

Recomendaciones para Seleccionar Índices para la Validación de Modelos

Recommendations to Select Indices for Model Validation

Juan R. Camarillo-Peñaranda¹
Andrés J. Saavedra-Montes²
Carlos A. Ramos-Paja³

1 Departamento de Energía Eléctrica y Automática,
Facultad de Minas, Universidad Nacional de Colombia,
Medellín-Colombia
jrcamarillo@unal.edu.co

2 Departamento de Energía Eléctrica y Automática,
Facultad de Minas, Universidad Nacional de Colombia,
Medellín-Colombia
ajsaaved@unal.edu.co

3 Departamento de Energía Eléctrica y Automática,
Facultad de Minas, Universidad Nacional de Colombia,
Medellín-Colombia
caramosp@unal.edu.co

Resumen

En este artículo se presentan recomendaciones para seleccionar índices de error para validar modelos. Las recomendaciones se basan en la comparación de los índices usados para validar sistemas dinámicos. Se presentan y definen matemáticamente nueve índices de error y un índice de ajuste. Con base en el tipo de datos reportados en la literatura para validar modelos, se seleccionan la función escalón y la función seno, como señales patrón, para evaluar los resultados de los índices. Como resultado relevante de este artículo se presentan recomendaciones para seleccionar e interpretar índices de error cuando se validan modelos de sistemas físicos.

Palabras clave

Índices de error; índice de ajuste; validación de modelos; sistemas físicos; criterios de selección.

Abstract

A set of recommendations to select error indices to model validation is presented in this paper. The recommendations are based on the comparison of indices used to validate dynamic systems. Nine error indices and one fit index are presented and mathematically defined. Based on the data reported in literature to validate models, step function and sine function are selected, as patron signals, to evaluate the index results. As a relevant contribution of this paper, recommendations to select and interpret error indices during the validation of physical system models are given.

Keywords

Error indices; fit index; model validation; physical system; selection criteria.

1. INTRODUCCIÓN

La validación de modelos es la etapa más importante tanto en el modelado como en la estimación de parámetros de modelos de sistemas físicos. Existen diferentes formas de validar modelos, una de ellas es comparar datos experimentales y estimaciones de un modelo frente a una misma entrada. Para este procedimiento se necesitan uno o varios índices de error que permitan observar que tan bien el modelo reproduce el comportamiento del sistema real. En la literatura hay reportados varios índices para validar modelos. Cada uno tiene características diferentes y su resultado se interpreta de forma diferente, por lo que es responsabilidad de la persona encargada de la validación escoger el índice apropiado.

Hyndman y Koehler (2006) presentan varios índices de error que se usan para medir la calidad de los pronósticos de una serie de tiempo, se clasifican y se hace una comparación teórica entre los distintos grupos de índices. También proponen que el índice *mean absolute scaled error* sea el usado para medir la calidad de los pronósticos de una serie de tiempo. Por otro lado Bergmeir y Benítez (2012) también hacen una presentación, clasificación e interpretación de varios índices de error, sin embargo el objetivo principal del artículo no es el estudio de los índices de error, si no el estudio de la validación cruzada para la evaluación de estimadores de series de tiempo. Aunque ambos artículos presentan una gran cantidad de índices y comentan algunas de sus características, ninguno presenta criterios de selección de los índices.

A la hora de validar modelos el investigador se enfrenta a la tarea de escoger dentro de la gran cantidad de indicadores que hay, cual o cuales son los más convenientes para realizar su validación. En muchos casos esta selección se hace sin tener en cuenta las características de los datos que se utilizan en la validación. En el presente artículo se presentan índices usados para validar modelos, su interpretación y algunas recomendaciones para su selección, enfocados a casos de ingeniería.

En la Sección 2 se presentan los índices más usados en la literatura consultada. Luego, en la Sección 3, se presentan tres conjuntos de comparación de señales para evidenciar las característi-

cas de los índices. Finalmente se presentan las conclusiones, agradecimientos y referencias.

