

COMPARACIÓN DE MEDIDAS DE DESVIACIÓN PARA VALIDAR MODELOS SIN SESGO, SESGO CONSTANTE O PROPORCIONAL

Comparison of deviance measures for validation of models without bias, constant or proportional bias

S Medina-Peralta ✉, L Vargas-Villamil, J Navarro-Alberto, C Canul-Pech, S Peraza-Romero

(SMP) Facultad de Matemáticas, Universidad Autónoma de Yucatán. Periférico Norte Tablaje 13615. Apdo. Postal 172 Cordemex, CP 97119. Tel. (999) 942-31-40. Mérida, Yucatán, México. mperalta@uady.mx

(LVV) Colegio de Postgraduados Campus Tabasco

(JNA) Facultad de Medicina Veterinaria y Zootecnia, UADY

(CCP)(SPR) Facultad de Matemáticas, UADY

Artículo recibido: 19 de febrero de 2008, **aceptado:** 9 de junio de 2010

RESUMEN. En la literatura se han expuesto diferentes enfoques y técnicas para validar modelos en predicción del sistema. Dos de ellas son: métodos gráficos y medidas de desviación que comparan los valores observados con los predichos. En este trabajo se comparan las medidas de desviación Eficiencia de Modelado (MEF), Coeficiente de Determinación del Modelo (CD), y el Coeficiente del Error (C) para validar modelos sin sesgo (SS), con sesgo constante (SC) o proporcional (SP). Las tres medidas indican un buen ajuste de los valores observados con los predichos para un modelo SS, y concuerdan con lo visto en los métodos gráficos. El CD contradice lo observado en los métodos gráficos para un modelo con SC. Las medidas MEF, CD y C contradicen a los métodos gráficos para un modelo con SP. Se recomienda utilizar conjuntamente medidas de desviación y métodos gráficos para validar modelos.

Palabras clave: Modelo, validación, medidas de desviación, sesgo.

ABSTRACT. Diverse approaches and techniques have been documented in the literature for model validation in the prediction of systems. Two techniques involve the comparison of observed and predicted values by 1) graphical methods and 2) deviance measures. In the present study, comparisons of three measures of deviation, Modeling Efficiency (MEF), Coefficient of Model Determination (CD), and Coefficient of Error (C), is performed for models without bias (SS), and with constant (SC) and proportional bias (SP). The three measures indicate a good fit between observed and predicted values for a SS model, in agreement to what is observed using graphical methods. For a SC model, CD contradicts the interpretations given by the graphical methods. MEF, CD and C all contradict what it is seen in the graphical methods for a SP model. According to these results, it is preferable a combination of graphical methods and deviance measures for the purpose of model validation.

Key words: Model, validation, deviance measures, bias.

INTRODUCCIÓN

Una de las etapas en el proceso de modelación es la validación del modelo, la cual presenta dificultades tanto conceptuales como prácticas (Barrales *et al.* 2004). La validación se define como la comparación de las predicciones del modelo con los valores observados del sistema real para determinar si el modelo es adecuado para el propósito estable-

cido (Mayer & Butler 1993; Mitchell 1997; Oberkamp & Trucano 2002; Montgomery *et al.* 2002; Halachmi *et al.* 2004). Las observaciones del sistema real son diferentes a las utilizadas para estimar los parámetros del modelo y cada valor predicho es una salida del modelo, resultado de ser evaluado o alimentado con los valores de las variables de entrada en el modelo, mismos que corresponden a la situación experimental y donde se conoce el valor

real observado.

En la literatura se han expuesto diferentes enfoques y técnicas para validar modelos. Hamilton (1991) recopiló una extensa lista de publicaciones (316) con respecto a la validación de modelos con énfasis en artículos de potencial interés para estadísticos, e incluyó para cada una un breve comentario acerca de que tratan y sus palabras clave. Para Mayer & Buttler (1993), las técnicas de validación se pueden agrupar en cuatro principales categorías: la evaluación subjetiva (involucra a un número de expertos en el campo de interés), las técnicas visuales (gráficas comparativas), las medidas de desviación (basadas en las diferencias entre valores observados y predichos) y las pruebas estadísticas. Por su parte Tedeschi (2006) realizó una revisión de varias técnicas para evaluar modelos matemáticos diseñados para propósitos predictivos. En su revisión expone las siguientes técnicas: análisis de regresión lineal, análisis de los errores ajustados, coeficiente de correlación de concordancia, diversas medidas para evaluación, el error cuadrado medio de predicción, análisis no paramétricos y la comparación de la distribución de los datos.

