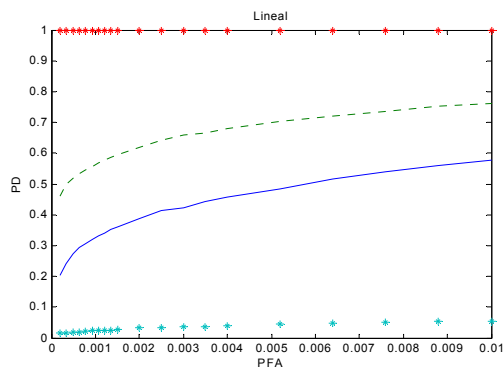
- a) Fuego mantenido.
- b) Alarma indeseada
- c) Fuego creciente

2.3. Validación del ajuste a los bloques A, B y C:

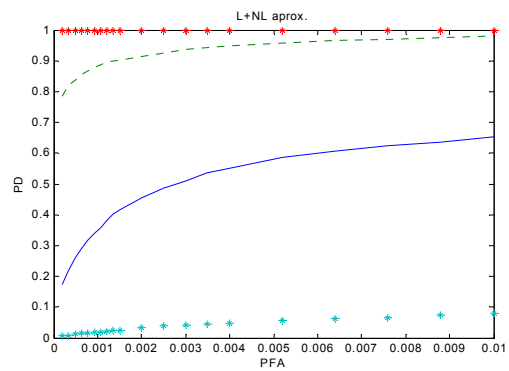
Una vez establecidos criterios para el ajuste de los parámetros β y p a partir de los diferentes análisis realizados, vamos a determinar la validez de los resultados obtenidos con los bloques de información usados a lo largo de la presente tesis. Se han usado los mismos modelos de señal considerados en el Capítulo IV para modelizar el fuego creciente y la alarma indeseada (ambas señales de energía 30). En el diseño del bloque detector se han considerado los siguientes valores de “betan” y p (en todos los casos $N=2, D=6$).

$$p = 1 \begin{cases} \text{betan} = 1 \\ \text{betan} = 2 \end{cases}$$

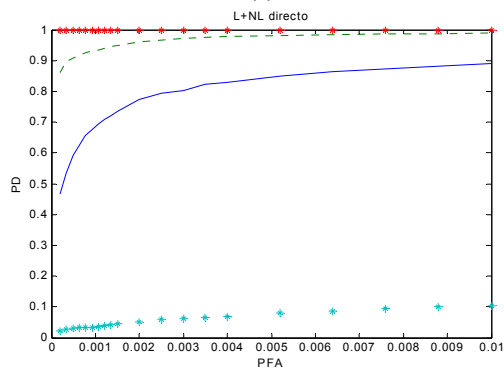
$$p = 2 \begin{cases} \text{betan} = 1 \\ \text{betan} = 2 \end{cases}$$



(a)

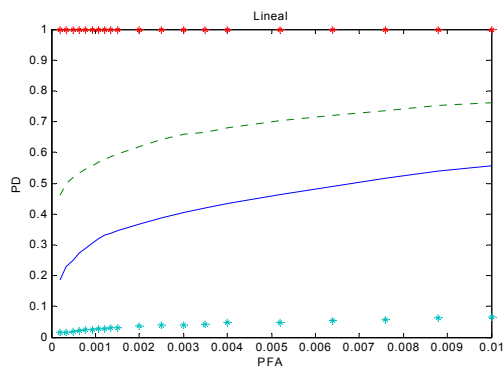


(b)

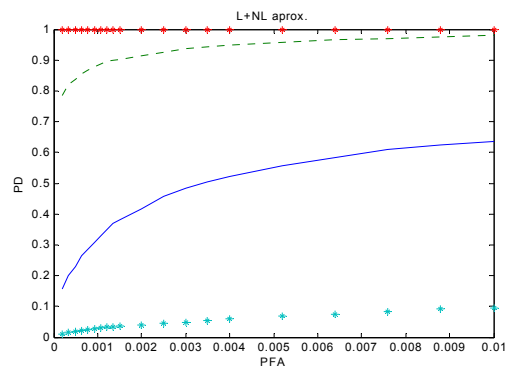


(c)

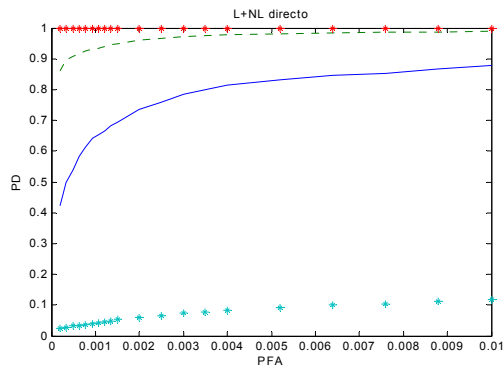
figura 14: PD frente a PFA, con $p=1$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque A.**



(a)

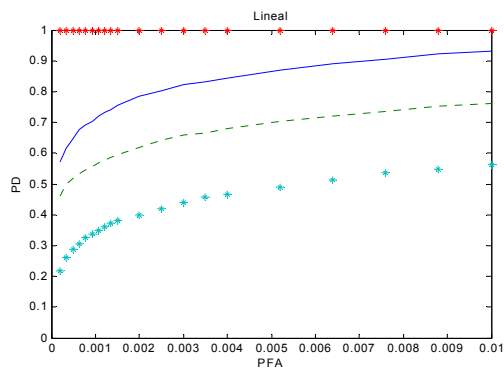


(b)

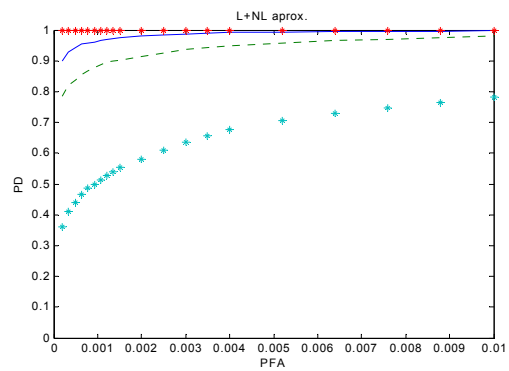


(c)

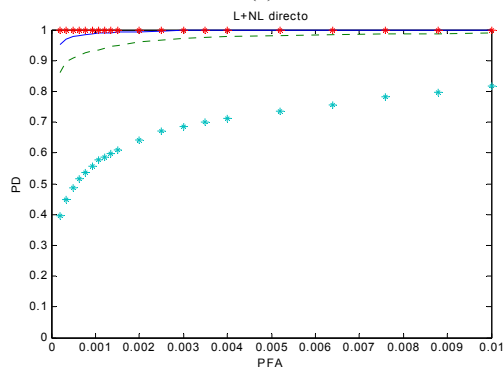
figura 15: PD frente a PFA, con $p=1$ y $\beta=2$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque A.**



(a)

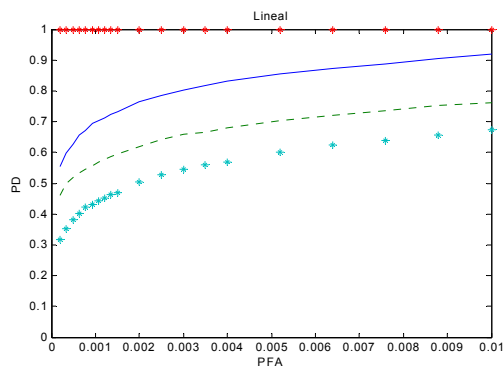


(b)

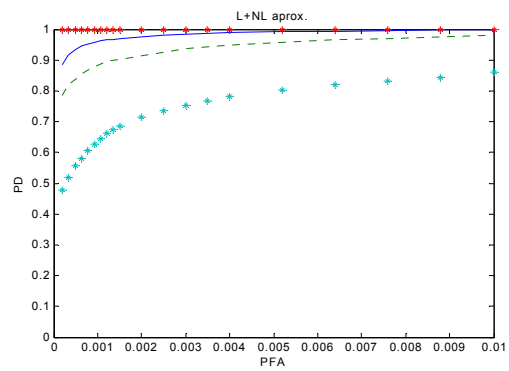


(c)

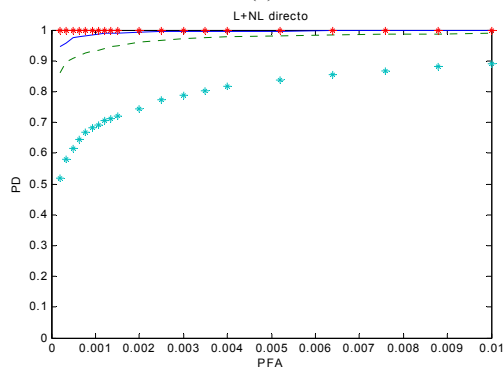
figura 16: PD frente a PFA, con $p=2$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque A.**



(a)

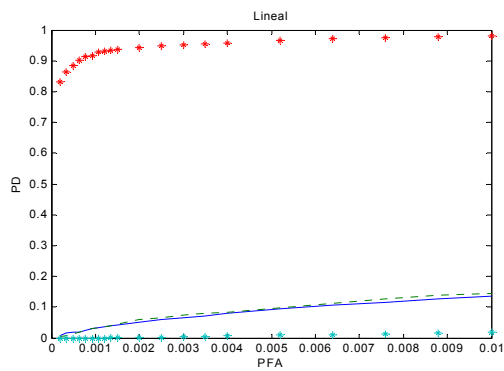


(b)

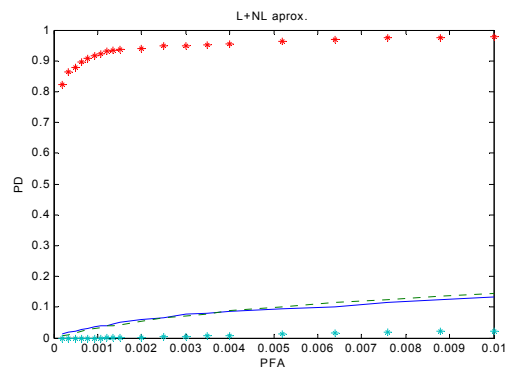


(c)

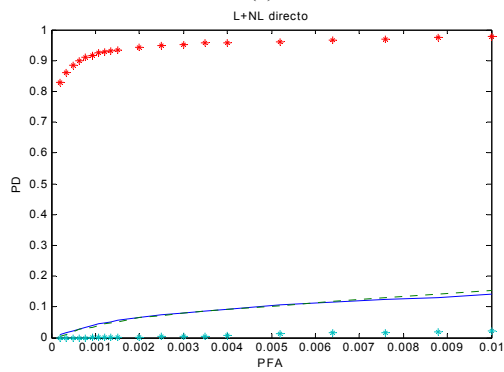
figura 17: PD frente a PFA, con $p=2$ y $\beta=2$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque A.**



(a)

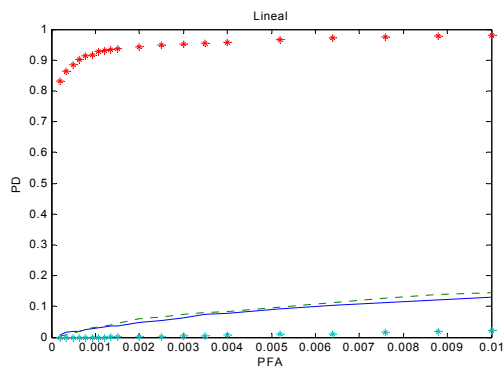


(b)

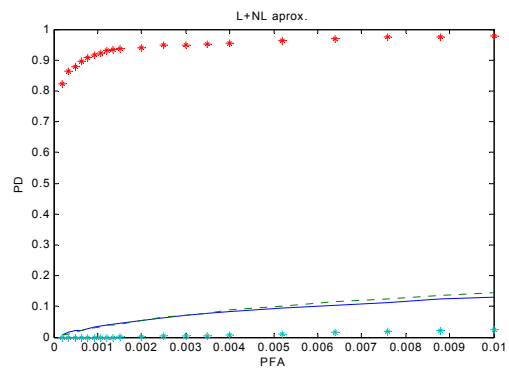


(c)

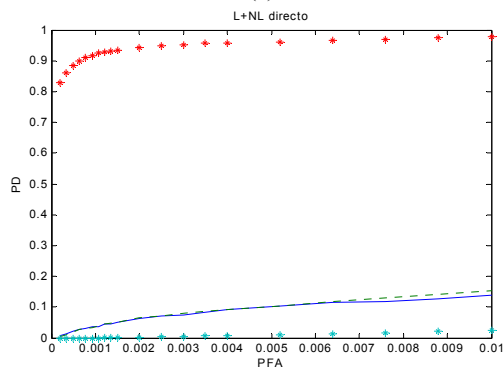
figura 18: PD frente a PFA, con $p=1$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque B.**



(a)

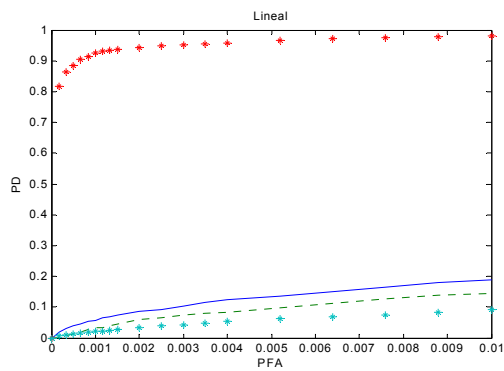


(b)

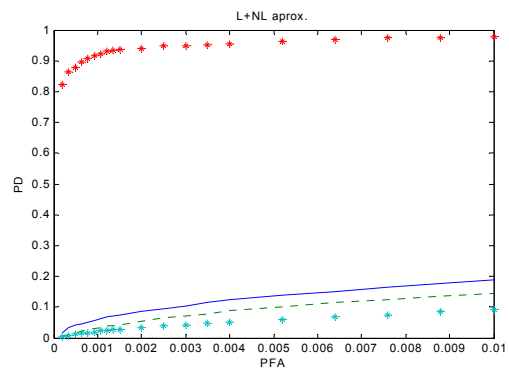


(c)

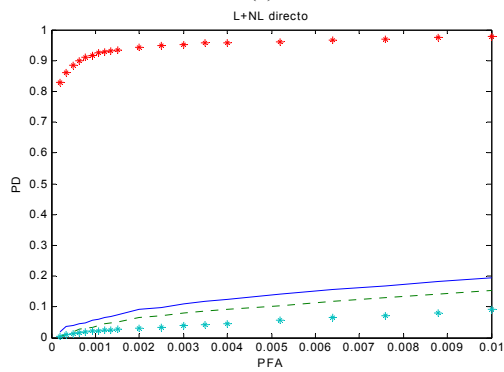
figura 19: PD frente a PFA, con $p=1$ y $\beta=2$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque B.**



(a)

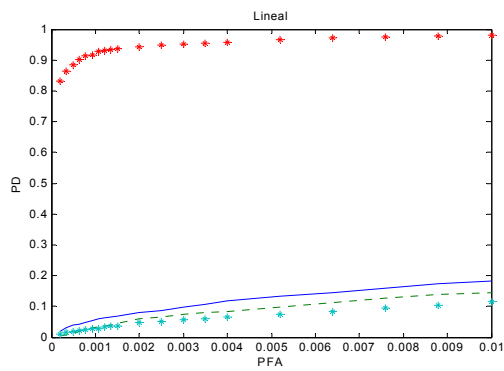


(b)

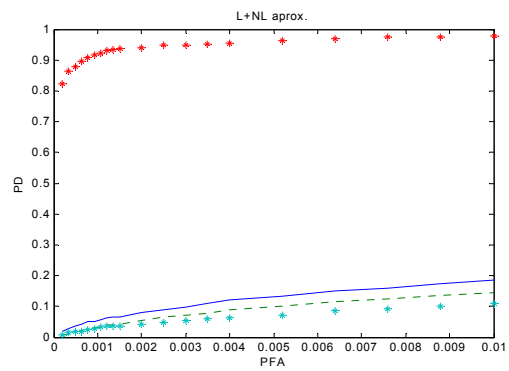


(c)

