



UNIVERSIDAD DE CHILE

UNIVERSIDAD DE CHILE

EL7012

CONTROL INTELIGENTE DE SISTEMAS, OTOÑO

Ejercicio N°1

Autor:

Elias Obrequé
Gustavo Ceballo
Maibeth Sánchez

15 de mayo de 2020

Índice

1. Introducción	1
1.1. Problema 1	1
2. Generación de Datos	3
3. Modelos de predicción	5
3.1. Modelo lineal	5
3.2. Modelo difuso Takagi-Sugeno Tipo-1	12
3.2.1. Predicciones a 1, 8, y 16 pasos	17
3.2.2. Intervalos Difusos	19
3.3. Modelo de red neuronal	24
3.3.1. Comentarios sobre los modelos	34
3.4. Intervalos de predicción por covarianza	39
3.4.1. Modelo lineal	39
3.4.2. Modelo difuso Takagi-Sugeno	39
3.4.3. Modelo de red neuronal	39
4. Conclusión	40
5. Anexos	41
5.1. Anexo 1	41

1. Introducción

La mayoría de los sistemas tienen un comportamiento no lineal, excepto en un determinado rango de operación donde pueden ser considerados lineales. En ocasiones un modelo lineal es insuficiente para explicar un fenómeno por lo que se debe recurrir a Modelos No Lineales.

Los modelos basados en redes neuronales y sistemas difusos han sido empleados para la identificación, dado que han mostrado una grancapacidad para aproximar funciones no lineales desconocidas, además de que ofrecen una estructura general tan compleja o sencilla como el problema lo demande. Por un lado, las redes neuronales son estructuras matemáticas inspiradas en las estructuras biológicas compuestas por neuronas, en donde a través de elementos relativamente simples (llamados neuronas) que efectúan operaciones y procesos muy sencillos logran, mediante interconexiones, realizar procesos complejos y procesamiento de información en paralelo. En general se consideran redes neuronales conformadas por “capas” de neuronas, donde una capa procesa la información recibida en paralelo y la envía a la siguiente capa de neuronas. Por otro lado, los sistemas difusos son también estructuras matemáticas, aunque estos se encuentran inspirados en los procesos de pensamiento y toma de decisiones llevados a cabo por los humanos, así como en la teoría de conjuntos difusos propuesta por Lofti Zadeh.

En este ejercicio se dará solución al Problema 1.

1.1. Problema 1

Considere la siguiente serie no lineal dinámica:

$$\begin{aligned} y(k) = & (0,8 - 0,5\exp\{-y^2(k-1)\})y(k-1) \\ & - (0,3 + 0,9\exp\{-y^2(k-1)\})y(k-2) \\ & + u(k-1) + 0,2u(k-2) + 0,1u(k-1)u(k-2) + e(k) \end{aligned} \quad (1)$$

donde el ruido del sistema

$$e(k) = 0,5\exp\{-y^2(k-1)\}\beta(k) \quad (2)$$

depende del estado previo de la salida del modelo, y $\beta(k)$ es un ruido blanco.

Como usted sabe existen varias técnicas que se pueden emplear para la modelación a partir de estos datos, por lo que debe seleccionar el tipo de modelo más adecua-

do para este tipo de sistema. Para este trabajo se le pide detallar la metodología utilizada para:

- a) Generar 600 datos a partir de esta serie. Considere 55 % para entrenamiento, 25 % test y 20 % validación.
- b) Obtener un modelo de predicción lineal, difuso tipo-1 (T&S) y neuronal para la salida. Evaluar las predicciones a 1, 8 y 16 pasos. Comparar el desempeño de todos los modelos a partir de las métricas más apropiadas tales como RMSE, MAPE, MAE, entre otras. Comente.
- c) Construir el intervalo de predicción de los modelos obtenidos en b) utilizando el método de la covarianza.
- d) Evaluar los intervalos de predicción obtenidos en b) realizando predicciones a 1, 8, y 16 pasos. Comparar el desempeño de los modelos a partir de las métricas más apropiadas tales como ancho del intervalo, probabilidad de cobertura, entre otras.
- e) Construir el intervalo de predicción del modelo difuso encontrado en a) con el método de optimización min-max. Compare este intervalo de predicción con el intervalo obtenido utilizando el método de la covarianza. Comente.
- f) Construir el intervalo de predicción neural utilizando el método de Joint Supervision. Compare con los métodos anteriores.
- g) Seleccione el modelo más apropiado y justifique.

2. Generación de Datos

En esta etapa es necesario generar datos que representen la dinámica del sistema en la mayor cantidad de rangos de operación posibles, ya que el modelo obtenido tiene un ancho de banda acotado, y por lo tanto las dinámicas definidas por fuera de dicha banda podrían no ser representadas adecuadamente. Para lo cual se debe diseñar una entrada $u(k)$ que excite a la planta en el rango de frecuencias en que se encuentran los fenómenos de interés.

En este trabajo se propone el uso de señales binarias pseudo aleatorias (Pseudo Random Binary Signal, PRBS), ya que es una de las señales más utilizadas en identificación de sistemas. Esta es una señal periódica, determinística y que posee principalmente propiedades similares al ruido blanco (contenido muy rico en frecuencias) [1].

Para generar la señal se suponen los siguientes parámetros de interés $f_{min} = 0,2Hz$, $f_{max} = 1Hz$ y tiempo de muestreo $T_s = 0,01$. Con los parámetros anteriores, y utilizando la expresión

$$n = \frac{\log(f_c/f_{min} + 1)}{\log(2)} \quad (3)$$

con $f_c = 2,5 * f_{max} = 2,5Hz$, se genera una PRBS de orden $n = 4$, por lo que el largo máximo corresponde a $N = 2^n - 1 = 15$. A su vez, la cantidad de muestras por bit son $N_s = 40$. Luego, el tiempo de un bit, $\Delta t = N_s * T_s = 0,4s$, por lo que la PRBS dura en total 6s y debe ser replicada 400 veces con diferentes condiciones iniciales para obtener los 6000 datos de interés. Finalmente se genera la APRBS variando la amplitud aleatoriamente de la PRBS generada, Fig.1 y se aplica a la serie no lineal como se muestra en la Fig.2.

Una vez obtenidos los datos experimentales de entrada-salida, éstos son clasificados en tres conjuntos con distinta información: datos de entrenamiento, datos de validación y datos de prueba; esto con el fin de evaluar adecuadamente los modelos generados. El conjunto de entrenamiento se utiliza para determinar los parámetros del modelo. El conjunto de prueba permite comparar distintas estructuras de los modelos generados. Finalmente, el conjunto de validación permite verificar el sobreajuste del modelo óptimo obtenido, evaluándolo en un nuevo conjunto de datos (distintos a los datos del conjunto de entrenamiento y validación), analizando su capacidad de generalización. En este caso se utiliza una división de 55 % de los datos para entrenamiento, 25 % para prueba y 20 % validación.

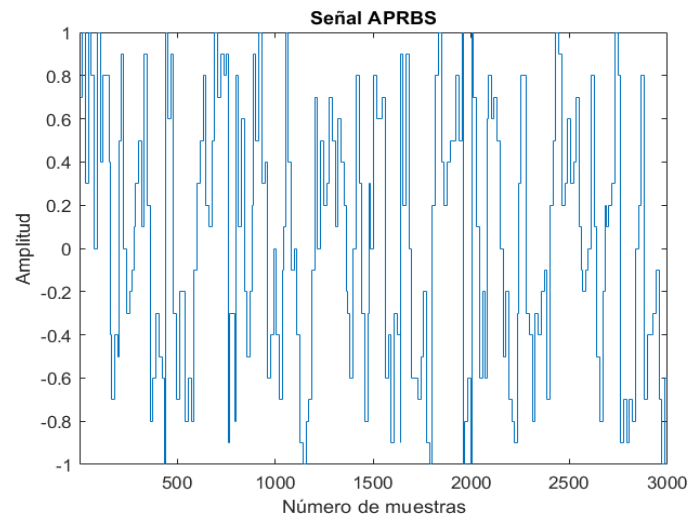


Figura 1: Señal APRBS con Amplitud entre -1 y +1.

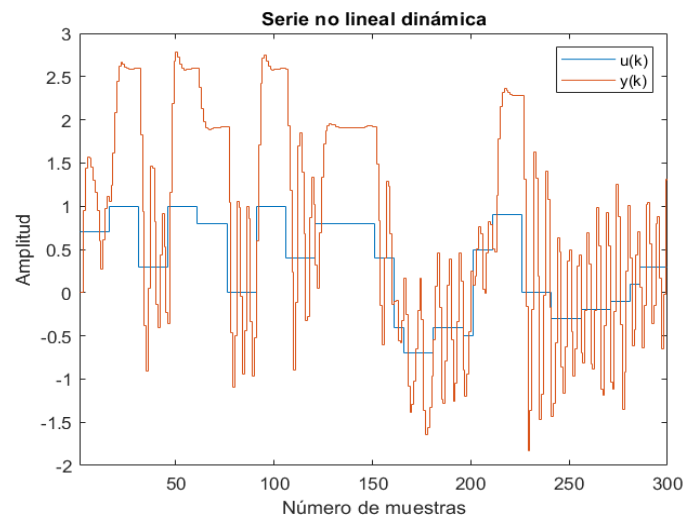


Figura 2: Respuesta de la serie no lineal.

3. Modelos de predicción

3.1. Modelo lineal

En este caso, supondremos que se ajustará un modelo lineal suponiendo que el sistema real es lineal con ruido blanco gaussiano aditivo, es decir,

$$y(k) = a_1y(k-1) + a_2y(k-2) + b_1u(k-1) + b_2u(k-2) + e(k) \quad (4)$$

Luego, se propone un modelo lineal para llevar a cabo la predicción a 1 paso, de modo tal que:

Predicción a 1 paso:

$$\hat{y}(k) = \hat{a}_1y(k-1) + \hat{a}_2y(k-2) + \hat{b}_1u(k-1) + \hat{b}_2u(k-2) \quad (5)$$

Este modelo no considera un valor constante o bias dado el supuesto que el sistema es lineal con ruido blanco aditivo. En caso que se sospechara que existe un bias o tendencia (trend) en el sistema, se puede agregar otro vector de unos a la matriz de regresores (o matriz de información).

Para llevar a cabo la estimación de los parámetros del modelo se utilizó la técnica de mínimos cuadrados, es decir:

$$\hat{\theta} = (Xent^T * Xent)^{-1} * Xent^T * \hat{y}(k) \quad (6)$$

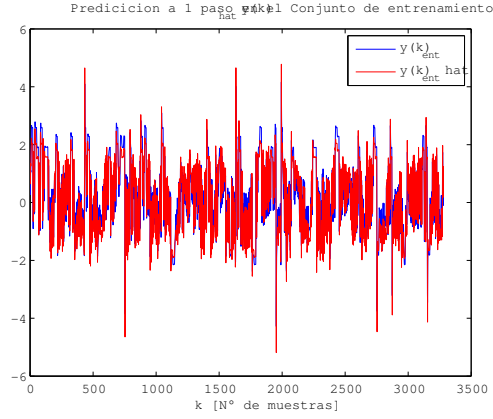
En que $\hat{\theta} = [\hat{a}_1 \quad \hat{a}_2 \quad \hat{b}_1 \quad \hat{b}_2]^T$ es el vector de parámetros y $Xent$ es la matriz de regresores con los valores de las n muestras ordenados por filas.

Los valores que se obtuvieron de los parámetros fueron los siguientes:

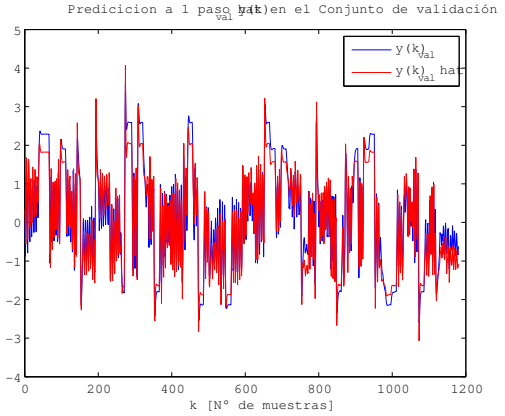
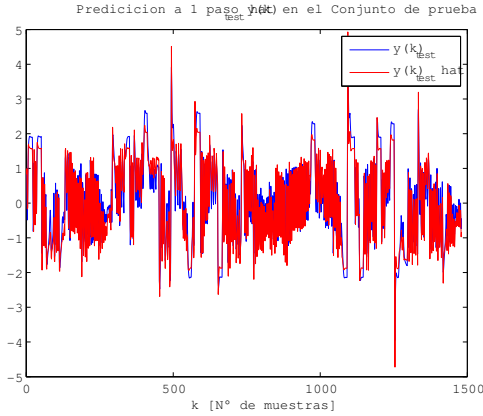
$$\hat{\theta} = \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} 0,8601 \\ -0,6930 \\ 0,9724 \\ 0,3486 \end{pmatrix} \quad (7)$$

En la Fig. 3, se muestran las gráficas de la predicción a un paso en cada conjunto de datos, a saber, conjunto de entrenamiento, prueba y validación. Esto es solo a modo de comparación puesto que finalmente es lo relevante de nuestro modelo, es el comportamiento o bondad de ajuste en el conjunto de validación. Evidentemente que los datos de los parámetros usados en las predicciones, para todos los conjuntos,

Ejercicio N°1



(a) Predicción a 1 paso en conjunto de entrenamiento



(b) Predicción a 1 paso en conjunto de prueba (c) Predicción a 1 paso en conjunto de validación

Figura 3: Predicción del modelo Lineal.

son aquellos obtenidos con el conjunto de entrenamiento, sin embargo, la matriz de reporesores y salidas cambian dependiendo del conjunto que se esté usando para predicción.

