NOTA CIENTÍFICA

VALIDACIÓN DE MODELOS: UN ENFOQUE APLICADO[1]

Model validation: an applied approach[1]

Luis Barrales V.[2] *, Iván Peña R.[2] y Pedro Fernández de la Reguera B.[3]

ABSTRACT

The varied statistical procedures used for model validation as well as fundamental and practical objections formulated by many authors to these procedures have led modelers to use non-inferential methods for empirical validation of models. The procedure presented by Freese (1960) to validate forestry models and the modification by Rennie and Wiant (1978) are perfectly applicable to other areas. These procedures and the proposed extensions allow covering a scenario with data coming from models that may or may not have biased estimates, where the maximum admissible error is expressed either or both in terms of the same units or as a percentage of the real value. The illustration of the methodologies used data from a perennial grassland growth system.

Key words: Freese chi-sq test, model validation, simulation models, bias correction, limiting error.

RESUMEN

Los variados procedimientos estadísticos usados en la validación de modelos y las objeciones tanto fundamentales como prácticas, formuladas por diversos autores a muchos de estos procedimientos, ha llevado a que los modeladores recurran a métodos no inferenciales para validar empíricamente sus modelos. El procedimiento presentado por Freese (1960) para validar modelos en el área forestal y la modificación presentada por Rennie y Wiant (1978), son perfectamente aplicables en otras áreas. Estos procedimientos y las extensiones propuestas, permiten cubrir un escenario de datos provenientes de modelos que presenten o no sesgos en sus estimaciones, donde el error máximo admisible se exprese en las mismas unidades o en términos porcentuales del valor real. La ilustración de las metodologías consideró datos asociados al crecimiento de una pradera perenne.

Palabras clave: dócima de χ^2 de Freese, validación de modelos, modelos de simulación, corrección de sesgos, error límite.

[1] Recepción de originales: 13 de noviembre de 2002.
[2] Pontificia Universidad Católica, Facultad de Agronomía e Ingeniería Forestal, Casilla 306, Santiago 22, Santiago, Chile.
E-mail: lbarrale@puc.cl *Autor para correspondencia.
[3] Universidad Federico Santa María, Facultad de Ingeniería, Casilla 29, Correo 35, Santiago.

INTRODUCCIÓN

En la modelación de sistemas, una etapa esencial y que presenta dificultades tanto conceptuales como prácticas, es la validación de los modelos. Una parte importante en este proceso es la validación empírica, que, según Reynolds (1984) y Mitchell (1997), se efectúa para comparar las predicciones del modelo con observaciones provenientes del mundo real.

Para Aguilar (1997) y Rauscher *et al.* (2000) estas comparaciones, idealmente, se deben efectuar usando métodos estadísticos adecuados, con un nivel de confianza aceptable, de tal modo que las inferencias sean correctas. Sin embargo, en muchos casos los nuevos modelos de simulación han sido presentados sin una adecuada validación. Según Reynolds (1984) una posible razón para esta falta de examen crítico, es el hecho que en la literatura científica ha habido relativamente poca discusión sobre la filosofía y procedimientos para este tipo de investigación. Mitchell (1997) coincidió con esta aseveración, y agregó que no es sorprendente, que para efectuar validación, los modeladores recurran a procedimientos simples, a su alcance, aparentemente adecuados, incluyendo gráficos de dispersión de predicciones y observaciones, algunas veces con regresión, la cual es pensada como método objetivo y cuantitativo para medir cuán bueno es un modelo.

Uno de estos procedimientos simples ha sido la dócima de "*t*" para observaciones pareadas, la cual prueba que la media de las desviaciones entre predicciones y observaciones no difiere significativamente de cero. Sin embargo, Freese (1960), Law (1983), Harrison (1990) y Mitchell (1997), la han cuestionado como procedimiento para probar exactitud.