En la práctica es común que para validar modelos en predicción del sistema se utilice diferentes medidas de desviación para comparar a los valores observados con los predichos (Loague & Green 1991; Mayer & Butler 1993; Analla 1998; Kobayashi & Salam 2000; Yang et al. 2000; Collao-Saenz et al. 2005; Tedeschi 2006). La medida de desviación Eficiencia de Modelado (MEF), ha sido reportada como la mejor medida de concordancia entre los valores observados y predichos (Mayer & Butler 1993). Tedeschi (2006) utiliza entre otras técnicas para validación de modelos, tanto a la MEF como al Coeficiente de Determinación del Modelo (CD). Yang et al. (2000) por medio de un análisis cluster determinó cuatro grupos formados por diferentes medidas de desviación y pruebas estadísticas, el grupo 1 incluye al Coeficiente del Error (C), el Error Absoluto Medio (MAE) y la Raíz del Error Cuadrado Medio (RMSE) e indica que sólo una estadística de cada grupo se necesita para validar el modelo. Por lo que en este trabajo se utilizaran MEF, CD y C.

Mitchell (1997) señala que no es sorprenden-

te que para efectuar la validación, los modeladores recurran a procedimientos simples a su alcance, aparentemente adecuados, incluyendo gráficos de dispersión de predicciones y observaciones, algunas veces utilizando regresión, la cual es pensada como un método objetivo y cuantitativo para medir cuán bueno es un modelo. Propuso un método que no requiere de los supuestos necesarios de los métodos estadísticos, en donde se grafica en el eje de las abscisas los valores predichos y en el eje de las ordenadas las desviaciones (predicho menos observado) y el porcentaje de puntos que caen dentro de un rango o precisión aceptable con centro en cero, es usado como un criterio de adecuación del modelo.

Un análisis gráfico exploratorio de los valores predichos (z) contra los observados (y) y de los predichos contra la desviación ($y-z$), es básico para visualizar la exactitud, precisión e identificación del tipo de sesgo. Adicionalmente pueden utilizarse medidas de desviación basadas en las diferencias entre los valores observados y predichos, para cuantificar la bondad del modelo en predicción del sistema.

Medina-Peralta (2006) señala que en ocasiones el CD contradice lo observado en los métodos gráficos para validar modelos. En este trabajo se comparan tres medidas de desviación para validar modelos que presentan o no sesgo en sus pronósticos.

MATERIALES Y MÉTODOS

Las medidas de desviación para validar modelos consideradas en este trabajo fueron: la Eficiencia de Modelado (MEF), el Coeficiente de Determinación del Modelo (CD) y el Coeficiente del Error (C).

La estadística MEF es similar al coeficiente de correlación de Pearson (r), el cual es interpretado como la proporción de la variación explicada por la recta de regresión ajustada, mientras que la MEF es la proporción de la variación explicada por la recta $y = z$, y en un ajuste perfecto ambas estadísticas tendrían un valor igual a uno (Tedeschi 2006). Sustituyendo \hat{y}_i por z_i en la expresión de r se obtiene MEF:

$$r = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$$\text{MEF} = 1 - \frac{\sum_{i=1}^n (y_i - z_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (d_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

donde y_i es el i -ésimo valor observado, \hat{y}_i es el i -ésimo valor predicho por la recta de regresión ajustada, \bar{y} es la media aritmética de los valores observados y z_i es el i -ésimo valor predicho por el modelo a validar. La cota superior de MEF es uno (Loague & Green 1991; Tedeschi 2006), pero puede ser negativo (Loague & Green 1991) y su cota inferior (teórica) es menos infinito (Tedeschi 2006). Si $\text{MEF} < 0$ los valores predichos por el modelo son peor que sencillamente usar la media observada \bar{y} (Loague & Green 1991).