A continuación, en la Tabla 1, se presentan las métricas de bondad del ajuste o errores en los diversos conjuntos de datos, a saber, conjunto de datos de entrenamiento, prueba o test y validación.

Es interesante notar que los ajustes del modelo, en todos los conjuntos, son bastante comparables entre sí, es decir, en general, los errores son similares en cuanto a RMSE y MAE. Esto también se puede apreciar al observar la Fig. 3 respectivamente. Si bien, el conjunto de entrenamiento presenta el menor MAPE, a su vez, el conjunto

Tabla 1: Errores o Métricas de bondad de ajuste a 1 paso

Métricas	Conjunto En- trenamiento	Conjunto de Prueba	Conjunto de Valida- ción
RMSE	0.0869	0.0559	0.0954
MAPE	162.2793	220.0863	238.2150
MAE	0.3569	0.3465	0.6020

de entrenamiento, presenta el menor RMSE.

Finalmente, para efectos de apreciar la robustez de este modelo, llevaremos a cabo la predicción a 8 y 16 pasos respectivamente, en el conjunto de validación del modelo lineal, para luego compararlos con los demás modelos, a saber, modelo de Takagi y Sugeno y Modelo Neuronal para el mismo conjunto de validación.

Construcción de Intervalos para el modelo lineal

En este caso, se pueden definir 2 modelos adicionales que den cuenta de un intervalo que contenga un porcentaje de las muestras al interior del mismo. A saber, podríamos determinar un modelo por arriba (y_u) y uno por abajo (y_l) del modelo de valor esperado y_{val} , con los datos del conjunto de validación, determinado en la sección anterior.

Estos modelos se podrían construir considerando la adición de la varianza de las salidas del conjunto de validación entre el valor real y el estimado, es decir, el intervalo tendría la siguiente expresión para ambos modelos:

$$\hat{y}_u(k) = X_{val}(k)\hat{\theta}(k) + \alpha[Var(y(k)_{val} - \hat{y}_{val}(k))]^{1/2} \quad (8)$$

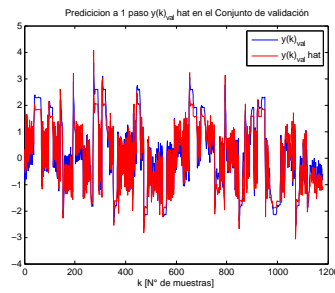
$$\hat{y}_l(k) = X_{val}(k)\hat{\theta}(k) - \alpha[Var(y(k)_{val} - \hat{y}_{val}(k))]^{1/2} \quad (9)$$

En que α es un parámetro de ajuste que tiene relación con el porcentaje de cobertura que tendría el intervalo, $Var()$ es el operador varianza de las salidas reales y estimadas en el conjunto de validación y $X_{val}(k)\hat{\theta}(k)$ es el modelo obtenido de valor esperado visto en la sección anterior, es decir,

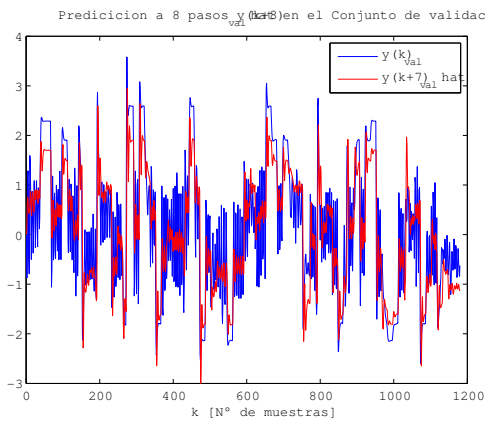
$$\hat{y}_{val}(k) = X_{val}(k)\hat{\theta}(k)$$

Así entonces, podemos determinar los modelos por arriba y por abajo del modelo clásico de valor esperado, los cuales nos definen el intervalo.

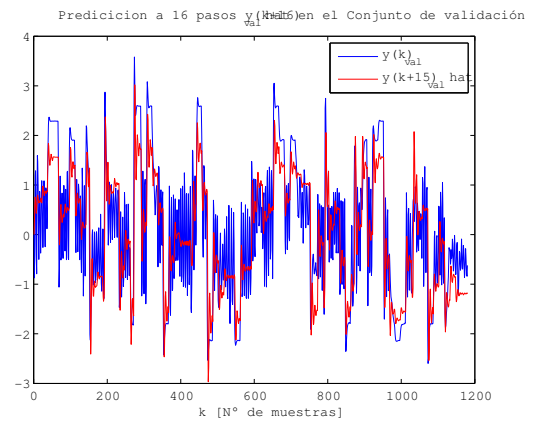
Ejercicio N°1



(a) Predicción a 1 paso en el conjunto de validación para el Modelo Lineal



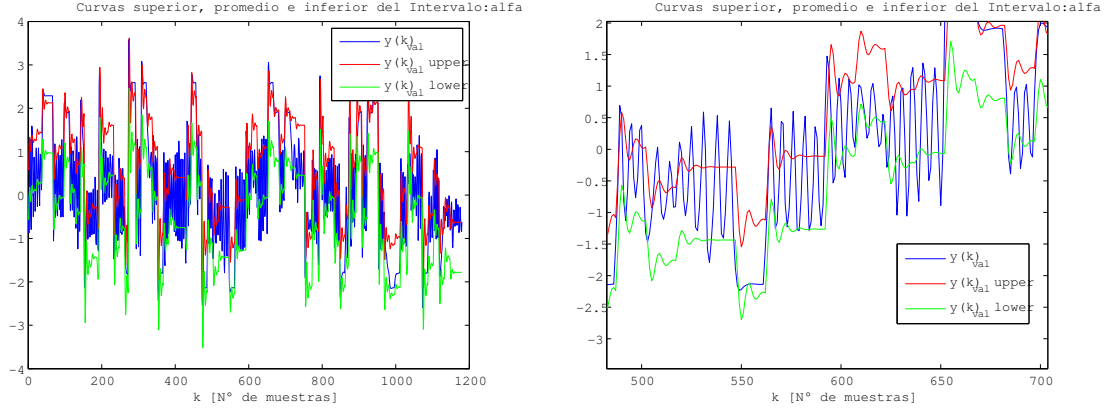
(b) Predicción a 8 pasos en el conjunto de validación para el Modelo Lineal.



(c) Predicción a 16 pasos en el conjunto de validación para el Modelo Lineal.

Figura 4: Predicción del modelo Lineal.

Ejercicio N°1



(a) Curvas para Intervalo del modelo lineal considerando predicción a 1 paso y $\alpha = 0,8$

(b) Porción del gráfico

Figura 5: Curvas para Intervalo del modelo lineal considerando predicción a 1 paso y $\alpha = 0,8$

A continuación, en la Fig. 5, se muestran las gráficas de las 3 curvas o modelos para la predicción a un paso con $\alpha = 0,8$.

Ahora bien, podemos ir ajustando el valor de α para aumentar o disminuir el porcentaje de cobertura de las muestras de validación.

En la Tabla 2, se muestran las métricas PICP y PINAW las cuales miden el porcentaje de cobertura y ancho del intervalo respectivamente, para diferentes valores de α de la predicción a 1 paso.

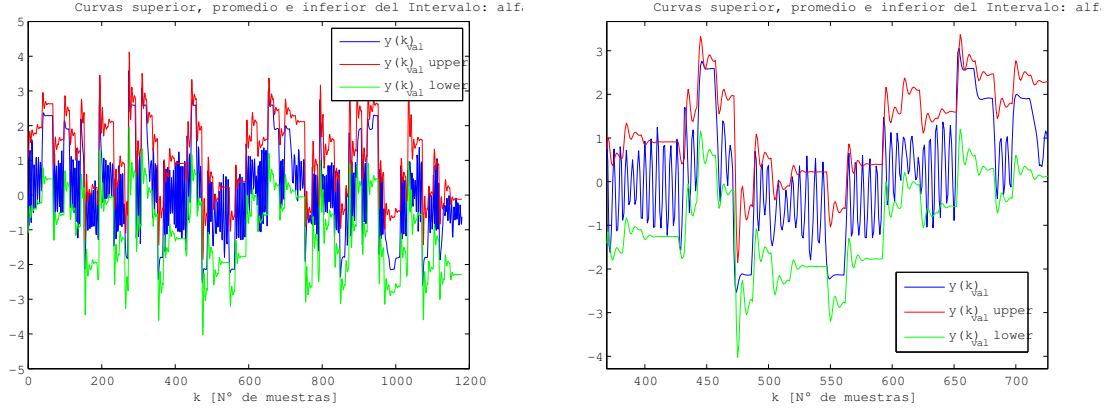
Tabla 2: Métricas de porcentaje de cobertura y ancho del intervalo para diferentes valores de α .

Parámetro α	0,5	0,8	0,9	1	1,5	2
PICP (%)	322,034	525,424	587,288	653,390	885,593	961,017
PINAW	11,687	186,992	210,366	237,339	350,609	467,479

A continuación, en el siguiente gráfico, se muestra el caso en que el parámetro α se aumenta a 1,5.

Es evidente, que los datos de salida de la curva promedio (color azul) están casi completamente contenidos en el intervalo (entre curva de color rojo y verde respec-

Ejercicio N°1



(a) Curvas para Intervalo del modelo lineal considerando predicción a 1 paso y $\alpha = 1,5$

(b) Porción del gráfico

Figura 6: Curvas para Intervalo del modelo lineal considerando predicción a 1 paso y $\alpha = 1,5$

tivamente), de hecho, está en un 88,56. Ahora bien, en el caso de la construcción del intervalo de predicción a 8 y 16 pasos, se procede de manera análoga al caso de 1 paso, con la salvedad que se debe usar el modelo de 8 pasos y 16 pasos respectivamente.

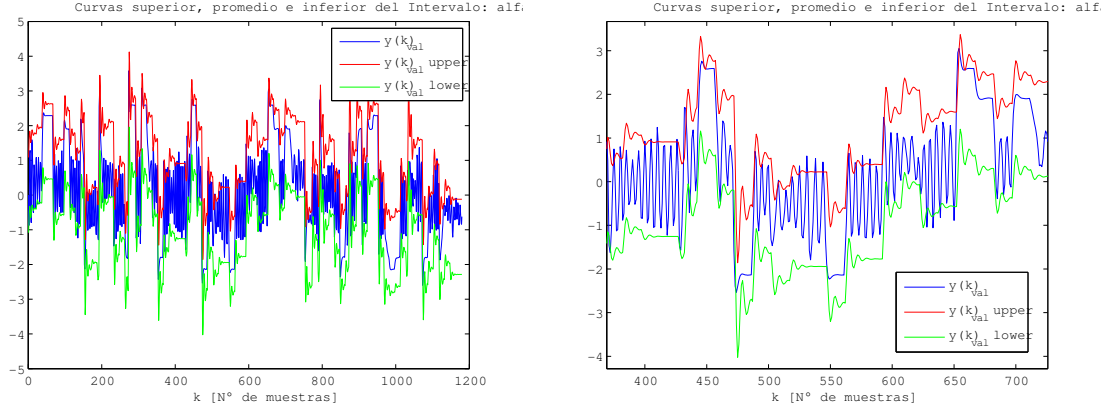
A continuación, en la Fig. ?? se muestran los intervalos para el caso de 8 y 16 pasos considerando un parámetro $\alpha = 0,8$, para efectos de comparación con el gráfico de la Fig. 5.

Tabla 3: Métricas de porcentaje de cobertura y ancho del intervalo para diferentes valores de $\alpha = 0,8$ considerando intervalos de predicción a 1, 8 y 16 pasos respectivamente.

