

Una librería en R para validación de modelos de simulación

An R library for validation of simulation models

Ablan, Magdiel*

Universidad de Los Andes, Facultad de Ingeniería,
Centro de Simulación y Modelos (CeSiMo), Mérida, Venezuela

Márquez, Renny

Universidad de Los Andes, Facultad de Ingeniería,
Departamento de Investigación de Operaciones, Mérida, Venezuela

Rivas, Yuraima y Molina, Ana

Universidad de Los Andes, Facultad de Ingeniería,
Postgrado en Modelado y Simulación de Sistemas, Mérida, Venezuela

Querales, José

Universidad de Los Andes, Facultad de Ingeniería,
Postgrado en Matemáticas Aplicadas a la Ingeniería, Mérida, Venezuela

Resumen

La validación es una fase importante del proceso de simulación que permite evaluar la calidad de un modelo. Específicamente en el caso de modelos de simulación continua se comparan datos u observaciones del sistema real con las predicciones generadas por el modelo. Para facilitar a los usuarios esta comparación, en este trabajo se desarrolló una librería de métodos para validar modelos de simulación continua. Se incluyeron varias de las técnicas más usadas para validar modelos: desde índices simples tales como el error cuadrático medio y sus derivados (índice de Theil, índice de eficiencia, etc.) hasta métodos estadísticos clásicos (regresiones predicciones de modelo versus datos, pruebas t-pareadas, entre otros). Se incluyeron además métodos estadísticos bayesianos y basados en información que permiten discriminar entre versiones alternativas de un modelo. La librería incluye además varios conjuntos de datos, ayuda y un manual para comprender mejor las diferentes técnicas y su aplicación. Fue desarrollada en el lenguaje estadístico R y se encuentra disponible libremente.

Palabras Clave: Librerías en R, Validación, Simulación.

Abstract

Validation is an important phase in simulation that helps to evaluate model's quality. Specifically, in continuous simulation models, real data and model's results are compared. In this paper we present a library of methods to validate continuous simulation models that facilitate this task. We included several of the most used models validation techniques: from simple index such as mean squared error and its variants (Theil statistics, model efficiency, etc.) to classical statistical methods (model vs data regressions, paired t-test and others). We also included Bayesian statistical methods and information criteria methods that allow choosing between different models. The library includes several data frames, help and a manual for better understanding of the techniques and its application. The library was developed in R language and is available.

Keywords: R libraries, Validation, Simulation.

*Correo Electrónico: mablan@ula.ve

1. Introducción

La validación es una fase fundamental de cualquier modelo de simulación. Aunque existe cierta controversia entre diferentes autores sobre la semántica del término (Barlas, 1996; Martis, 2006) se puede definir de manera bastante amplia la validación como el examen de la calidad del modelo con respecto a los objetivos planteados. Específicamente este trabajo se basa en métodos de validación que sirven para comparar los resultados de series temporales de datos reales con los resultados o salidas obtenidos de una simulación para el mismo lapso temporal, como los que surgen en muchos problemas de simulación, tal como es el caso de la simulación ecológica (Pansu et al., 2010) o la simulación de procesos industriales (Márquez y Ramírez, 2009), entre otros.

La forma más común de validar un modelo consiste en examinar gráficamente el ajuste de los resultados del modelo a los datos correspondientes para el mismo período. Este proceso se hace de forma iterativa y se repite para versiones diferentes de un mismo modelo obtenidas por un cambio en el conjunto de parámetros usado o por cambios en la estructura del modelo. Muchas veces, no es fácil decidir, basado sólo en la evaluación visual, cuál versión del modelo es la mejor. Si la comparación se establece entre un mismo conjunto de datos y varios modelos alternativos, este proceso de evaluación visual puede ser aún más complicado y sujeto a errores. Es por esta razón que se requieren métodos y criterios estadísticos objetivos que permitan comparar los resultados arrojados por un modelo y los datos, y más aún, que permita evaluar modelos alternativos para un mismo conjunto de datos. En este sentido, los objetivos de este trabajo son:

1. Mostrar algunos de los criterios o índices de validación más importantes que pueden ser usados para comparar datos y resultados de un modelo.
2. Facilitar el uso de estos índices mediante el desarrollo de una librería del programa estadístico R.