El Coeficiente de Determinación del Modelo (CD)

$$\text{CD} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (z_i - \bar{y})^2} \quad (3)$$

es el cociente de la variación total de los datos observados entre el total de las diferencias al cuadrado de los valores predichos respecto a la media de los datos observados. CD indica la proporción de la variación total de los datos observados explicada por los datos predichos (Loague & Green 1991). Al igual que la MEF, en un ajuste perfecto CD valdría uno. Un valor de CD cercano a uno indica una mejora en las predicciones del modelo, $\text{CD} > 1$ es un indicador de baja predicción y si $\text{CD} < 1$ de sobrepredicción (Tedeschi 2006).

La estadística llamada Coeficiente del Error (C)

$$C = \frac{\sum_{i=1}^n |y_i - z_i|}{n \bar{y}} = \frac{\text{MAE}}{\bar{y}} \quad (4)$$

es una medida promedio relativa de las diferencias absolutas, la cual es expresada como una proporción de la media de los valores observados (\bar{y}) (Klepper & Rouse 1991). Un valor de C muy cercano a cero indica que el modelo cumple con su objetivo. Ha sido utilizada para describir el ajuste del modelo a los datos observados y para comparar la exactitud de las salidas de un modelo con diferentes variables (Yang *et al.* 2000).

Para comparar las medidas de desviación descritas anteriormente, se utilizaron datos observados y predichos por un modelo: (i) de simulación estocástica sin sesgo (SS) en sus pronósticos para 21 áreas basales de parcelas de eucalipto de Tasmania (Tabla 1), correspondiente al Ejemplo 1 en Reynolds (1984); y (ii) dinámico mecanístico inédito llamado Wakax POS, con sesgo constante (SC) en sus predicciones de ganancia de peso promedio (GPP) por día de bovinos en una zona tropical de México (Tabla 2). Este modelo fue desarrollado por el Dr. Luís Vargas Villamil del Colegio de Postgraduados Campus Tabasco en una estancia posdoctoral en el Departamento de Ciencia Animal de la Universidad de California, Davis (UCD). El modelo Wakax POS describe las relaciones biológicas (digestión, crecimiento bacteriano, fermentación y absorción) durante la nutrición de bovinos alimentados con caña de azúcar (CZ) y predice la ganancia de peso promedio (GPP) por día de bovinos en pastoreo suplementado con CZ, maíz quebrado y/o melaza en una zona tropical de México. Consta de 119 variables de estado que describen el sistema compuesto por cinco submodelos: Concentrado, Pasto, Caña de azúcar, Melaza y Crecimiento animal. Las variables de entrada del modelo son: a) peso vivo; b) consumo de materia seca de maíz, melaza y pasto; c) fracción soluble de pasto y CZ; d) fracción degradable de pasto y CZ; y e) razón de degradación de pasto y CZ (Medina-Peralta 2006). En el caso de un modelo con sesgo proporcional (SP) en sus pronósticos, se seleccionaron 30 datos del ejercicio 2 del capítulo 5 de Kleinbaum *et al.* (1998), que trata sobre la presión arterial sistólica (SBP), tamaño corporal (QUET), edad y si fuma o no para una muestra hipotética de hombres mayores de 40 años de una

ciudad. Para efecto de ilustrar la aplicación sólo se utilizaron las variables SBP y QUET, ambas divididas entre 10, por presentar tendencia lineal positiva. Así, $(SBP/10, QUET/10) = (z, d) = (z, y - z)$ donde z denota un valor predicho, y un valor observado, $d = y - z$ la desviación entre ambos, siendo $y = z + d$ (Tabla 3). En los dos primeros casos de estudio, para el propósito de validación, los valores observados fueron distintos a los utilizados para estimar los parámetros de cada modelo. Para el tercero, esto se asumió.

Tabla 1. Valores observados y predichos de 21 áreas basales ($m^2 ha^{-1}$) de parcelas de eucalipto de Tasmania.

Table 1. Observed and predicted values of basal areas ($m^2 ha^{-1}$) of Tasmanian eucalypt forest measured in 21 plots.

Observado (y_i)	Predicho (z_i)	Diferencia ($d_i = y_i - z_i$)
62	61	1
40	39	1
60	56	4
52	52	0
75	73	2
39	41	-2
50	49	1
54	54	0
68	64	4
37	39	-2
56	64	-8
69	68	1
59	52	7
48	44	4
74	71	3
60	59	1
56	55	1
65	70	-5
43	50	-7
52	60	-8
57	59	-2

Se elaboró una tabla comparativa en donde las filas fueron los diferentes tipos de datos, y las columnas las medidas de desviación.