Parámetro α	0,8		
	1 paso	8 pasos	16 pasos
PICP (%)	525,424	527,966	527,966
PINAW	186,992	184,244	187,348

Analizando las figuras anteriores y de la Tabla 3, se puede concluir que las métricas de bondad para los intervalos son bastante similares. Por una parte, entendemos que esto se debe a que como el número de muestras del conjunto es muy elevado (1.200 muestras aproximadamente), respecto a las predicciones que se piden (1, 8

Ejercicio N°1



(a) Curvas para Intervalo del modelo lineal considerando predicción a 8 pasos y $\alpha = 0,8$.

(b) Curvas para Intervalo del modelo lineal considerando predicción a 16 pasos y $\alpha = 0,8$.

Figura 7: Curvas para Intervalo del modelo lineal

y 16 pasos), es más o menos evidente que no exista diferencia en los porcentajes de cobertura. La diferencia podría notarse en caso que los intervalos de predicción fueran a más muestras, por ejemplo, a 200 muestras en adelante. Otro factor que favorece la similitud de las métricas anteriores, es el hecho que la entrada (señal APRBS) es bastante aleatoria y por tanto el ajuste no se ve más favorecido en un conjunto que en otro (conjunto de prueba o validación, por ejemplo). Es decir, los parámetros de ajuste que se obtienen son similares, no importando que conjunto que usemos para entrenamiento. Es evidente que el caso del conjunto de entrenamiento, lo favorece el hecho de que cuenta con más datos, pero dada la aleatoriedad de la señal de entrada, este factor no es tan incidente a la hora de comparar las métricas.

3.2. Modelo difuso Takagi-Sugeno Tipo-1

Los modelos difusos de Takagi-Sugeno son estructuras basadas en la lógica difusa que permiten representar procesos con dinámicas no lineales mediante la combinación de información otorgada por modelos locales. Estos tipos de modelos pueden ser expresados a partir de una base de reglas del tipo “Si-Entonces”, de la forma

$$R_r : \text{Si } z_1(k) \text{ es } MF_1^r \text{ y...y } z_p(k) \text{ es } MF_p^r \text{ entonces } y_r(z(k)) = f_r(z(k)) \quad (10)$$

donde R_r denota la r -ésima regla del modelo difuso, con $r \in 1, \dots, N_r$ y N_r el número total de reglas; $y_r(z(k))$ es su consecuencia o modelo local; $z(k) = [z_1(k), \dots, z_p(k)]$ es el vector de premisas en el tiempo k , las cuales por lo general son regresores de la entrada y/o salida del sistema; $f_r(z(k))$ es una función de las premisas del modelo; y MF_i^r es el conjunto difuso (función de pertenencia) de la i -ésima premisa correspondientes a la r -ésima regla.

Sea $\mu_r(z_i(k))$ el grado de pertenencia de la i -ésima premisa $z_i(k)$ al conjunto difuso MF_i^r , donde $\mu_r(z_i(k)) \in [0, 1]$, siendo 0 cuando la premisa no pertenece en ningún grado al conjunto MF_i^r , y siendo 1 si pertenece completamente a dicho conjunto. Luego, se define el grado de activación de la r -ésima regla, $w_r(z(k))$, como

$$w_r(z(k)) = \text{oper}(\mu_r(z_1(k)), \dots, \mu_r(z_p(k))) \quad (11)$$

donde $\text{oper}(\cdot)$ puede ser el operador mínimo o el producto. Se denota $h_r(z(k))$ al grado de activación normalizado de la r -ésima regla, es decir,

$$h_r(z(k)) = \frac{w_r(z(k))}{\sum_{l=1}^{N_r} w_l(z(k))} \quad (12)$$

Ya definido el grado de activación de cada regla, la salida del modelo difuso, $y_{fuzzy}(k)$, está dada por una suma ponderada de cada modelo local por su grado de activación normalizado, de la forma

$$y_{fuzzy}(k) = \sum_{r=1}^{N_r} h_r(z(k)) * y_r(z(k)) \quad (13)$$

Los modelos difusos TS son una clase de sistemas no lineales, cuya formulación requiere definir una serie de variables que no se conocen a priori, por lo que una estructura adecuada para representar un sistema es desconocida y, como consecuencia,

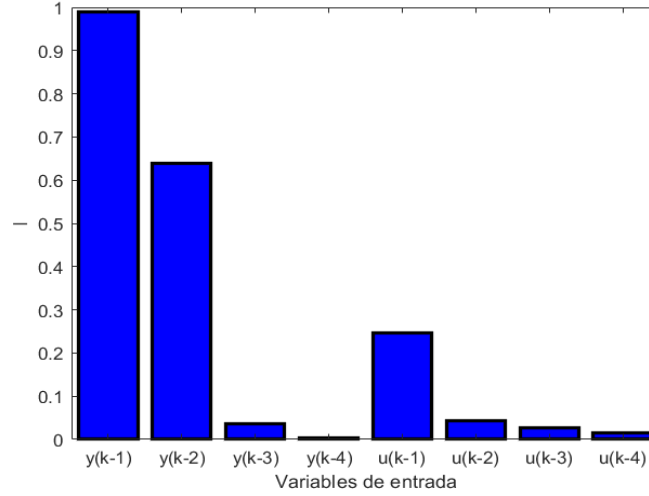


Figura 8: Índice de Sensibilidades.

un proceso de identificación debe ser llevado a cabo para determinar la estructura y cada uno de los parámetros del modelo [2].

- **Selección de variables:** Para seleccionar las variables que actúan como entrada al sistema difuso, se realiza un análisis de sensibilidad. Suponiendo una estructura del modelo inicial difuso con 8 variables de entrada $y(k-1), \dots, y(k-4), u(k-1), \dots, u(k-4)$. En la Fig. 8 se muestran los índices de las sensibilidades del modelo inicial para las 8 variables de entrada, comprobándose que las variables $y(k-4)$, $u(k-3)$ y $u(k-4)$ presentan menores índices de las sensibilidades, por lo cual no son incluidos en el modelo difuso. A pesar de que los regresores $y(k-3)$ y $u(k-2)$ presentan una sensibilidad similar se decidió escoger $u(k-1)$ ya que el mismo es un parámetro de la serie no lineal dinámica.

La Tabla 4 indica el valor de la Raíz del Error Cuadrático Medio (RMSE) para los modelos con 8 y 4 regresores, y se puede notar que son muy semejantes, y por consiguiente se selecciona el modelo más sencillo, que coincide con el número de regresores de la serie no lineal original.

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - y_{fuzzy}(k))^2} \quad (14)$$

donde N es la cantidad total de datos, $y(k)$ es la salida de la planta real en

el instante k , e $y_{fuzzy}(k)$ es la predicción realizada por el modelo difuso en el instante k .

Tabla 4: Índices de Error para el Análisis de Sensibilidades.

Modelo	Variables de entrada	RMSE
1	$y(k-1), y(k-2), y(k-3), y(k-4),$ $u(k-1), u(k-2), u(k-3), u(k-4)$	0.2177
2	$y(k-1), y(k-2), u(k-1), u(k-2)$	0.2109

- **Optimización de la estructura:** La optimización de la estructura del modelo difuso consiste principalmente en determinar el número óptimo de reglas del modelo difuso. En este caso se definió un número máximo de 20 clusters y se entrenó el modelo para cada uno de los posibles valores de clusters, utilizando como algoritmo de clustering el Fuzzy C-Means.

La Fig. 9 muestra el RMSE para los conjuntos de entrenamiento y prueba. Si no importa la complejidad, el mejor modelo es aquel que tiene menor RMSE. Sin embargo, es posible que un modelo con peor índice de desempeño, pero menos complejo que el modelo óptimo, pueda obtener resultados aceptables bajo un estándar de rendimiento definido preliminarmente. Por lo antes expuesto para este problema se escoge como número de clusters 5, por lo que el modelo difuso contará con 5 reglas, Fig 10.

Una vez realizado el clustering es posible proyectar las agrupaciones en el espacio de entrada y ajustar funciones paramétricas para describir los conjuntos difusos MF_r^i . En particular, se consideraron funciones de pertenencia gaussianas dadas por

$$MF_r^i(z_i(k)) = \exp(-0,5(a_{r,i} * (z_i(k) - b_{r,i}))^2) \quad (15)$$

donde $MF_r^i(z_i(k))$ es la función de pertenencia de la i -ésima premisa $z_i(k)$ al r -ésimo cluster, $a_{r,i}$ es el inverso de la desviación estándar de los datos ajustados por la gaussiana, $b_{r,i}$ representa su media. Para este modelo difuso se tiene:

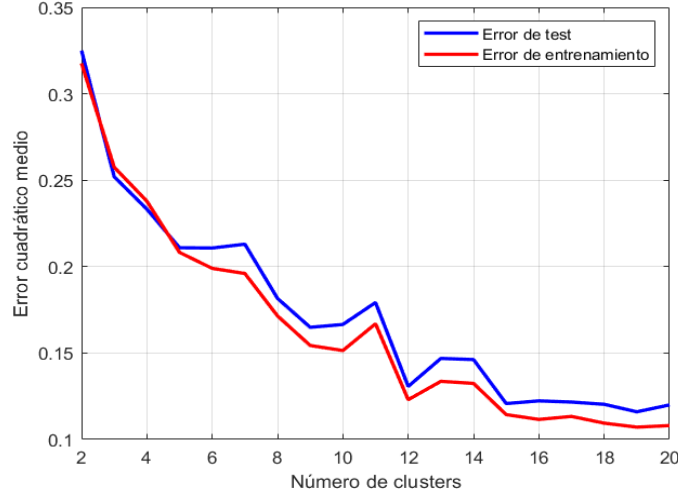


Figura 9: Índice de Sensibilidades.

$$a = \begin{pmatrix} 1,9799 & 1,4049 & 4,31509 & 4,9400 \\ 1,8300 & 1,4483 & 4,6810 & 6,1218 \\ 2,4547 & 1,8383 & 5,2820 & 8,5600 \\ 2,0375 & 1,773 & 4,5619 & 4,7213 \\ 1,6123 & 1,4681 & 4,9096 & 5,0399 \end{pmatrix} \quad (16)$$

$$b = \begin{pmatrix} 0,6742 & 0,4811 & 0,4425 & 0,4375 \\ -1,4657 & -1,4022 & -0,7580 & -0,7632 \\ 2,1665 & 2,1262 & 0,8501 & 0,86403 \\ -0,43602 & -0,5047 & -0,4782 & -0,4821 \\ -0,0353 & 0,2984 & 0,0744 & 0,07290 \end{pmatrix} \quad (17)$$

Además del RMSE, se pueden definir el Error Porcentual Absoluto Medio (MAPE, Mean Absolute Percentage Error) y Error Absoluto Medio (MAE, Mean Absolute Error) para comprobar el modelo difuso obtenido.