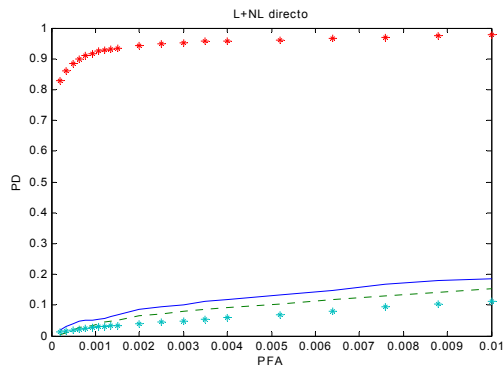
figura 20: PD frente a PFA, con $p=2$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque B.**



(a)

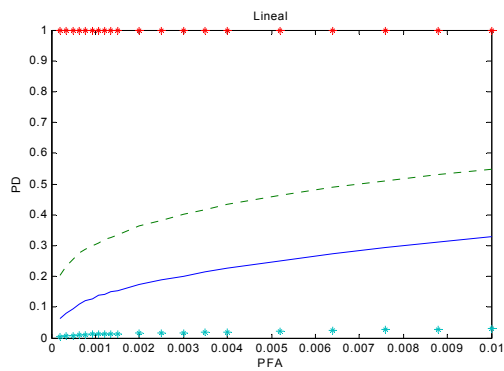


(b)

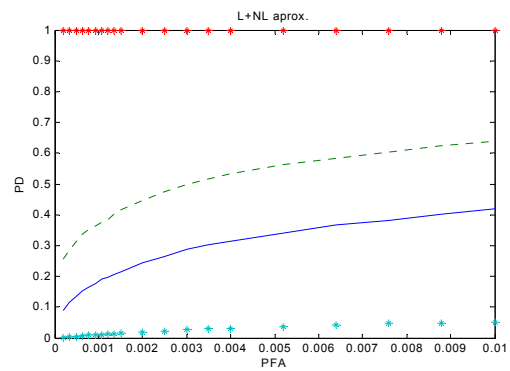


(c)

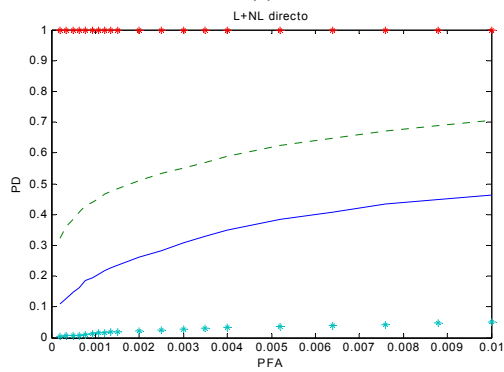
figura 21: PD frente a PFA, con $p=2$ y $\beta=2$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque B.**



(a)

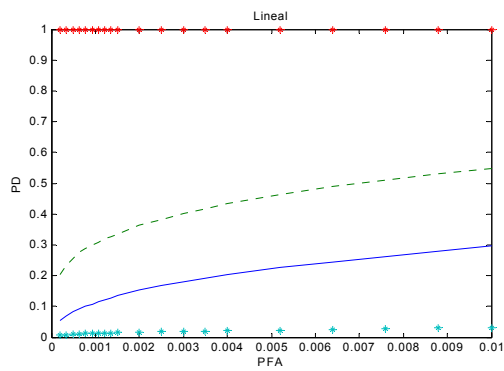


(b)

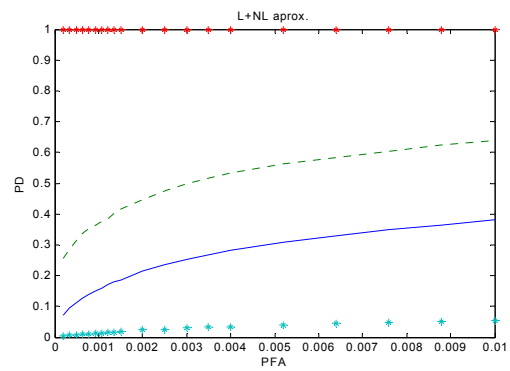


(c)

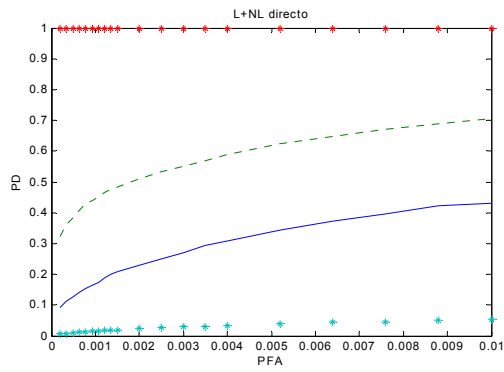
figura 22: PD frente a PFA, con $p=1$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque C.**



(a)

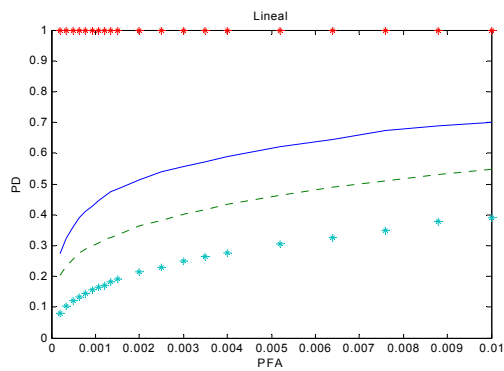


(b)

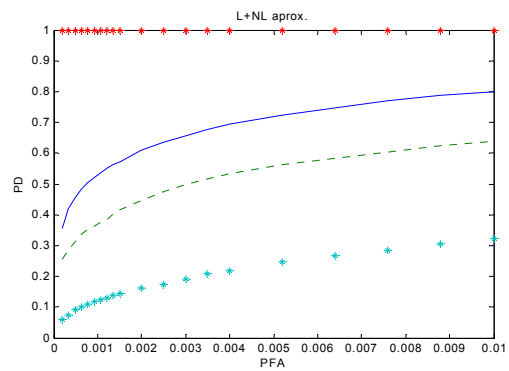


(c)

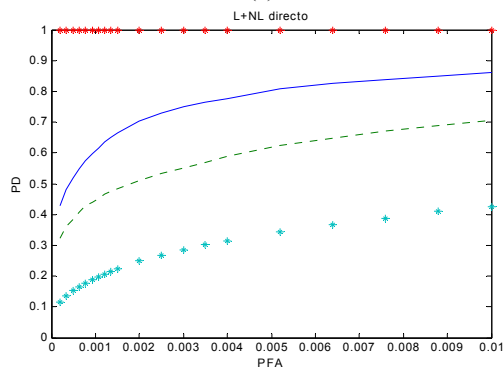
figura 23: PD frente a PFA, con $p=1$ y $\beta=2$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque C.**



(a)

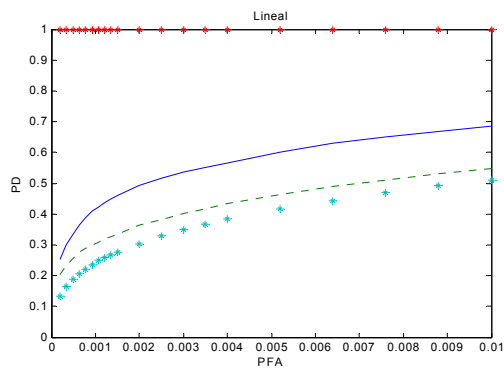


(b)

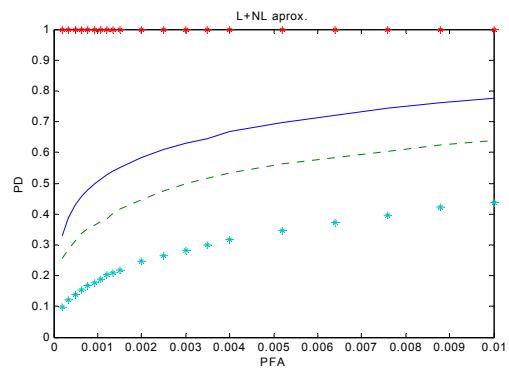


(c)

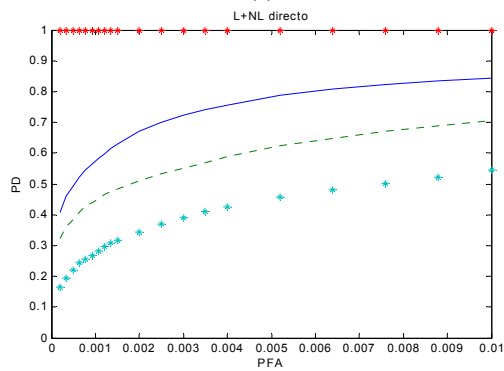
figura 24: PD frente a PFA, con $p=2$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque C.**



(a)



(b)



(c)

figura 25: PD frente a PFA, con $p=2$ y $\beta=2$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, detector adaptado al subespacio (línea asteriscos inferior), detector de energía (línea asteriscos superior). ($N=2, D=6$) **Bloque C.**

Si analizamos con detalle las figuras 14, 15, 18, 19, 22 y 23 ($p=1$) observamos, como para el modelo de fuego creciente, en todos los casos, el considerar un solo grado de libertad en la construcción de la matriz de proyección \mathbf{P} , empeora la PD (línea sólida) frente al uso del detector de energía (línea de puntos), si bien también reduce drásticamente la PDI (curva asteriscos inferior) frente al uso del detector de energía. Por otro lado, si comparamos la PDI obtenida en estas figuras ($p=1$) con las figuras donde $p=2$ (figuras 16, 17, 20, 21, 24, 25) la reducción parece notable, aunque es aparente. La explicación es sencilla, si recordamos la ecuación (19), la $GSNR_p$ depende del valor j -ésimo de la matriz de proyección (P_{jj}) (valor que corresponde a la posición donde se sitúa la muestra de amplitud 1 en el vector de señal correspondiente a una alarma indeseada, expresión (17)). En nuestro caso al haber considerado $j=5$, resulta que, $P_{55}^2 \gg P_{55}^1$, donde P_{jj}^p indica el elemento j -ésimo de la matriz \mathbf{P} con p grados de libertad (ver figura 26). Observándose incluso (figura 26a) como este hecho puede provocar que la PDI sea mayor que la PD, lo que nos lleva a tener que incluir otros esquemas de detección con los que corregir estas posibles situaciones.

En la figura 27, se compara, para el **bloque C**, el resultado de aplicar el diseño seleccionado a una señal decreciente (modelo alarma indeseada) frente al modelo de señal creciente usado para simular un fuego creciente, en ambos casos la energía de la señal es 30. Del resultado obtenido se deduce, que el esquema propuesto para la detección no es capaz de diferenciar estos dos tipos de señal. Este es el motivo, por el que a continuación se analice el detector de crecimiento (figura 2) como esquema de detección que nos permita distinguir tendencias crecientes de la señal (posibles fuegos) frente a tendencias decrecientes (posibles alarmas indeseadas) en el vector de muestras a la entrada del filtro adaptado.

Finalmente, de las diferentes gráficas obtenidas se observa como al incrementar el parámetro betan (tal como vimos en las simulaciones teóricas) tanto la PD como la PDI empeoran. Como conclusión, deducimos nuevamente que los valores adecuados son: $p=2$ y $\text{betan}=1$.

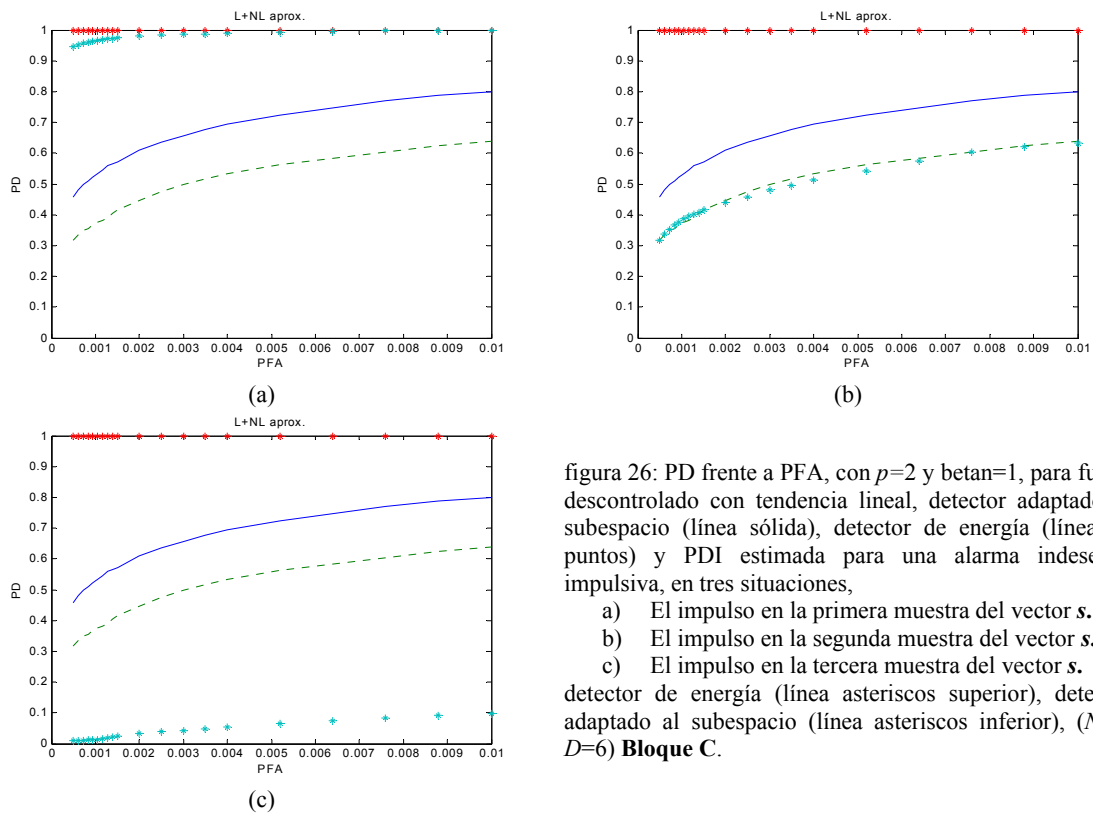


figura 26: PD frente a PFA, con $p=2$ y $\beta=1$, para fuego descontrolado con tendencia lineal, detector adaptado al subespacio (línea sólida), detector de energía (línea de puntos) y PDI estimada para una alarma indeseada impulsiva, en tres situaciones,

- a) El impulso en la primera muestra del vector s .
- b) El impulso en la segunda muestra del vector s .
- c) El impulso en la tercera muestra del vector s .

detector de energía (línea asteriscos superior), detector adaptado al subespacio (línea asteriscos inferior), ($N=2$, $D=6$) **Bloque C.**

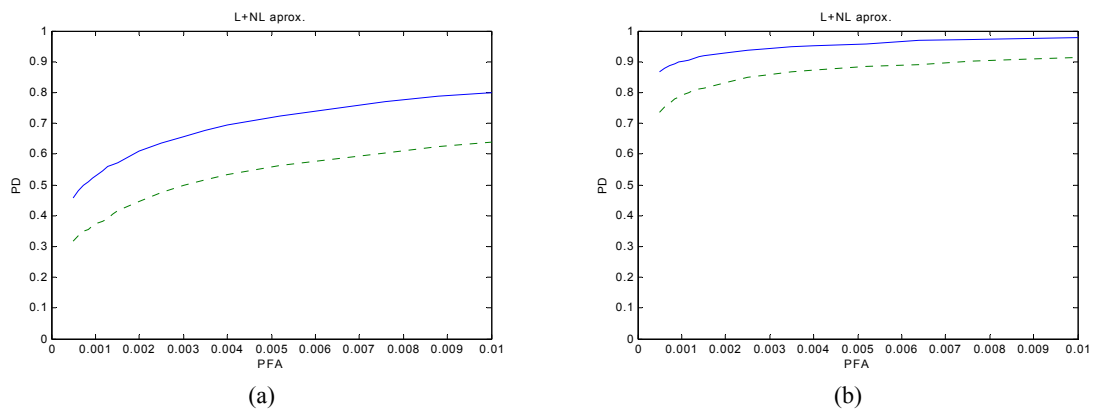


figura 27: PD frente a PFA, con $p=2$ y $\beta=1$, para fuego descontrolado con tendencia lineal, ($N=2$, $D=6$) **Bloque C.**

- a) Tendencia lineal creciente. (Modelo fuego descontrolado).
- b) Tendencia lineal decreciente (Modelo alarma indeseada).