Adicionalmente se construyeron los gráficos de los valores predichos contra las desviaciones ($d_i = y_i - z_i$) y el de dispersión de los valores observados (y_i) contra los valores predichos (z_i), junto con la recta 45° ($y = z$) (Flavelle 1992; Mayer & Butler 1993; Mitchell 1997; Yang et al. 2000; Hayirli

et al. 2003; Tedeschi 2006; Medina-Peralta 2006), para corroborar los resultados obtenidos en dicha tabla.

Tabla 2. Promedios observados de ganancia de peso (kg) de 34 experimentos y sus correspondientes predichos con el modelo inédito Wakax POS.

Table 2. Observed average weight gain (kg) of 34 experiments and their corresponding predicted values using the unpublished model Wakax POS.

Observado (y_i)	Predicho (z_i)	Diferencia ($d_i = y_i - z_i$)
0.366	0	0.366
0.398	0	0.398
0.493	0.39	0.103
0.63	0.43	0.2
0.43	0.13	0.3
0.4	0.28	0.12
0.49	0.23	0.26
0.6	0.29	0.31
0	0.28	-0.28
0.43	0.09	0.34
0.7	0.32	0.38
0.69	0.33	0.36
0.5	0.02	0.48
0.68	0	0.68
0.62	0	0.62
0.52	0.26	0.26
0.64	0	0.64
0.45	0	0.45
0.43	0.33	0.1
0.76	0.39	0.37
0.036	0	0.036
0.019	0	0.019
0.412	0	0.412
0.052	0	0.052
0.414	0	0.414
0.05	0	0.05
0.292	0.35	-0.058
0.054	0	0.054
0.308	0.39	-0.082
0.054	0	0.054
0.037	0	0.037
0.051	0	0.051
0.062	0	0.062
0.364	0	0.364

RESULTADOS

Medidas de desviación para un modelo sin sesgo, sesgo constante y proporcional

La Tabla 4 contiene los valores de las medi-

das de desviación MEF, CD y C para los casos de un modelo SS, otro con SC y un tercero con SP.

Gráficas del sesgo y de dispersión de los valores predichos contra observados. Gráficas para un modelo sin sesgo

Reynolds (1984) probó que $D = Y - Z$ se distribuye normal con media igual a cero, por lo tanto el modelo no presenta sesgo. Esto también puede observarse en la Figura 1, en donde los puntos (z_i, d_i) de la Tabla 1 forman prácticamente una banda horizontal centrada alrededor de la recta $d = 0$.

Tabla 3. Valores observados y predichos de un modelo con sesgo proporcional.

Table 3. Observed and simulated predicted values for a model with proportional bias.

Observado (y_i)	Predicho (z_i)	Diferencia ($d_i = y_i - z_i$)
13.79	13.5	0.29
14.33	14	0.33
13.31	13	0.31
15.18	14.8	0.38
14.9	14.6	0.3
13.18	12.9	0.28
16.57	16.2	0.37
16.36	16	0.36
16.99	16.6	0.39
15.3	14.9	0.4
15.61	15.2	0.41
14.17	13.8	0.37
14.36	14	0.36
13.7	13.4	0.3
14.84	14.5	0.34
14.5	14.2	0.3
13.82	13.5	0.32
14.54	14.2	0.34
15.36	15	0.36
14.78	14.4	0.38
14.03	13.7	0.33
13.52	13.2	0.32
15.23	14.9	0.33
13.5	13.2	0.3
12.28	12	0.28
12.9	12.6	0.3
16.48	16.1	0.38
17.41	17	0.41
15.6	15.2	0.4
16.8	16.4	0.4

El gráfico de dispersión de los valores predichos (z_i) vs. observados (y_i) (Figura 2), permite

visualizar que tan alejados están los puntos de la recta $y = z$, así como, tendencias de los puntos a dicha recta que representa la exactitud ideal. La distribución de los puntos en la Figura 2 muestra que éstos se encuentran “ceranos” y alrededor de la exactitud ideal ($y = z$), esto indica gráficamente que el modelo no presenta sesgo en sus pronósticos.