$$MAPE = \frac{1}{N} \sum_{k=1}^N \frac{y(k) - y_{fuzzy}(k)}{y(k)} \quad (18)$$

$$MAE = \frac{1}{N} \sum_{k=1}^N |y(k) - y_{fuzzy}(k)| \quad (19)$$

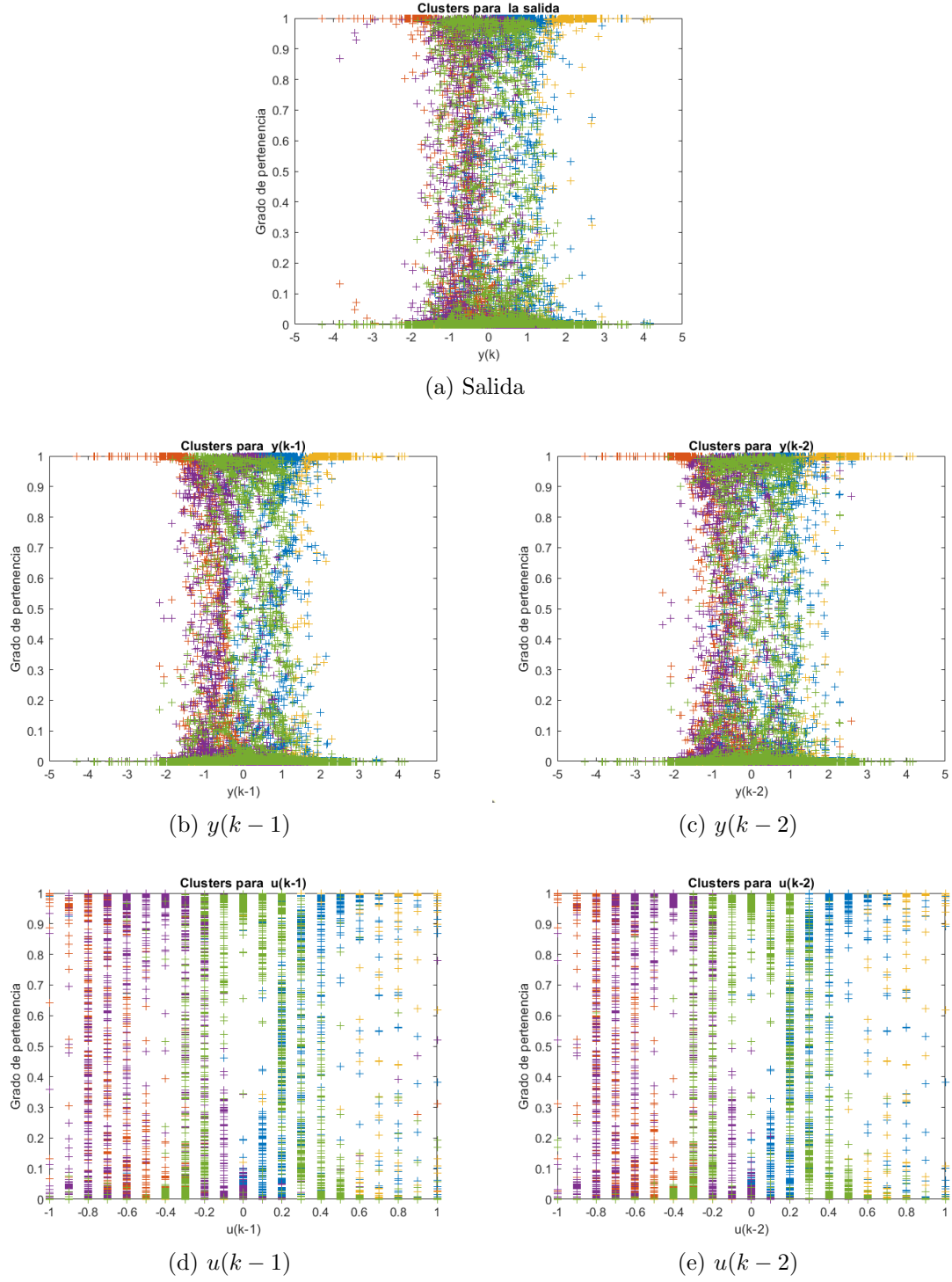


Figura 10: Clusters del Modelo Difuso Tipo 1.

Tabla 5: Índices de Error para el Modelo Difuso

Índices	Conjunto		
	Entrenamiento	Prueba	Validación
RMSE	0.2082	0.2109	0.1938
MAPE	70.92	65.65	67.71
MAE	0.1454	0.1505	0.1344

En la Tabla 5 se puede ver como el RMSE se mantiene de similar para los conjuntos de prueba y validación lo que indica la capacidad de generalización del modelo. Por su parte el índice MAPE, mide el tamaño del error (absoluto) en términos porcentuales, indicando que el error porcentual promedio del modelo se encuentra alrededor del 70 %, si bien este es un valor grande, es válido destacar que se está trabajando con una señal cuya salida se encuentra entre $[-2,3]$. El MAE tampoco presenta variaciones entre los diferentes conjuntos de datos. Considerando un compromiso entre complejidad y desempeño se determina que el modelo propuesto con 4 regresores y 5 reglas es suficiente.

- **Optimización de los parámetros:** Se utiliza el algoritmo de clustering difuso Fuzzy C-Means obteniéndose siguientes parámetros de los consecuentes $\theta^T = [\theta_{1,1}, \dots, \theta_{N_r,1}, \dots, \theta_{1,n}, \dots, \theta_{N_r,1}]$

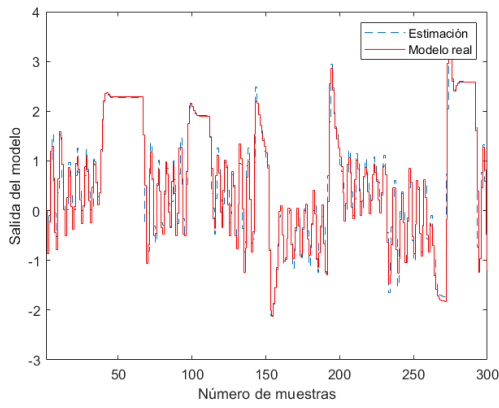
$$\theta^T = \begin{pmatrix} -0,1625 & 0,99755 & -0,7451 & 0,8430 & 0,2983 \\ 0,4561 & 1,1500 & -0,3126 & 0,5930 & 0,1362 \\ -0,2139 & 0,8892 & -0,2883 & 1,0148 & 0,2207 \\ -0,10934 & 0,8182 & -0,9810 & 0,6261 & 0,1752 \\ 0,0032 & 0,5939 & -0,9615 & 1,0893 & 0,3707 \end{pmatrix} \quad (20)$$

3.2.1. Predicciones a 1, 8, y 16 pasos

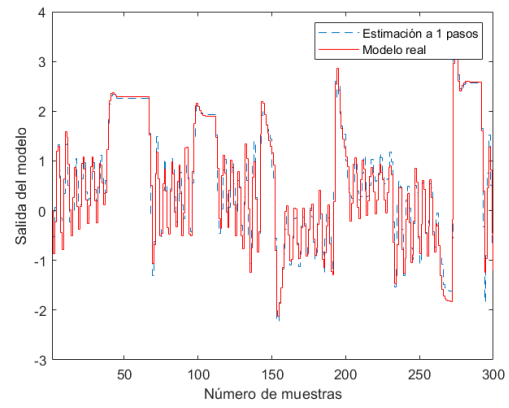
Las Fig. 11 muestra la salida del modelo difuso y la estimación para 1, 8 y 16 pasos respectivamente.

En la Tabla 6 se establece una comparación entre los modelos obtenidos, como se puede comprobar a medida que aumenta el paso de la estimación el desempeño del modelo difuso se deteriora.

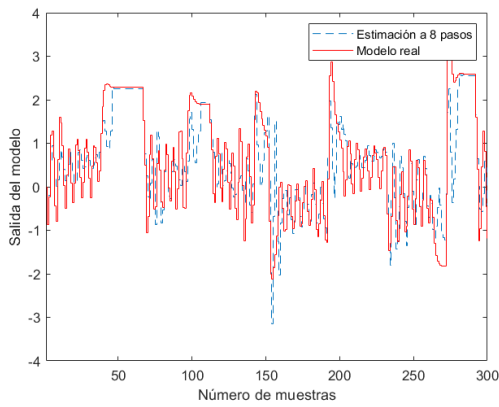
Ejercicio N°1



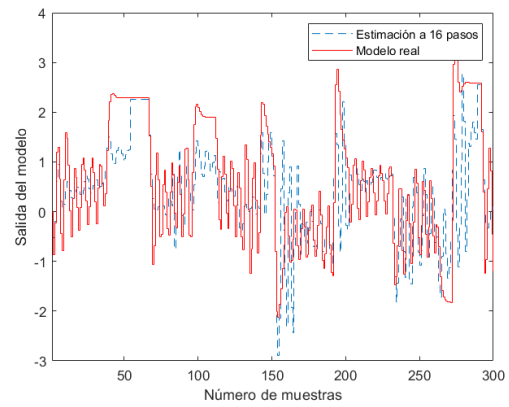
(a) Salida del Modelo



(b) Salida del modelo a 1 paso



(c) Salida del modelo a 8 pasos



(d) Salida del modelo a 16 pasos

Figura 11: Respuesta del Modelo Difuso Tipo 1.

Tabla 6: Desempeño del Modelo difuso a j-pasos

Índices	Modelos			
	Estimado	Predicción a 1 pasos	Predicción a 8 pasos	Predicción a 16 pasos
RMSE	0.1938	0.2432	0.7077	0.8847
MAPE	67.71	108.72	159.78	169.81
MAE	0.1344	0.1816	0.5076	0.6814

3.2.2. Intervalos Difusos

El modelo de intervalo difuso corresponde a un modelo difuso con parámetros superiores e inferiores que contiene un porcentaje de todos los valores medidos. Se debe encontrar una función difusa superior \bar{f} y una función difusa inferior \underline{f} , tal que se satisfaga

$$\underline{f}(z_k) \leq g(z_k) \leq \bar{f}(z_k) \quad (21)$$

en donde $z_k \in Z$ es un conjunto de entradas, $Y = y_1, \dots, y_N$ contiene los valores medidos de la salida, y se tiene que $y_k = g(z_k)$, $k = 1, \dots, N$.

Para encontrar las funciones \underline{f} y \bar{f} se utilizarán dos métodos: el método de la covarianza, y el método min-max.

El principal requerimiento al definir la banda del intervalo es que sea lo mas estrecha posible y que contenga un cierto porcentaje de datos, llamado nivel de confianza para lo cual se definen los índices Prediction Interval Coverage Probability (PICP) para medir el porcentaje de cobertura y el Prediction Interval Normalized Average Width (PINAW) para medir el ancho promedio.

$$PICP = \frac{1}{N} \sum_{k=1}^N c \quad c = \begin{cases} 1, & \hat{y}_l(k) \leq y(k) \leq \hat{y}_u(k) \\ 0, & \text{caso contrario.} \end{cases} \quad (22)$$

$$PINAW = \frac{1}{NR} \sum_{k=1}^N (\hat{y}_u(k) - \hat{y}_l(k)) \quad \text{donde} \quad R = \max(y(k)) - \min(y(k)) \quad (23)$$

■ **Método de la Covarianza:**

Este método se basa en utilizar la covarianza del error entre los datos reales y la estimación de los modelos locales del sistema difuso, de tal manera de determinar los parámetros de las funciones difusas limitantes a partir de cada consecuencia [3, 4].

$$Var(\hat{y}_j - h_j(Z^*)y_j) = \hat{\sigma}_j^2(1 + \Psi_j^{*T}(\Psi_j \Psi_j^T)^{-1} \Psi_j^*) \quad (24)$$

con la que se puede definir un intervalo difuso para cada modelo local

$$\hat{y}^j(k) = Z^{*T}(k)\hat{\theta}_j + \alpha[Var(\hat{y}_j - h_j(Z^*)y_j)]^2 \quad (25)$$

$$\hat{y}^j(k) = Z^{*T}(k)\hat{\theta}_j - \alpha[Var(\hat{y}_j - h_j(Z^*)y_j)]^2 \quad (26)$$

y obtener el modelo de intervalo total

$$\hat{y}_u(k) = \sum_{j=1}^{N_r} h_j(Z^*(k)) * \hat{y}_u^j(k) \quad (27)$$

$$\hat{y}_l(k) = \sum_{j=1}^{N_r} h_j(Z^*(k)) * \hat{y}_l^j(k) \quad (28)$$

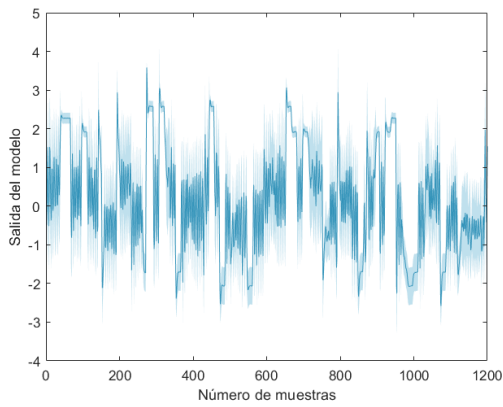
En la Fig. 12 se muestra los intervalos difusos obtenidos para el método de la Covarianza, donde se puede observar y se corrobora con la Tabla 7 que el modelo estimado presenta un porcentaje de cobertura inicial de 100 % y que comienza a deteriorarse a medida que aumenta el paso de estimación del modelo. Además esto se logra con un ancho promedio porcentual del intervalo que se mantiene constante para los diferentes modelos obtenidos.