Reckhow *et al.* (1990), Flavelle (1992) y Mayer *et al.* (1994) concordaron en la necesidad de un método objetivo y cuantitativo para evaluar modelos. Propusieron la regresión como técnica de validación; sin embargo, coincidiendo con Harrison (1990), reconocieron los problemas de esta metodología para satisfacer los supuestos cuando es empleada con este fin. Posteriormente, Mitchell (1997) objetó la regresión por inapropiada para la validación empírica de modelos. Por su parte Analla (1998), usando un fundamento diferente, coincidió con Harrison (1990) quien demostró que una prueba F simultánea de intercepto cero y pendiente uno, puede ser engañosa, y por lo tanto, no aceptable de aplicar en la validación de modelos. A pesar de lo señalado, la regresión por su atracción intuitiva (Harrison, 1990), familiaridad (Flavelle, 1992) y disponibilidad en paquetes computacionales (Webster, 1989) es frecuente y erróneamente usada para validar modelos de simulación.

Según Prodan *et al.* (1997) los antecedentes bibliográficos sobre validación de modelos, en general, no recomiendan la realización de inferencia estadística al realizar una validación debido a que las pruebas adecuadas son paramétricas, requiriendo supuestos que generalmente no se cumplen en los datos bajo análisis. Con anterioridad, una argumentación similar llevó a Harrison (1990) a sugerir que los métodos estadísticos para estudios de validación se confinen al uso de instrumentos descriptivos (y no inferenciales) combinados con comparaciones gráficas, y que la calificación del modelo en cuanto a su validez sea efectuada por especialistas. Sin embargo, para Flavelle (1992) este enfoque subjetivo no es una base apropiada para evaluar los resultados de una validación. Mitchell (1997) propuso un método que no requiere de los tradicionales supuestos de los métodos estadísticos, que se concentra en graficar las desviaciones (predicción menos observación) junto con la precisión aceptable expresada en términos absolutos. Lo adecuado del modelo es calificado por el modelador de acuerdo a su criterio y propósitos.

Prodan *et al.* (1997) propusieron indicadores, como la raíz cuadrada del error medio cuadrático (REMC), expresado en unidades absolutas o porcentuales, el error medio absoluto (EMA) y la desviación estándar de los residuos (DSR), para medir la exactitud con la cual un modelo predice los valores observados, mientras que sugieren la diferencia agregada (DA), como medida de sesgo del modelo. Para Analla (1998) una prueba adecuada para la validación de modelos se basa en el cuadrado medio del error de la regresión de los modelos evaluados, magnitud que, expresada en términos de la media, entregaría una indicación de la calidad de la validación. Sin embargo, estos índices 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 procedimiento, lleguen a las mismas conclusiones.

Freese (1960) aborda el problema de validación de modelos en el área forestal, y propone un procedimiento estadístico inferencial para determinar si la exactitud de un modelo o técnica de estimación es adecuada para cumplir los requerimientos del modelador. Reynolds (1984), al revisar los supuestos y la derivación del procedimiento de Freese, señaló explícitamente las condiciones bajo las cuales puede ser usado, condiciones que permiten aplicar este procedimiento de validación en otras áreas.

En este trabajo se presenta y extiende el procedimiento de validación de modelos desarrollado por Freese (1960), y la modificación sugerida por Rennie y Wiant (1978). Se desarrolla un ejemplo numérico para ilustrar la aplicación e interpretación de ambos enfoques inferenciales en un área distinta a la forestal, como es la disponibilidad de materia seca de una pradera perenne.

MATERIALES Y MÉTODOS

La propuesta de Freese para la validación de modelos

Freese (1960), bajo la filosofía de prueba de hipótesis y sin establecer explícitamente los supuestos básicos, aborda el problema de validación y establece que puede ser resuelto por un procedimiento que considera tres elementos: una definición de la exactitud requerida, una cuantificación de la exactitud alcanzada, y un método objetivo para decidir si esta última exactitud corresponde con aquella requerida. Partiendo de la usual dócima de χ^2 de bondad de ajuste, propone que, si la variable aleatoria x_i se distribuye normal, las siguientes dos expresiones generales para probar la magnitud de las diferencias (d_i) entre valores estimados (x_i) y reales (μ_i), se distribuyen como χ^2 con n grados de libertad:

$$\sum_{i=1}^{n}(x_i-\mu_i)^2/\sigma^2=\sum_{i=1}^{n}d_i^2/\sigma^2=(Z^2_{(1-\alpha_2/2)}/E^2)\sum_{i=1}^{n}d_i^2 \quad (1)$$

$$\sum_{i=1}^{n}(x_i-\mu_i)^2/\sigma_i^2=\sum_{i=1}^{n}d_i^2/\sigma_i^2=(Z^2_{(1-\alpha_2/2)}/p^2)\sum_{i=1}^{n}(d_i/\mu_i)^2 \quad (2)$$

En estas expresiones n es el tamaño de la muestra y σ^2 es la varianza hipotética, especificada como un cuociente entre E^2 y Z^2 definidos a continuación.