Tabla 4. Medidas de desviación para modelos SS, SC y SP.

Table 4. Deviance measures for SS, SC and SP models.

Caso	Medidas de desviación		
	MEF	CD	C
SS	0.861	1.122	0.003
SC	-0.837	0.698	0.637
SP	0.927	0.98	0.023

Gráficas para un modelo con sesgo constante

Los datos de la Tabla 2 presentan SC por las siguientes razones: (i) la estimación de $\hat{B} = \bar{D}$ es diferente de cero, $E(D) = \mu_D \neq 0$ (Prueba T t = 6.092; $p < 0.0001$), y (ii) los puntos (z_i, d_i) en la Figura 3 forman prácticamente una banda horizontal centrada alrededor de la recta $d = \bar{d} = 0,233$ kg (Medina-Peralta 2006).

La distribución de los puntos en la Figura 4 muestra que éstos se encuentran “alejados” de la exactitud ideal ($y = z$), aunque si se trasladan una cierta cantidad perpendicularmente al eje de las abscisas, la exactitud mejorará. Lo anterior refuerza la percepción de que el modelo presenta SC en sus pronósticos.

Gráficas para un modelo con sesgo proporcional

Los datos de la Tabla 3 presentan sesgo proporcional (SP), los puntos (z_i, d_i) en la Figura 5 tienen una tendencia lineal positiva significativa, $d = -0,0354 + 0,0263z$ (ANDEVA F = 51.23; $P < 0.0001$).

La distribución de los puntos en la Figura 6 muestra que éstos se encuentran casi en la exactitud ideal ($y = z$), aunque por encima de dicha recta e indicando sobrepredicción. Si se trasladan una cierta cantidad perpendicularmente al eje de las abscisas, los puntos “coincidirán” con $y = z$.

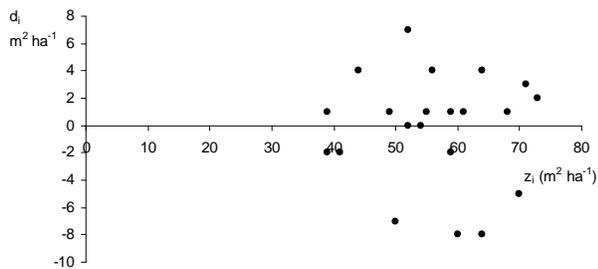


Figura 1. Relación entre los valores simulados (z_i) y el sesgo (d_i). Caso: modelo sin sesgo.
Figure 1. Relation between simulated values (z_i) and bias (d_i). Case: model without bias.

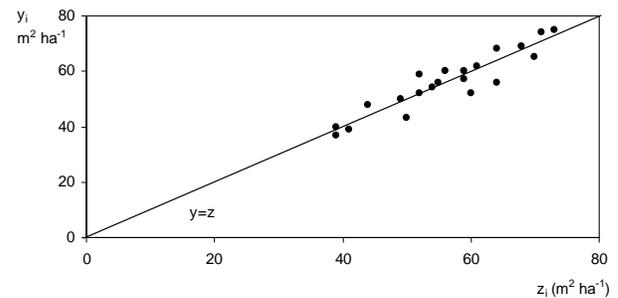


Figura 2. Relación entre los valores simulados (z_i) y los valores observados (y_i) para un modelo sin sesgo. $y = z$ es la recta que representa la exactitud ideal.
Figure 2. Relation between simulated (z_i) and observed values (y_i) for a model without bias. $y = z$ is the straight line representing the ideal accuracy.

