

# Predictor Intervalar basado en hiperplano soporte

J.M. Bravo, M.J. Vasallo y E. Cojocarú

Departamento de Ingeniería Electrónica, de Sistemas Informáticos y Automática, Universidad de Huelva  
email:{caro,manuel.vasallo}@uhu.es

T. Alamo

Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla  
email:talamo@us.es

## Resumen

En este trabajo se aborda el problema de construir un modelo predictor. Dicho modelo debe ser capaz de predecir la salida futura de un sistema dinámico, haciendo uso de información pasada guardada en históricos. Este trabajo se centra en el estudio de modelos predictores que proporcionan una predicción de tipo intervalar. Una predicción intervalar está compuesta por un límite inferior y otro superior. La salida del sistema debe estar comprendida entre ambos límites del intervalo. Dado un conjunto de datos obtenidos del sistema dinámico, la predicción intervalar se calcula utilizando hiperplanos soporte a dichos datos. Cualquier punto dentro del intervalo se puede utilizar como predicción puntual. En concreto, en este trabajo se propone la utilización del punto central entre hiperplanos soporte. La principal bondad del predictor propuesto es su capacidad de proporcionar predicciones donde el tamaño del intervalo y el error de la predicción puntual están balanceados. Un parámetro de diseño incluido en el predictor permite compensar dichos objetivos. Además el trabajo propone una metodología basada en validación cruzada para realizar el ajuste de dicho parámetro. El trabajo incluye un ejemplo con datos reales con el fin de ilustrar el comportamiento del predictor.

**Palabras clave:** Predicción, error acotado, identificación de sistemas.

## 1. INTRODUCCIÓN

La mayoría de los métodos de predicción se basan en un modelo matemático del sistema. Este modelo se puede obtener aplicando los principios físicos que definen el funcionamiento del sistema. Sin embargo, el modelado de un sistema requiere los conocimientos de un experto y el ajuste de los parámetros que se incluyan en el modelo. Otra opción es aplicar técnicas de identificación de sistemas [13]. Los métodos de identificación de sistemas infieren modelos dinámicos a partir de datos. Uno de los métodos más utilizados es el basado en error de predicción. Dada una familia de mo-

delos parametrizada y un conjunto de datos, estos métodos seleccionan el valor de los parámetros que minimizan cierto criterio, por ejemplo, el error cuadrático de predicción. Una vez identificado el modelo, aplicando nuevas medidas se pueden obtener predicciones del futuro comportamiento del sistema. Nótese que con este esquema de trabajo, una vez seleccionado el modelo de sistema, el conjunto de datos usado para la identificación ya no es necesario para realizar las predicciones.

Una alternativa a lo expuesto en el párrafo anterior son los métodos no paramétricos [9]. Cuando se necesita una predicción, los métodos paramétricos acceden al conjunto histórico de datos del sistema y lo utilizan para construir la predicción requerida. En este contexto se han propuesto predictores basados en los vecinos más cercanos, linealización local o funciones kernel [8].

Los sistemas de control automático utilizan sistemas predictores para optimizar su rendimiento. Ejemplos en este sentido los podemos encontrar en los sistemas de control predictivo [5] y en los sistemas de detección y diagnosis de fallos [11]. Una extensión interesante de los sistemas de predicción son los predictores intervalares. Dichos predictores proporcionan una predicción intervalar basada en una cota inferior y otra superior. La predicción intervalar acota externamente el futuro comportamiento del predictor teniendo en cuenta perturbaciones, ruido o dinámicas no modeladas. Además, un punto del intervalo se puede utilizar como predicción puntual. Los predictores intervalares son componentes importantes de sistemas de control robusto [3] o sistemas de detección de fallos en presencia de incertidumbres [10, 6].

Una forma de obtener predictores intervalares es utilizar métodos de identificación de sistemas basados en error acotado [14]. En este caso se considera que la salida del sistema está contaminada por un ruido desconocido pero acotado. Teniendo en cuenta una familia paramétrica de modelos, un conjunto de medidas y la acotación del ruido, estos métodos obtienen el conjunto de parámetros que es consistente con las medidas, la familia de modelos y el error acotado. Para acotar dicho conjunto de parámetros se han utilizado diferen-

tes figuras geométricas. Elipsoides, paralelotos y zonotos son las más conocidas [7, 16, 1]. Los métodos de acotación también se pueden formular desde el punto de vista no paramétrico. En [15] se asume que el modelo pertenece a un conjunto de funciones con constante de Lipschitz acotada. En [4] se propone una familia de predictores intervalares que utiliza datos almacenados y norma-1 para acotar la predicción.

En este trabajo se propone un predictor intervalar basado en el concepto de hiperplano soporte. Dado un conjunto de datos, el predictor calcula planos soporte superior e inferior a dicho conjunto y los utiliza para proporcionar la predicción intervalar. Dichos planos se obtienen resolviendo un problema de optimización con restricciones convexas. El predictor incluye un parámetro que permite balancear su rendimiento. En un límite del valor del parámetro, el método proporciona la predicción intervalar de menor anchura proporcionada por los hiperplanos soporte. En el límite opuesto, el predictor minimiza los errores cuadráticos de la predicción puntual. Valores intermedios generan predicciones compensadas entre dichos extremos. En el artículo se propone una metodología basada en validación cruzada para estimar el valor adecuado de dicho parámetro.