Las técnicas de validación utilizadas en la librería son: el error cuadrático medio (Stermán, 1984; Mayer y Butler, 1993; Stermán, 2000; Haefner, 2005; Tedeschi, 2006); el estadístico de Theil (Stermán, 1984; Power, 1993; Mayer y Butler, 1993; Stermán, 2000); el error absoluto medio (Mayer y Butler, 1993; Stermán, 2000; Haefner, 2005; Warnock et al., 2005; Álvarez et al., 2007); el índice de acuerdo de Willmott (Willmott, 1981; Willmott et al., 1985; Ferrer et al., 2000); la eficiencia del modelo de Nash Sutcliffe, el coeficiente de correlación de Pearson y el coeficiente de determinación (Ferrer et al., 2000;

Kaur et al., 2003; Waller et al., 2003; Tedeschi, 2006; Sourisseau et al., 2008), la prueba F y la prueba t pareada (Mayer y Butler, 1993; Tedeschi, 2006); el estadístico Hotelling T^2 , el criterio de información de Akaike y la inferencia bayesiana (Chick, 2000; Burnham y Anderson, 2002; Timm, 2002; Haefner, 2005; Orhan et al., 2010).

Se ha seleccionado el programa estadístico R por ser una herramienta muy potente y versátil para análisis y visualización de datos y cálculo científico y estadístico. Posee además una gran comunidad de usuarios y es software libre, disponible para las principales plataformas de cómputo.

La estructura de este documento es como sigue: primero se describen algunas características de las series temporales de datos y resultados de simulación que conviene tener presentes al utilizarlas durante el proceso de validación. Enseguida se realiza una breve reseña de las técnicas de validación seleccionadas. Seguidamente, se describe la librería, con sus funciones y conjuntos de datos. Finalmente, se muestran algunos ejemplos que ilustran el uso de las técnicas y la librería. Para finalizar, se discuten los logros de este trabajo y sus limitaciones.

2. Datos observados y resultados del modelo

Existen ciertos atributos que deben ser considerados cuando se comparan las series temporales de datos observados o reales y los resultados o salidas de la simulación. Estos atributos determinan el tipo de validación que es posible. Los atributos a considerar son: independencia, número de respuestas del modelo, número de puntos de comparación y réplicas (Haefner, 2005).

1. Independencia: una de las primeras lecciones en el desarrollo y uso de modelos es que para que pueda realizarse una validación rigurosa, los datos que se usan para la misma deben ser diferentes e independientes de los datos usados para formular el modelo y calcular sus parámetros.
2. Respuestas simples o múltiples: casi cualquier modelo tiene varias variables de estado que pueden ser de interés. La validación debe decidir cuáles variables van a ser sujetas a comparación y si se va a hacer de manera individual o tratando de generar un índice que mida el ajuste agregado de las distintas salidas.
3. Comparaciones en un tiempo en particular o en varios instantes de tiempo: puede que sólo sea necesario realizar las comparaciones en un instante de tiempo en particular que revista de importancia especial en el sistema modelizado o en varios puntos a lo largo del tiempo.

4. Uso de réplicas: ciertas técnicas estadísticas requieren alguna forma de variabilidad, ya sea en los datos (réplicas experimentales) o en los resultados de la simulación (resultados de simulaciones estocásticas con diferentes semillas del generador).

3. Técnicas de validación de modelos de simulación

En la sección 1 se hizo referencia a la importancia del uso de técnicas de validación para evaluar modelos de simulación. A continuación se da una breve descripción de cada una de las técnicas utilizadas en la librería **Validacion** elaborada en el programa estadístico R.

3.1. Error cuadrático medio

El error cuadrático medio (MSE, por sus siglas en inglés *mean square error*) consiste de las diferencias entre los valores observados $Y_1, Y_2, \dots, Y_i, \dots, Y_n$ y las predicciones del modelo $X_1, X_2, \dots, X_i, \dots, X_n$, respectivamente (Tedeschi, 2006). La Ec. (1) muestra el cálculo de esta medida.

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \quad (1)$$

En este caso se ha considerado indistinto que el origen de la medida sea de predicciones provenientes de una regresión o de los resultados obtenidos de un modelo de simulación; sin embargo, existen algunos autores como (Haefner, 2005) y (Tedeschi, 2006) que para diferenciar ambas, describen las predicciones del modelo como error cuadrático medio de las predicciones (MSEP, por sus siglas en inglés *mean square error of prediction*).

Cuando cada par de datos (X_i, Y_i) es mutuamente independiente y el modelo es independiente (los parámetros fueron obtenidos de experimentos independientes y no fueron ajustados al experimento actual) el estimado del MSE es una medida confiable de exactitud del modelo. Si los parámetros del modelo fueron ajustados a los datos observados, el MSE subestimaré el verdadero valor de la medida porque el modelo tenderá a reproducir los datos reales (Tedeschi, 2006).