DISCUSIÓN

Mayer & Butler (1993) clasifican a las técnicas de validación en: evaluación subjetiva, técnicas visuales, medidas de desviación y pruebas estadísticas. La primera carece de una cuantificación objetiva, y la última, cada método inferencial se encuentra principalmente sujeto a las dificultades para satisfacer sus supuestos. Por lo que las técnicas gráficas aplicadas a la par con las medidas de desviación, pueden complementarse al validar un modelo. Harrison (1990) señala que cuando no se satisfacen los supuestos del método inferencial, se debe recurrir a métodos descriptivos y a expertos en el área de estudio para que proporcionen una opinión acerca de la validación del modelo, o lo adecuado del modelo sea calificado por el modelador de acuerdo a su criterio y propósitos (Mitchell 1997). Así, un análisis gráfico exploratorio de los valores predichos contra los observados y de los predichos contra las desviaciones, es fundamental para visualizar la exactitud, precisión e identificación del tipo de sesgo. Con más razón cuando el tamaño de muestra es relativamente pequeño, debido a que pueden ser demasiado costosas o difíciles de obtener en campo o laboratorio. Para Barrales *et al.* (2004), los índices o medidas para efectuar validación no presentan el carácter objetivo que se demanda de las pruebas o métodos estadísticos en el sentido que para un mismo conjunto de datos, todos los modeladores, usando el mismo pro-

cedimiento, lleguen a las mismas conclusiones. Para Yang *et al.* (2000), las medidas de desviación tienen ventaja sobre las pruebas estadísticas, ya que son fáciles de interpretar y no necesitan de la aplicación de alguna transformación a los datos.

Tedeschi (2006) señala que en un ajuste perfecto MEF tendría un valor igual a uno. La cota superior de MEF es uno (Loague & Green 1991; Tedeschi 2006), pero puede ser negativo (Loague & Green 1991) y su cota inferior (teórica) es menos infinito (Tedeschi 2006). Por su parte Loague & Green (1991), indican que si $MEF < 0$ los valores predichos por el modelo son peor que sencillamente usar la media observada \bar{y} .

Tedeschi (2006) indica que al igual que la MEF, en un ajuste perfecto CD valdría uno y que un valor de CD cercano a uno indica una mejora en las predicciones del modelo, $CD > 1$ es un indicador de baja predicción y si $CD < 1$ de sobrepredicción. Yang *et al.* (2000) señalan que un valor de C muy cercano a cero indica que el modelo ajusta a los datos observados. Para el ejemplo de un modelo SS en sus pronósticos, el valor de MEF es cercano a la unidad, indicando un buen ajuste de los valores observados con los predichos; también el valor de CD es próximo a uno y refleja de nuevo un buen ajuste del modelo; el valor de la estadística C es casi cero y coincide con lo señalado con MEF y CD. En este caso (SS), concuerdan los resultados tanto con las medidas de desviación como con los métodos grá-

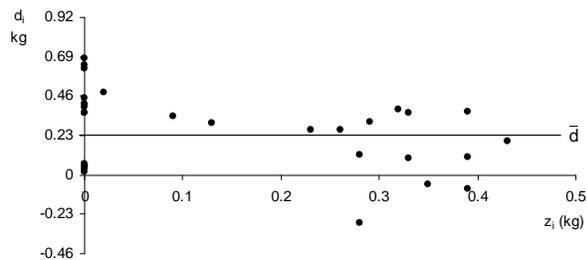


Figura 3. Relación entre los valores simulados (z_i) y el sesgo (d_i). Caso: modelo con sesgo constante.

Figure 3. Relation between simulated values (z_i) and bias (d_i). Case: model with constant bias.

ficos (Figuras 1 y 2). Para el caso de un modelo con SC en sus pronósticos, el valor de MEF es negativo, por lo que sería mejor usar la media de los valores observados que los valores predichos por el modelo; $CD < 1$ indica sobrepredicción, contrario a lo observado en la Figura 4; el valor de C se aleja de cero señalando falta de ajuste de los valores predichos con los observados. Lo anterior se debe a que el modelo presenta SC en sus pronósticos, la cual se observó en las Figuras 3 y 4. En este caso (SC), CD contradice lo visto en los métodos gráficos. Medina-Peralta (2006) señala que en ocasiones el CD contradice lo observado en los métodos gráficos para validar modelos y que es debido a que con el CD se comparan las distancias al cuadrado de los valores observados y predichos respecto a la media de los valores observados (\bar{y}), por lo que depende de que tan lejos se encuentran dichos valores de \bar{y} . Para el caso de un modelo con SP en sus pronósticos, el valor de MEF es cercano a la unidad, indicando un buen ajuste de los valores predichos con los observados; también el valor de CD es próximo a uno e indica de nuevo un buen ajuste del modelo; el valor de C es cercano a cero y concuerda con lo señalado con MEF y CD . En este caso (SP), las medidas de desviación contradicen a los métodos gráficos al indicar un buen ajuste de los valores observados con los predichos, cuando el modelo presenta SP en sus pronósticos (Figuras 5 y 6).