El método propuesto presenta una serie de ventajas respecto a otros métodos conocidos para obtener predictores intervalares. A diferencia de los métodos basados en error acotado, no es necesario conocer a priori una cota del error en las medidas. Tampoco es necesario asumir como conocida la distribución estadística del error asociado a las medidas, información sí necesaria para obtener intervalos de confianza en el contexto estocástico. Finalmente, el método propuesto está relacionado con la regresión cuantil [12]. En regresión basada en cuantiles, se resuelve un problema de optimización que pondera la penalización de errores de signo positivo o negativo. La principal ventaja del método propuesto respecto a la regresión cuantil, es la posibilidad antes comentada de poder balancear el rendimiento de la predicción puntual y la intervalar.

En la Sección 2 se formulará el problema. El predictor propuesto se presenta en la Sección 3. Algunas consideraciones prácticas se pueden consultar en la Sección 4. La Sección 5 muestra un ejemplo que ilustra el método propuesto. Finalmente algunas conclusiones y trabajo futuro se pueden ver en las Secciones 6 y 7.

## 2. FORMULACIÓN DEL PROBLEMA

Considérese un sistema dinámico en tiempo discreto. En el instante de tiempo  $k$ ,  $y_k \in \mathbb{R}$  denota una salida medida del sistema y  $x_k \in X \subseteq \mathbb{R}^{n_x}$ , siendo  $X$  un conjunto compacto, un vector de medidas del sistema. Este vector puede incluir, por ejemplo, medidas pasadas de entradas y salidas del sistema. La formulación matemática que relaciona  $y_k$  y  $x_k$  es desconocida. Dado el vector  $x_k$ , el problema planteado en este trabajo es obtener un intervalo  $IP_{\mathcal{D}}(x_k) = [\underline{y}_k, \bar{y}_k]$  tal que  $\underline{y}_k \leq y_k \leq \bar{y}_k$ .

El predictor intervalar propuesto se basa en tomar un conjunto  $\mathcal{D}$  de datos de entrada y salida del sistema e identificar un modelo de predicción. La Hipótesis 1 establece la necesidad de disponer de dicho conjunto.

**Hipótesis 1** *Se dispone de un conjunto  $\mathcal{D} = \{(x_i, y_i), \text{ con } i = 1, \dots, N\}$  de medidas del sistema obtenido con un nivel de excitación adecuado.*

La futura capacidad de predicción del modelo está relacionada con la similitud del comportamiento futuro y pasado del sistema al que se aplica. Todo sistema de inferencia se basa en ese principio. Por ejemplo, si estamos en el contexto estocástico, se asume que las propiedades de dos muestras aleatorias de una misma variable deben ser muy similares. En el presente trabajo, la Hipótesis 2 concreta el concepto de similitud que se asume a la hora de aplicar el predictor intervalar propuesto.

**Hipótesis 2** *Sean  $\mathcal{D}_A$  y  $\mathcal{D}_B$  dos conjuntos de datos de entrada-salida obtenidas del sistema a partir de una excitación adecuada. La envolvente convexa del conjunto  $\mathcal{D}_A$  tiende a ser la misma que la envolvente convexa del conjunto  $\mathcal{D}_B$  cuando el número de muestras crece.*

## 3. PREDICTOR PROPUESTO

Tal y como se comentó en la Introducción, el predictor intervalar se basa en el concepto de hiperplano soporte [2].

**Definición 1 (Hiperplano soporte)** *Dado un conjunto  $\mathcal{C}$  y un punto  $z_o \in \text{bd}\mathcal{C}$  en la frontera de  $\mathcal{C}$ . Si  $\beta \neq 0$  satisface  $\beta^T z \leq \beta^T z_o$  para todo  $z \in \mathcal{C}$ , entonces el hiperplano  $\{z : \beta^T z = \beta^T z_o\}$  es un hiperplano soporte a  $\mathcal{C}$  en el punto  $z_o$ .*

Existen muchas formas de obtener un hiperplano soporte a un conjunto dado. En la siguiente definición se establece un tipo de hiperplano que tal y como se muestra en la Propiedad 1 es soporte al

conjunto  $\mathcal{D}$ . Este hiperplano soporte es una de las ideas en las que se apoya el predictor intervalar desarrollado.

**Definición 2** Dado el conjunto  $\mathcal{D}$  se define el hiperplano  $\mathcal{H}_{\mathcal{D}}$  como

$$\mathcal{H}_{\mathcal{D}} = \{ [x^T \ y]^T : [x^T \ 1]\theta^* - y = 0 \}$$

donde  $\theta^*$  se obtiene resolviendo el problema de optimización

$$\begin{aligned} \min_{\theta} \quad & \sum_{i=1}^N (y_i - [x_i^T \ 1]\theta)^2 \quad (1) \\ \text{s.t.} \quad & [x_i^T \ 1]\theta \leq y_i \quad i = 1, \dots, N. \end{aligned}$$

**Propiedad 1** Una propiedad interesante del hiperplano  $\mathcal{H}_{\mathcal{D}}$  es que es un hiperplano soporte a  $\mathcal{D}$ .