3. Detector de crecimiento:

En esta sección vamos a caracterizar el denominado “detector de crecimiento”, (figura 2) con el propósito de determinar tendencias crecientes (lineales o no lineales) en el estadístico c (entrada filtro adaptado) frente a otros tipos de tendencias, de esta forma pretendemos mejorar el esquema detector presentado en la presente tesis, al incluir en la decisión los resultados de extraer información sobre el crecimiento o no en el vector e , con ello pretendemos mejorar los problemas del detector adaptado al subespacio analizados en las figuras 26 y 27.

Tal como podemos ver en la figura 2, el detector de crecimiento se basa en aplicar al vector e un número determinado de derivadas (n) con el fin de que a la salida del mismo tengamos para la componente de señal una respuesta continua (por ejemplo, para un fuego modelado con una señal de tendencia lineal creciente, la matriz diferenciadora Q debe realizar simplemente una derivación de la señal de entrada, para obtener la componente continua, en este caso, $n=1$).

Debido a que el ruido z' presente en el $Q^{(n)}e$ presenta una cierta matriz de correlación $R_{z'z'}$, de [SCH91] hemos de aplicar el filtro adaptado a los vectores previamente pre-blanqueados por la matriz de autocorrelación, tal como se observa en la figura 2, de forma similar a como hacíamos en el esquema del detector adaptado al subespacio de señal.

La señal usada para aplicar el filtro adaptado corresponde a la señal continua de energía 1, previamente pre-blanqueada, en ésta se ha tenido en cuenta que según el orden de derivación (n), la dimensión del vector r será $D-n$.

3.1. Determinación de la distribución del estadístico:

Para el diseño del detector de crecimiento hemos de establecer la matriz diferenciadora $Q^{(n)}$, para ello definimos la matriz Q_D ($D \times D-1$), como

$$\mathbf{Q}_D = \underbrace{\begin{bmatrix} -1 & 1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 1 & \dots & \dots & 0 \\ & & \vdots & & & \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}}_D \Bigg\}^{D-1} \quad (21)$$

De tal forma, al pasar el vector \mathbf{e} por la matriz \mathbf{Q}_D obtenemos su derivada primera ($\mathbf{r}_d = \mathbf{Q}_D \cdot \mathbf{e}$). A partir de la matriz (21) construimos $\mathbf{Q}^{(n)}$ para que su salida corresponda a la derivada n -ésima del vector $\mathbf{e} = \mathbf{s} + \mathbf{z}$. El orden (n) de la matriz diferenciadora se fijará para conseguir a la salida de la misma una componente continua.

$$\mathbf{Q}^{(n)} = \mathbf{Q}_{D-n+1} \cdots \mathbf{Q}_{D-1} \cdot \mathbf{Q}_D \quad (n \text{ veces}) \quad (22)$$

de esta forma para conseguir la componente continua, n tomará el valor 1 para un crecimiento lineal, 2 para un crecimiento cuadrático, etc.

Si evaluamos la expresión (22), para distintos valores de n , encontramos una expresión general para determinar el elemento $q_{ij}^{(n)}$ de la matriz diferenciadora en función de los parámetros D y n .

$$q_{ij}^{(n)} = \begin{cases} \binom{n}{j-i} (-1)^{n+j-i} & \text{para } i \leq j \leq n+i \\ 0 & \text{resto} \end{cases} \quad (23)$$

con i (fila i -ésima) variando de 1 a $D-n$, y j (columna j -ésima) variando de 1 a D , donde,

$$\binom{n}{j-i} = \frac{n!}{(j-i)! (n-j+i)!} \quad (24)$$

En consecuencia, finalmente estaremos interesados en detectar una componente continua s' en un fondo de ruido z' (figura 2).

$$\mathbf{r}_d = \mathbf{Q}^{(n)} \cdot \mathbf{e} = \mu \mathbf{Q}^{(n)} \mathbf{s} + \mathbf{Q}^{(n)} \mathbf{z} = \mu \mathbf{s}' + \mathbf{z}' \quad (25)$$

Para ello el detector debe estar adaptado a la señal s_e (componente continua, de dimensión $D-n$).

$$\mathbf{s}_e = \frac{1}{\sqrt{D-n}} \underbrace{[1, 1, \dots, 1]}_{D-n}^T \quad (26)$$

Previo al filtro adaptado a la componente continua, el vector \mathbf{r}_d debe ser preblanqueado por la matriz de autocorrelación $\mathbf{R}_{z'z'}$ [SCH91].

$$\mathbf{R}_{z'z'} = E[\mathbf{z}' \cdot \mathbf{z}'^T] = E[\mathbf{Q}^{(n)} \mathbf{z} \cdot \mathbf{z}^T \mathbf{Q}^{(n)T}] = \mathbf{Q}^{(n)} E[\mathbf{z} \cdot \mathbf{z}^T] \mathbf{Q}^{(n)T} = \mathbf{Q}^{(n)} \mathbf{R}_{zz} \mathbf{Q}^{(n)T} \quad (27)$$

El detector decidirá entre alarma o no alarma, en función del estadístico c ,

$$c = \mathbf{r}_d^T \mathbf{R}_{z'z'}^{-1} \mathbf{R}_{z'z'}^{-1} \mathbf{s}_e = \mathbf{e}^T \mathbf{Q}^{(n)T} \cdot (\mathbf{Q}^{(n)} \mathbf{R}_{zz} \mathbf{Q}^{(n)T})^{-1} \cdot \mathbf{s}_e \quad (28)$$

Si z (ruido residual) presenta una distribución multivariable Gaussiana de media cero y varianza unidad, el estadístico c presenta una distribución multivariable Gaussiana de media \sqrt{SNR} y varianza 1.

$$c : N[\sqrt{SNR}, 1] \quad (29)$$

siendo SNR ,

$$SNR = \frac{\mathbf{s}^T \mathbf{Q}^{(n)T} \mathbf{R}_{zz}^{-1} \mathbf{Q}^{(n)} \mathbf{s}_e}{\mathbf{s}_e^T \mathbf{R}_{z'z'}^{-1} \mathbf{s}_e} \quad (30)$$

Lo anterior implica que podemos ajustar fácilmente el umbral del detector de crecimiento para conseguir una cierta PFA.

3.2. Experimentos:

El esquema de la figura 2 ha sido aplicado a los tres bloques de datos reales **A**, **B** y **C**. En la realización de los experimentos se han tenido en cuenta las siguientes consideraciones, tal como hicimos para el detector adaptado al subespacio de señal. En primer lugar hemos simulado fuego descontrolado sumando a todos los posibles vectores z (ruido de fondo) una tendencia lineal creciente de energía 30. Para la simulación de alarmas indeseadas hemos considerado dos situaciones, en primer lugar, una tendencia lineal decreciente de energía 30, sumada a todos los posibles vectores z (estamos interesados en resolver la problemática surgida en la figura 27), esta tendencia decreciente es simulada por el mismo vector que el considerado para la simulación de la tendencia creciente pero invirtiéndolo temporalmente y en segundo lugar hemos simulado una alarma indeseada impulsiva añadido solamente al quinto elemento del vector z un valor igual a la $\sqrt{30}$ (con la que pretendemos validar el detector ante situaciones en las que no tengamos ningún tipo de tendencia).

El valor del umbral lo fijamos para obtener una determinada probabilidad de falsa alarma. Esto es sencillo si consideramos lo analizado en la sección 3.2, donde si asumimos que el ruido presente tiene una distribución Gaussiana de media cero y varianza unidad, el estadístico c presenta la misma distribución.

En las figuras 28, 29 y 30 se observa como la PDI correspondiente a la tendencia lineal decreciente, es prácticamente cero (en el **bloque B**, empieza a incrementarse a partir de una PFA del orden de 0.5). De esta forma con el modelo de detector de crecimiento eliminamos drásticamente las alarmas indeseadas del tipo tendencias decrecientes que el detector adaptado al subespacio de energía era incapaz de resolver (ver figura 27).

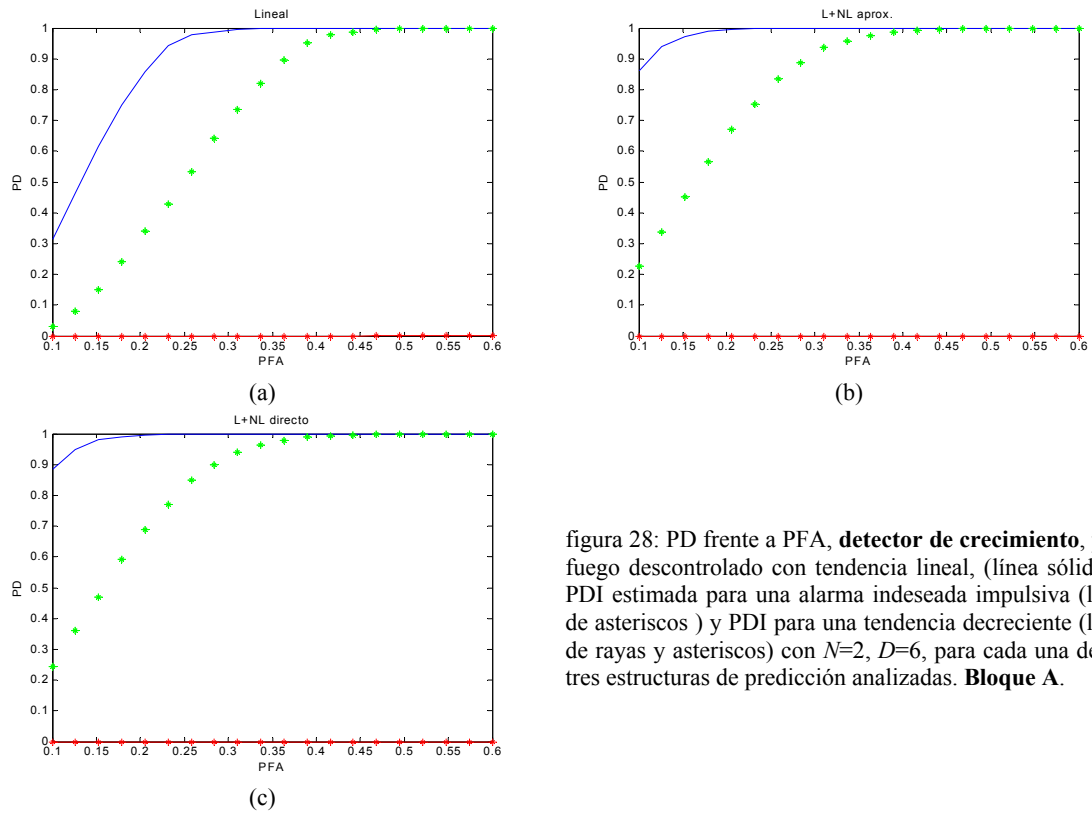


figura 28: PD frente a PFA, **detector de crecimiento**, para fuego descontrolado con tendencia lineal, (línea sólida) y PDI estimada para una alarma indeseada impulsiva (línea de asteriscos) y PDI para una tendencia decreciente (línea de rayas y asteriscos) con $N=2$, $D=6$, para cada una de las tres estructuras de predicción analizadas. **Bloque A.**

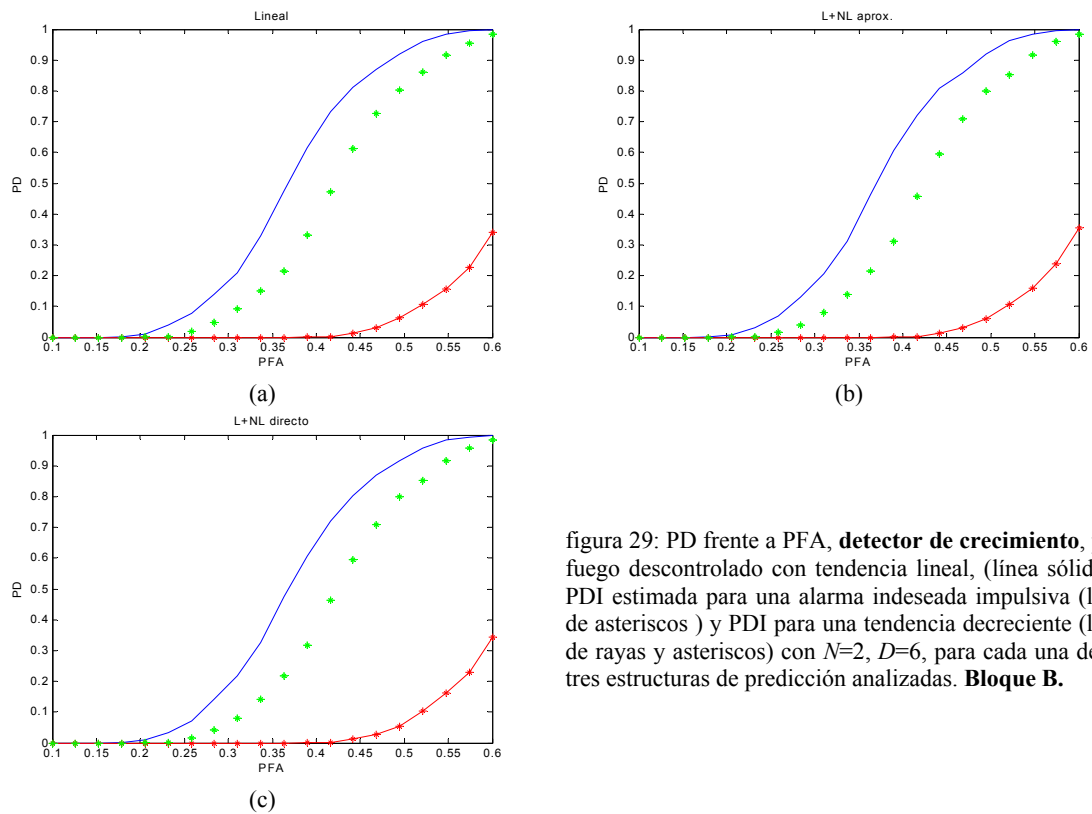


figura 29: PD frente a PFA, **detector de crecimiento**, para fuego descontrolado con tendencia lineal, (línea sólida) y PDI estimada para una alarma indeseada impulsiva (línea de asteriscos) y PDI para una tendencia decreciente (línea de rayas y asteriscos) con $N=2$, $D=6$, para cada una de las tres estructuras de predicción analizadas. **Bloque B.**

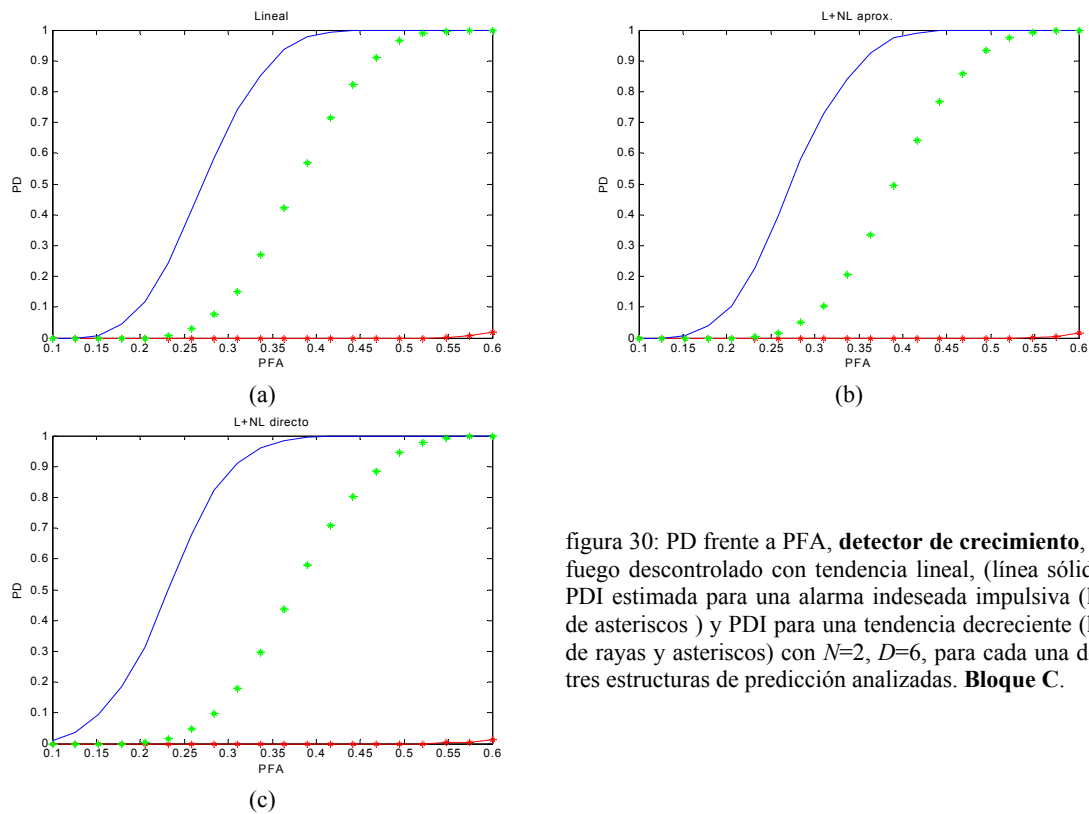


figura 30: PD frente a PFA, **detector de crecimiento**, para fuego descontrolado con tendencia lineal, (línea sólida) y PDI estimada para una alarma indeseada impulsiva (línea de asteriscos) y PDI para una tendencia decreciente (línea de rayas y asteriscos) con $N=2$, $D=6$, para cada una de las tres estructuras de predicción analizadas. **Bloque C**.