2. ÍNDICES DE ERROR E ÍNDICE DE AJUSTE PARA VALIDAR MODELOS

A continuación se muestra una clasificación de los índices según su dependencia con la escala de las señales comparadas. Para cada índice se presenta su definición matemática, su interpretación y sus unidades.

2.1 Errores que Dependen de la Escala

Estos indicadores son de utilidad cuando se validan diferentes modelos con el mismo conjunto de datos y se quiere seleccionar cual método es el mejor. No se deben usar cuando se comparan datos con diferentes escalas, pues ponderan el error de la escala más grande.

2.1.1 Error Medio Absoluto

El error medio absoluto o MAE por sus siglas en inglés, mide la magnitud promedio del error entre los datos medidos y los datos estimados por el modelo (1). El MAE es presentado por Hyndman y Koehler (2006) y es usado por Karamirad et al. (2013) y Gaiser et al. (2010). Su valor mínimo es cero y ocurre cuando los datos medidos y las estimaciones son iguales en todo el rango de muestras. Conserva las unidades de los datos medidos.

$$MAE = (1/n) \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

Dónde y_i es el valor medido, \hat{y}_i el valor estimado y n el número de muestras.

2.1.2 Error Medio

El error medio o ME por sus siglas en inglés, mide si el modelo sobreestimó o subestimó los datos medidos (2). Es usado por Gaiser et al. (2010). Un valor de cero no necesariamente indica ajuste

perfecto, pues al no usar el valor absoluto en su cálculo, los errores positivos compensan a los negativos. Éste índice conserva las unidades de los datos medidos.

$$ME = (1/n) \sum_{i=1}^n (\hat{y}_i - y_i) \quad (2)$$

2.1.3 Error Cuadrático Medio

El error cuadrático medio o MSE por sus siglas en inglés, pondera las estimaciones que están más alejadas del valor medido (3). Es usado por Karamirad et al. (2013). Su resultado está en unidades al cuadrado de la variable medida.

$$MSE = (1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

2.1.4 Raíz del Error Cuadrático Medio

La raíz del error cuadrático medio o RMSE por sus siglas en inglés, mide el valor medio cuadrático del error (4). Es presentado por (Hyndman & Koehler, 2006; Bergmeir & Benítez, 2012) y usado en (Liu et al., 2010; Karamirad et al., 2013; Gaiser et al., 2010; Wu et al., 2008). Éste índice es el más usado para la validación de modelos de sistemas físicos en la literatura revisada. Su resultado tiene las unidades de la variable medida y pondera los pronósticos que están más alejados del valor medido.

$$RMSE = \sqrt{(1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

2.2 Errores que no Dependen de la Escala

Tienen la ventaja de ser independientes de la escala. Algunos están definidos porcentualmente siendo de fácil interpretación para los ingenieros, dado que en ingeniería se suelen usar valores porcentuales. Cuando el conjunto de datos tiene cruce por cero o

cercanos al mismo, algunos índices toman valores altos o indeterminados.

2.2.1 Error Medio Absoluto Porcentual

El error medio absoluto porcentual o MAPE por sus siglas en inglés, mide el porcentaje de error promedio de las estimaciones (5). Es presentado por Bergmeir y Benítez (2012) y usado por Liu *et al.* (2010). Su valor mínimo es cero, y significa que las estimaciones y los datos medidos son iguales en todo el conjunto de datos.

$$MAPE[\%] = (100/n) \sum_{i=1}^n |(y_i - \hat{y}_i)/y_i| \quad (5)$$

2.2.2 Suma Normalizada de Errores Cuadráticos Modificada

La suma normalizada de errores cuadráticos modificada o MNSSE por sus siglas en inglés, es una modificación del NSSE con el fin de que su valor esté expresado en porcentaje (6). Es usado por (Camarillo-Peñaranda & Saavedra-Montes, 2012; Choi *et al.*, 2006). Su valor mínimo es cero, lo cual indica un ajuste perfecto.