*Prueba.* Para realizar la demostración hay que probar que el hiperplano  $\mathcal{H}_{\mathcal{D}}$  incluye al menos uno de los puntos del conjunto  $\mathcal{D}$  y que el resto de puntos están en uno de los semiespacios creados por el hiperplano. Se procede a la demostración por reducción al absurdo. Supongamos que  $\theta^*$  cumple las restricciones del problema y no contiene a ningún punto. Esto es equivalente a decir que las restricciones del problema de optimización se cumplen de forma estricta:  $[x_i^T \ 1]\theta^* < y_i, i = 1, \dots, N$ . Esto se puede expresar diciendo que  $[x_i^T \ 1]\theta^* = y_i - \epsilon_i$ , donde  $\epsilon_i > 0, i = 1, \dots, N$ . Entonces, tomando  $s = \arg \min_{i=1, \dots, N} \epsilon_i$ , se puede definir

$$\hat{\theta} = \theta^* + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \epsilon_s.$$

Por definición:

$$\begin{aligned} [x_i^T \ 1]\hat{\theta} &= [x_i^T \ 1]\theta^* + \epsilon_s \\ &= y_i - \epsilon_i + \epsilon_s \leq y_i - \epsilon_i + \epsilon_i = y_i. \end{aligned}$$

De lo anterior se deduce que  $\hat{\theta}$  es una solución factible al problema. Por otro lado, nótese que

$$\begin{aligned} (y_i - [x_i^T \ 1]\hat{\theta})^2 &= (y_i - [x_i^T \ 1]\theta^* - \epsilon_s)^2 \\ &= (\epsilon_i - \epsilon_s)^2 \\ &< \epsilon_i^2 \\ &= (y_i - [x_i^T \ 1]\theta^*)^2. \end{aligned}$$

Finalmente sumando la anterior desigualdad desde  $i = 1$  hasta  $i = N$  obtenemos que:

$$\sum_{i=1}^N (y_i - [x_i^T \ 1]\hat{\theta})^2 < \sum_{i=1}^N (y_i - [x_i^T \ 1]\theta^*)^2.$$

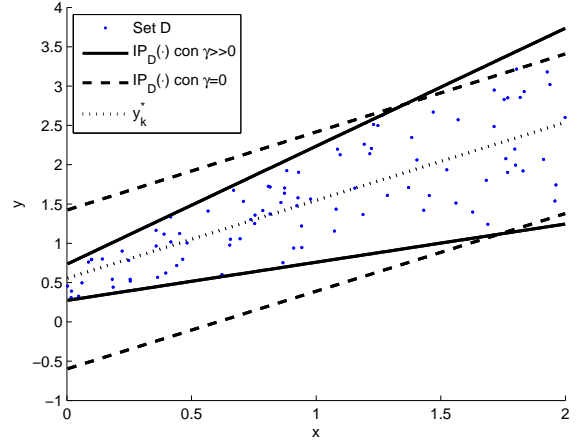


Figura 1: Ejemplo de aplicación de  $IP_{\mathcal{D}}(\cdot)$ .

Pero esto contradice la optimalidad de  $\theta^*$ . Se ha demostrado pues que

$$[x_i^T \ 1]^T \theta^* \leq y_i, i = 1, \dots, N.$$

donde la desigualdad se hace igualdad al menos en un índice  $s$ . Es decir, existe  $s$  tal que  $[x_s^T \ 1]\theta^* = y_s$ . Si restamos esta igualdad a las desigualdades anteriores obtenemos

$$[(x_i - x_s)^T \ 0]^T \theta^* \leq y_i - y_s, i = 1, \dots, N.$$

Denotando  $\xi^*$  a las primeras  $n$  componentes de  $\theta^*$  se obtiene

$$(x_i - x_s)^T \xi^* \leq y_i - y_s, i = 1, \dots, N.$$

Finalmente hacemos

$$\beta = \begin{bmatrix} \xi^* \\ -1 \end{bmatrix}.$$

Con esta definición

$$[x_i^T \ y_i]\beta \leq [x_s^T \ y_s]\beta, i = 1, \dots, N.$$

■

**Observación 1** Por definición se tiene que  $[x_i^T \ 1]\theta^* \leq y_i$  para  $i = 1, \dots, N$ . Por tanto, el hiperplano  $\mathcal{H}_{\mathcal{D}}$  acota inferiormente a la envolvente convexa del conjunto  $\mathcal{D}$  en la componente  $y$ . Además, por la Propiedad 1, esta cota es ajustada pues existe un índice  $s$  tal que  $[x_s^T \ 1]\theta^* = y_s$ .

El problema de optimización (2) se puede modificar para obtener un hiperplano que acote superiormente el conjunto  $\mathcal{D}$  en la componente  $y$ . Bastaría con cambiar el sentido de las desigualdades,

esto es,  $[x_i^T \ 1]\theta \geq y_i$  con  $i = 1, \dots, N$ . Esta idea nos permite proponer un predictor intervalar para la salida del sistema. El predictor obtiene sendos hiperplanos que acotan superiormente e inferiormente la envolvente convexa del conjunto  $\mathcal{D}$  en la componente  $y$ .