El valor E es la discrepancia aceptada entre la predicción x_i y el valor real μ_i, es decir, el valor máximo de las desviaciones $|x_i-\mu_i|$. El valor E se define en las mismas unidades de μ para (1) y como una proporción p del valor real μ_i para (2), en cuyo caso $E = p\mu_i$, luego σ_i^2 es la varianza hipotética de x_i cuando el error se expresa como proporción de μ_i. El valor Z es el percentil $(1-\alpha_2/2)$ de la distribución normal estándar, asociado a la probabilidad α_2 que $|x_i-\mu_i|$ exceda el valor E.

En el trabajo que comentamos, Freese compara un valor de χ^2 calculado para (1) y (2) según corresponda, con un valor tabular $\chi^2_{(1-\alpha 1)}$ (n), para decidir si una nueva técnica o modelo tiene el deseado nivel de exactitud. Enfatizamos que esta decisión es una dócima realizada a nivel crítico α_1. Esta prueba de χ^2 rechazaría los modelos o técnicas inexactas sin importar la fuente de la inexactitud (sesgo, falta de precisión, o ambas). Sin embargo, algunos modelos o técnicas podrían ser capaces de proporcionar una excelente exactitud si el sesgo del modelo o técnica pudiese ser removido, ya sea éste constante (asumido igual para todos los valores μ_i) o proporcional (que se incrementa o decrece directamente con los valores μ_i). Freese (1960) presenta una corrección por sesgo de la fórmula general (1) y proporciona pruebas aproximadas de exactitud para validar modelos después de eliminar un sesgo constante o eliminar un sesgo proporcional, fórmulas 1.a y 1.b, respectivamente. Para el primer caso, sustituye en la fórmula general el término d_i por $(d_i - B)$, donde el sesgo constante $B = (\bar{x} - \bar{\mu})$. Para el segundo caso, reemplaza el término d_i de la fórmula general por e_i, que corresponde al residual del ajuste de la regresión lineal de valores x_i sobre los respectivos valores reales μ_i. Las correcciones presentadas por Freese (1960) a la ecuación (1) son:

$$(Z^2_{(1-\alpha_2/2)} / E^2) \sum_{i=1}^{n} (d_i - B)^2 \tag{1.a}$$

$$(Z^2_{(1-\alpha_2/2)} / E^2) \sum_{i=1}^{n} e_i^2 \tag{1.b}$$

Cada una de estas expresiones se distribuye aproximadamente como χ^2 con (n - 1) y con (n - 2) grados de libertad, respectivamente.

El procedimiento sugerido por Freese (1960) para la aplicación de estas dócimas es, primero, decidir si las diferencias observadas, sin importar su fuente, pueden ser toleradas. Si así ocurriese, la nueva técnica o modelo puede sustituir a la estándar. En caso contrario, la fuente de la inexactitud debe ser identificada y removida o la nueva técnica descartada. Esta forma de realizar la validación implica efectuar un detallado examen de los datos, con el propósito de detectar el tipo de inexactitud presente en las estimaciones obtenidas a partir de la nueva técnica y así usar la adecuada prueba modificada que permita remover el tipo de sesgo presente.

La modificación de Rennie y Wiant a la propuesta de Freese

Rennie y Wiant (1978) modifican la propuesta de Freese (1960) y proponen un enfoque de límites de confianza. Con esta modificación, el resultado de la validación se presenta como un error límite E*. Conceptualmente, el error límite anterior E y el actual E* representan lo mismo, pero con una diferencia. Ésta es, que E se establece *a priori* por el modelador, mientras que E* se calcula *a posteriori*. El modelador decide acerca de la nueva técnica o modelo al comparar el error límite E* con la exactitud que necesita. A partir de las ecuaciones generales (1) y (2) se despeja el error límite E*, el cual se expresa en las mismas unidades de μ_i o en términos porcentuales, como se indica en las ecuaciones 3 y 4, respectivamente.