$$MNSSE[\%] = 100 * \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n y_i^2} \quad (6)$$

2.2.3 Error Medio Relativo

El error medio relativo o MRE por sus siglas en inglés, mide el bias relativo entre las estimaciones y los datos medidos (7). Es usado en Gaiser *et al.* (2010) y es una modificación del ME. Resulta de dividir el ME en cada dato por el dato medido, lo cual lo independiza de la escala. Su resultado es un número adimensional.

$$MRE = (1/n) \sum_{i=1}^n (\hat{y}_i - y_i)/y_i \quad (7)$$

2.2.4 Suma Normalizada de Errores Cuadráticos

La suma normalizada de errores cuadráticos o NSSE por sus siglas en inglés, mide el porcentaje de error cuadrático entre las estimaciones y los datos medidos (8). Se usa en Saavedra-Montes et al. (2011). Su resultado es un número adimensional.

$$NSSE = 100 \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) / \left(\sum_{i=1}^n y_i^2 \right) \quad (8)$$

2.3 Otros Indicadores

En este grupo están los índices que no se pueden clasificar en los grupos ya mencionados, dichos índices son los siguientes:

2.3.1 Mejor Ajuste

El índice de mejor ajuste o FIT por su nombre en inglés, es una comparación entre los datos medidos y estimados con respecto al promedio de los datos medidos (9). Se usa en Wibowo y Saad (2010). Su resultado es un porcentaje que llega hasta 100%, donde 100% indica un ajuste perfecto y 0% muestra que, en promedio, los datos medidos duplican a los datos estimados. Puede tomar valores negativos, en cuyo caso la capacidad de réplica del modelo estimado es nula.

$$FIT[\%] = 100 \left(1 - \sum_{i=1}^n |y_i - \hat{y}_i| / \sum_{i=1}^n |y_i - \bar{y}| \right) \quad (9)$$

Dónde \bar{y} es el valor promedio de los datos medidos.

3. RESULTADOS Y DISCUSIÓN

Para realizar una comparación de los índices de error y el índice de ajuste, se generaron conjuntos de datos que tienen las principales características de los datos que se usan para validar modelos y procedimientos experimentales en ingeniería. Las señales de referencia se comparan con una señal a la cual se le introduce un

error conocido. Las características principales de los conjuntos de datos estudiados son dos: datos que tienen diferentes escalas en el rango de muestras y series con datos que cruzan por cero. Con base en estas dos características, se generan los conjuntos de comparación.

Se presentan tres conjuntos de comparación que incluyen una señal de referencia y una señal de comparación. Las señales en los conjuntos de comparación se basan en la función escalón y la función seno: La función escalón se define como (10) y es presentada en Oppenheim et al. (1997):

$$u(t) = \begin{cases} a & \rightarrow t < c \\ b & \rightarrow t > c \end{cases} \quad (10)$$

Dónde a es el valor inicial de la función, b el valor final de la función y c es el tiempo en [s] en el que ocurre el escalón. Para $t=c$, el cambio de valor ocurre instantáneamente. La función seno se define como (11) y es presentada en Oppenheim et al. (1997):

$$u(t) = A \sin(\omega t + \varphi) \quad (11)$$

Dónde A es la amplitud de la función, ω es la frecuencia angular de la función en [rad/s] y φ es el desfase de la función en [rad]. A continuación se definen los conjuntos de comparación propuestos y se discuten los resultados de los índices de error e índice de ajuste.

3.1 Conjunto de Comparación 1

Para el conjunto de comparación 1, se toman como señales de referencia un escalón con $a = 0$, $b = 100$ y $c = 0$ y una sinusoidal con $A = 100$, $\omega = 2$ y $\varphi = 0$. Las señales de comparación Señal 1 y Señal 2 son señales iguales a las de referencia. Este conjunto de comparación se hizo para verificar los valores mínimos de los índices de error y el índice de ajuste. En cada caso, la señal de referencia hace las veces de los datos medidos experimentalmente y las señales de comparación son los datos estimados por el modelo.