**Definición 3** Dado un vector  $x_k$  se define el predictor intervalar  $IP_{\mathcal{D}}(x_k) = [\underline{y}_k, \bar{y}_k]$  donde

$$\underline{y}_k = [x_i^T \ 1]\underline{\theta}^* - \underline{\alpha}^*$$

$$\bar{y}_k = [x_i^T \ 1]\bar{\theta}^* + \bar{\alpha}^*$$

y siendo  $\underline{\theta}^*, \bar{\theta}^*, \underline{\alpha}^*$  y  $\bar{\alpha}^*$  las soluciones óptimas de los problemas de optimización

$$\begin{aligned} \min_{\underline{\theta}, \underline{\alpha}} \quad & \sum_{i=1}^N (y_i - [x_i^T \ 1]\underline{\theta})^2 + \gamma \underline{\alpha} \\ \text{s.t.} \quad & [x_i^T \ 1]\underline{\theta} \leq y_i + \underline{\alpha} \quad i = 1, \dots, N \\ & \underline{\alpha} \geq 0 \end{aligned} \quad (2)$$

$$\begin{aligned} \min_{\bar{\theta}, \bar{\alpha}} \quad & \sum_{i=1}^N (y_i - [x_i^T \ 1]\bar{\theta})^2 + \gamma \bar{\alpha} \\ \text{s.t.} \quad & [x_i^T \ 1]\bar{\theta} \geq y_i - \bar{\alpha} \quad i = 1, \dots, N \\ & \bar{\alpha} \geq 0 \end{aligned} \quad (3)$$

donde  $\gamma > 0$  es un parámetro de diseño. Así mismo se propone como predicción puntual

$$y_k^* = \frac{1}{2} [x_k^T \ 1] (\bar{\theta}^* + \underline{\theta}^*).$$

En la figura 1 se muestra un ejemplo numérico del predictor propuesto. El conjunto de puntos representa el conjunto  $\mathcal{D}$ . Las acotaciones intervalares obtenidas para  $\gamma = 0$  y  $\gamma \gg 0$  se muestra con líneas discontinuas y continuas respectivamente. La línea central punteada es la predicción puntual obtenida con  $\gamma = 0$ .

**Observación 2** Nótese que las variables  $\bar{\alpha}, \underline{\alpha}$  son redundantes en (2) y (3), ya que aplicando las igualdades  $\bar{\alpha} = \max_{i=1, \dots, N} \{0, y_i - r_i^T \bar{\theta}\}$  y  $\underline{\alpha} = \max_{i=1, \dots, N} \{0, r_i^T \underline{\theta} - y_i\}$  es posible reformular los problemas de optimización (2) y (3) como problemas de optimización convexos pero sin restricciones. En este caso, es sencillo ver que cuando  $\gamma = 0$ , los problemas de optimización (2) y (3) son equivalentes a un solo problema de optimización lineal cuadrático sin restricciones. Por otro lado, las formulaciones (2) y (3) son muy cercanas a la formulación estándar de los problemas de optimización QPs, lo que facilita su implementación en un computador.

El predictor propuesto tiene un conjunto de características que se comentan a continuación. Lo primero que hay que destacar es que se puede considerar que la Definición 3 proporciona una familia de predictores intervalares parametrizada en  $\gamma$ . Si  $\gamma = 0$  entonces  $\underline{\theta}^* = \bar{\theta}^* = \theta^{LS}$  donde  $\theta^{LS} = (\Omega^T \Omega)^{-1} \Omega^T Y$  con  $\Omega = [r_1 \ r_i \ \dots \ r_N]^T$ ,  $r_i = [x_i \ 1]$ ,  $i = 1, \dots, N$  y  $Y = [y_1 \ y_2 \ \dots \ y_N]$ . De lo anterior se deduce que cuando  $\gamma = 0$ , la predicción puntual  $y_k^*$  es óptima desde el punto de vista del error cuadrático de predicción, de hecho, es la predicción obtenida mediante mínimos cuadrados. Así mismo, teniendo en cuenta que  $\underline{\alpha}^* = \max_{i=1, \dots, N} \{0, r_i^T \theta^{LS} - y_i\}$  y  $\bar{\alpha}^* = \max_{i=1, \dots, N} \{0, y_i - r_i^T \theta^{LS}\}$ , se tiene que  $\bar{y}_k - \underline{y}_k \geq \underline{\alpha}^* + \bar{\alpha}^*$ . Esta propiedad es importante porque se garantiza que para todo  $x_k \in X$  siempre existe un valor de  $\gamma \geq 0$  tal que  $\underline{y}_k \leq y_k^* \leq \bar{y}_k$ , asegurándose de esta forma la consistencia de los límites en la predicción.

Por otro lado, cuando  $\gamma \rightarrow \infty$  se tiene que  $\underline{\alpha}^*, \bar{\alpha}^* \rightarrow 0$ . La consecuencia es que los problemas de optimización (2) y (3) pertenecen a la clase presentada en la Definición 2. Por tanto, las soluciones óptimas  $\underline{\theta}^*$  y  $\bar{\theta}^*$  proporcionan sendos hiperplanos soporte que acotan externa y ajustadamente el conjunto  $\mathcal{D}$ . En este caso la predicción intervalar tiende a ser más ajustada que cuando  $\gamma \rightarrow 0$ , pero se pierde la optimalidad de la predicción puntual.