Tabla 7: Intervalos de predicción. Método de la Covarianza

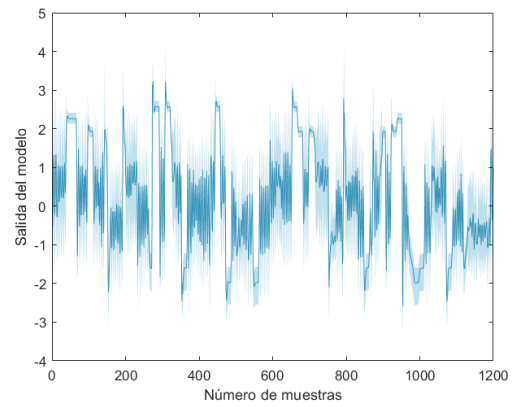
Índices	Modelos			
	Estimado	Predicción a 1 pasos	Predicción a 8 pasos	Predicción a 16 pasos
PICP	100	99.33	84.67	72
PINAW	27.59	29.2	29.55	28.1

- **MinMax:** A diferencia del anterior no se fija un porcentaje de datos para el cual se desea el mínimo intervalo de confianza, sino que se realiza una optimización para que se contenga la mayor cantidad de datos dentro del intervalo de confianza y al mismo tiempo optimizando que su ancho sea el menor. Por lo que asumiendo la misma estructura de modelo obtenida anteriormente (número de reglas, número de regresores, y parámetros de los antecedentes), se calculan los parámetros de las consecuencias de ambos modelos (θ_u, θ_l) minimizando el máximo error absoluto de modelación mediante los siguientes problemas de optimización

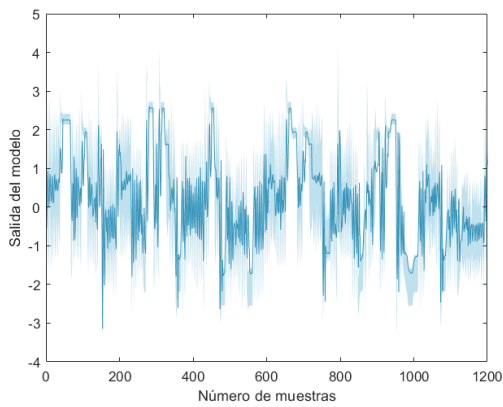
$$\min_{\theta_u} \max |y(k) - \sum_{j=1}^{N_r} h_j(Z(k)) * (\theta_u^j)^T Z(k)| \quad (29)$$



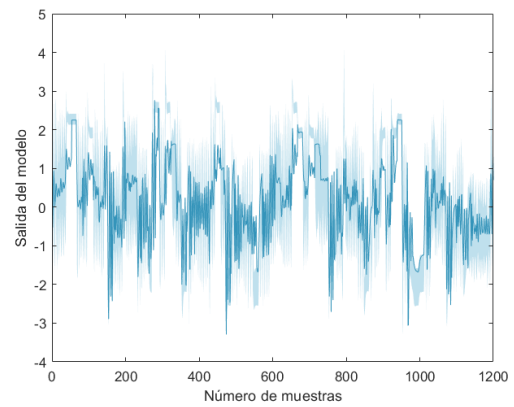
(a) Salida del modelo. Método de la Covarianza



(b) Salida del modelo a 1 paso



(c) Salida del modelo a 8 pasos



(d) Salida del modelo a 16 pasos

Figura 12: Intervalo difuso. Método de la Covarianza.

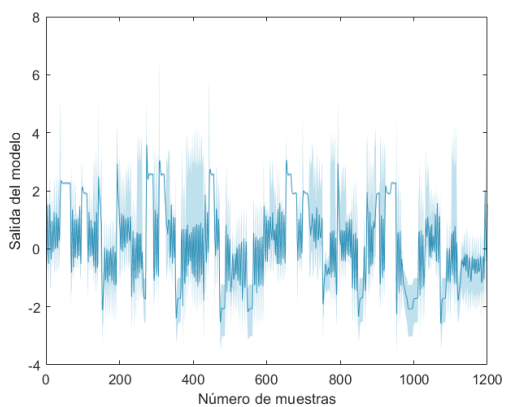
$$\begin{aligned}
s.a \quad & y(k) - \sum_{j=1}^{N_r} h_j(Z(k)) * (\theta_u^j)^T Z(k) \leq 0 \\
\min_{\theta_l} \max & |y(k) - \sum_{j=1}^{N_r} h_j(Z(k)) * (\theta_l^j)^T Z(k)| \\
s.a \quad & y(k) - \sum_{j=1}^{N_r} h_j(Z(k)) * (\theta_l^j)^T Z(k) \geq 0
\end{aligned} \tag{30}$$

En la Fig. 13 se muestra el resultado de la aplicación del Método MinMax, para los modelos difusos obtenidos, en este caso se tiene un porcentaje de cobertura inicial de 96.92%, al obtenerse a partir de un algoritmo de optimización no se puede predefinir este valor; para los modelos de predicción a j-pasos el porcentaje de cobertura disminuye. El ancho promedio porcentual del intervalo que se mantiene constante para los diferentes modelos obtenidos.

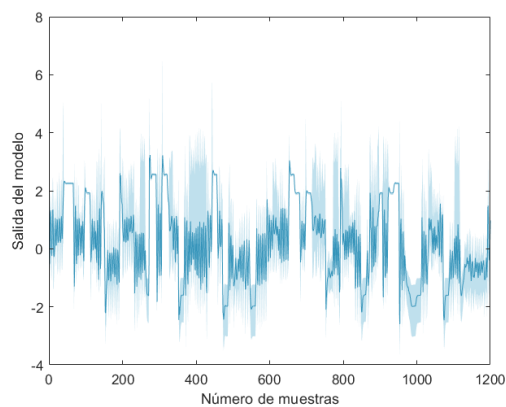
Tabla 8: Intervalos de predicción. Método MinMax

Índices	Modelos			
	Estimado	Predicción a 1 pasos	Predicción a 8 pasos	Predicción a 16 pa- sos
PICP	96.92	92.92	73.50	63.33
PINAW	30.91	32.71	31.11	31.49

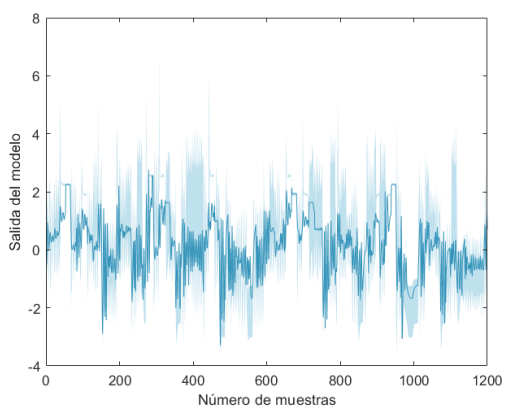
Si se comparan los métodos de intervalos difusos utilizados anteriormente, se puede determinar que ambos métodos presentan un PINAW similar, pero el método MinMax garantiza un mejor PICP para los modelos de predicción a 8 y 16 pasos, mientras que el método de la Covarianza para el modelo difuso y para la predicción a 1 paso.



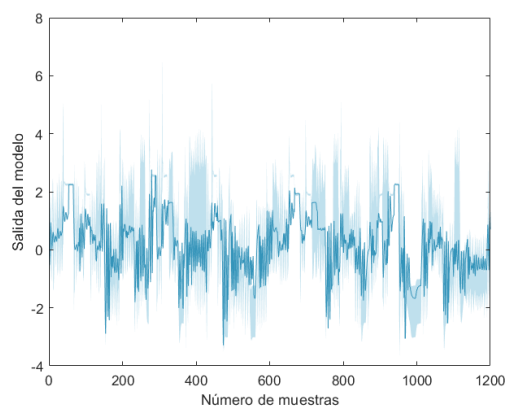
(a) Salida del modelo.



(b) Salida del modelo a 1 paso



(c) Salida del modelo a 8 pasos



(d) Salida del modelo a 16 pasos

Figura 13: Intervalo difuso. Método MinMax

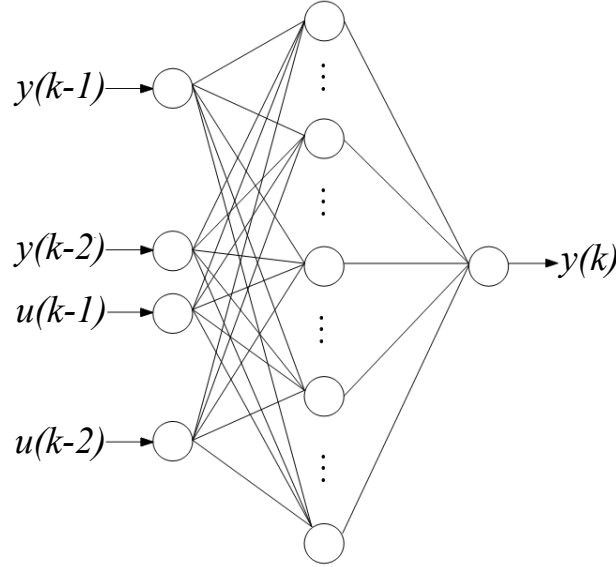


Figura 14: Estructura de la red neuronal perceptrón

3.3. Modelo de red neuronal

Para encontrar un buen modelo de red neuronal se propone seguir los pasos de identificación.

- **Obtención de datos:** Para ello se utiliza el set de datos creado en el punto a) del problema.
- **Selección de datos:** Análogo a los modelos anteriores, se utilizan los datos repartido con un 55 % en el conjunto de entrenamiento, 25 % en el conjunto de prueba y 20 % en el conjunto de validación.
- **Definición de la estructura de la red:** Se propone una red con una capa oculta, función de activación \tanh en la salida de la capa oculta y algoritmo de aprendizaje Levenberg-Marquardt. Las variables de entradas son iguales al número de regresores (4 entradas) como muestra la Figura 14. Por otro lado, el modelo matemático de la red, queda expresado como se muestra en la ecuación 31.

$$\hat{y}(k) = \sum_{i=1}^{N_h} r w_i \left(\tanh \left(\sum_{j=1}^{N_I} l w_{ji} x_j + b_i \right) \right) + c \quad (31)$$

donde, $x = [y(k-1), y(k-2), u(k-1), u(k-2)]$ es el vector de entrada, N_h es el número de neuronas en la capa oculta, N_I es el número de variables en la

entrada, rw_i es el peso que conecta la i -ésima neurona de la capa oculta con el nodo de salida y lw_{ji} corresponde al peso que una la entrada j con la i -ésima neurona en la capa oculta. Los sesgos de cada neuronas en la capa oculta y para el nodo de salida son b_i y c respectivamente.

- **Selección de entradas relevantes:** De los 4 regresores presente en el sistema se debe analizar cuál tiene mayor peso en el modelo. Un método para encontrar dichos regresores es mediante un análisis de sensibilidad evaluando la derivada de la salida de la red por cada premisa de nuestros datos, es decir,

$$\xi_j = \frac{\partial \hat{y}(k)}{\partial x_j} \quad (32)$$

Como la funcion de activacion es \tanh y en la salida es lineal, se tiene,

$$\begin{aligned} \xi_j &= \frac{\partial \hat{y}(k)}{\partial x_j} \\ &= \sum_{i=1}^{N_h} rw_i \left(1 - \tanh \left(\sum_{m=1}^{N_l} lw_{mi} x_m + b_i \right)^2 \right) lw_{ji} \end{aligned} \quad (33)$$

Como se tendrá un valor de ξ_j para cada dato vector de entrada, se genera un vector $\boldsymbol{\xi}_j$ del mismo largo que el número de datos de cada variable.

Luego, se hace uso de un *indicador* I_j para cada entrada j definido como,

$$I_j = \mu^2(\boldsymbol{\xi}_j) + \sigma^2(\boldsymbol{\xi}_j) \quad (34)$$

donde μ es la media del vector de datos y σ^2 es la varianza para cada entrada j .

- **Optimización paramétrica y estructural:** Para encontrar los valores óptimos de las parámetros peso y sesgo de la red neuronal se utiliza el algoritmo de Levenberg-Marquardt backpropagation [5]. Por otro lado, para encontrar el óptimo de la estructura se analiza cuantas neuronas debe tener la capa oculta. Para ello se evalúa el RMSE (Raíz del error cuadrático medio) del conjunto de prueba en la salida de la red para un número de neuronas entre $[2 - 41]$. Los resultados de sensibilidad para cada neurona en la capa oculta se muestran en las Figura 16 y 17 y el RMSE evaluado en los 3 conjuntos se muestra en la Figura 15. Se puede ver que el mínimo RMSE para el conjunto de prueba es para 6 neuronas y que el modelo es menos sensible a la entrada $u(k-2)$ el cual es eliminado del entrenamiento.

El entrenamiento se configura a una velocidad inicial de aprendizaje de la red de 0,05 y un valor de épocas de 5000 con evaluación de *Overfitting* de 200 épocas de validación. Es importante señalar que durante el experimento se probaron diferentes épocas de validación para cuantificar el efecto del sobre ajuste en el número óptimo de neuronas con la hipótesis de que, independiente del RMSE obtenido, el mínimo global se da con la misma cantidad de neuronas óptimas afectando solo a los mínimos locales que pudieran aparecer. En la Tabla 9 se muestran los resultados de varios experimentos demostrando la hipótesis, es más, el valor RMSE obtenido con todo los modelos es casi el mismo y solo cambia el número de neuronas, donde, 6 demuestra ser un mínimo global y los demás un mínimo local. Por otro lado, se observo que tras varios experimentos el mejor valor RMSE del conjunto de prueba se observa cuando las épocas de validación por sobre ajuste se encuentran entre $[50 - 300]$, de esta manera, de aquí en adelante se utilizan 50 épocas de validación por sobre ajuste para disminuir los costos de calculo.

Tabla 9: Experimento para ver la influencia del sobre ajuste.