Determinar comportamientos extraños entre los pronósticos de un modelo y los valores obser-

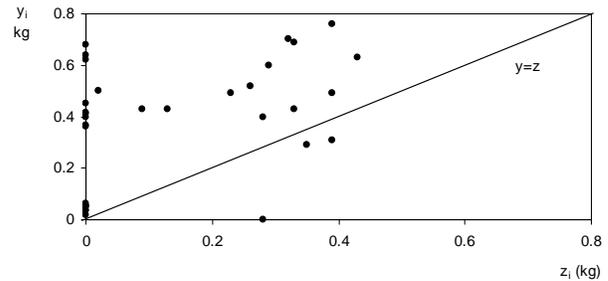


Figura 4. Relación entre los valores simulados (z_i) y los valores observados (y_i) para un modelo con sesgo constante. $y = z$ es la recta que representa la exactitud ideal.

Figure 4. Relation between simulated (z_i) and observed values (d_i). $y = z$ is the straight line representing the ideal accuracy.

vados del sistema real, como por ejemplo el tipo de sesgo, es fundamental para el mejoramiento del modelo a través de cuestionar desde su estructura hasta los datos y métodos empleados en todos los procesos. Según McCarthy *et al.* (2001), probar un modelo ayuda a identificar sus debilidades para que sea mejorado su desempeño predictivo por medio de un proceso iterativo de desarrollo del modelo, probarlo, modificarlo y probarlo nuevamente. Para Tedeschi (2006), la identificación y aceptación de inexactitudes de un modelo es un paso hacia la evolución de un modelo más exacto y de más confianza.

Para Mayer & Butler (1993) la complejidad de los modelos y del tipo de datos, origina que no haya un conjunto combinado de técnicas de validación aplicable en todas las situaciones de modelación, y señalan que en la mayoría de los casos, un número de medidas de validación son necesarias para apreciar "la foto completa". Yang *et al.* (2000), indican que con varios grupos formados con medidas de desviación o pruebas estadísticas, puede obtenerse la misma conclusión y que deberían utilizarse gráficas para la comparación entre los valores predichos y observados, así como para la interpretación de las estadísticas usadas en la evaluación. Por su parte Tedeschi (2006), indica que la valoración de la adecuación de un modelo solamente es posible por medio de una combinación de varios análisis estadísticos y propios al propósito para la cual el modelo matemático fue inicialmente conceptualizado y desarrollado.

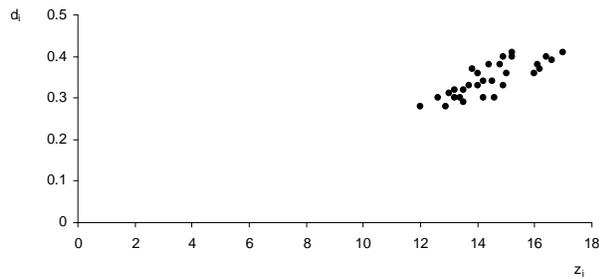


Figura 5. Relación entre los valores simulados (z_i) y el sesgo (d_i). Caso: modelo con sesgo proporcional.
Figure 5. Relation between simulated (z_i) and bias (d_i). Case: model with proportional bias.

Para los casos de estudio presentados en este trabajo, la MEF, CD y C indican un buen ajuste de los valores predichos con los observados para un modelo SS, y concuerdan con lo visto en los métodos gráficos. El CD contradice lo observado en los métodos gráficos para un modelo con SC y las medidas MEF, CD y C contradicen a los métodos gráficos, al indicar un buen ajuste de los valores observados con los predichos para un modelo con SP en sus

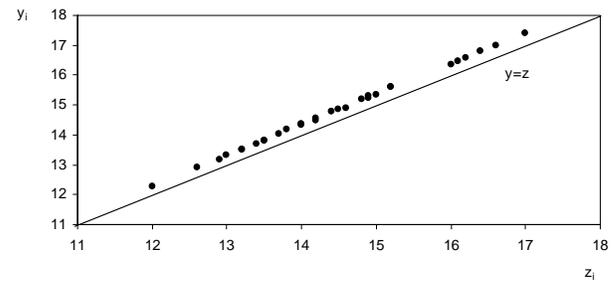


Figura 6. Relación entre los valores simulados (z_i) y los valores observados (y_i) para un modelo con sesgo proporcional. $y = z$ es la recta que representa la exactitud ideal.
Figure 6. Relation between simulated (z_i) and observed values (y_i) for a model with proportional bias. $y = z$ is the straight line representing the ideal accuracy.

pronósticos.