Tabla 1. Resultados de los índices para el conjunto de comparación 1.

Fuente: Autores

Índice	Señal 1	Señal 2
MAE [Unidades]	0,00	0,00
ME [Unidades]	0,00	0,00
MSE [Unidades^2]	0,00	0,00
RMSE [Unidades]	0,00	0,00
MAPE [%]	0,00	NaN
MNSSE [%]	0,00	0,00
MRE [-]	0,00	NaN
NSSE [-]	0,00	0,00
FIT [%]	NaN	100,00

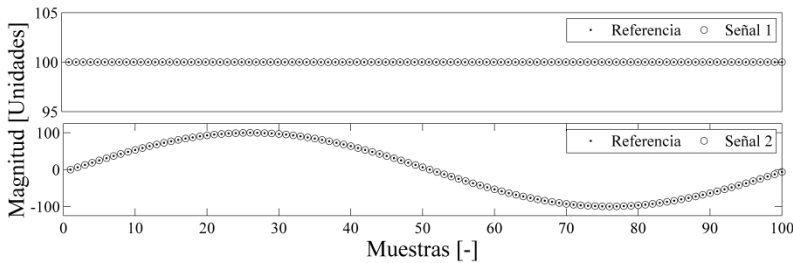


Fig. 1. Conjunto de comparación 1. Fuente: Autores

El resultado NaN corresponde a una indeterminación de la forma 0/0. Aquí se observa el problema mencionado en la Sección 2.2. Cuando las series de datos tiene valores en cero como en la Señal 2, algunos índices que no dependen de la escala se indeterminan como es el caso del MAPE y el MRE. La indeterminación para el FIT se da porque el valor promedio y el valor medido para la Señal 1 son iguales en cada. En los demás casos, se verifican los valores mínimos de los índices.

3.2 Conjunto de Comparación 2

En este conjunto de comparación, se toman como señales de referencia un escalón con $a = 0$, $b = 100$ y $c = 0$ y una senoide con $A = 100$, $\omega = 2$ y $\varphi = 0$. Las señales Señal 3, Señal 4 y Señal 5 son funciones escalón con $a = 0$, $b = 110$ y $c = 0$; $a = 0$, $b = 90$ y $c = 0$; $a = 110$, $b = 90$ y $c = 0.5$, respectivamente. La Señal 6 es una función

seno con $A = 110$, $\omega = 2$ y $\varphi = 0$. En este conjunto de comparación las señales de comparación tienen un error constante y conocido del 10% en cada punto. La representación gráfica del conjunto se muestra en la Fig. 2.

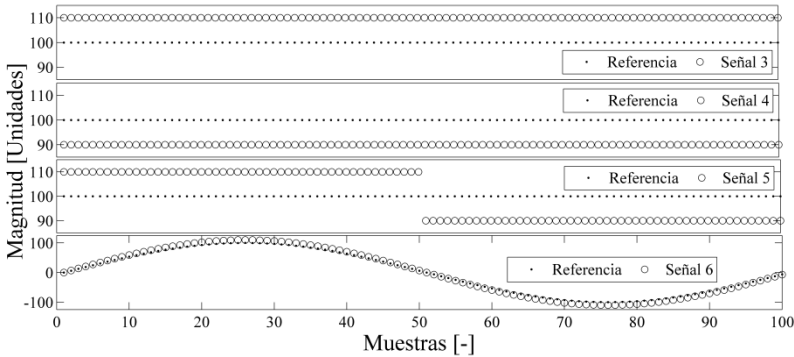


Fig. 2. Conjunto de comparación 2. Fuente: Autores

Tabla 2. Resultados de los índices para el conjunto de comparación 2.