Número de épocas de validación	N_h óptimo	Valor RMSE
1	11	0.0025392
10	7	0.0025384
50	6	0.0025342
100	6	0.0025295
200	6	0.0025324
500	6	0.0025326
1000	10	0.0025299
5000	17	0.0025383

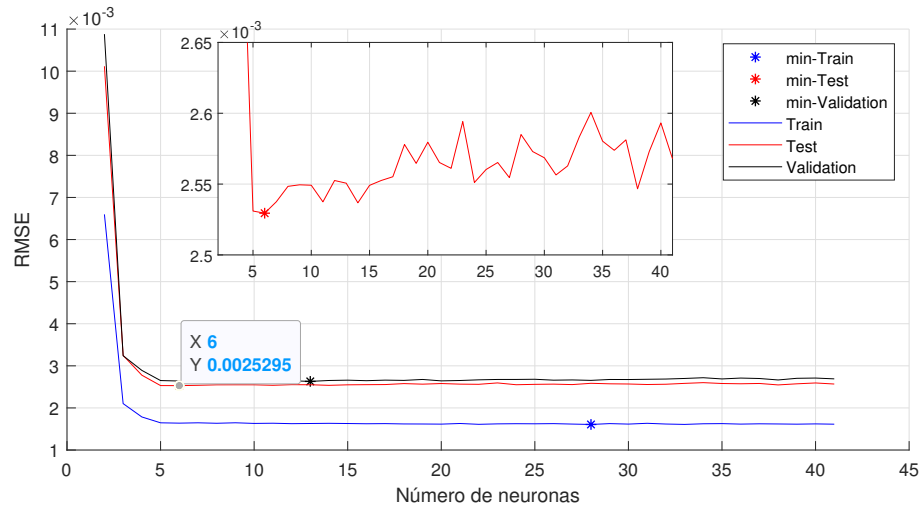


Figura 15: RMSE para diferente número de neuronas.

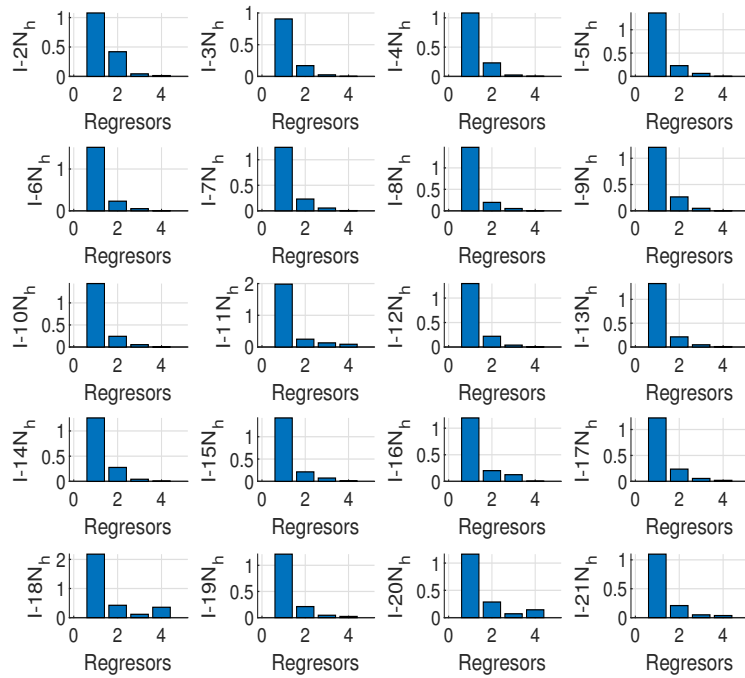


Figura 16: Sensibilidad para un número de neuronas entre $[2 - 21]$ en la capa oculta.

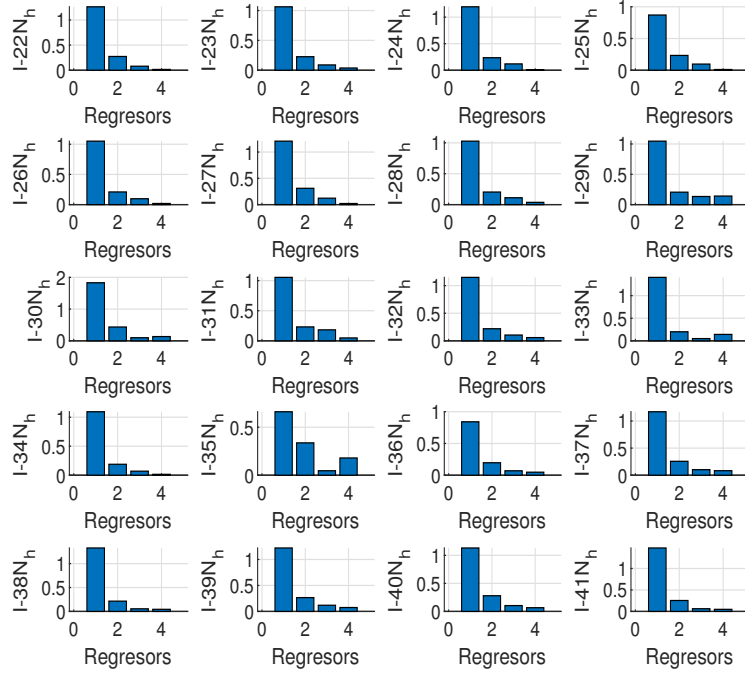


Figura 17: Sensibilidad para un número de neuronas entre $[22 - 41]$ en la capa oculta.

- **Desempeño de la red definida:** Se procede a evaluar el desempeño de la red con 8 neuronas en la capa oculta utilizando las cuatro entradas y luego solo con 3, $[y(k-1), y(k-2), u(k-1)]$, para notar las diferencias. Los resultados se muestran en las Figuras 18 hasta la 20 y los datos numéricos se agrupan en la Tabla 10

Tabla 10: Valores MSE para los 3 conjuntos de datos evaluados en una red con 4 entradas y otra con 3 entradas.

Número de entradas	MSE - Entrenamiento	MSE - Prueba	MSE - Validación
4	0.0089	0.0097	0.0084
3	0.0108	0.0118	0.0108

Se puede notar de la Figura 18 que al quitar la entrada $u(k-2)$ aumenta el indicador I_1 de la entrada $y(k-1)$ para compensar la falta del autoregresor $u(k-2)$ manteniendo casi al mismo valor las otras dos entradas. Por otro lado, no se eliminan más entradas dado que el ancho del histograma de la Figura 20(b) comienza a aumentar de valor mostrando que más datos tienen

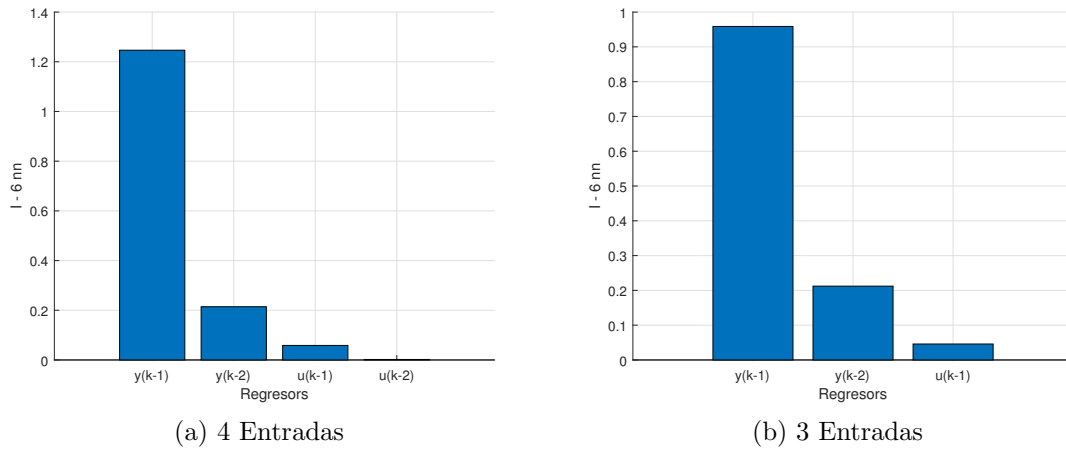
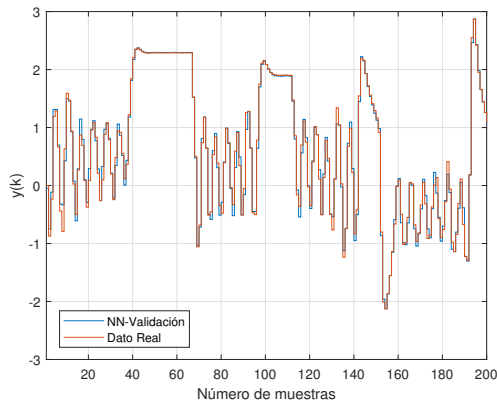


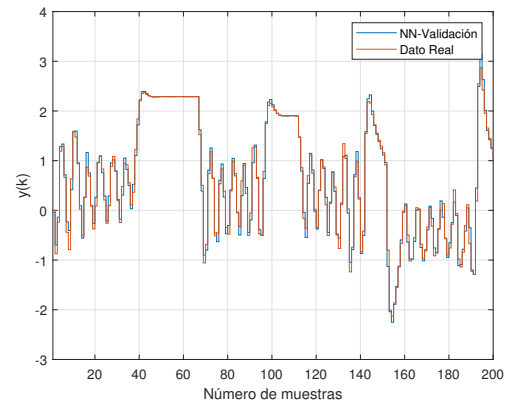
Figura 18: Indice de sensibilidad para 4 y 3 entradas.

errores grandes. Este error puede tener solución al utilizar un nuevo número de neuronas en la capa oculta, por lo tanto, en el siguiente ítem se procede nuevamente a encontrar el número óptimo de neuronas con 3 entradas.

Ejercicio N°1

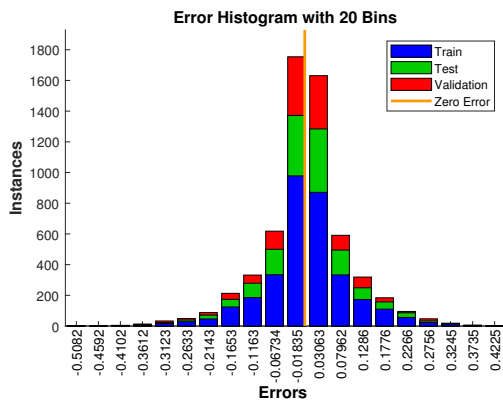


(a) 4 Entradas

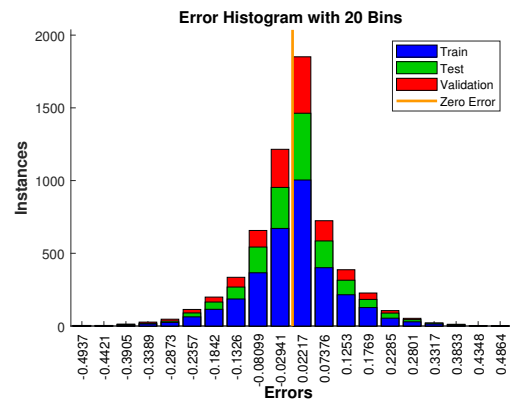


(b) 3 Entradas

Figura 19: Comparación entre la salida de la red neuronal y el valor real del conjunto de validación.



(a) 4 Entradas



(b) 3 Entradas

Figura 20: Histograma del error de cada conjunto de datos.

- **Optimizar estructura 2:** Se busca el número de neuronas óptimo en un rango de $[2 - 41]$. El resultado de la curva RMSE se muestra en la Figura 21 y el análisis de sensibilidad para cada neurona en las Figuras 22 y 23.

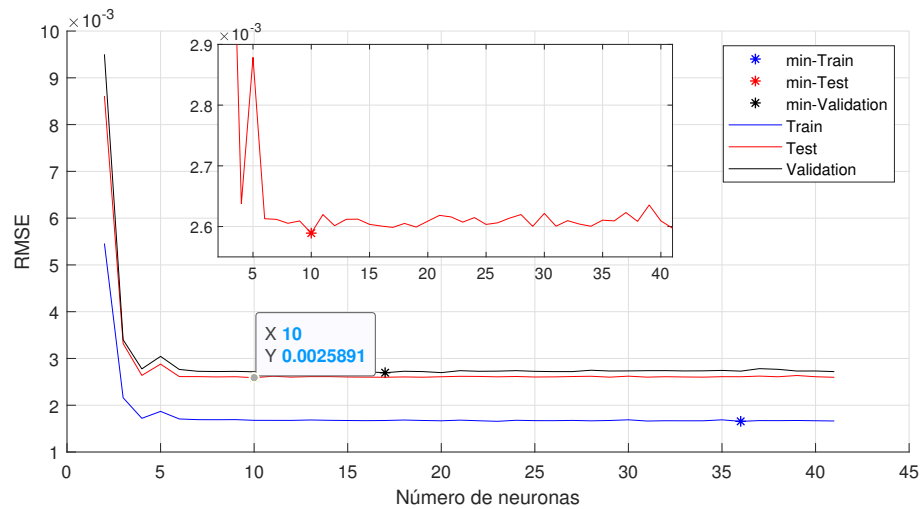


Figura 21: RMSE para diferente número de neuronas.