Con frecuencia se utiliza la raíz del cuadrado medio del error (RMSE), mostrada en la Ec. (2), la cual provee una medida del error en las mismas unidades que la variable bajo consideración. Igualmente existe otra medida que puede ser más conveniente, esta es el porcentaje de la raíz del cuadrado medio del error (RMSPE), que se muestra en la Ec. (3), donde el error es normalizado con el uso del promedio de las observaciones, \bar{Y} (Stermann, 1984).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \times \frac{100}{\bar{Y}}} \quad (3)$$

3.2. Estadístico de Theil

El estadístico de Theil se basa en la descomposición del MSE en tres componentes, asociados con las diferencias entre las medias del modelo y del sistema denotadas por U^M , diferencias entre la varianza de la salida del modelo y la varianza de las observaciones denotadas por U^S , y correlación de los resultados del modelo y los valores de las observaciones denotada como U^C (Haefner, 2005). La Ec. (4) muestra la descomposición del MSE en tres partes (Stermann, 1984; Stermann, 2000).

$$MSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} = (\bar{X} - \bar{Y})^2 + (S_X - S_Y)^2 + 2(1 - r)S_X S_Y \quad (4)$$

Donde S_X es la desviación estándar de los resultados del modelo, S_Y es la desviación estándar de las observaciones y r es el coeficiente de correlación entre los resultados del modelo y las observaciones.

Si se divide ambos lados de la Ec. (4) entre MSE , se obtiene que la suma de los tres componentes es igual a 1. En la Ec. (5), Ec. (6) y Ec. (7), se muestra el cálculo de U^M , U^S y U^C , respectivamente.

$$U^M = \frac{(\bar{X} - \bar{Y})^2}{MSE} \quad (5)$$

$$U^S = \frac{(S_X - S_Y)^2}{MSE} \quad (6)$$

$$U^C = \frac{2 \times (1 - r) \times S_X \times S_Y}{MSE} \quad (7)$$

Un sesgo grande viene dado con un valor de MSE y de U^M grandes; los errores de este tipo son potencialmente serios y usualmente se deben a errores en la estimación de los parámetros. Las diferencias en la varianza pueden ser sistemáticas; cuando U^S domina el MSE , el

modelo se ajusta a las observaciones y están altamente correlacionados, pero la variación alrededor de la media es distinta. Si la mayor parte del error está concentrado en U^C , el modelo captura la media y las tendencias en forma correcta, pero difiere en el ajuste punto a punto; generalmente, se quiere que U^C sea grande, porque este tipo de errores tienden a no ser sistemáticos y se deben a la presencia de ruido y aleatoriedad en el modelo (Sterman, 2000).

El estadístico de Theil permite definir la fuente del error. El criterio para decidir si un error es grave o sistemático depende del propósito que tenga el modelo. Si los errores surgen de formas de comportamiento que no son de interés para el modelo, no comprometen la integridad del mismo. Para una revisión más detallada de algunos comportamientos particulares y la interpretación del estadístico de Theil, puede revisarse (Sterman, 2000, p. 876).

3.3. Error absoluto medio

Otra técnica que puede utilizarse, similar al MSE es el error absoluto medio (MAE, por sus siglas en inglés *mean absolute error*), reflejado en la Ec. (8).

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| \quad (8)$$

MAE mide el error linealmente. Al igual que para el error cuadrático medio, una medida ponderada del error pudiera ser útil, por lo cual se propone el porcentaje del error absoluto medio (MAPE), como se define en la Ec. (9), que provee una medida adimensional robusta (Mayer y Butler, 1993).

$$MAPE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| \times \frac{100}{\bar{Y}} \quad (9)$$

Algebraicamente, $RMSE > MAE$ (debido a la influencia de valores cuadrados grandes), tendiendo estas medidas a ser iguales si las diferencias absolutas son de magnitud similar (Mayer y Butler, 1993).