Índice	Fuente: Autores			
	Señal 3	Señal 4	Señal 5	Señal 6
MAE [Unidades]	10,00	10,00	10,00	6,36
ME [Unidades]	10,00	-10,00	0,00	0,00
MSE [Unidades ²]	100,00	100,00	100,00	50,00
RMSE [Unidades]	10,00	10,00	10,00	7,07
MAPE [%]	10,00	10,00	10,00	NaN
MNSSE [%]	10,00	10,00	10,00	10,00
MRE [-]	0,10	-0,10	0,00	NaN
NSSE [-]	1,00	1,00	1,00	1,00
FIT [%]	Inf	Inf	Inf	90,00

El resultado Inf se debe a una división por cero, que para el FIT se da por la misma razón mencionada en el conjunto de comparación 1, al igual que el resultado NaN para el MAPE y el MRE. Se observa que los índices ME y MRE tienen valores negativos que son consistentes con su definición, lo que están indicando es que el modelo subestima los datos medidos, también que para la Señal 5 y la Señal 6 los errores positivos compensan a los negativos, obteniendo como resultado un valor de cero. Asimismo se observa que

el MNSSE y el NSSE tienen un valor consistente en todas las señales y que los errores que dependen de la escala cambian su valor en la Señal 6, la cual es la señal sinusoidal, pues a lo largo de la serie de datos la escala va cambiando, tiene valores en la escala de 10^{-1} , 10^0 , 10^1 y 10^2 .

3.3 Conjunto de comparación 3

En este conjunto de comparación, se toman tres señales de referencia escalón con diferentes escalas. En todas tres $a = 0$ y $c = 0$; pero b es 0.1, 1 y 10, respectivamente. Las señales Señal 7, Señal 8 y Señal 9 tienen un error con respecto a la señal de referencia del 10%. La representación gráfica del conjunto se muestra en la Fig. 3.

Tabla 3. Resultados de los índices para el conjunto de comparación 1.

Fuente: Autores

Índice	Señal 7	Señal 8	Señal 9
MAE [Unidades]	0,01	0,10	1,00
ME [Unidades]	0,01	0,10	1,00
MSE [Unidades ²]	0,0001	0,01	1,00
RMSE [Unidades]	0,01	0,10	1,00
MAPE [%]	10,00	10,00	10,00
MNSSE [%]	10,00	10,00	10,00
MRE [-]	0,10	0,10	0,10
NSSE [-]	1,00	1,00	1,00
FIT [%]	Inf	Inf	Inf

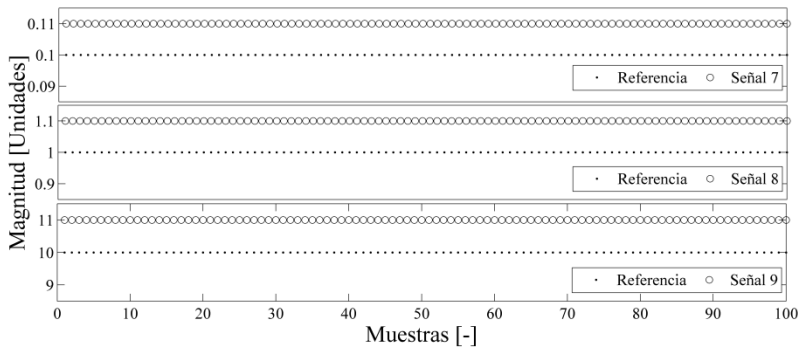


Fig. 3. Conjunto de comparación 3. Fuente: Autores

Aquí se observa el mismo fenómeno que en los conjuntos de comparación 1 y 2 para el FIT. Los errores que no dependen de la escala tienen un único valor para todas las comparaciones y lo contrario ocurre para los errores que dependen de la escala. Aquí se observa una desventaja de los errores que dependen de la escala pues si se usan datos de diferentes escalas para validar el mismo modelo, el resultado de los índices será menor cuando la escala sea menor.

3.4 Recomendaciones

Cuando se validan modelos con señales que varían en diferentes escalas, e.g. señales que toman valores entre 0 y 100 [-], se recomienda usar los errores que no dependen de la escala, puesto que si se utilizan errores dependientes de la escala, estos errores ponderan las desviaciones de la escala mayor generando un resultado ponderado.