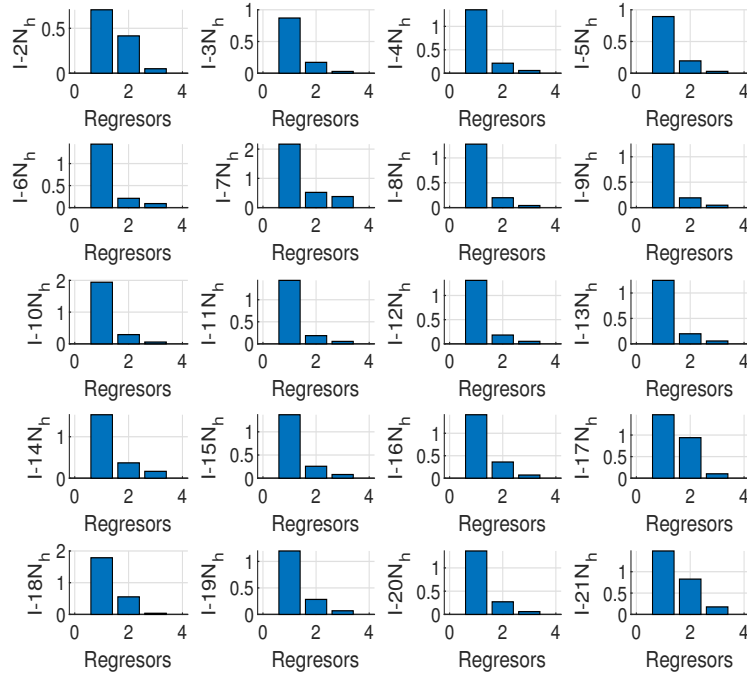


Figura 22: Sensibilidad para un número de neuronas entre $[2 - 21]$ en la capa oculta con 3 entradas.

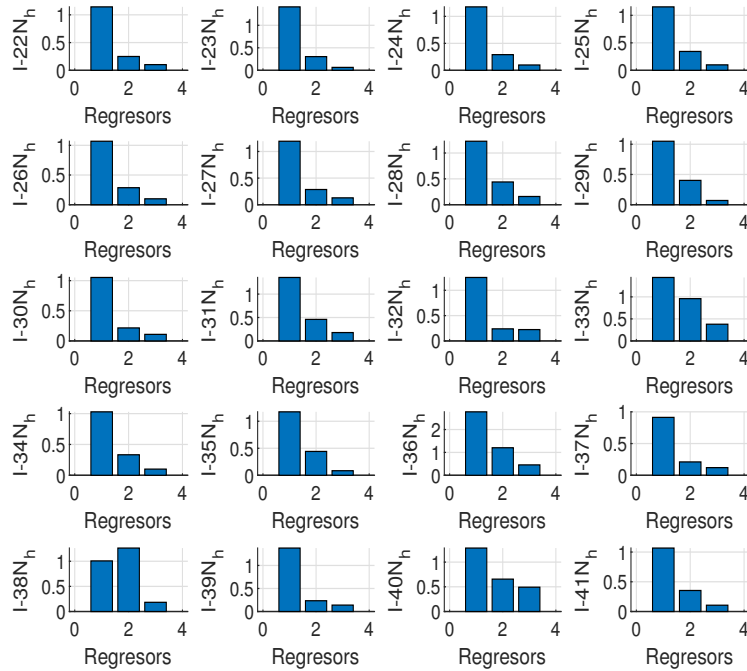


Figura 23: Sensibilidad para un número de neuronas entre $[22 - 41]$ en la capa oculta con 3 entradas.

Validación final de la estructura: Para finalizar con la identificación del modelo de predicción por red neuronal, se evalúan los parámetros de la red, MAE, MAPE, RMSE, MSE reagrupados en la Tabla ??.

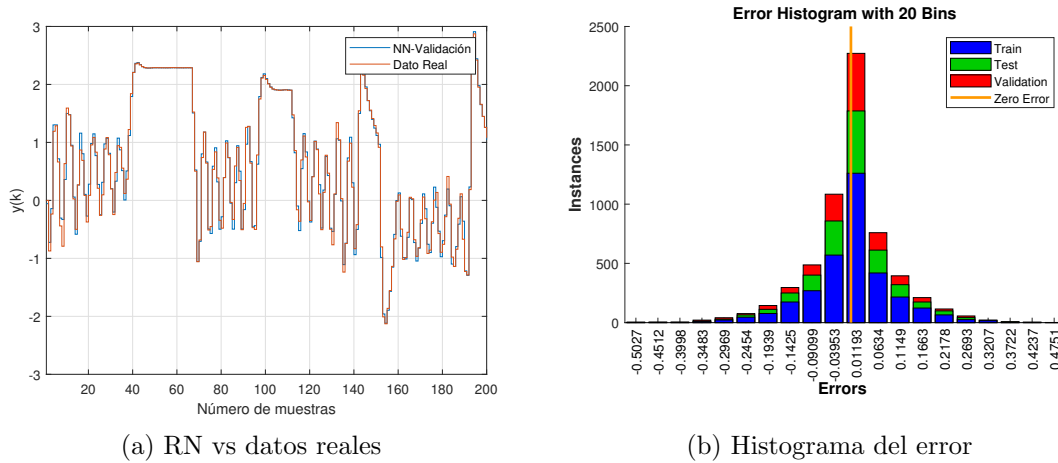


Figura 24: Resultados del modelo neuronal con 10 neuronas en la capa oculta en comparación con los datos reales.

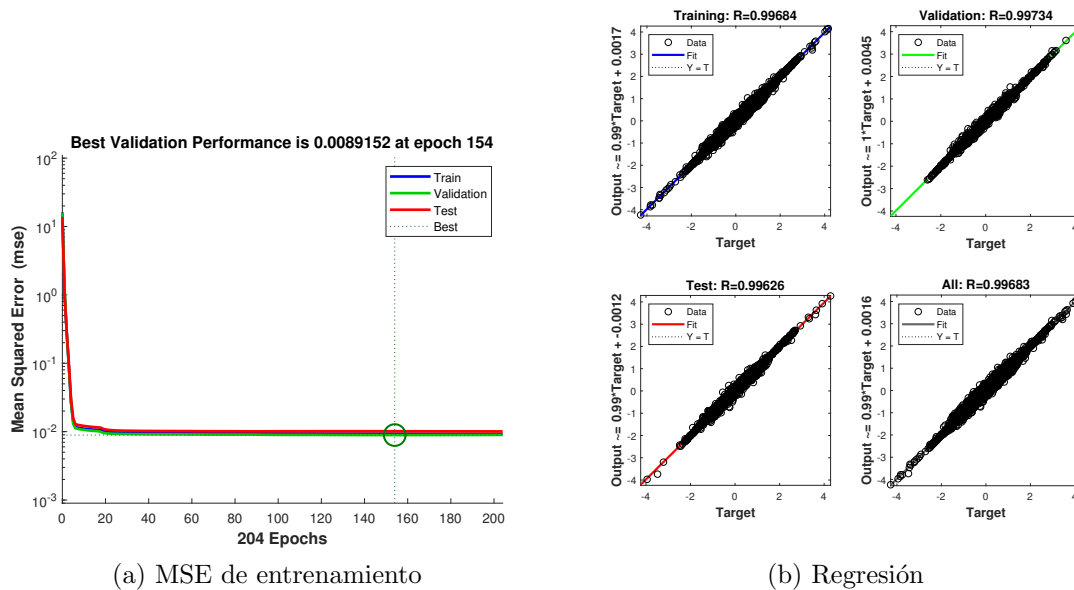


Figura 25: Resultados de entrenamiento mostrando la regresión de cada conjunto.

Tabla 11: Métricas finales del modelo neuronal

Métricas	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validación
RMSE	0.001681	0.002599	0.002726
MSE	0.009324	0.01013	0.008915
MAE	0.06538	0.06852	0.06117
MAPE	45.25	60.20	35.55

Modelo predictivo: Se comienza de la premisa que se conocen todos los valores hasta $\hat{y}(k)$.

- **A 1 paso:** Para realizar un modelo predictivo a 1 paso (un paso más y aparte del generado en sí por el modelo neuronal para pasar de $(k-1)$ a (k)), se tiene,

$$\hat{y}(k+1) = \sum_{i=1}^{N_h} r w_i \left(\tanh \left(\sum_{m=1}^{N_I} l w_{mi} x_m(k+1) + b_i \right) \right) + c \quad (35)$$

recordando la definición del vector x , se tiene, que $x(k+1) = [\hat{y}(k), y(k-1), u(k)]$. Existe una forma de determinar la señal de control futura $u(k)$ mediante modelo de control predictivo (MPC) [6], sin embargo, dado que se tiene un número grande de datos y que la señal APBRS se mantiene constante por lo menos en 40 puntos por bits antes de cambiar, se dará por conocido el control $u(k)$ requerido hasta el total de datos menos j . Los resultados para un paso se muestran en la Figura 26 para todos los conjuntos de datos.

- **A 8 paso:** Para este escenario, junto al de 16 pasos, se requiere conocer entradas no conocidas. Estas entradas deben ser estimadas de los datos conocidos como muestra la Figura 27, donde hasta $j = 1$ se conocen los datos y para $j \geq 2$ se deben usar las estimaciones anteriores. Los resultados para 8 pasos se muestran en la Figura 28.
- **A 16 paso:** Los resultados se muestran en la Figura 29. Reagrupando las métricas, se tiene el resultado de la Tabla 12.

3.3.1. Comentarios sobre los modelos

Ejercicio N°1

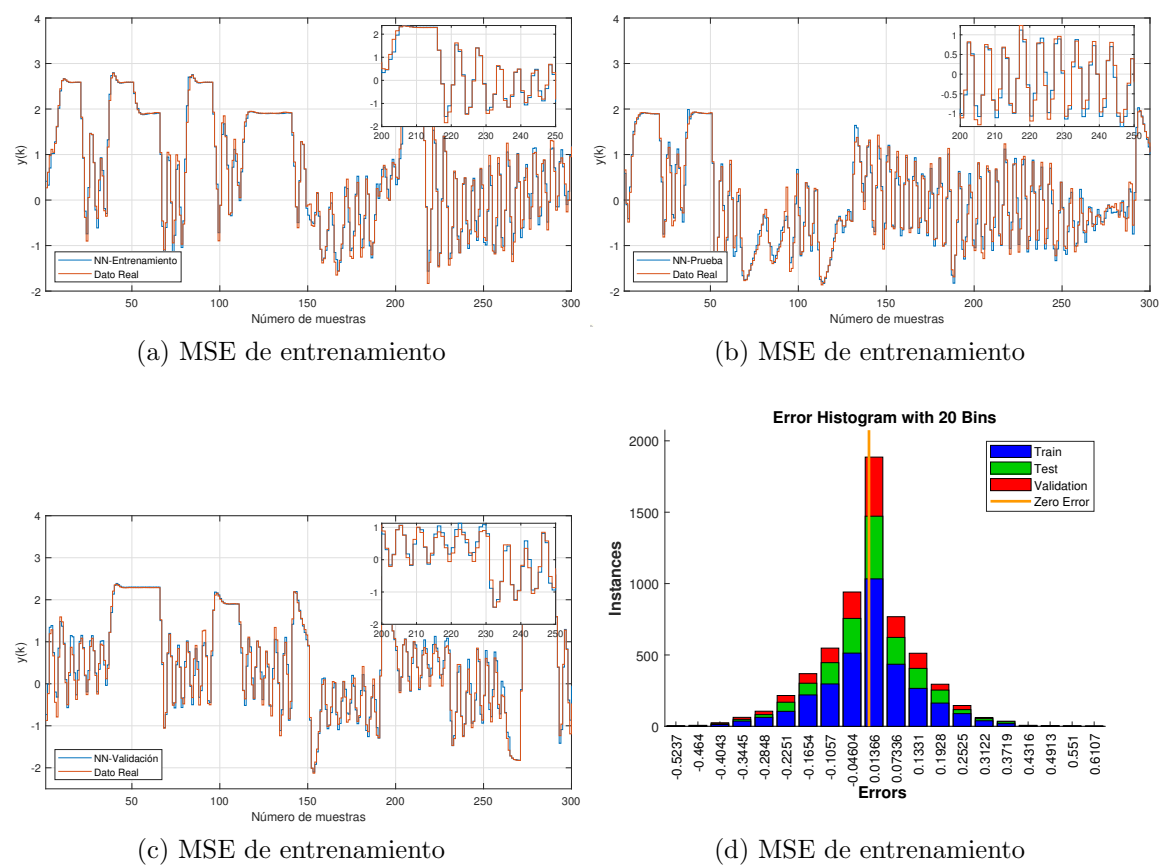


Figura 26: Predicción a 1 paso con modelo neuronal.