$$E^* = \sqrt{\frac{Z^2_{(1-\alpha_2/2)} \sum_{i=1}^{n} (d_i)^2}{\chi^2_{(n,\alpha_1)}}} \tag{3}$$

$$E^*\% = \sqrt{\frac{Z^2_{(1-\alpha_2/2)}}{\chi^2_{(n,\alpha_1)}} \sum_{i=1}^{n} \left(\frac{d_i}{\mu_i}\right)^2} \, (100) \tag{4}$$

La interpretación de E* a partir de la expresión (3) es que $(1 - \alpha_2)$ por ciento de las desviaciones entre los valores proporcionados por la nueva técnica o modelo y los valores desde la estándar o reales presentarán valores menores que o iguales a ese error límite, a menos que los datos constituyan una muestra extraña y la probabilidad asociada a esta situación es igual a α_1. La correspondiente interpretación E*% a partir de la expresión (4) es que se espera que un $(1 - \alpha_2)$ por ciento de las desviaciones estén dentro de ± E*%, a menos que los datos constituyan una muestra extraña y que la probabilidad que así ocurra es α_1.

Extensión de la propuesta de Freese

Una de las formas comunes de definir el máximo error tolerado es expresarlo como un porcentaje de los valores reales a ser determinados. Sin embargo, Freese (1960) consideró este criterio en la expresión (2), pero no presentó en forma explícita los procedimientos de prueba para modelos cuyas estimaciones presenten ya sea un sesgo fijo o proporcional. Sin embargo, la extensión es directa. Así para evaluar la exactitud de un modelo después de eliminar la presencia de un sesgo constante se sustituye en (2) el valor d_i por $(d_i - B)$. De igual forma, para el caso de eliminar la presencia de un sesgo proporcional, se sustituye en (2) el término d_i por e_i, quedando ambas expresiones de la siguiente forma:

$$(Z^2_{(1-\alpha_2/2)}/p^2)\sum_{i=1}^{n}\left(\frac{d_i - B}{\mu_i}\right)^2 \qquad (2.a)$$

$$(Z^2_{(1-\alpha_2/2)}/p^2)\sum_{i=1}^{n}\left(\frac{e_i}{\mu_i}\right)^2 \qquad (2.b)$$

Cada una de éstas sigue aproximadamente una distribución χ^2 con (n - 1) y (n - 2) grados de libertad.

Para ilustrar la aplicación e interpretación de la metodología descrita, se usó la información de Castellaro (2003), quien desarrolló un modelo de simulación de crecimiento de praderas. Los datos simulados por el modelo y los reales se presentan en el Cuadro 1.

Cuadro 1. MS[1] real (μ_i), MS pronosticada (x_i), diferencias (d_i), residuales (e_i)* y relaciones entre estos valores.
Table 1. Observed dry matter (μ_i) model predicted dry matter (x_i), differences (d_i) and residual (e_i)* and relations between these values.

Observación	MS real (kg ha^{-1}) μ_i	MS pronosticada (kg ha^{-1}) x_i	Diferencias d_i	e_i*	$(d_i/\mu_i)^2$	$((di-B)^2/\mu_i)^2$	$(e_i/\mu_i)^2$
1	1.872,4	1.790,5	-81,9	-149,99	0,00191	0,00549	0,00642
2	1.766,9	1.535,7	-131,2	-301,61	0,01710	0,02657	0,02914
3	1.857,0	1.737,5	-121,5	-189,93	0,00430	0,00922	0,01046
4	3.494,6	2.528,6	-966,0	-998,48	0,07640	0,08566	0,08164
5	2.913,8	2.632,1	-281,7	-326,93	0,00930	0,01350	0,01259
6	3.397,4	3.136,9	-260,5	-195,12	0,00590	0,00872	0,00755
7	3.485,0	4.402,9	917,9	885,21	0,06940	0,06105	0,06452
8	2.423,0	3.711,6	1.288,6	1.232,59	0,28280	0,25845	0,25878
9	2.070,1	3.413,3	1.343,2	1.279,45	0,42100	0,38616	0,38200
10	1.831,8	1.933,1	101,3	32,31	0,00310	0,00059	0,00031
11	2.622,2	2.458,2	-164,0	-215,63	0,00390	0,00709	0,00676
12	2.009,0	2.139,9	130,9	65,8	0,00420	0,00136	0,00107
13	1.661,7	1.691,8	30,1	-42,62	0,00030	0,00026	0,00066
14	1.651,5	1.664,1	12,6	-60,34	0,00010	0,00072	0,00133
15	1.654,1	1.879,8	255,7	152,81	0,01860	0,01043	0,00853
16	2.166,7	2.477,2	310,5	248,87	0,02050	0,01371	0,01319
17	2.366,0	2.196,7	-163,3	-226,56	0,00510	0,00913	0,00917
18	2.571,1	2.379,3	-191,8	-244,56	0,00560	0,00935	0,00905
10	2.918,0	2.518,9	-399,1	-444,24	0,01870	0,02441	0,02318
20	3.478,2	3.349,9	-128,3	-161,14	0,00140	0,00283	0,00215
21	3.509,3	4.219,0	709,7	677,54	0,04090	0,03461	0,03728
22	1.966,1	2.155,8	189,7	123,66	0,00930	0,00457	0,00396
23	1.763,0	1.615,6	-147,4	-219,90	0,00700	0,01342	0,01528
24	1.909,2	1.153,3	-755,9	-823,16	0,15680	0,18120	0,18591
Totales					1,18361	1,16850	1,17093