Como resumen de lo expuesto anteriormente, el predictor intervalar propuesto incluye un parámetro  $\gamma$  a definir por el usuario que permite balancear el tamaño del intervalo obtenido y el error de la predicción puntual. En la siguiente sección se proporciona un método empírico que permite ajustar dicho parámetro a partir del conjunto de datos  $\mathcal{D}$ .

## 4. Consideraciones Prácticas

El predictor intervalar propuesto  $IP_{\mathcal{D}}(\cdot)$  asume el contenido de la Hipótesis 2. Sin embargo, al igual que en cualquier sistema de inferencia, la comprobación de las hipótesis de trabajo no es sencilla. Sin embargo, hay otra forma de proceder. Dado un conjunto de datos, lo que sí se puede comprobar es si dichos datos son compatibles con las hipótesis asumidas. En este artículo se propone un esquema de trabajo basado en validación cruzada que permite rechazar la validez de lo asumido. El esquema de chequeo se basa en la  $\mathcal{D}^\mu$ -consistencia del parámetro de diseño  $\gamma$ . Básicamente la idea es, fijado un valor de  $\gamma$ , comprobar qué número de salidas  $y_i$  del conjunto de datos  $\mathcal{D}$  se predicen de forma correcta, esto es  $y_i \in IP_{\mathcal{D}_i}(x_i)$ , utilizando el predictor propuesto y el conjunto de puntos  $\mathcal{D}_i = \mathcal{D} \setminus \{(x_i, y_i)\}$ . Nótese que la pareja  $(x_i, y_i)$

se ha eliminado del conjunto  $\mathcal{D}$  para realizar la validación cruzada. La Definición 4 formaliza esta idea.

**Definición 4** Dado el predictor intervalar  $IP_{\mathcal{D}}(\cdot)$  y un valor  $\mu \in [0, 1]$ , el parámetro  $\gamma$  es  $\mathcal{D}^\mu$ -consistente si

$$\frac{1}{N} \sum_{i=1}^N \mathbb{I}_{\{IP_{\mathcal{D}_i}(x_i)\}}(y_i) \geq \mu$$

donde

$$\mathbb{I}_{\{A\}}(y) = \begin{cases} 1 & \text{si } y \in A \\ 0 & \text{si } y \notin A \end{cases}$$

es la función indicador y  $\mathcal{D}_i = \mathcal{D} \setminus \{(x_i, y_i)\}$ .

La comprobación de la  $\mathcal{D}^\mu$ -consistencia de cierto valor del parámetro  $\gamma$  se puede realizar mediante el Algoritmo 1.

**Algoritmo 1:**  $\mathcal{D}^\mu$ -consistencia( $\mathcal{D}, \gamma, \mu$ )

```

cont = 0;
for i = 1:N
    Obtener  $\mathcal{D}_i = \mathcal{D} \setminus \{(x_i, y_i)\}$ 
    if  $y_i \in IP_{\mathcal{D}_i}(x_i)$ 
        cont = cont+1
    endif
endfor
if  $\frac{cont}{N} \geq \mu$ 
    return true
else
    return false
endif

```

End

Nótese que un valor  $\gamma$  que sea  $\mathcal{D}^1$ -consistente implica que para todas las medidas  $y_i$  del conjunto  $\mathcal{D}$  se cumple que  $y_i \in IP_{\mathcal{D}_i}(x_i)$ . Sin embargo, exigir  $\mu = 1$  puede ser demasiado estricto para ciertos problemas, provocando una degradación de los resultados. Por ejemplo, puede haber dinámicas puntuales no recogidas por el modelo, o bien, los datos pueden tener valores extraños, outliers, que invaliden las predicciones intervalares. Este problema u otros similares se pueden atenuar permitiendo valores de  $\mu$  menores a 1.

Por último, el usuario debe seleccionar un valor concreto de  $\gamma$ . Una leve ampliación del Algoritmo 1 nos puede permitir realizar dicha elección. Fijado un valor  $\mu$ , se puede tomar como  $\gamma$  el valor  $\mathcal{D}^\mu$ -consistente que proporcione el mejor rendimiento en algún sentido. Por ejemplo se puede tomar el valor de  $\gamma$  que sea  $\mathcal{D}^\mu$ -consistente y minimice el tamaño medio de la anchura de la predicción intervalar. En este artículo se propone la utilización de un valor de  $\gamma$  que balancee el tamaño del intervalo y el error de la predicción puntual. En la siguiente sección se presenta un ejemplo de dicha elección.

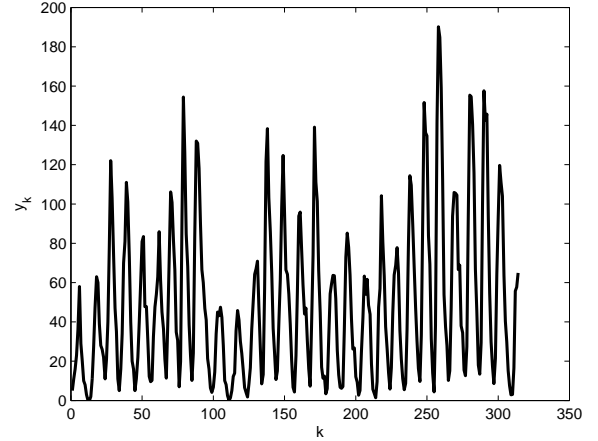


Figura 2: Serie Temporal Sunspot Numbers.