En las figuras 28, 29, 30 observamos como el detector de crecimiento no elimina los efectos de la alarma indeseada impulsiva. Si analizamos las alarmas indeseadas impulsivas con detalle observamos que su PDI depende del instante en que se produce el impulso, efectos que analizamos en los ejemplos de la figura 31. Estos ejemplos han sido realizados para el **bloque C**, considerando la estructura no lineal en el predictor diseñada por el método directo, en tres situaciones según el instante del impulso.

En consecuencia el detector de crecimiento no elimina adecuadamente las alarmas indeseadas del tipo impulsivo, cosa que en mayor grado, tal como hemos analizado en la sección 2, sí realizaba el detector adaptado al subespacio de señal.

De los resultados obtenidos concluimos que con PFA mayor de 0.5, el detector de crecimiento detecta siempre las situaciones de fuego descontrolados en las tres estructuras de bloques considerados. Por otro lado, el límite superior será fijado para

que la PDI de alarmas indeseadas de tendencias decrecientes este por debajo de un cierto valor. Con estas consideraciones fijamos la PFA adecuada a la aplicación.

Los resultados nos inclinan a pensar que ambos detectores (adaptado al subespacio y el de crecimiento) deben ser usados conjuntamente para resolver adecuadamente los dos tipos de alarmas indeseadas consideradas.

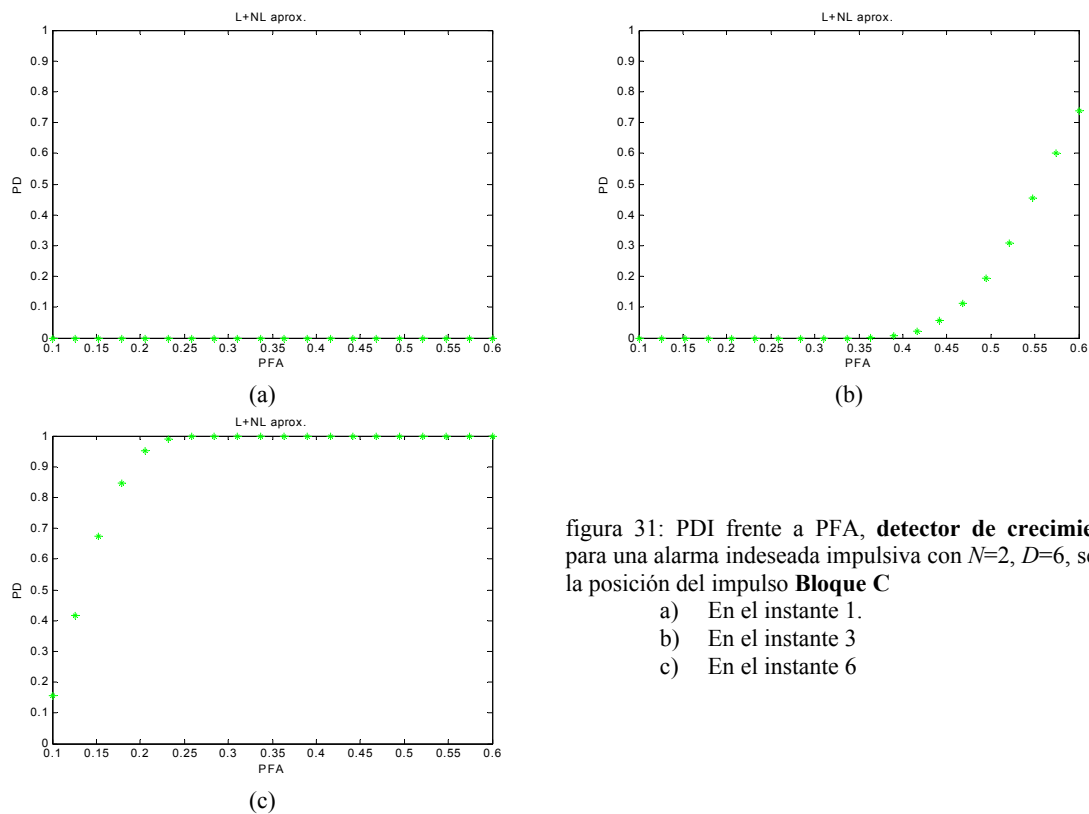


figura 31: PDI frente a PFA, **detector de crecimiento**, para una alarma indeseada impulsiva con $N=2$, $D=6$, según la posición del impulso **Bloque C**

- a) En el instante 1.
- b) En el instante 3
- c) En el instante 6

4. Análisis conjunto de ambos tipos de detectores:

De los resultados anteriores concluimos que podemos mejorar la decisión de alarma o no alarma, considerando un bloque de detección formado por el detector de crecimiento más el detector adaptado al subespacio, de esta forma, el sistema implementado considerará la existencia de alarmas cuando éstas se hayan producido en los dos detectores al mismo tiempo.

Suponiendo que ambos procesos son independientes, la probabilidad de detección será el producto de cada una de las probabilidades de detección obtenidas al aplicar por separado el detector de crecimiento y el detector adaptado al subespacio de señal.

De esta forma podemos representar gráficas de PD versus PFA a la salida del sistema completo. La probabilidad de falsa alarma del sistema de detección conjunto será también el producto de la PFA de cada uno de los detectores.

A partir de las figuras 28, 29 y 30, fijamos una PFA en el detector de crecimiento donde la PD de fuegos descontrolados sea prácticamente 1, y la PDI de tendencias decrecientes sea pequeña, mientras que en el detector adaptado al subespacio de señal fijamos un PFA entre 0.001 y 0.1 (con $p=2$ y $\beta_{tan}=1$), en todos los casos.

Con esta consideración hemos fijado, para representar la PD frente a la PFA en el esquema de detección conjunto, en el **bloque A** una PFA entre 0.5 y 0.7 (figura 32), en el **bloque B** una PFA entre 0.4 y 0.5 (figura 33), en el **bloque C** una PFA entre 0.4 y 0.6.

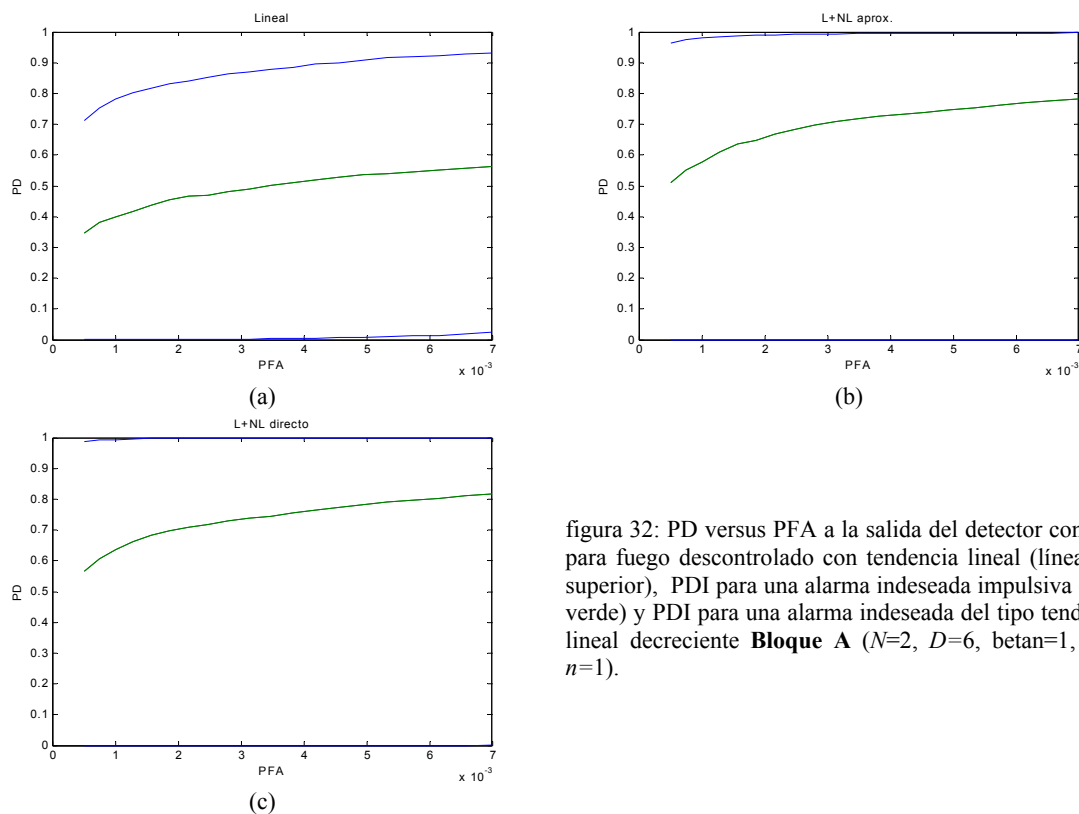
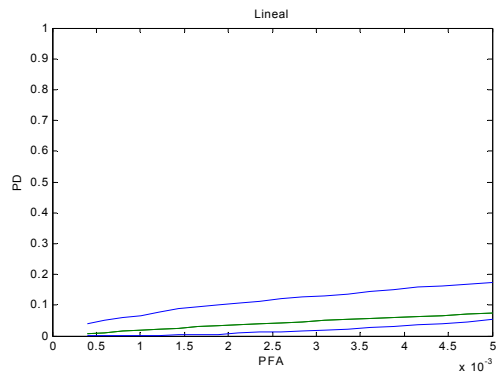
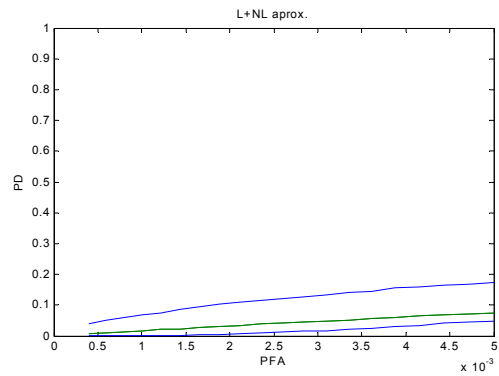


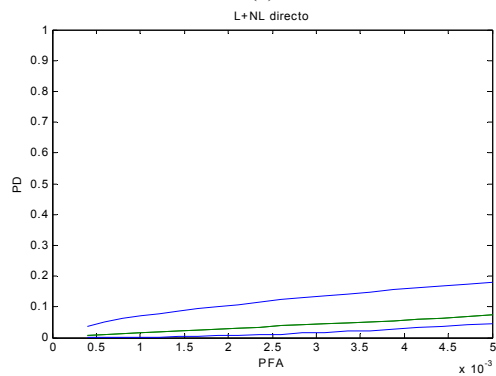
figura 32: PD versus PFA a la salida del detector conjunto para fuego descontrolado con tendencia lineal (línea azul superior), PDI para una alarma indeseada impulsiva (línea verde) y PDI para una alarma indeseada del tipo tendencia lineal decreciente **Bloque A** ($N=2$, $D=6$, $\beta_{tan}=1$, $p=2$, $n=1$).



(a)

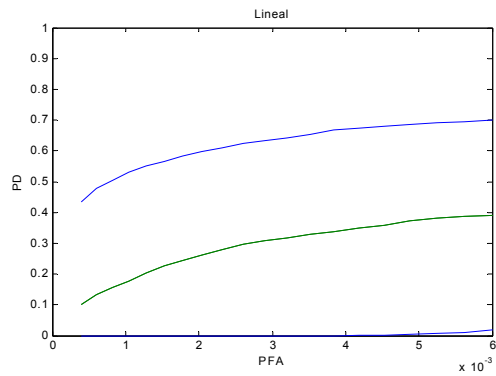


(b)

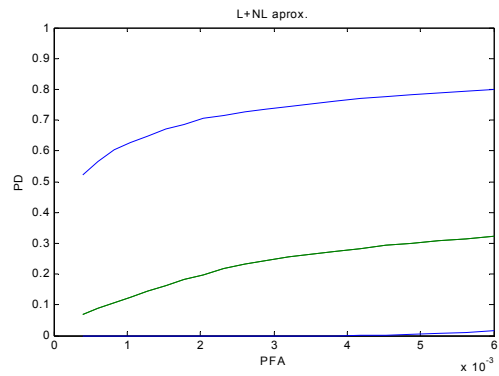


(c)

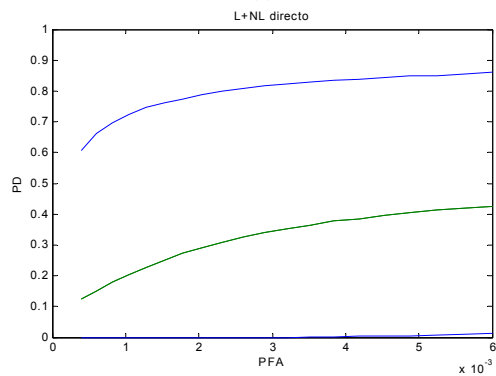
figura 33: PD versus PFA a la salida del detector conjunto para fuego descontrolado con tendencia lineal (línea azul superior), PDI para una alarma indeseada impulsiva con $j=5$ (línea verde) y PDI para una alarma indeseada del tipo tendencia lineal decreciente **Bloque B** ($N=2$, $D=6$, $\beta_{tan}=1$, $p=2$, $n=1$).



(a)



(b)



(c)

figura 34: PD versus PFA a la salida del detector conjunto para fuego descontrolado con tendencia lineal (línea azul superior), PDI para una alarma indeseada impulsiva con $j=5$ (línea verde) y PDI para una alarma indeseada del tipo tendencia lineal decreciente **Bloque C** ($N=2$, $D=6$, $\beta_{tan}=1$, $p=2$, $n=1$).

A partir de los resultados observados en las figuras 32, 33 y 34, concluimos que la incorporación del detector de crecimiento nos ha permitido eliminar las tendencias decrecientes, que por sí sólo el detector adaptado al subespacio de señal no era capaz de eliminar, mientras las alarmas indeseadas del tipo impulsivo, cuya problemática fue presentada en las figuras 26 y 31 (detector adaptado al subespacio de señal y detector de crecimiento respectivamente) son más o menos compensadas, aunque en el caso que estas se produzcan en los instantes finales del vector de decisión a la entrada de cada uno de los detectores, la PD puede ser alta, ver figura 35, en la figura 34 se puede ver la PD para el caso $j=5$.