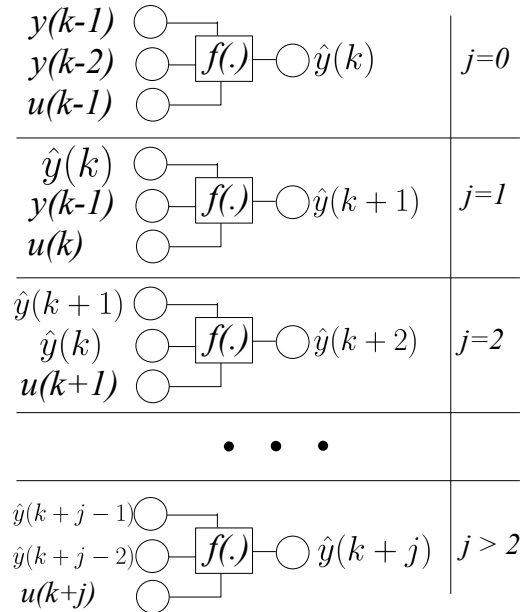
Figura 27: Modelo de predicción para j pasos del modelo neuronal.

Tabla 12: Métricas de evaluación para 1, 8 y 16 pasos adelante.

Predicción a 1 paso			
Métricas	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validación
RMSE	0.0022	0.0034	0.0036
MSE	0.0162	0.0172	0.0157
MAE	0.0917	0.0948	0.0877
MAPE	61.1581	75.2533	37.2537
Predicción a 8 paso			
Métricas	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validación
RMSE	0.0050	0.0082	0.0082
MSE	0.0811	0.1012	0.0809
MAE	0.2006	0.2243	0.1900
MAPE	150.8271	172.2750	89.0900
Predicción a 16 paso			
Métricas	Conjunto de entrenamiento	Conjunto de prueba	Conjunto de validación
RMSE	0.0072	0.0122	0.0116
MSE	0.1695	0.2191	0.1582
MAE	0.2792	0.3272	0.2621
MAPE	209.8643	263.4764	125.4239

Ejercicio N°1

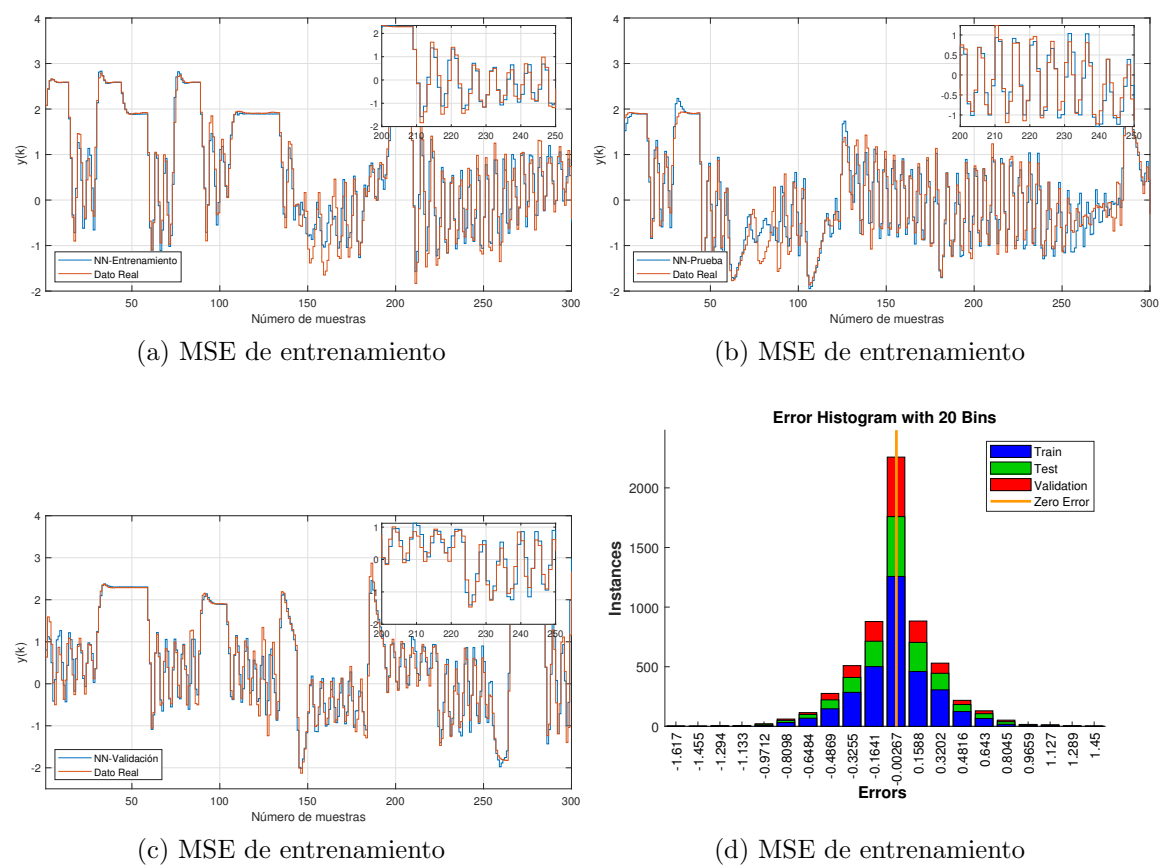


Figura 28: Predicción a 8 paso con modelo neuronal.

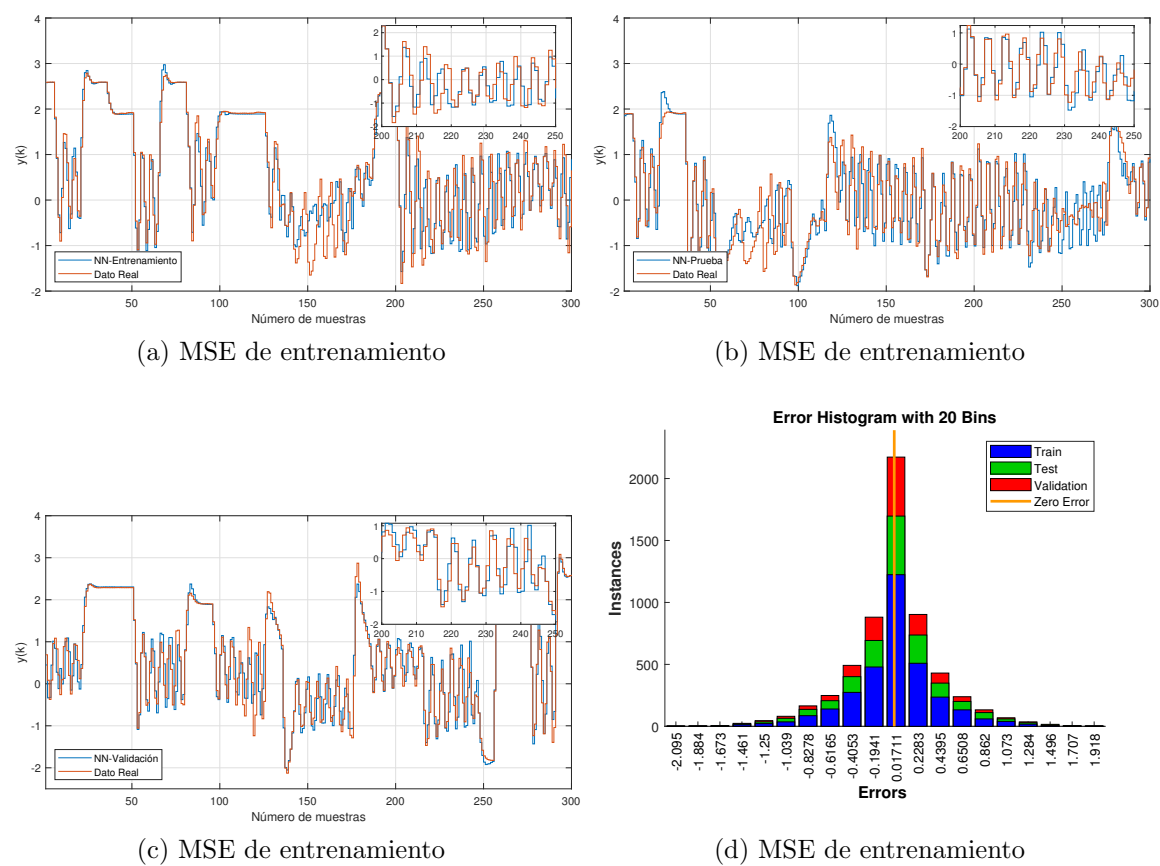


Figura 29: Predicción a 16 paso con modelo neuronal.

3.4. Intervalos de predicción por covarianza

3.4.1. Modelo lineal

3.4.2. Modelo difuso Takagi-Sugeno

3.4.3. Modelo de red neuronal

Del modelo matemático de la ecuación 31, es posible reescribir la ecuación separando las capas, es decir,

$$\hat{y}(k) = \sum_{i=1}^{N_h} rw_i \tilde{Z}_i(k) + c \quad (36)$$

donde $\tilde{Z}_i(k)$ es la de la i -ésima neurona de la capa oculta, tal que,

$$\begin{aligned} \tilde{Z}_i(k) &= \tanh \left(\sum_{j=1}^{N_I} lw_{ji} x_j + b_i \right) \\ \tilde{Z} &= L W X + B \end{aligned} \quad (37)$$

Donde \tilde{Z} es la matriz de todos los datos a la salida de la capa oculta, X es la matriz que contiene todos los datos de la entrada, RW es la matriz de los pesos internos entre la entrada y la capa oculta y B es el vector de sesgo de cada neurona en la capa oculta. Luego, el método de la covarianza nos da los límites superiores $\hat{y}_u(k)$ y límite inferior $\hat{y}_l(k)$ como sigue,

$$\begin{aligned} \hat{y}_u(k) &= [RW \tilde{Z}^{*T}(k) + c] + \alpha \cdot \sigma_e^2 \left(1 + \tilde{Z}^{*T}(k) \left(\tilde{Z} \tilde{Z}^T \right)^{-1} \tilde{Z}^{*T}(k) \right)^{1/2} \\ \hat{y}_l(k) &= [RW \tilde{Z}^{*T}(k) + c] - \alpha \cdot \sigma_e^2 \left(1 + \tilde{Z}^{*T}(k) \left(\tilde{Z} \tilde{Z}^T \right)^{-1} \tilde{Z}^{*T}(k) \right)^{1/2} \end{aligned} \quad (38)$$

Donde,

- $\tilde{Z}^{*T}(k)$: Es el vector/matriz del conjunto de prueba
- \tilde{Z} : Vector/matriz de los datos de entrenamiento.
- RW : Matriz de pesos que conectan la capa oculta con la capa de salida de la red.
- α : Hiperparámetro para ajustar el ancho del intervalo.
- σ_e^2 : Varianza del error en el conjunto de entrenamiento $Var(y(k) - \hat{y}(k))$.

4. Conclusión

5. Anexos

5.1. Anexo 1

```
1  
clear all  
clc  
4%-----Generar BPRS-----  
Nd=6010;
```

Referencias

- [1] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Springer Science & Business Media, 2013.
- [2] G. Alvarez, “Metodología de Identificación Difusa Basada en el Estudio de Controlabilidad de Sistemas Dinámicos,” p. 172.
- [3] I. Škrjanc, “Confidence interval of fuzzy models: An example using a waste-water treatment plant,” *Chemometrics and Intelligent Laboratory Systems*, vol. 96, no. 2, pp. 182–187, Apr. 2009. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0169743909000124>
- [4] —, “Fuzzy confidence interval for pH titration curve,” *Applied Mathematical Modelling*, vol. 35, no. 8, pp. 4083–4090, Aug. 2011. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0307904X11000904>
- [5] Matlab, “trainlm, levenberg-marquardt backpropagation.” [Online]. Available: <https://www.mathworks.com/help/deeplearning/ref/trainlm.html>
- [6] M. Nørgaard, O. Ravn, N. K. Poulsen, and L. K. Hansen, *Neural Networks for Modelling and Control of Dynamic Systems*, ser. Advanced Textbooks in Control and Signal Processing, M. J. Grimble and M. A. Johnson, Eds. London: Springer London, 2000. [Online]. Available: <http://link.springer.com/10.1007/978-1-4471-0453-7>