* Residuales e_i desde el modelo $\hat{x}_1 = 109,1975 + 0,9780\mu_i$

[1] MS: materia seca.

RESULTADOS

Siguiendo el procedimiento de validación delineado por Freese (1960), la primera etapa consistió en determinar si las diferencias observadas, sin importar su fuente, pueden ser toleradas; esto requirió decidir la modalidad y magnitud de expresar el error máximo permitido E y los valores de probabilidad de α_1 y de α_2. La modalidad más usada por los modeladores de expresar el valor E es en porcentaje del valor μ_i, que en esta ilustración se fijó en 15%, α_1 en 0,05 y α_2 en 0,10. Haciendo uso de la fórmula general (2) se calculó el valor de χ^2:

$$\chi^2 = \frac{(1{,}64)^2}{(0{,}15)^2}(1{,}1836) = 141{,}49$$

Este resultado, de acuerdo con las condiciones impuestas, entregó una evidencia con respecto al modelo, en el sentido que los valores pronosticados para el crecimiento de la pradera no concuerdan con los obtenidos experimentalmente (valor $p < 0{,}0001$).

La segunda etapa del procedimiento consistió en identificar si la inexactitud presente puede atribuirse a un sesgo constante o a uno proporcional. En general, la presencia de sesgo puede identificarse por la magnitud de B y por la forma en que se distribuyen los puntos provenientes de graficar en un plano de coordenadas, los valores d_i y μ_i. Un sesgo constante es reconocido por un valor de B muy diferente de cero y adicionalmente por una graficación, en que los puntos conforman una banda horizontal centrada alrededor de B, con una distribución sistemática a ser positivos o negativos. Un sesgo proporcional es esperado cuando los puntos del gráfico configuren una banda con una tendencia lineal positiva o negativa.

Para la situación analizada, se concluyó que la fuente de la inexactitud no es debida a un sesgo constante, puesto que el valor de B igual a 56,8 es comparativamente pequeño con respecto a los valores de las desviaciones absolutas ($|d_i|$), por lo tanto, considerado cercano a cero. Adicionalmente, la distribución de los puntos en el gráfico (Figura 1), no mostró la distribución definida para ese tipo de sesgo y tampoco la distribución de puntos característica de la presencia de un sesgo proporcional. Así, ninguna fuente de inexactitud debida a sesgo pudo ser identificada para ser removida, implicando que la inexactitud presente es propia del modelo. Sólo con un fin ilustrativo, se procedió a corregir primero por la presencia de un supuesto sesgo constante, haciendo uso de la expresión (2.a) y posteriormente por la presencia de un supuesto sesgo proporcional, utilizando la expresión (2.b). Para hacer comparables estos valores de χ^2 con el obtenido anteriormente, se usaron las mismas magnitudes ya definidas para E, α_1 y α_2. Los resultados obtenidos para estos casos son:

$$\chi^2_{(23\,gl)} = \frac{(1{,}64)^2}{(0{,}15)^2}(1{,}16850) = 139{,}68$$

$$\chi^2_{(22\,gl)} = \frac{(1{,}64)^2}{(0{,}15)^2}(1{,}17093) = 139{,}97$$

Ambos resultados corroboran que la inexactitud presente, no es consecuencia de un potencial sesgo en las estimaciones, sino que, de acuerdo a las exigencias impuestas, el modelo no cumple con los niveles de exactitud requerido.