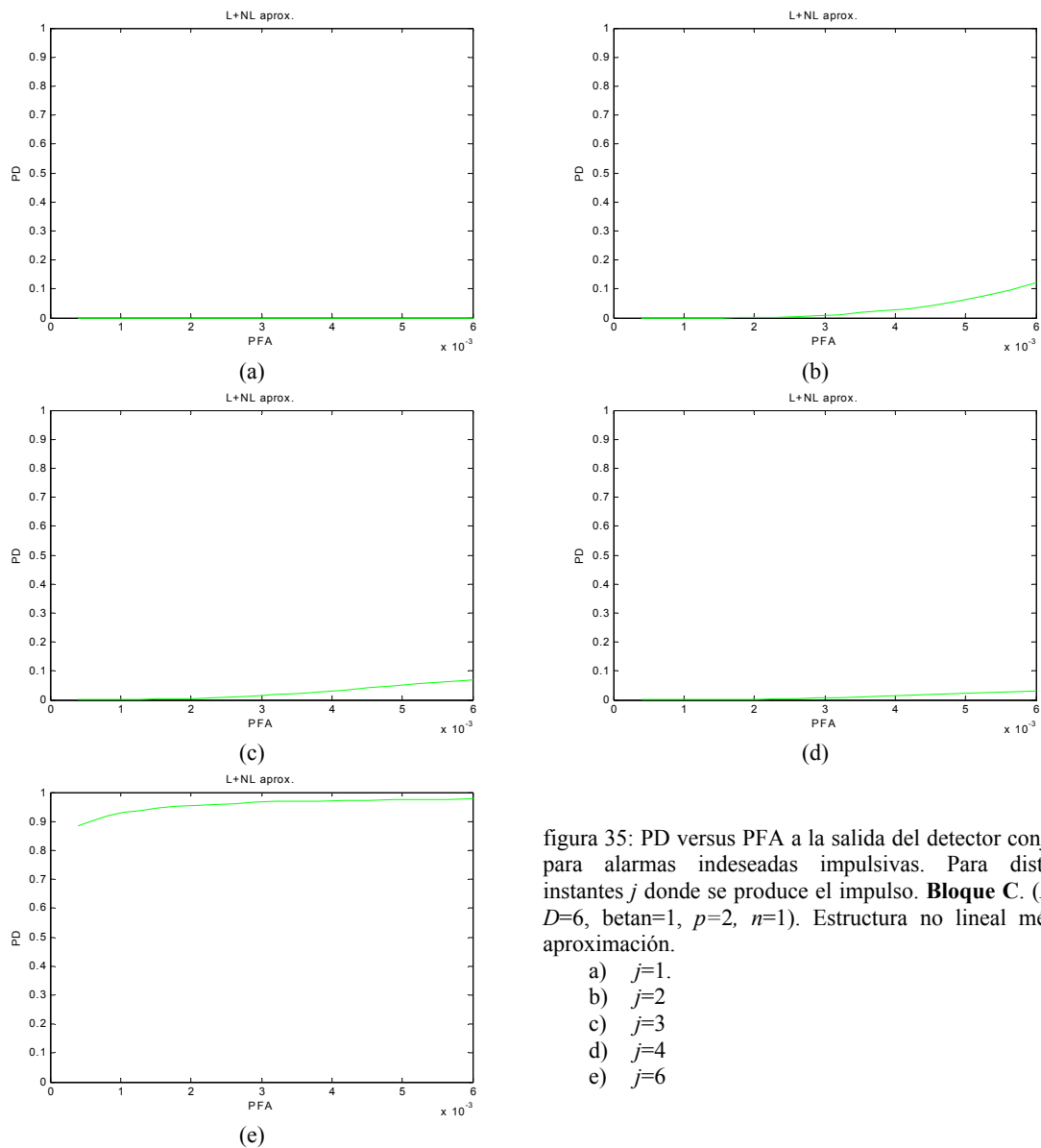


figura 35: PD versus PFA a la salida del detector conjunto para alarmas indeseadas impulsivas. Para distintos instantes j donde se produce el impulso. **Bloque C.** ($N=6$, $D=6$, $\beta_{tan}=1$, $p=2$, $n=1$). Estructura no lineal método aproximación.

- a) $j=1$.
- b) $j=2$
- c) $j=3$
- d) $j=4$
- e) $j=6$

Esto nos lleva a tener que considerar mecanismos que nos permiten controlar esta situación a priori poco probable y que puede afectar al sistema.

En definitiva la combinación conjunta de ambos detectores nos ha permitido minimizar la PDI de alarmas impulsivas con el impulso situado en los puntos iniciales del vector, pero por el contrario la combinación de ambos detectores, provoca un PD alta en el caso en que el impulso se produzca en el último elemento del vector de decisión, (figura 35e).

5. Conclusiones:

Hemos verificado el interés de incluir un detector basado en el subespacio de señal para distinguir fuegos descontrolados de variación razonablemente lenta, frente a efectos ocasionales de carácter impulsivo.

El diseño del subespacio implica fijar dos parámetros β y p . Una regla práctica deducida del estudio realizado es elegir $\beta \cong \frac{1}{2D}$ y $p=2$. En todo caso la rapidez del crecimiento del fuego descontrolado, caso de poder ser conocida a priori podría llevar a ajustes más aproximados.

Hemos propuesto el denominado detector de crecimiento para distinguir los casos de fuego creciente frente a fuego decreciente, que no pueden ser discernidos por el detector en subespacio de señal. Un único parámetro aparece en este caso, el valor n relativo al número de veces que resulta preciso diferenciar para convertir la evolución creciente de fuego en una componente continua. Nuevamente, dicho parámetro depende de la evolución esperada en un fuego creciente que, deseablemente debería ser conocida a priori para conseguir optimizar el diseño.

Capítulo VI:

Conclusiones y Líneas Futuras

Capítulo VI

CONCLUSIONES Y LÍNEAS FUTURAS

1. Conclusiones:

1.1. Relativas a la predicción:

Hemos presentado una revisión de técnicas no lineales de predicción (Capítulo III) que creemos de interés en sí misma para poner en perspectiva las técnicas no lineales concretas presentadas en esta tesis y facilitar el camino hacia variantes o extensiones más complejas.

Se han propuesto dos técnicas nuevas de diseño de la parte no lineal de la estructura de Wiener, una, denominada método directo, estima de forma directa la media condicionada de los valores a predecir, la otra utiliza una aproximación polinómica cuyos coeficientes deben ser calculados a partir de estimas de momentos cruzados de orden superior. La parte lineal del predictor se diseña de forma convencional mediante la ecuación de Wiener-Hopf.

Las técnicas propuestas se han aplicado a los tres bloques seleccionados, obteniéndose algunas conclusiones que permiten cierto grado de generalización a saber:

- Situaciones de no-Gaussianidad, pero de comportamiento bi o multimodal en el que para cada modo hay Gaussianidad, pueden llevar a reducciones aparentemente muy poco significativas en la potencia media del error de predicción al incluir una no linealidad, si el coeficiente de correlación en la dirección de predicción es el mismo en cada modo. Esto es especialmente cierto en situaciones de alta correlación, en las que el predictor lineal consigue ya una elevada reducción de error de predicción (caso del **bloque A**).

- Situaciones de Gaussianidad no precisan en principio de la inclusión de una no linealidad, sin embargo si se combinan con la presencia de baja correlación pueden darse amplificaciones del error de predicción lineal si el orden del predictor lineal y/o la distancia a predecir toman valores altos. Al incorporar la no linealidad “contenemos” el error de predicción, dotando al sistema de mayor robustez (caso **bloque B**).
- La presencia de la no linealidad resulta conveniente también en situaciones altamente no estacionarias, producidas por el entrenamiento a partir de bloques de datos que abarcan zonas espaciales muy amplias, tanto desde el punto de vista de la reducción del error de predicción, como de la robustez frente a amplificaciones de error de predicción (caso **bloque C**).
- El método directo de ajuste de la no linealidad es siempre el que da mejores resultados a cambio de su mayor complejidad de diseño e implementación. La aproximación polinómica de orden 3 da prácticamente los mismos resultados cuando la salida del predictor lineal puede considerarse Gaussiana (**bloque B**), dado que entonces se satisfacen las hipótesis asumidas en la derivación de la ecuación correspondiente.
- Las probabilidades de falsas alarmas admiten un razonable ajuste dado que en todos los casos el error de predicción (ruido de entrada al detector) pueden considerarse Gaussiano.
- La probabilidad de detección mejora al incluir la no linealidad junto al predictor lineal, siempre que la distribución de los datos sea no Gaussiana. Esto ocurre así incluso en casos como el del **bloque A** en que la reducción del error de predicción es aparentemente poco significativa. Debe tenerse en cuenta que la detección se hace sobre las D muestras, y

que pequeñas mejoras en la predicción de cada elemento, pueden llevar a una mejora global significativa en la detección.

Se ha propuesto una técnica concreta de ajuste del orden del predictor lineal, que tiene en cuenta la conveniencia de no colocar órdenes altos dada la no estacionaridad de los datos.

Se ha señalado un procedimiento para el cálculo de la cota mínima de potencia mínima del error de predicción que puede resultar útil para evaluar cómo de cerca estamos de dicha cota con el predictor lineal diseñado y, caso de estar lejos, optar por un predictor IIR, previo a la no linealidad.

1.2. Relativas al detector:

Hemos verificado el interés de incluir un detector basado en el subespacio de señal para distinguir fuegos descontrolados de variación razonablemente lenta, frente a efectos ocasionales de carácter impulsivo.

El diseño del subespacio implica fijar dos parámetros β y p . Una regla práctica deducida del estudio realizado es elegir $\beta \cong \frac{1}{2D}$ y $p=2$. En todo caso la rapidez del crecimiento del fuego descontrolado, caso de poder ser conocida a priori podría llevar a ajustes más aproximados.

Hemos propuesto el denominado detector de crecimiento para distinguir los casos de fuego creciente frente a fuego decreciente, que no pueden ser discernidos por el detector en subespacio de señal. Un único parámetro aparece en este caso, el valor n relativo al número de veces que resulta preciso diferenciar para convertir la evolución creciente de fuego en una componente continua. Nuevamente, dicho parámetro depende de la evolución esperada en un fuego creciente que, deseablemente debería ser conocida a priori para conseguir optimizar el diseño.

2. Líneas futuras:

- ✓ Experimentación con otras estructuras de predicción no lineal entre las revisadas en el Capítulo III, y en particular de las estructuras alcanzables mediante la combinación de varios sistemas de Wiener.
- ✓ Implementaciones adaptativas del sistema de Wiener para facilitar la calibración del predictor.
- ✓ Obtención de métodos para la determinación de la cota mínima de error de predicción, como ayuda a la decisión de incluir o no la no linealidad, o incluir estructuras no lineales más o menos complicadas.
- ✓ Estudio de filtros predictores lineales IIR.
- ✓ Análisis de métodos de detección que incorporen más conocimiento sobre la señal a detectar, como por ejemplo mediante modelos estocásticos de las señales (podemos incorporar información de la matriz de covarianza de la señal si existe la posibilidad de estimarla mediante un calibración previa del sistema con fuegos reales controlados).
- ✓ Implementación adaptativa de la decorrelación del error de predicción para facilitar la calibración del detector.
- ✓ Aplicación del esquema de propuesto en otras situaciones, sean de detección de fuegos (por ejemplo con cámara infrarroja), sean de otro tipo (sistemas automáticos de detección radar, sonar o de ultrasonidos, por ejemplo).

Apéndice A:

Obtención de la Fórmula para la Aproximación Polinómica de la Media Condicional

APÉNDICE A:**OBTENCIÓN DE LA FÓRMULA PARA LA APROXIMACIÓN POLINÓMICA DE LA MEDIA CONDICIONAL.**

En un contexto estadístico, el mínimo error cuadrático medio de predicción corresponde a la media condicional de la variable aleatoria correspondiente a la muestra predicha a partir del valor de las muestras anteriores al instante de predicción.

$$E[x(n)/x(n-1)x(n-2)\dots x(n-N+1)] \quad (A1)$$

En el esquema planteado, Capítulo IV, una parte fundamental es el diseño de la no linealidad por la técnica denominada de aproximación polinómica, con la que pretendemos encontrar una expresión cerrada y de coste computacional bajo para estimar la media condicional. Para ello, vamos a establecer una fórmula general para la media condicional usando estadísticos de orden superior.

En este apéndice, pretendemos demostrar la expresión usada en el Capítulo IV, a partir de la fórmula general presentada en [ERT97] [ERT00].

En primer lugar vamos a determinar la fórmula general multidimensional, y a partir de ella, demostrar la expresión unidimensional usada en la presente tesis.

1) Fórmula General para estimar la media condicional multidimensional.

Nuestro objetivo es obtener una fórmula general para la media condicional de una variable aleatoria escalar y , a partir de un vector de variables aleatorias $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, para ello usaremos estadísticos de orden superior. Vamos a introducir por un lado, la función generadora de momentos (FGM) del vector \mathbf{x} como la transformada multidimensional de Laplace de la función densidad de probabilidades (fdp) de \mathbf{x} .

$$\phi_x(\mathbf{s}) = E(e^{s^T \mathbf{x}}) = \int e^{s^T \mathbf{x}} p(\mathbf{x}) d\mathbf{x} \quad (A2)$$

y por otro lado la función generadora de momentos conjunta de \mathbf{x} e y

$$\phi(u, \mathbf{s}) = E(e^{uy} e^{s^T \mathbf{x}}) = \int \int e^{uy} e^{s^T \mathbf{x}} p(\mathbf{x}, y) dx dy \quad (A3)$$

donde u representa una variable de Laplace con el fin de distinguirla de las otras variables. A partir de la expresión (A3) caracterizamos la función generadora de los cumulantes conjuntos, la cual se define como el logaritmo natural de la función generadora de momentos conjunta.

$$\Psi(u, \mathbf{s}) = \ln \phi(u, \mathbf{s}) \quad (A4)$$

En definitiva, la expresión (A2) la podemos caracterizar mediante la expresión,

$$\phi(u, s_1, s_2, \dots, s_N) = E[e^{(uy + s_1 x_1 + s_2 x_2 + \dots + s_N x_N)}] \quad (A5)$$

de tal forma que el momento conjunto de orden $r = l + k_1 + k_2 + \dots + k_N$ lo obtenemos a partir de (A5) como:

$$E[y^l x_1^{k_1} x_2^{k_2} \dots x_N^{k_N}] = \left. \frac{\partial^r \phi(u, s_1, s_2, \dots, s_N)}{\partial u^l \partial s_1^{k_1} \partial s_2^{k_2} \dots \partial s_N^{k_N}} \right|_{u=s_1=s_2=\dots=s_N=0} \quad (A6)$$

Si asumimos que todos los cumulantes (o momentos) conjuntos de \mathbf{x} e y son finitos, entonces podemos definir la función generadora de cumulantes de la siguiente forma:

$$\Psi(u, \mathbf{s}) = \sum_{l, k_1, k_2, \dots, k_N} \frac{\text{cum}[y^l x_1^{k_1} x_2^{k_2} \dots x_N^{k_N}]}{l! k_1! k_2! \dots k_N!} u^l s_1^{k_1} s_2^{k_2} \dots s_N^{k_N} \quad (A7)$$

expresión que conviene demostrar.

Demostración:

A partir de (A4) y (A6). Para facilitar los cálculos supongamos que partimos de una variable aleatoria $\{x_1\}$.