## 5. Ejemplo

Para ilustrar el nuevo predictor intervalar se ha considerado el problema de predecir una serie temporal denominada Sunspot Numbers (ver Figura 2). Esta serie temporal es interesante ya que es un ejemplo del mundo real, los datos son públicos y su predicción es relevante en muchos campos de aplicación. La serie está compuesta por 314 datos, desde el año 1700 hasta 2013 y reflejan el número de manchas solares presentes. Los primeros 219 datos se han utilizado para construir el conjunto  $\mathcal{D}$ . Los restantes 70 datos se han incluido en un conjunto de test. Se ha considerado un predictor a un paso donde  $x_k = [y_{k-1} \ y_{k-2} \ \dots \ y_{k-9}]^T$ .

En la Sección 4 se ha propuesto un método para estimar el valor del parámetro  $\gamma$  a utilizar. Con el fin de comprobar la efectividad del método se utilizará el conjunto  $\mathcal{D}$  y el método basado en validación cruzada presentado en la Sección 4 para estimar el parámetro  $\gamma$ . Finalmente se probará el rendimiento de diferentes valores de  $\gamma$  en el conjunto de test y se comprobará que el mejor valor de  $\gamma$  es muy cercano al estimado mediante validación cruzada y  $\mathcal{D}$ .

Se propone una primera métrica que mide el error cometido por la predicción puntual al predecir la serie real, se utiliza la raíz del error cuadrático medio ( $RMSE$ ).

$$RMSE_\gamma = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^* - y_i)^2}. \quad (4)$$

Por otro lado se considera una segunda métrica que mide la anchura de los intervalos de predicción

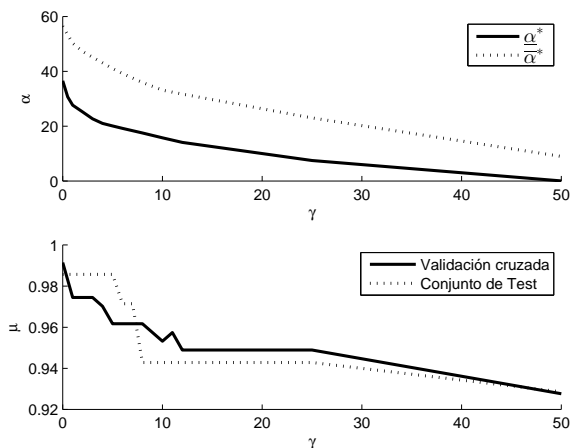


Figura 3: Coeficiente  $\alpha^*$  y parámetro  $\mu$ .

obtenidos. El índice  $INT$  se define mediante

$$INT_\gamma = \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \underline{y}_i).$$

Se propone una métrica  $BM_\gamma$  que mide el rendimiento balanceado de las dos métricas  $INT_\gamma$  y  $RMSE_\gamma$ . Esta métrica viene definida por la expresión :

$$BM_\gamma = \left( \frac{INT_\gamma - \underline{INT}}{\overline{INT} - \underline{INT}} + \frac{RMSE_\gamma - \underline{RMSE}}{\overline{RMSE} - \underline{RMSE}} \right)$$

donde  $\overline{INT}$ ,  $\underline{INT}$ ,  $\overline{RMSE}$  y  $\underline{RMSE}$  son los valores máximo y mínimo alcanzado por las métricas  $INT_\gamma$  y  $RMSE_\gamma$ .

En la Figura 3 superior se puede ver las evoluciones de los coeficientes óptimos  $\underline{\alpha}^*$  y  $\bar{\alpha}^*$  respecto los posibles valores de  $\gamma$ . Cuando  $\gamma = 0$ ,  $\underline{\alpha}^*$  y  $\bar{\alpha}^*$  tienen su valor máximo pues en ese caso la predicción puntual viene dada por  $\underline{\theta}^* = \bar{\theta}^* = \theta^{LS}$  y son  $\underline{\alpha}^*$  y  $\bar{\alpha}^*$  quienes definen la anchura del intervalo. En el otro extremo  $\underline{\alpha}^*$  y  $\bar{\alpha}^*$  tienden a cero y son los vectores óptimos  $\underline{\theta}^*$  y  $\bar{\theta}^*$  quienes definen el intervalo de predicción. En la Figura inferior se representa el valor  $\mu$  máximo alcanzado para cada valor de  $\gamma$ . Se puede observar que el perfil obtenido por validación cruzada en  $\mathcal{D}$  y el obtenido en el conjunto de test es muy parecido. Por otro lado, al incrementarse  $\gamma$ , la predicción intervalar se ajusta más por lo que normalmente se decremanta la  $\mathcal{D}^\mu$ -consistencia obtenida.