Se recomienda utilizar conjuntamente medidas de desviación y métodos gráficos para validar modelos.

La identificación de inexactitudes del modelo como el tipo de sesgo, es fundamental para el mejoramiento del modelo a través de cuestionar desde su estructura hasta los datos y métodos empleados en todos los procesos.

LITERATURA CITADA

- Analla M (1998) Model validation through the linear regression fit to actual versus predicted values. *Agricultural Systems* 57: 115-119.
- Barrales VL, Peña RI, Fernández RB (2004) Model validation: an applied approach. *Agric. Téc.* 64(1): 66-73.
- Collao-Saenz EA, Dijkstra J, Aguiar PPC, Bannink A, Arcuri PB, Teixeira JC, Olalquiaga PJR, David FM (2005) Simulation model for particle dynamics in rumen of cattle fed sugarcane diet. *Sci. Agric. (Piracicaba, Braz.)* 62(2): 102-110.
- Flavelle P (1992) A quantitative measure of model validation and its potential use for regulatory purpose. *Advances in Water Resources* 15: 5-13.
- Halachmi I, Edan Y, Moallem U, Maltz E (2004) Predicting feed intake of the individual dairy cow. *Journal of Dairy Science* 87: 2254-2267.
- Hamilton MA (1991) Model validation: an annotated bibliography. *Communications in Statistics: Theory & Methods* 20: 2207-2266.
- Harrison SR (1990) Regression of a model on real-system output: an invalid test of model validity. *Agricultural Systems* 34: 183-190.
- Hayirli A, Grummer RR, Nordheim EV, Crump PM (2003) Models for predicting dry matter intake of Holsteins during the prefresh transition period. *Journal of Dairy Science* 86: 1771-1779.

- Kleinbaum DG, Kupper LL, Muller KE, Nizam A (1998) Applied regression analysis and other multivariate methods. 3a Ed. Duxbury Press, Pacific Grove CA USA. 798 pp.
- Klepper O, Rouse DI (1991) A procedure to reduce parameter uncertainty for complex models by comparison with real system output illustrated on a potato growth model. *Agricultural Systems* 36: 375-395.
- Kobayashi K, Salam MU (2000) Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal* 92: 345-352.
- Loague K, Green RE (1991) Statistical and graphical methods for evaluating solute transport models: Overview and applications. *Journal of Contaminant Hydrology* 7: 51-73.
- Mayer DG, Butler DG (1993) Statistical validation. *Ecological Modelling* 68: 21-32.
- McCarthy MA, Possingham HP, Day JR, Tyre AJ (2001) Testing the accuracy of population viability analysis. *Conservation Biology* 15(4): 1030-1038.
- Medina-Peralta S (2006) Validación de modelos mecanísticos basada en la prueba ji-cuadrada de Freese, su modificación y extensión. Tesis de Maestría. Colegio de Postgraduados, México. 150 pp.
- Mitchell PL (1997) Misuse of regression for empirical validation of models. *Agricultural Systems* 54(3): 313-326.
- Montgomery DC, Peck EA, Vining GG (2002) Introducción al análisis de regresión lineal. 3a. Ed. CECSA. México, DF. 588 pp.
- Oberkampf WL, Trucano TG (2002) Verification y validation in computacional fluid dynamics. *Progress in Aerospace Sciences* 38: 209-272.
- Reynolds MRJr (1984) Estimating the error in model predictions. *Forest Sci.* 30(2): 454-469.
- Tedeschi LO (2006) Assessment of the adequacy of mathematical models. *Agricultural Systems* 89: 225-247.
- Yang J, Greenwood DJ, Rowell DL, Wadsworth GA, Burns IG (2000) Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. *Agricultural Systems* 64: 37-53.