De la expresión (A5) su FGM será,

$$\phi(s_1) = E[e^{s_1 x_1}] = \int e^{s_1 x_1} p(x_1) dx_1 \tag{A8}$$

con lo que de (A6) sus momentos de orden k_1 son:

$$m_{k_1} = E[x_1^{k_1}] = \left. \frac{\partial^{k_1} \phi(s_1)}{\partial s_1^{k_1}} \right|_{s_1=0} \tag{A9}$$

con lo que de (A4) su correspondiente función generadora de cumulantes será:

$$\Psi(s_1) = \ln \phi(s_1) \tag{A10}$$

expresión a la que vamos a realizar su correspondiente desarrollo de McLaurin, para lo cual debemos saber que:

$$\phi(s_1) \Big|_{s_1=0} = \int e^{s_1 x_1} p(x_1) dx_1 \Big|_{s_1=0} = \int p(x_1) dx_1 = 1 \tag{A11}$$

$$\begin{aligned} \Psi(s_1) \Big|_{s_1=0} &= \ln \phi(s_1) \Big|_{s_1=0} = \ln(1) = 0 \\ \Psi^{(I)}(s_1) \Big|_{s_1=0} &= \left. \frac{\partial \Psi(s_1)}{\partial s_1} \right|_{s_1=0} = \left. \frac{\phi^{(I)}(s_1)}{\phi(s_1)} \right|_{s_1=0} = \frac{m_1}{1} = cum(x_1) \\ \Psi^{(II)}(s_1) \Big|_{s_1=0} &= \left. \frac{\partial^2 \Psi(s_1)}{\partial s_1^2} \right|_{s_1=0} = \left. \frac{\phi^{(II)}(s_1)\phi(s_1) - \phi^{(I)}(s_1)\phi^{(I)}(s_1)}{\phi^2(s_1)} \right|_{s_1=0} = \frac{m_2 - m_1^2}{1} = cum(x_1^2) \\ &\vdots \\ \Psi^{(k_1)}(s_1) \Big|_{s_1=0} &= \left. \frac{\partial^{k_1} \phi(s_1)}{\partial s_1^{k_1}} \right|_{s_1=0} = \dots = cum(x_1^{k_1}) \end{aligned}$$

con lo que su desarrollo de McLaurin corresponde a la expresión,

$$\psi(s_1) = s_1 cum(x_1) + \frac{s_1^2}{2!} cum(x_1^2) + \dots + \frac{s_1^{k_1}}{k_1!} cum(x_1^{k_1}) = \sum_{k_1=1} \frac{cum(x_1^{k_1})}{k_1!} s_1^{k_1} \quad (A12)$$

Queda demostrada la expresión (A7) sin más que generalizar la ecuación anterior para el caso multidimensional, obteniendo así la expresión generadora de cumulantes conjuntos.

A partir de lo visto con anterioridad, si nosotros manipulamos la ecuación (A3), y teniendo en cuenta que la función densidad de probabilidades de la distribución condicional puede expresarse como: $p(y/x) = p(x)p(x, y)$,

$$\begin{aligned} \phi(u, s) &= E(e^{uy} e^{s^T x}) = \iint e^{uy} e^{s^T x} p(x, y) dx dy = \iint e^{uy} e^{s^T x} p(x) \cdot p(y/x) dx dy = \\ &= \int \left(\int e^{uy} p(y/x) dy \right) e^{s^T x} p(x) dx = \int \phi_{y/x}(u) \cdot e^{s^T x} p(x) dx = E[\phi_{y/x}(u) e^{s^T x}] \end{aligned} \quad (A13)$$

donde $\phi_{y/x}(u)$ es la función generadora de momentos condicionales, cuyo desarrollo de McLaurin son los momentos condicionales de y dado x

$$\phi_{y/x}(u) = 1 + \sum_{l=1} \frac{E[y^l/x]}{l!} u^l \quad (A14)$$

Si realizamos la $\frac{\partial \phi(u, s)}{\partial u}$ en la expresión (A13) nos queda

$$\left. \frac{\partial \phi(u, s)}{\partial u} \right|_{u=0} = \int \left. \frac{\partial \phi_{y/x}(u)}{\partial u} \right|_{u=0} e^{s^T x} p(x) dx = E \left[\left. \frac{\partial \phi_{y/x}(u)}{\partial u} \right|_{u=0} e^{s^T x} \right] \quad (A15)$$

a partir de la expresión (A14)

$$\left. \frac{\partial \phi_{y/x}(u)}{\partial u} \right|_{u=0} = E[y/x] \quad (A16)$$

con lo que:

$$\left. \frac{\partial \phi(u, s)}{\partial u} \right|_{u=0} = E[E[y/x]e^{s^T x}] = \int E[y/x]e^{s^T x} p(\mathbf{x}) dx \quad (A17)$$

de esta forma la esperanza condicional y la derivada de la FGM en $u=0$, constituyen un par transformado. Por otro lado, teniendo en cuenta que, $\phi = e^\Psi$ podemos también escribir,

$$\left. \frac{\partial \phi(u, s)}{\partial u} \right|_{u=0} = \left. \frac{\partial \Psi(u, s)}{\partial u} \phi(u, s) \right|_{u=0} = \left. \frac{\partial \Psi(u, s)}{\partial u} \right|_{u=0} \phi_x(s) \quad (A18)$$

Si derivamos la expresión (A7) con respecto a u , y evaluamos en $u=0$

$$\left. \frac{\partial \Psi(u, s)}{\partial u} \right|_{u=0} = \sum_{k_1, k_2, \dots, k_N=0} \frac{cum(y, x_1^{k_1}, x_2^{k_2}, \dots, x_N^{k_N})}{k_1! k_2! \dots k_N!} s_1^{k_1} s_2^{k_2} \dots s_N^{k_N} \quad (A19)$$

Ahora, si realizamos la Transformada Inversa de Laplace de la ecuación (A19), usando para ello la siguiente propiedad de la Transformada de Laplace,

$$(-1)^{k_1+k_2+\dots+k_N} \frac{\partial^{k_1+k_2+\dots+k_N}}{\partial x_1^{k_1} \partial x_2^{k_2} \dots \partial x_N^{k_N}} p(\mathbf{x}) \leftrightarrow s_1^{k_1} s_2^{k_2} \dots s_N^{k_N} \phi_x(s)$$

y usando la ecuación (A18), encontramos una fórmula general para expresar la media condicional,

$$E[y/x] = \frac{1}{p(\mathbf{x})} \sum_{k_1, k_2, \dots, k_N=0} \frac{cum(y, x_1^{k_1}, x_2^{k_2}, \dots, x_N^{k_N})}{k_1! k_2! \dots k_N!} (-1)^{k_1+k_2+\dots+k_N} \frac{\partial^{k_1+k_2+\dots+k_N}}{\partial x_1^{k_1} \partial x_2^{k_2} \dots \partial x_N^{k_N}} p(\mathbf{x}) \quad (A20)$$

A partir de esta expresión podemos enunciar el siguiente teorema: “Para una variable aleatoria escalar y , y un vector de variables aleatorias $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ cuyos cumulantes (o momentos) conjuntos son finitos, la media condicional $E[y/x]$ viene dada por la expresión (A20)”.

En la literatura sobre estadísticos de orden superior, los cumulantes que tienen variables con potencia unidad (posiblemente con repeticiones de las variables) son

usados mayoritariamente. De esta forma podemos usar una formulación alternativa a la ecuación (A14).

$$E[y/\mathbf{x}] = cum(y) + \frac{1}{p(\mathbf{x})} \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k_1=1}^N \sum_{k_2=1}^N \cdots \sum_{k_n=1}^N cum(y, x_{k_1}, x_{k_2}, \dots, x_{k_n}) \times (-1)^n \frac{\partial^n}{\partial x_{k_1} \partial x_{k_2} \cdots \partial x_{k_n}} p(\mathbf{x}) \quad (A21)$$

Consideremos un simple ejemplo para ilustrar la expresión (A21), en la que y , x e y son variables escalares, y con distribución Gaussiana tanto marginalmente como conjuntamente. Entonces la ecuación (A20) o (A21) queda reducida a la expresión (A22).

$$E[y/x] = cum(y) - cum(y, x) \frac{p^{(1)}(x)}{p(x)} + \frac{1}{2!} cum(y, x, x) \frac{p^{(2)}(x)}{p(x)} + \dots \quad (A22)$$

donde el exponente de $p(x)$ indica el grado de derivación con respecto a x . Como sabemos, para una distribución Gaussiana los cumulantes de orden superior a dos, son cero y teniendo en cuenta que para una distribución Gaussiana, [CHR93].

$$cum(y) = E[y], \quad cum(y, x) = cov ar(y, x), \quad \frac{p^{(1)}(x)}{p(x)} = -\frac{x - E[x]}{var(x)}$$

tenemos,

$$E[y/x] = E[y] + cov ar(y, x) \frac{x - E[x]}{var(x)} \quad (A23)$$

Donde podemos observar que la media condicional es una expresión lineal con respecto a x .

Esta es pues, la mejor expresión conocida para la media condicional de variables Gaussianas. A partir de esta expresión y como conclusión, tenemos que la ecuación (A20) es una generalización de la ecuación (A23) para variables aleatorias cuyos momentos conjuntos son finitos. Los efectos de los elementos no-Gaussianos que

aparecen en la media condicional se incorporan a partir de los cumulantes conjuntos de orden superior y el resultado de evaluar la derivada *m-ésima* de la expresión,

$$\frac{p^{(m)}(x)}{p(x)} \tag{A24}$$

2) Obtención expresión polinómica para estimar la media condicional.

Estamos interesados en demostrar a partir de la fórmula general (A21), la expresión polinómica usada en el Capítulo IV (A25), que para un orden 3 quedaba.

$$g(x_{pl}) = \frac{1}{6}(a_3 - 3a_1)x_{pl}^3 + \frac{1}{2}a_2x_{pl}^2 + \left(\frac{5}{2}a_1 - \frac{1}{2}a_3\right)x_{pl} - \frac{1}{2}a_2 \tag{A25}$$

para caracterizar de forma cerrada la función no lineal sin memoria, de tal forma que la salida sea la media condicional de los datos originales a partir de las predicciones obtenidas al aplicar el predictor lineal (figura 1),

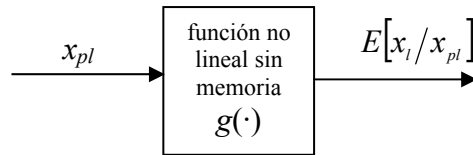


Figura 1: Esquema entrada/salida de la expresión polinómica.

Si consideramos ahora que x_{pl} y x_l son variables aleatorias escalares (caso unidimensional), entonces la expresión (A21) se simplifica como hemos visto en el ejemplo anterior en la ecuación (A22). En ese caso, sustituyendo las variables x e y por x_{pl} y x_l respectivamente, obtenemos la ecuación (A26), a partir de la cual vamos a demostrar la validez de la expresión (A25), cuando x_{pl} es una variable aleatoria con función densidad de probabilidad Gaussiana normalizada y x_l variable aleatoria de media cero y varianza unidad y con una cierta función densidad de probabilidades.

$$E[x_l/x_{pl}] = cum(x_l) - cum(x_l, x_{pl}) \frac{p'(x_{pl})}{p(x_{pl})} + \frac{1}{2!} cum(x_l, x_{pl}, x_{pl}) \frac{p''(x_{pl})}{p(x_{pl})} + \dots = g(x_{pl}) \tag{A26}$$

Si suponemos que x_{pl} es de distribución Gaussiana, de media μ y varianza σ su función densidad de probabilidades vendrá dada por:

$$p(x_{pl}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{pl}-\mu}{\sigma}\right)^2} \quad (\text{A27})$$

de tal forma que en este caso la expresión (A24) evaluada para $m=1, 2, 3$ equivale a:

$$p^{(1)}(x_{pl}) = -\frac{1}{\sigma\sqrt{2\pi}} \left(\frac{x_{pl}-\mu}{\sigma^2}\right) e^{-\frac{1}{2}\left(\frac{x_{pl}-\mu}{\sigma}\right)^2} = p(x_{pl}) \frac{x_{pl}-\mu}{\sigma^2} \quad (\text{A28})$$

$$\text{con lo que: } -\frac{p^{(1)}(x_{pl})}{p(x_{pl})} = \frac{x_{pl}-\mu}{\sigma^2}$$

$$p^{(2)}(x_{pl}) = -p^{(1)}(x_{pl}) \left(\frac{x_{pl}-\mu}{\sigma^2}\right) - p(x_{pl}) \frac{1}{\sigma^2} = p(x_{pl}) \left(\frac{x_{pl}-\mu}{\sigma}\right)^2 - p(x_{pl}) \frac{1}{\sigma^2} \quad (\text{A29})$$

$$\text{con lo que: } \frac{p^{(2)}(x_{pl})}{p(x_{pl})} = \left(\frac{x_{pl}-\mu}{\sigma^2}\right)^2 - \frac{1}{\sigma^2}$$

$$\begin{aligned} p^{(3)}(x_{pl}) &= p^{(1)}(x_{pl}) \left[\frac{(x_{pl}-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2} \right] + p(x_{pl}) \cdot \frac{2(x_{pl}-\mu)}{\sigma^4} = \\ &= -p(x_{pl}) \cdot \left(\frac{x_{pl}-\mu}{\sigma^2}\right) \left[\frac{(x_{pl}-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2} \right] + p(x_{pl}) \cdot \frac{2(x_{pl}-\mu)}{\sigma^4} \end{aligned} \quad (\text{A30})$$

$$\text{con lo que: } -\frac{p^{(3)}(x_{pl})}{p(x_{pl})} = \left(\frac{x_{pl}-\mu}{\sigma^2}\right)^3 - \frac{3}{\sigma^2} \left(\frac{x_{pl}-\mu}{\sigma^2}\right)$$

de tal forma que podemos generalizar la expresiones anteriores como

$$(-1)^m \frac{p^{(m)}(x_{pl})}{p(x_{pl})} = H_m \left(\frac{x_{pl}-\mu}{\sigma^2}\right) \quad (\text{A31})$$

donde $H_m\left(\frac{x_{pl} - \mu}{\sigma^2}\right)$ es el polinomio de Hermite de orden m .

Los polinomios de Hermite pueden ser generados usando la ecuación de recursión, podemos demostrarla a partir de los resultados obtenidos de (A28-A30).

$$H_{m+1}\left(\frac{x_{pl} - \mu}{\sigma^2}\right) = \left(\frac{x_{pl} - \mu}{\sigma^2}\right)H_m\left(\frac{x_{pl} - \mu}{\sigma^2}\right) - \frac{m}{\sigma^2}H_{m-1}\left(\frac{x_{pl} - \mu}{\sigma^2}\right)$$

con $H_0\left(\frac{x_{pl} - \mu}{\sigma^2}\right) = \frac{1}{\sigma^2}$, y $H_1\left(\frac{x_{pl} - \mu}{\sigma^2}\right) = \frac{x_{pl} - \mu}{\sigma^2}$

En nuestro caso, tal como hemos planteado en el Capítulo IV, sección 2.2, x_{pl} es de media cero y varianza unidad, por lo que las ecuaciones anteriores quedan simplificadas a las expresiones.

$$(-1)^1 \frac{p^{(1)}(x_{pl})}{p(x_{pl})} = x_{pl}, \quad (-1)^2 \frac{p^{(2)}(x_{pl})}{p(x_{pl})} = x_{pl}^2 - 1, \quad (-1)^3 \frac{p^{(3)}(x_{pl})}{p(x_{pl})} = x_{pl}^3 - 3x_{pl} \quad (A32)$$

$$(-1)^m \frac{p^{(m)}(x_{pl})}{p(x_{pl})} = H_m(x_{pl}) \quad (A33)$$

$$H_{m+1}(x_{pl}) = (x_{pl})H_m(x_{pl}) - mH_{m-1}(x_{pl})$$

con $H_0(x_{pl}) = 1$, y $H_1(x_{pl}) = x_{pl}$

De esta forma podemos escribir la ecuación (A26) de la forma:

$$g(x_{pl}) = \sum_{m=1}^{\infty} \frac{1}{m!} C_m(x_l, x_{pl}) H_m(x_{pl}) \quad (A34)$$

donde $C_m(x, x_p) = cum\left(x, \overbrace{x_p, \dots, x_p}^{m \text{ veces}}\right)$ caracteriza al cumulante de orden m .

A partir de esta definición nuestro objetivo se centra ahora en obtener el valor de los cumulantes para cada m .