En la Figura 4 se comparan las métricas  $INT_\gamma$  y  $RMSE_\gamma$  obtenidas por validación cruzada en  $\mathcal{D}$  y las obtenidas en el conjunto de test. Como se puede apreciar los perfiles son muy parecidos. Por otro lado es normal que los valores obtenidos en el conjunto de entrenamiento  $\mathcal{D}$  sean mejores que los obtenidos en el conjunto de test, a pesar de usar

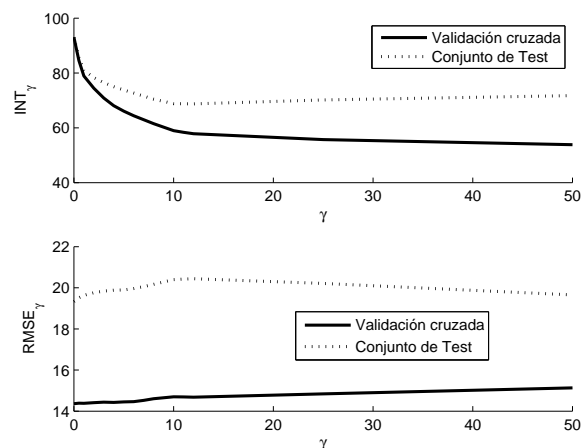


Figura 4: Métricas  $INT_\gamma$  y  $RMSE_\gamma$ .

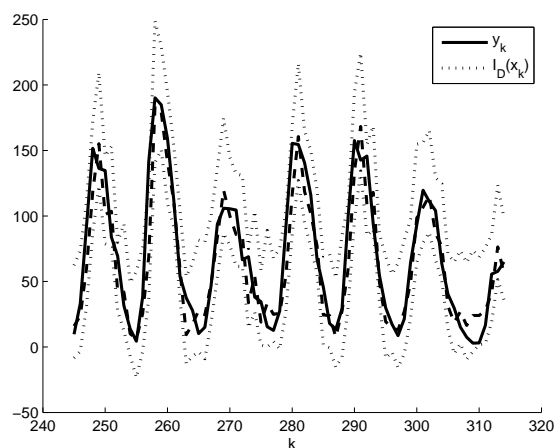


Figura 5: Predicción intervalar con  $\gamma = 4$ .

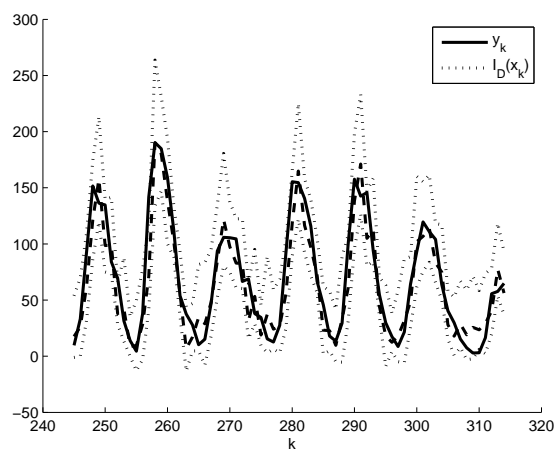
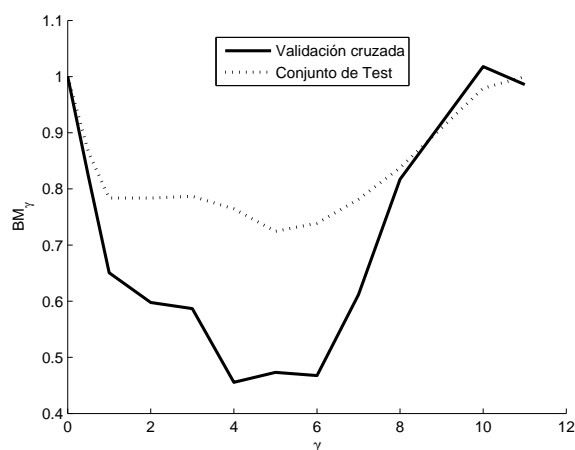


Figura 6: Predicción intervalar con  $\gamma = 11$ .

Figura 7: Métrica  $BM_\gamma$ .

validación cruzada en entrenamiento. También parece razonable que cuando el parámetro de diseño  $\gamma$  crece, la métrica  $INT_\gamma$  mejora y por otro lado  $RMSE_\gamma$  empeora. Con el fin de balancear el comportamiento de ambas métricas, se propone  $BM_\gamma$ . Nótese que la métrica  $RMSE_0$  sería la obtenida por un modelo autoregresivo AR(9) ajustado mediante mínimos cuadrados.

Fijado  $\mu = 0,95$ , en la figura 7 se compara la métrica  $BM_\gamma$  obtenida por validación cruzada en  $\mathcal{D}$  y las obtenidas en el conjunto de test para los valores  $\mathcal{D}^{0,95}$ -consistentes de  $\gamma$ . Como se puede observar el mínimo se obtiene para  $\gamma = 4$  en validación cruzada, siendo ese valor también muy cercano al mínimo en el conjunto de test. Por tanto se pueden extraer dos conclusiones. El mejor valor de  $\gamma$  para obtener un balanceo óptimo es un valor intermedio (en este caso  $\gamma$  en torno a 4) y por tanto queda justificado el predictor propuesto. Y la segunda conclusión es que el esquema de validación cruzada propuesto en este artículo, se puede utilizar en este ejemplo para inferir un buen valor de  $\gamma$ .

En la Tabla 1 se muestran los valores extremos de  $INT_\gamma$  y  $RMSE_\gamma$  en validación cruzada y test para los valores  $\mathcal{D}^{0,95}$ -consistentes de  $\gamma$ .