Se aplicó también el enfoque de límites de confianza, en el cual el modelador decide acerca de la utilidad del modelo, al comparar el error límite E* obtenido desde la expresión (4), con el nivel de exactitud impuesto al modelo. Para la ilustración, se utilizó el correspondiente valor del Cuadro 1, y los mismos valores ya definidos para α_1 y α_2. Desde la expresión 4 se obtuvo el valor para E*%:

$$E^*\% = \sqrt{\frac{(1{,}64)^2}{36{,}4}(1{,}18361)}\;(100) = 29{,}57\%$$

Según este resultado, se espera que el 90% de las desviaciones entre los valores simulados por el modelo y los reales, se encuentren dentro de un error límite de 29,57%, a menos que los datos constituyan una muestra extraña y la probabilidad de que esto ocurra es de 5%. El modelador, de acuerdo a la exactitud por él requerida, decide

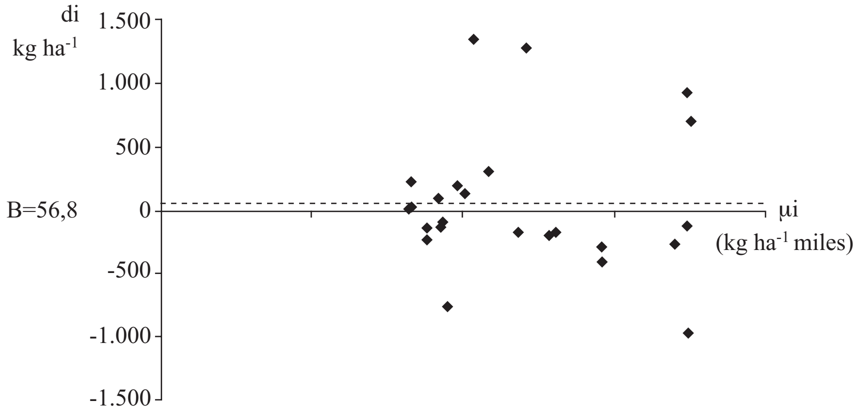

Figura 1. Relación de las desviaciones d_i y valores reales μ_i.
Figure 1. Relation between deviations d_i and real values μ_i.
B= Sesgo constante.

acerca de la utilidad del modelo. En este caso particular, ¿estará dispuesto a aceptar un límite de error de hasta un 30%, entre los valores pronosticados por el modelo y los verdaderos valores, en el 90% de los casos?

Rennie y Wiant (1978) no presentaron la estimación del error límite E* para la situación de corregir por la presencia de un sesgo fijo o proporcional. Sin embargo, la estimación es directa al despejar el término E* desde las expresiones (1.a) y (1.b) o desde (2.a) y (2.b), cuando el error límite se exprese en las mismas unidades del valor real o en términos porcentuales del valor real, respectivamente.

DISCUSIÓN

Las variadas causas por las cuales los procedimientos estadísticos comunes son considerados inapropiados para establecer la conformidad de los resultados, obtenidos de la modelación de sistemas con respecto a los de un escenario real, ha llevado a los modeladores a validar modelos recurriendo a procedimientos simples, aparentemente adecuados y que les parecen apropiados, como son gráficos de dispersión de predicciones y observaciones, algunas veces con regresión, la cual es pensada como método objetivo y cuantitativo.

Sin dejar de reconocer que lo señalado constituye un procedimiento de validación, la dócima χ^2 de varianza hipotética de Freese (1960), se presenta como una vía alternativa inferencial para la validación de modelos de simulación. Para la aplicación de este procedimiento, aparentemente es necesario conocer las varianzas poblacionales de antemano. Sin embargo, la expresión particular de las varianzas que se usa en las dócimas, permite su aplicación sin necesidad de conocerlas. Por otra parte, requiere de un compromiso por parte del usuario, que es decidir el nivel de exactitud asignado al modelo para su aceptación y adicionalmente, un estudio de la información generada por el modelo, que le permita seleccionar de las metodologías propuestas, aquella más adecuada a las características de los datos. Detectar en el modelo la presencia de un sesgo, ya sea constante o proporcional, en función de los valores del sistema real, le permitiría al usuario identificar en su modelo, la o las causas que lo producen, corregir deficiencias en el comportamiento predictivo, que llevaría a disminuir las discrepancias entre lo estimado por el modelo y los valores proporcionados por el escenario real, permitiendo que las conclusiones que se obtengan acerca de la confiabilidad del modelo cumplan con los objetivos establecidos.