En [NIK93] se indica la relación existente entre los momentos de $\{x_1, x_2, \dots, x_n\}$ y los cumulantes conjuntos de orden $r=n$ dada por,

$$cum[x_1, x_2, \dots, x_n] = \sum (-1)^{p-1} (p-1)! \cdot E \left[\prod_{i \in s_1} x_i \right] \cdot E \left[\prod_{i \in s_2} x_i \right] \cdots E \left[\prod_{i \in s_p} x_i \right] \quad (A35)$$

donde la suma se extiende sobre todas las posibles particiones (s_1, s_2, \dots, s_p) , para $p=1, 2, \dots, n$, de los enteros $(1, 2, \dots, n)$, por ejemplo, la clase de enteros $(1, 2, 3)$ puede particionarse como:

$$\begin{aligned} p=1 \quad s_1 &= \{1, 2, 3\} \\ p=2 \quad s_1 &= \{1\} \quad s_2 = \{2, 3\} \\ & \quad s_1 = \{2\} \quad s_2 = \{1, 3\} \\ & \quad s_1 = \{3\} \quad s_2 = \{1, 2\} \\ p=3 \quad s_1 &= \{1\} \quad s_2 = \{2\} \quad s_3 = \{3\} \end{aligned}$$

En nuestro caso ($m=3$) y de (A35) tenemos que:

$$\begin{aligned} cum(x_l, x_{pl}) &= E[x_l x_{pl}] - E[x_l]E[x_{pl}] \\ cum(x_l, x_{pl}, x_{pl}) &= E[x_l x_{pl}^2] - 2E[x_l x_{pl}]E[x_{pl}] - E[x_{pl}^2]E[x_l] + 2E[x_l]E[x_{pl}]E[x_{pl}] \\ cum(x_l, x_{pl}, x_{pl}, x_{pl}) &= E[x_l x_{pl}^3] - 3E[x_{pl}^2]E[x_l x_{pl}] - E[x_l]E[x_{pl}^3] - 3E[x_{pl}]E[x_l x_{pl}^2] + \\ & \quad + 4E[x_l]E[x_{pl}]E[x_{pl}^2] + 6E[x_{pl}]^2 E[x_l x_{pl}] - 6E[x_l]E[x_{pl}]^3 \end{aligned} \quad (A36)$$

si además normalizamos x_l (media cero y varianza unidad), y definimos $a_i = E[x_l x_{pl}^i]$ nos queda (A37)

$$\begin{aligned}cum(x_l, x_{pl}) &= a_1 \\cum(x_l, x_{pl}, x_{pl}) &= a_2 \\cum(x_l, x_{pl}, x_{pl}, x_{pl}) &= a_3 - 3a_1\end{aligned}\tag{A37}$$

de tal forma que la aproximación de tercer orden de la ecuación (A34) es:

$$g_3(x_p) = a_1 H_1(x_{pl}) + \frac{1}{2!} a_2 H_2(x_{pl}) + \frac{1}{3!} (a_3 - 3a_1) H_3(x_{pl})\tag{A38}$$

y teniendo en cuenta (A32) y (A33) queda demostrada (A25).

$$g(x_{pl}) = \frac{1}{6} (a_3 - 3a_1) x_{pl}^3 + \frac{1}{2} a_2 x_{pl}^2 + \left(\frac{5}{2} a_1 - \frac{1}{2} a_3 \right) x_{pl} - \frac{1}{2} a_2$$

Apéndice B:

Cálculo de la Descomposición de Wold

APÉNDICE B

CÁLCULO DE LA DESCOMPOSICIÓN DE WOLD

Podemos evaluar la respuesta impulsional $h(n)$ de la descomposición de Wold a partir de los coeficientes cepstrum (Transformada de Fourier Inversa del logaritmo del espectro).

Los coeficientes cepstrum son definidos como,

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_x(\omega) \exp(-jn\omega) d\omega \quad (B1)$$

de tal forma que podemos definir la transformada de Fourier de la parte causal de la secuencia $\{c_n\}$.

$$A(\omega) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \exp(-jn\omega) = \frac{c_0}{2} + C(\omega) \quad (B2)$$

La transformada de Fourier de la respuesta impulsional de nuestro filtro IIR puede relacionarse con la transformada de Fourier de la descomposición de Wold [PRI88].

$$H(\omega) = \sum_{n=0}^{\infty} h(n) \exp(-jn\omega) = \exp(A(\omega)) = \exp(c_0/2) \exp(C(\omega)) \quad (B3)$$

con lo que si aplicamos a la expresión (B3) la transformada de Fourier inversa,

$$h(n) = \exp(c_0/2) F^{-1}[\exp(C(\omega))] \quad (B4)$$

Si nosotros ahora realizamos el desarrollo de McLaurin de la función exponencial

$$\exp(C(\omega)) = 1 + C(\omega) + \frac{1}{2!} C^2(\omega) + \dots + \frac{1}{k!} C^k(\omega) + \dots \quad (B5)$$

y aplicamos el teorema de la convolución y la ecuación (B2), tenemos que:

$$F^{-1}[\exp(C(\omega))] = \delta(n) + c_n u(n-1) + \frac{1}{2!} \{c_n u(n-1) * c_n u(n-1)\} + \dots + \frac{1}{k!} \left\{ \underbrace{c_n u(n-1) * \dots * c_n u(n-1)}_{k \text{ times}} \right\} + \dots \quad (\text{B6})$$

donde $u(n-1)$ es la función escalón discreta desplazada una muestra hacia la derecha, a partir de la ecuación (B6) podemos describir la expresión (B4) de la forma

$$h(n) = \exp(c_0 / 2) \left\{ \delta(n) + \sum_{k=1}^{\infty} \frac{1}{k!} (c_n u(n-1))^{*(k-1)} \right\} \quad (\text{B7})$$

con lo que hemos obtenido la respuesta impulsional del sistema causal de duración infinita a partir de los coeficientes cepstrum de los datos de entrada, donde $\delta(n)$ es la función impulso unidad discreta y $(c_n u(n-1))^{*(k-1)}$ representa el valor de la convolución $(k-1)$ veces de la función $c_n u(n-1)$ consigo misma.

En nuestro caso estamos interesados en evaluar $h(n)$ en un instante particular n_0 , por ejemplo,

$$h(n_0) = \exp(c_0 / 2) \left\{ \delta(n_0) + \sum_{k=1}^{n_0} \frac{1}{k!} (c_n u(n-1))^{*(k-1)} \Big|_{n=n_0} \right\} \quad (\text{B8})$$

donde tenemos que tener en cuenta que $(c_n u(n-1))^{*(k-1)} \Big|_{n=n_0} = 0 \quad \forall k > n_0$. El valor para $n_0=0$ será $h(0) = \exp(c_0 / 2)$, el cual es un resultado bastante conocido [PRI88] de la teoría de predicción clásica, pero de hecho,

$$h^2(0) = \exp(c_0) = \exp\left(\frac{1}{2\pi} \int_0^{2\pi} \log(S(\omega)) d\omega\right) = \sigma_1^2 \quad (\text{B9})$$

corresponde a la varianza del primer paso de predicción de un predictor lineal causal. Bien, si consideramos ahora un caso general $n_0 \geq 0$ y teniendo en cuenta que,

$$c_n u(n-1) = \sum_{l=1}^{\infty} c_l \delta(n-l) \quad (\text{B10})$$

y que

$$\begin{aligned} \left(\sum_{l=1}^{\infty} c_n \delta(n-l) \right)^{*(k-1)} &= \left(\sum_{l_1=1}^{\infty} c_{l_1} \delta(n-l_1) \right) * \dots * \left(\sum_{l_k=1}^{\infty} c_{l_k} \delta(n-l_k) \right) = \\ &= \sum_{l_1=1}^{\infty} \dots \sum_{l_k=1}^{\infty} c_{l_1} \delta(n-l_1) * \dots * c_{l_k} \delta(n-l_k) = \\ &= \sum_{l_1=1}^{\infty} \dots \sum_{l_k=1}^{\infty} c_{l_1} \dots c_{l_k} \delta(n-l_1 - \dots - l_k) \end{aligned} \quad (\text{B11})$$

donde hemos hecho uso de la propiedad distributiva de la convolución. Si evaluamos la ecuación (B11) para el caso particular $n=n_0$, tenemos que,

$$\left(\sum_{l=1}^{\infty} c_n \delta(n-l) \right)^{*(k-1)} \Big|_{n=n_0} = \sum_{l_1} \dots \sum_{l_k} c_{l_1} \dots c_{l_k} \quad \text{con } l_1 + \dots + l_k = n_0 \quad (\text{B12})$$

donde $\{l_i\}$ son números enteros mayores que 1.

De tal forma que si sustituimos la ecuación (B12) en el valor correspondiente de la ecuación (B11), encontramos una ecuación que nos permite determinar la respuesta impulsional $h(n)$ a partir de los coeficientes cepstrum de los datos de entrada $x(n)$ para la relación entrada-salida definida en la expresión (B3).

$$h(n_0) = \exp(c_0/2) \left\{ \delta(n_0) + \sum_{k=1}^{n_0} \frac{1}{k!} \sum_{l_1} \dots \sum_{l_k} c_{l_1} \dots c_{l_k} \right\} \quad (\text{B13})$$

sujeto a la restricción $l_1 + \dots + l_k = n_0$.

De tal forma que a partir de (B13) podemos evaluar los k coeficientes de la descomposición de Wold, por ejemplo hasta $n_o=4$.

$$\begin{aligned}
 h(0) &= \exp\left(\frac{c_0}{2}\right) & h(1) &= \exp\left(\frac{c_0}{2}\right)c_1 & h(2) &= \exp\left(\frac{c_0}{2}\right)\left[c_2 + \frac{1}{2!}c_1^2\right] \\
 h(3) &= \exp\left(\frac{c_0}{2}\right)\left[c_3 + \frac{1}{2!}2c_1c_2 + \frac{1}{3!}c_1^3\right] \\
 h(4) &= \exp\left(\frac{c_0}{2}\right)\left[c_4 + \frac{1}{2!}(2c_1c_3 + c_2^2) + \frac{1}{3!}3c_1^2c_2 + \frac{1}{4!}c_1^4\right]
 \end{aligned} \tag{B14}$$

A partir de las diferentes expresiones vamos a encontrar una expresión recursiva que nos permitirá evaluar de manera sencilla la expresión (B13), para ello, y partiendo de los diferentes resultados de (B14), los cuales los podemos expresar a partir de los coeficientes cepstrum y de $h(0)$.

$$\begin{aligned}
 h(0) &= \exp\left(\frac{c_0}{2}\right) \\
 h(1) &= h(0) \cdot c_1 \\
 h(2) &= h(0) \cdot c_2 + \frac{1}{2}h(0) \cdot c_1^2 \\
 h(3) &= h(0) \cdot c_3 + h(0) \cdot c_1 \cdot c_2 + \frac{1}{6}h(0) \cdot c_1^3 \\
 h(4) &= h(0) \cdot c_4 + h(0) \cdot c_1 \cdot c_3 + \frac{1}{2}h(0) \cdot c_2^2 + \frac{1}{2}h(0) \cdot c_1^2 \cdot c_2 + \frac{1}{24}h(0) \cdot c_1^4
 \end{aligned} \tag{B15}$$

De tal forma que $h(2)$, puede expresarse a partir de $h(1)$ y $h(0)$, como

$$h(2) = h(0) \cdot c_2 + \frac{1}{2}h(1) \cdot c_1 \tag{B16}$$

y $h(3)$, teniendo en cuenta (B19), lo podemos expresar a partir de $h(2)$, $h(1)$ y $h(0)$, como

$$\begin{aligned}
 h(3) &= h(0) \cdot c_3 + \underbrace{h(0) \cdot c_1}_{h(1)} \cdot c_2 + \frac{1}{6} h(0) \cdot c_1^3 = \\
 &= h(0) \cdot c_3 + h(1) \cdot c_2 + \frac{1}{3} \left(\underbrace{\frac{1}{2} h(1) \cdot c_1}_{\text{de (B19)}} \right) \cdot c_1 = \\
 &= h(0) \cdot c_3 + h(1) \cdot c_2 + \frac{1}{3} (h(2) - h(0) \cdot c_2) \cdot c_1 = \\
 &= h(0) \cdot c_3 + h(1) \cdot c_2 + \frac{1}{3} h(2) \cdot c_1 - \frac{1}{3} \overbrace{h(0) \cdot c_1}^{h(1)} \cdot c_2 = h(0) \cdot c_3 + \frac{2}{3} h(1) \cdot c_2 + \frac{1}{3} h(2) \cdot c_1
 \end{aligned} \tag{B17}$$

Veamos si ahora si somos capaces de encontrar una relación entre $h(4)$ y los valores anteriores $h(3)$, $h(2)$, $h(1)$ y $h(0)$.

$$\begin{aligned}
 h(4) &= h(0) \cdot c_4 + h(0) \cdot c_1 \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{1}{2} h(0) \cdot c_1^2 \cdot c_2 + \frac{1}{24} h(0) \cdot c_1^4 = \\
 &= h(0) \cdot c_4 + h(1) \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{1}{2} h(1) \cdot c_1 \cdot c_2 + \frac{1}{12} \left(\frac{1}{2} h(1) \cdot c_1 \right) \cdot c_1^2 = \\
 &= h(0) \cdot c_4 + h(1) \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{1}{2} h(1) \cdot c_1 \cdot c_2 + \frac{1}{12} (h(2) - h(0) \cdot c_2) \cdot c_1^2 = \\
 &= h(0) \cdot c_4 + h(1) \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{1}{2} h(1) \cdot c_1 \cdot c_2 + \frac{1}{12} h(2) \cdot c_1^2 - \frac{1}{12} h(1) \cdot c_1 \cdot c_2 = \\
 &= h(0) \cdot c_4 + h(1) \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{5}{12} h(1) \cdot c_1 \cdot c_2 + \frac{1}{4} \left(\frac{1}{3} h(2) \cdot c_1 \right) \cdot c_1 = \\
 &= h(0) \cdot c_4 + h(1) \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{5}{12} h(1) \cdot c_1 \cdot c_2 + \frac{1}{4} \left(h(3) - h(0) \cdot c_3 - \frac{2}{3} h(1) \cdot c_2 \right) \cdot c_1 = \\
 &= h(0) \cdot c_4 + h(1) \cdot c_3 + \frac{1}{2} h(0) \cdot c_2^2 + \frac{5}{12} h(1) \cdot c_1 \cdot c_2 + \frac{1}{4} h(3) \cdot c_1 - \frac{1}{4} h(1) \cdot c_3 - \frac{1}{6} h(1) \cdot c_1 \cdot c_2 = \\
 &= h(0) \cdot c_4 + \frac{3}{4} h(1) \cdot c_3 + \frac{1}{2} \left(\overbrace{h(0) \cdot c_2 + \frac{1}{2} h(1) \cdot c_1}^{h(2)} \right) \cdot c_2 + \frac{1}{4} h(3) \cdot c_1 = \\
 &= h(0) \cdot c_4 + \frac{3}{4} h(1) \cdot c_3 + \frac{2}{4} h(2) \cdot c_2 + \frac{1}{4} h(3) \cdot c_1
 \end{aligned} \tag{B18}$$

de los resultados obtenidos ya estamos en condiciones de encontrar la expresión recursiva con la que obtener los diferentes coeficientes de la descomposición de Wold.

$$h(n) = \sum_{k=0}^{n-1} \frac{k+1}{n} h(n-k-1) \cdot c_{k+1} \quad (\text{B19})$$

Hemos encontrado por tanto dos procedimientos para calcular $h(n)$ $\{n=0\dots m\}$. En ambos casos debemos estimar el cepstrum del proceso estocástico, cosa que podemos hacer a partir de una estimación del espectro $\hat{S}(\omega)$, como transformada de Fourier inversa de $\log \hat{S}(\omega)$. Conocidos los coeficientes del cepstrum podemos aplicar (B13) (forma cerrada) o la recursión (B19) [CRO83].