El resultado de los índices ME, MAE, y RMSE coinciden con las unidades de las señales comparadas. Cuando se requiere la desviación en las unidades de las señales, estos índices son la mejor opción, cuidando el uso del RMSE cuando hay variación de la escala. En el caso de señales que cruzan por cero, el MAPE y el MRE aumentaran su resultado cuando el dato medido sea cercano a cero y la desviación aumente. Esta condición dará un valor alto del MAPE y el MRE incluso cuando las señales comparadas estén cercanas. Cuando el dato medido coincide en cero, los errores se indeterminan.

El MNSSE mostró resultados consecuentes en todas las condiciones evaluadas, i.e. su resultado fue cero cuando las señales comparadas eran iguales, su resultado mostró el error introducido aun cuando los datos variaron de escala y también cuando los datos cruzaron por cero. Además el resultado de este índice es un porcentaje, facilitando la interpretación por parte de ingenieros. Dado que no muestra la sobre o subestimación del modelo, se recomienda su uso en conjunto con el ME, el cual muestra la sobre o subestimación incluso en señales que crucen por cero.

4. CONCLUSIONES

En este artículo se presentó una comparación de los índices de error MAE, ME, MSE, RMSE, MAPE, MNSSE, MRE y NSSE y el índice de ajuste FIT. De los resultados de la comparación se hicieron recomendaciones para seleccionar los índices de error con base en las características de los datos utilizados en la validación de modelos.

5. AGRADECIMIENTOS

Este artículo fue soportado por la beca doctoral 095-2005 de COLCIENCIAS y por la Universidad Nacional de Colombia a través de los proyectos IDENT-EXGEN y SMART- ALEN.

6. REFERENCIAS

- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- Camarillo-Peñaranda, J. R., Saavedra-Montes, A. J. (2012). Dynamic model validation via error indexes. In 2012 IEEE 4th Colombian Workshop on Circuits and Systems (pp. 1–6).
- Choi, B. K., Chiang, H. D., Li, Y., Li, H., Chen, Y. T., Huang, D. H., & Lauby, M. G. (2006). Measurement-Based Dynamic Load Models: Derivation, Comparison, and Validation. *IEEE Transactions on Power Systems*, 21(3), 1276–1283.
- Gaiser, T., de Barros, I., Sereke, F., & Lange, F. M. (2010). Validation and reliability of the EPIC model to simulate maize production in smallholder farming systems in tropical sub-humid West Africa and semi-arid Brazil. *Agriculture, Ecosystems & Environment*. 135(4), 318–327.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*. 22(4), 679–688.
- Karamirad, M., Omid, M., Alimardani, R., Mousazadeh, H., & Heidari, S. N. (2013). ANN based simulation and experimental verification of analytical four- and five-parameters models of PV modules. *Simulation Modelling Practice and Theory*, 34, 86–98.

- Liu, H., Tian, H. Q., Chen, C., & Li, Y. (2010). A hybrid statistical method to predict wind speed and wind power. *Renewable Energy*, 35(8), 1857–1861.
- Oppenheim, A. V., Willsky, A. S., & Nawab, S. H. (1997). *Signals and Systems*. Prentice Hall.
- Saavedra-Montes, A. J., Ramirez-Scarpetta, J. M., & Malik, O. P. (2011). Methodology to estimate parameters of an excitation system based on experimental conditions. *Electric Power Systems Research*, 81(1), 170–176.
- Wibowo, T. C. S., & Saad, N. (2010). MIMO model of an interacting series process for Robust MPC via System Identification. *ISA Transactions*, 49(3), 335–347.
- Wu, X., Zhu, X., Cao, G., & Tu, H. (2008). Dynamic modeling of SOFC based on a T–S fuzzy model. *Simulation Modelling Practice and Theory*, 16(5), 494–504.