El enfoque de límites de confianza, es de mayor aplicabilidad, pero requiere por parte del modelador un conocimiento de las variabilidades asociadas a la particular aplicación, de manera que le permita decidir, con respecto a la magnitud del error límite de acuerdo a la exactitud necesitada para los fines del modelo.

CONCLUSIONES

Las dócimas de exactitud presentadas originalmente por Freese (1960), las modificaciones efectuadas por Rennie y Wiant (1978) y las extensiones presentadas en este trabajo, son propuestas como metodologías inferenciales para determinar cuán bien un modelo se comporta en predecir los valores que son observados desde un mundo real. El escenario de los datos que cubren estas metodologías, proviene de modelos desarrollados en cualquier área, que presenten o no sesgos en sus pronósticos, y donde el máximo error admisible es expresado, ya sea en las mismas unidades del valor real, o como porcentaje de éste.

Los procedimientos de prueba de hipótesis tienen la misma hipótesis nula, esto es, que el modelo de simulación estima los valores reales con exactitud. Valores de χ^2 extremos, indican falta de exactitud, al igual como lo indica un valor extremo para el error límite, bajo el enfoque de límites de confianza.

LITERATURA CITADA

Aguilar, C. 1997. Simulación de sistemas, aplicaciones en producción animal. 241 p. Colección en Agricultura. Pontificia Universidad Católica de Chile, Facultad de Agronomía, Santiago, Chile.

Analla, M. 1998. Model validation through the linear regression fit to actual versus predicted values. Agric. Syst. 57:115-119.

Castellaro, G.L. 2003. Crecimiento de praderas mesofíticas a largo plazo, en respuesta a factores edafoclimáticos y modalidades de defoliación. 132 p. Tesis Magister en Ciencias Animales. Pontificia Universidad Católica de Chile, Facultad de Agronomía e Ingeniería Forestal, Santiago, Chile.

Flavelle, P. 1992. A quantitative measure of model validation and its potential use for regulatory purpose. Adv. Water Resour. 15:5-13.

Freese, F. 1960. Testing accuracy. For. Sci. 6:139-145.

Harrison, S.R. 1990. Regression of a model on real-system output: an invalid test of model validity. Agric. Syst. 34:183-190.

Law, A.M. 1983. Statistical analysis of simulation output data. Operations Res. 31:983-1029.

Mayer, D.G., M.A. Stuart, and A.J. Swain. 1994. Regression of real world data on model output: an appropriate overall test of validity. Agric. Syst. 45:93-104.

Mitchell, P.L. 1997. Misuse of regression for empirical validation of models. Agric. Syst. 54:313-326.

Prodan, M., R. Peters, F. Cox, y P. Real. 1997. Mensura forestal. 586 p. Serie Investigación y Educación en Desarrollo Sostenible. Instituto Interamericano de Cooperación para la Agricultura (IICA), San José, Costa Rica.

Rauscher, H.M., M.J. Young, C.D. Webb, and D.J. Rohison. 2000. Testing the accuracy of growth and yields models for Southern hardwood forests. South. J. Appl. For. 24:176-185.

Reckhow, K.H., J.T. Clements, and R.C. Dodds. 1990. Statistical evaluation of mechanistic water-quality models. J. Environ. Eng. 116:250-268.

Rennie, F.C., and H.V. Wiant, Jr. 1978. Modification of Freese's chi-square test of accuracy. USDI Bur. Land Manage. Research. Inventory. Note 14:1-3. Denver Service Center, Denver, Colorado.

Reynolds, M.R. Jr. 1984. Estimating the error in model predictions. For. Sci. 30:454-469.

Webster, R. 1989. Is regression what you really want? Soil Use Manage. 5:47-53.