Referencias:

REFERENCIAS

- [ATY96] A. Atyia, El-Shoura, S. Shaheen, El-Sherif: "River Flow Forecasting Using Neuronal Networks". *Proc. World Congress on Neuronal Networks*, pp. 461-464, San Diego 1996.
- [BER01] Neil J. Bershad, Patrick Celka, and Stephen McLaughlin: "Analysis of Stochastic Gradient Identification of Wiener-Hammerstein Systems for Nonlinearities with Hermite Polynomial Expansions". *IEEE Transactions on Signal Processing*, vol. 49, n° 5, pp. 1060-1072, May 2001.
- [BER98] H.P. Bernhard: "A tight upper bound on the gain of linear and nonlinear predictors for stationary stochastic processes". *IEEE Transactions on Signal Processings*, vol. 46, No. 11, November 1998.
- [BIL77] S.A. Billings and S.Y. Fakhouri: "Identification of nonlinear systems using the Wiener model". *Electronic Letters*, Vol. 13, pp. 502-504, 1977.
- [BIL79] S.A. Billings and S.Y. Fakhouri: "Nonlinear system identification using the Hammerstein model". *Int. J. Syst. Sci.*, Vol. 10, pp. 567-578, June 1979.
- [BOU93] M. Boutayeb, M. Darouach, H. Rafaralahy, and G. Krzakala: "A new technique for identification of MISO Hammerstein model". *Proc. Amer. Control Conf.*, pp. 1991-1992, California 1993.
- [BOU95] M. Boutayeb and M. Darouach: "Recursive identification method for MISO Wiener-Hammerstein model". *IEEE Trans. Automatic Control*, Vol. 40, No. 2, pp. 287-291, February 1995.
- [BUS74] J.J. Bussgang, L. Ehrman, and J.W. Graham: "Analysis of nonlinear systems with multiple inputs". *Proceeding IEEE*, Vol.62, No.8, pp. 1088-1119, August 1974.
- [CAR97] Mark J. Carlotto: "Detection and Analysis of Change in Remotely Sensed Imagery with application to wide area surveillance". *IEEE Transactions on Image Processing*, vol. 6, n°1, pp. 189-202, January1997.
- [CAS89] M. Casdagli: "Nonlinear prediction of chaotic time series". *Physica D*, vol. 35, pp. 335-356, North-Holland, Amsterdam, 1989.

REFERENCIAS

- [CEL01] Patrick Celka, Neil J. Bershad, and Jean-Marc Vesin: "Stochastic Gradient Identification of Polynomial Wiener Systems: Analysis and Application". *IEEE Transactions on Signal Processing*, vol. 49, n° 2, pp. 301-313, February 2001.
- [CHA71] F.H.I. Chan and R. Luus: "A non-iterative method for identification using Hammerstein model". *IEEE Trans. Autom. Control*, Vol. AC-16, pp. 464-468, 1971.
- [CHA98] S.L. Chang and T.Ogunfunmi: "LMS/LMF and RLS Volterra system identification based on nonlinear Wiener model". *In Proc. ISCAS*, vol. 5, pp. 206-209, 1998.
- [CHE90] S. Chen, A. Billings, and P.M. Grant: "Non-linear system identification using neural networks". *Int. J. Contr.*, vol. 51, pp. 1191-1214, 1990.
- [CHE97] G. Chen, Y. Chen, and H.Ogmen: "Identifying chaotic systems via a Wiener-type cascade model". *IEEE Contr. Sys. Mag.*, vol. 17, pp. 29-36, 1997.
- [CHN96] E.S Chng, S. Chen, and B. Mulgrew: "Gradient Radial Basis Function Networks for Nonlinear and Nonstationary Time Series Prediction". *IEEE Transactions on Neuronal Networks*, vol. 7, n° 7, pp. 190-194, January 1996.
- [CHU88] H. Chun and S.Y. York: "Analysis and parameter estimation of nonlinear systems with Hammerstein model using Taylor series approach". *IEEE Transactions on Circuits and Systems*, Vol. 35, No. 12, pp. 1539-1541, December 1988.
- [CRO83] R.E. Crochiere and L.R. Rabiner: "Multirate Signal Processing". *Englewood Cliff, NJ: Prentice Hall*, 1983.
- [DZW99] M. Dzwoncyk and T.H.Y. Meng: "How good is your predictor?. Expanding confidence intervals to define probability densities on adaptive parameters". *Pro. ICASSP'1999*, 1999.
- [ERT00] M. Ertugrul Çelebi: "A general procedure for nonlinear system identification using higher order statistics". *Proc. ICASSP'2000*, Istanbul, Turkey, June 2000.
- [ERT97] M. Ertugrul Çelebi: "General formula for conditional mean using higher-order statistics". *Electronics Letters*, Vol. 33, pp. 2097-2099, 1997.

REFERENCIAS

- [ESK91] E. Eskinat, S.H. Johnson, and W.L. Luyben: "Use of the Hammerstein models in identification of nonlinear systems". *Amer. Inst. Chem. E . J.*, Vol. 37, pp. 255-268, 1991.
- [FRE76] A. S. French: "Practical nonlinear system analysis by Wiener kernel estimation in the frequency domain". *Biol. Cybern.*, Vol. 24, No. 2, pp. 111-119, 1976.
- [GAL76] P.G. Gallman: "A comparison of two Hammerstein model identification algorithms". *IEEE Transactions Autom. Control*, Vol. AC-21, pp. 124-126, 1976.
- [GAR94] Lee M. Garth, H. Vincent Poor: "Detection of Non-Gaussian Signals: A Paradigm for Modern Statistical Signal Processing". *Proceedings of the IEEE*, vol. 82, n° 7, pp. 1061-1095, July 1994.
- [GOL89] G.H.Golub, C.F.Van Loan: "Matrix Computations". *Johns Hopkins University Press, Baltimore*, 1989.
- [GUE97] D. Guegan, L. Mercier: "Prediction in chaotic time series: Methods and comparisons using simulations". *Proc. European Conference on Signal Analysis and Prediction (ECSAP'97)*, pp. 222-227, Prague 1997.
- [HAB87] R. Haber: "Structure identification of the simple Hammerstein and Wiener cascade models from impulse and step response". *Report, Institute of Machine and Process Automation, Technical University of Vienna, Austria*, 1987.
- [HAL69] A.V. Hall: "Generalisation of Wiener 's theory of nonlinear systems for process identification". *IEEE Trans. Auto. Control*, Vol. AC-14, pp. 312-313, June 1969.
- [HAY91] S. Haykin: "Advances in Spectrum Analysis and Array Processing". *Prentice Hall*, vol. 1, 1991.
- [HAY94a] S. Haykin: "Neural Networks: A Comprehensive Foundation". *Macmillan College Publishing Company, New York*, 1994.
- [HAY94b] S. Haykin: "Blind deconvolution". *Prentice-Hall, Englewood Cliffs, NJ*, 1994.

REFERENCIAS

- [HAY95] Simon Haykin, Liang Li: "Nonlinear Adaptive Prediction of Nonstationary Signals". *IEEE Transactions on Signal Processing*, vol. 43, n° 2, pp. 525-535, February 1995.
- [HAY96] S. Haykin: "Neural Networks Expand SP's Horizons" *IEEE Signal Processing Magazine*, vol. 13, pp 24-49, March 1996.
- [HUA98] A. Huang, J.M.A. Tanskanen, and O.I. Hartimo: "Design of optimum power estimator based on Wiener model applied to mobile transmitter power control". *In Proc. ISCAS*, vol. 5, pp. 249-252, 1998.
- [HUS93] D.R. Hush, and B.G. Horne: "Progress in supervised neural networks: What's new since Lippmann?". *IEEE Signal Processing Magazine*, vol. 10, pp. 8-39, 1993.
- [IEE97] "Special Issue on Automatic Target Detection". *IEEE Trans. Image Process.* 6 vol. 1 .January 1997.
- [IEE98] S.Haykin, J.Principe; "Using Neural Networks to dynamically model chaotic events such as sea clutter: Making sense of a complex World". *IEEE Signal Processing Magazine*, vol. 15, May 1998.
- [KAN98] H.W. Kang, Y.S. Cho, and D.H. Youn: "Adaptive precompensation of Wiener systems". *IEEE Trans. Signal Processing*, vol. 46, pp. 2825-2829, October 1998.
- [KAY81] S.M. Kay, and S.L. Marple: "Spectrum Analysis: A modern Perspective". *Proc. IEEE* vol. 69, pp. 1380-1419, 1981.
- [KAY88] S.M. Kay: "Modern Spectral Estimation: Theory and Applications". *Prentice Hall*, 1988.
- [KAY98] S.M. Kay: "Fundamentals of Statistical Signal Processing". vol. 2, *Detection Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1998.
- [KOR86] M.J. Korenberg and W. Hunter: "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models". *Biological Cybernetics*, Vol. 55, pp. 125-134, 1986.

REFERENCIAS

- [KOR91] M.J. Korenberg: "Parallel cascade identification and Kernel estimation for nonlinear systems". *Ann. Biomed. Eng.*, vol. 19, pp. 429-455, 1991.
- [KUB98] G.Kubin: "On the Nonlinearity of Linear Prediction". *Proc. IX European Signal Processing Conference (EUSIPCO'98)*, pp. 557-560, Rodas 1998.
- [LEE93] J. Lee and V.J. Mathews: "On the extended RLS adaptive bilinear filters". *In Proceedings 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 428-431, Minneapolis, Minnesota, 1993.
- [LEE94] Junghsi Lee, V. John Mathews: "Adaptive Bilinear Predictors". *Proc. ICASSP-94*, pp. 489-492, Adelaida 1994.
- [LIP87] R.P. Lippmann: "An introduction to computing with neural nets". *IEEE ASSP Magazine*, vol 4, pp. 4-22, 1987.
- [LIU97] Zhenqing Liu, Mingda Lu, Moan Wei: "Structure noise reduction of ultrasonic signals using artificial neural network adaptive filter". *Elsevier, Ultrasonics* 35 pp. 325-328, 1997.
- [LJU87] L. Ljung: "System Identification Theory for the User". *Englewood Cliffs, NJ: Prentice-Hall*. 1987.
- [MAH98] M. Mahfouf and D.A. Linkens: "Non-linear generalized predictive control (NLPGPC) applied to muscle relaxant anesthesia". *Int. J. Contr.*, vol. 71, pp. 239-257, 1998.
- [MAK98] G.N. Maksyn, R.E. Kearney, and J.H.T. Bates: "Nonparametric block-structured modeling of lung tissue strip mechanics". *Ann. Biomed. Eng.*, vol. 26, pp. 242-252, 1998.
- [MAR87] S.L. Marple: "Digital Spectral Analysis with applications". *Prentice Hall, Englewood Cliffs, NJ*, 1987.
- [MAT96] V.J. Mathews: "Adaptive Volterra filters using orthogonal structures". *IEEE Signal Processing Lett.*, vol. 3, pp307-309, November 1996.

REFERENCIAS

- [MUL91] C.T. Mullis, L.L. Scharf: "Quadratic Estimators of the Power Spectrum" , in *S. Haykin (Ed.), Advances in Spectrum Analysis and Array Processing I, Prentice-Hall, Englewood Cliffs*, pp. 1-16, 1991.
- [MUL96] Bernard Mulgrew: "Applying Radial Basis Functions". *IEEE Signal Processing Magazine*, vol. 13, pp 50-64, March 1996.
- [NEU97] André Neubauer: "Self-Adaptive Evolution Strategies for the Adaptation of Bilinear Predictors in Time Series Analysis". *Proc. European Conference on Signal Analysis and Prediction (ECSAP'97)*, pp. 222-227, Prague 1997.
- [NIK93] C.L. Nikias, A.P. Petropolu: "Higher-order spectra analysis: A nonlinear signal processing framework". *Prentice Hall*, 1993.
- [NIS96] Kenneth C. Nisbert, Bernat Mulgrew, Stephen McLaughlin: "Reduced State Methods in Nonlinear Prediction". *Signal Processing* vol. 48, pp. 37-49, 1996.
- [NOR01] Anders E. Nordsjö and Lars H.Zetterberg: "Identification of Certain Time-Varying Nonlinear Wiener and Hammerstein Systems". *IEEE Transactions on Signal Processing*, vol 49, n° 3, pp. 577-592 March 2001.
- [NOR98] A.E. Nordsjö and L.H. Zetterber: "Algorithms for identification of certain classes of time-varying nonlinear Wiener and Hammerstein systems". *Technical Report IR-S3-SB-9840, Royal Institute of Technology, Stockholm, Sweden*, 1998.
- [OZD96] M.T. Özden, A.H. Kayran and E. Panayirci: "Adaptive Volterra filtering with complete lattice orthogonalization". *IEEE Trans. Signal Processing*, vol. 44, pp. 2092-2098, August 1996.
- [PAW91] M. Pawlak: "On the series expansion approach to the identification of Hammerstein systems". *IEEE Transactions on Autom. Control*, Vol. 36, pp. 736-767, 1991.
- [PIT90] I. Pitas, A.N. Venetsanopoulos: "Nonlinear Digital Filters". *Kluwer Academic Publishers*, 1990.

REFERENCIAS

- [PON98] A. Poncet, G.S. Moschytz: "Estimating the predictability and the linearity of a process by Kernels". *Proc. IX European Signal Processing Conference (EUSIPCO'98)*, Rodas 1998.
- [PRI88] M.B. Priestly: "Non-linear and Non-stationary Times Series Analysis". *Academic Press, London*, 1988.
- [PRO92] J.G. Proakis, C.M. Rader, F. Ling, C.L. Nikias: "Advanced digital signal processing". *MacMillan Publishing Company*, 1992.
- [PRO96] Ales Procházka, Martina Mudrová, Jaroslav Fiala: "Nonlinear Time-Series Modelling and Prediction". *Neuronal Network World*, vol. 22, pp. 215-221, February, 1996.
- [PRO97] Ales Procházka, Martin Storek: "Neuronal Networks in Signal Prediction". *Proc. First European Conference on Signal Analysis and Prediction (ECSAP'97)*, pp. 228-231, Praga 1997.
- [RAL95] J. Ralston and A. Zoubir: "Identification of time-varying Hammerstein models". *Proc. ICASSP '95*, pp. 1685-1688, 1995.
- [RAL97] J. C. Ralston, A. M. Zoubir, and B. Boashash: "Identification of a Class of Nonlinear Systems Under Stationary Non-Gaussian Excitation". *IEEE Transactions on Signal Processing*, vol 45, n° 3, pp. 719-735 March 1997.
- [RUD79] M. Rudko and D.D. Weiner: "Optimum nonlinear Wiener filters". *J. Franklin Inst.*, Vol. 308, No.1, pp. 57-65, July 1979.
- [SCH81] M. Schetzen: "Nonlinear System Modeling Based on the Wiener Theory". *Proceedings of IEEE*, Vol. 69, pp. 1557-1573, December 1981.
- [SCH91] Louis L. Scharf: "Statistical Signal Processing: Detection, Estimation, and Times Series Analysis". *Addison-Wesley*, 1991.
- [SCO97] I. Scott and B. Mulgrew: "Nonlinear system identification and prediction using orthonormal functions". *IEEE Trans. Signal Processing*, vol. 45, pp. 1842-1853, July 1997.

REFERENCIAS

[SHI94] J. Shi and H.H. Sun: "Advanced Methods of Physiological System Modeling- Identification of Nonlinear System with Feedback Structure". *V.Z. Marmarelis, Ed. New York: Plenum, 1994.*

[SUG90] G. Sughira, R.M. May: "Nonlinear forecasting as a way of distinguishing chaos from measurement error in times series". *Nature*, vol. 344, 1990.

[SUY96] J.A.K. Suykens, J.P.L. Vandewalle, and B.L.R. De Moor: "Artificial Neural Networks for Modelling and Control of Non-Linear Systems". *Boston, MA: Kluwer, 1996.*

[SUY98] Johan A.K. Suykens, Joos Vandewalle: "Nonlinear Modeling". *Kluwer Academic Publisher, 1998.*

[THO82] D.J. Thomson: "Spectrum Estimation and Harmonic Analysis". *Proceedings IEEE* vol. 70, pp. 1055-1096, 1982.

[WIE58] N. Wiener: "Nonlinear Problems in Random Theory". *M.I.T. Press, Cambridge, 1958.*

Publicaciones y Congresos

PUBLICACIONES:

L. Vergara, P. Bernabeu: “Automatic signal detection applied to fire control by infrared digital signal processing”. *Signal Processing*, vol. 80 (4) pp. 659-669, 2000.

L. Vergara, P. Bernabeu: “Simple approach to nonlinear prediction”. *Electronics Letters*, vol. 37 (14) pp. 926-928, July 2001.

CONGRESOS:

L. Vergara, P. Bernabeu, J. Igual: “Wide Area Fire Surveillance by Infrared Digital Signal Processing”. *Proceedings of the third International Conference on Information Fusion*, 2000.

P. Bernabeu, I. Bosch, L. Vergara: “Nonlinear prediction of infrared data by the Wiener system”. (Remitida a EUSIPCO-2002)