Cuadro 1: Valores extremos

	$\overline{INT}$	$\underline{INT}$	$\overline{RMSE}$	$\underline{RMSE}$
V.C.	92.8	58.35	14.69	14.36
Test	93.0	68.7	20.42	19.33

Finalmente, en las figuras 5 y 6 se pueden observar las predicciones intervalares (líneas punteadas) obtenidas en el conjunto de test para el valor óptimo  $\gamma = 4$  asociado a  $BM_\gamma$  y el valor  $\gamma = 11$  que está cerca del límite de la  $\mathcal{D}^{0,95}$ -consistencia. Co-

mo se puede observar la predicción intervalar para  $\gamma = 11$  tiene menor anchura a costa de incluir cierto desvío en la predicción puntual (líneas discontinuas). Los valores reales se muestran en líneas continuas.

## 6. Conclusiones

Este trabajo ha presentado una nueva metodología para diseñar predictores intervalares que permitan acotar externamente la salida futura de un sistema dinámico. El método está basado en el concepto de hiperplano soporte. La predicción intervalar se obtiene aplicando un modelo de predicción que se ha identificado resolviendo un problema de optimización convexo. El método incluye un parámetro que permite balancear el tamaño de la predicción intervalar y el error de la predicción puntual. El valor de dicho parámetro se puede estimar utilizando validación cruzada en un conjunto de datos obtenidos del sistema. Como ventajas objetivas del método presentado frente a otras metodologías, se tiene que no es necesario conocer a priori una cota máxima del ruido asociado a una medida, ni tampoco asumir ciertas funciones de distribución aleatorias en dicho ruido. Es importante indicar que ese conocimiento a priori condiciona claramente los resultados obtenidos en esas metodologías alternativas. Finalmente indicar que el método incluye la identificación de modelos lineales mediante mínimos cuadrados como caso particular.

## 7. Trabajo futuro

En esta sección se establecen algunas líneas de trabajo futuras para mejorar el funcionamiento del predictor. Una primera línea sería incluir pesos en la función de coste de los problemas de optimización (2) y (3). El objetivo de los pesos es obtener planos soporte ajustados en un entorno al vector de interés  $x_k$ . La ventaja de esa ampliación sería un mejor ajuste de la predicción intervalar, siendo la principal desventaja el tener que resolver los problemas de optimización (2) y (3) cada vez que se realiza una predicción. Este esquema de trabajo no captura comportamientos locales cuando el conjunto de puntos no es convexo. En ese caso se puede añadir a las variables  $\underline{\alpha}$  y  $\bar{\alpha}$  un componente local del tipo  $\|x_i - x_k\|$  que permite ajustar los planos soporte localmente a zonas de interés. Por otro lado, si se conoce una cota superior del posible ruido presente en las medidas  $y_i$ , se puede añadir dicha cota al predictor intervalar para obtener planos soporte más robustos y por tanto ampliar las predicciones intervalares. Estas ampliaciones y otras serán abordadas en próximos

trabajos.

### Agradecimientos

Esta investigación ha sido financiada por DPI2013-48243-C2-2-R, DPI2016-76493-C3-1-R y DPI2016-76493-C3-2-R del MEC de España.

### Referencias

- [1] T. Alamo, J. M. Bravo, and E. F. Camacho. Guaranteed state estimation by zonotopes. *Automatica*, 41(6):1035–1043, 2005.
- [2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] J. M. Bravo, T. Alamo, and E. F. Camacho. Robust MPC of constrained discrete-time nonlinear systems based on approximated reachable sets. *Automatica*, 42(10):1745–175, 2006.
- [4] J. M. Bravo, T. Alamo, M. Vasallo, and M. E. Gegundez. A general framework for predictors based on bounding techniques and local approximation. *IEEE Transactions on Automatic Control*, In press, 2017.
- [5] E.F. Camacho and C. Bordons. *Model Predictive Control*. Springer Verlag, 2 edition, 1999.
- [6] Christophe Combastel. An extended zonotopic and gaussian kalman filter (ezgkf) merging set-membership and stochastic paradigms: Toward non-linear filtering and fault detection. *Annual Reviews in Control*, 42:232–243, 2016.
- [7] E. Fogel and Y. F. Huang. On the value of information in system identification - bounded noise case. *Automatica*, 18(2):229–238, 1982.
- [8] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002.
- [9] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, 1992.
- [10] A. Ingimundarson, J.M. Bravo, V. Puig, T. Alamo, and P. Guerra. Robust fault detection using zonotope-based set-membership consistency test. *Int. J. Adapt. Control Signal Process*, 23:311–330, 2009.
- [11] Rolf Isermann. *Fault-Diagnosis systems. An Introduction from Fault Detection to Fault Tolerance*. Springer, 2006.
- [12] Roger Koenker and Jr. Gilbert Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.
- [13] L. Ljung. *System Identification*. Prentice Hall, Upper Saddle River, NJ, 1999.
- [14] M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter. *Bounding Approaches to System Identification*. Plenum Press, New York, 1996.
- [15] M. Milanese and C. Novara. Set membership prediction of nonlinear time series. *IEEE Transactions on Automatic Control*, 50(11):1655–1669, 2005.
- [16] A. Vicino and G. Zappa. Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Transactions on Automatic Control*, 41(6):774–785, 1996.