3.4. Índice de acuerdo

El índice de acuerdo de Willmott d , definido por la Ec. (10), refleja el grado en el cual las observaciones son estimadas en forma correcta por el modelo. No es una medida de correlación como tal sino una medida del grado en el cual las predicciones del modelo están libres de errores (Willmott, 1981; Wainwright y Mulligan, 2003).

$$d = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (|X_i - \bar{Y}| + |Y_i - \bar{Y}|)^2} \quad (10)$$

Se utiliza con el fin de evitar problemas asociados con r y R^2 . Varía entre 0 (completa discordancia entre el valor estimado por el modelo y el valor observado) y 1 (un ajuste perfecto). Es un estadístico adimensional y su valor debe ser evaluado en base al fenómeno estudiado, la medición de exactitud y el modelo elaborado. Se hace significativo luego de su uso repetido en distintos problemas (Willmott, 1981; Wainwright y Mulligan, 2003).

3.5. Eficiencia del modelo

La eficiencia del modelo (EF, del inglés *modelling efficiency*), medida propuesta por Nash y Sutcliffe, es una estadística adimensional que relaciona directamente las predicciones del modelo con las observaciones. La Ec. (11) refleja el cálculo de este índice (Tedeschi, 2006).

$$EF = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (11)$$

En un ajuste perfecto $EF = 1$, el cual es su límite superior. Si el error es de la misma magnitud que la varianza observada $EF = 0$ y la media de las observaciones es una representación del sistema tan buena como el modelo (con lo cual se ha perdido el tiempo invertido en el modelo). Conforme el error se incrementa, los valores del índice se hacen cada vez más negativos, con lo cual los valores que predice el modelo son peores que la media de las observaciones, siendo el peor escenario teórico $EF \rightarrow -\infty$ (Wainwright y Mulligan, 2003; Tedeschi, 2006).

3.6. Coeficiente de correlación de Pearson

El coeficiente de correlación r , descrito por la Ec. (12), mide la fuerza de la relación lineal entre las observaciones y el modelo.

$$r = \frac{S_{x,y}}{S_x S_y} \quad (12)$$

Pese a que existen análisis estadísticos para probar la hipótesis $r = 0$ (no existe correlación), no hay valores de r distintos de cero a priori, contra los cuales probar. Por ejemplo, no hay razón para probar $r > 0,6$, al menos

que este valor haya formado parte de los objetivos del modelo (Haefner, 2005).

El coeficiente r , así como R^2 , requieren una interpretación cuidadosa. Existen varios malentendidos alrededor de estos estadísticos; por ejemplo, a) un coeficiente de correlación alto no indica que puedan hacerse predicciones útiles con el modelo porque mide precisión y no exactitud, b) un coeficiente de correlación alto no implica que la regresión lineal estimada permita un buen ajuste a las predicciones del modelo, porque la relación puede ser no lineal, y c) un valor de r cercano a cero no indica que los valores observados y las predicciones del modelo no están correlacionados, porque podrían tener una relación no lineal (Tedeschi, 2006).

3.7. Coeficiente de determinación

Una de las medidas de ajuste utilizada con mayor frecuencia es el coeficiente de determinación R^2 , reflejado por la Ec. (13), que es el cuadrado del coeficiente de correlación de Pearson, mostrado en la Ec. (12) (Stermán, 2000).

$$R^2 = r^2 = \frac{S_{x,y}^2}{S_x^2 S_y^2} \quad (13)$$

El coeficiente de determinación mide la fracción de la varianza en los datos que es explicada por el modelo. Si el modelo replica exactamente la serie actual, $R^2 = 1$; si la salida del modelo es constante, sin reflejar de ninguna manera la varianza de las observaciones, $R^2 = 0$ (Stermán, 2000; Wainwright y Mulligan, 2003).

3.8. Prueba F

En este enfoque se hace una regresión de las observaciones (eje y) sobre las predicciones (eje x). Si el modelo es perfecto, todos los puntos caerán en la línea $y = x$, y la pendiente de la regresión será 1 y el intercepto 0. Las predicciones del modelo que caen cerca de esta línea también estarán altamente correlacionadas con las observaciones, pero existen casos en los que un modelo no se ajusta correctamente a las observaciones, pero está altamente correlacionado con las mismas; un ejemplo de esto se muestra en (Haefner, 2005, p. 155).

Con base en las dificultades mencionadas, se propone realizar una prueba con la hipótesis nula de que la pendiente es 1 y el intercepto es 0. Para ello, se hace uso del estadístico F, reflejado en la Ec. (14) (Tedeschi, 2006).

$$F = \frac{n \times a^2 + 2 \times a \times (b-1) \times \sum_{i=1}^n X_i}{2 \times S_{X,Y}^2} + \frac{(b-1)^2 \times \sum_{i=1}^n X_i^2}{2 \times S_{X,Y}^2} \quad (14)$$

Donde a es el intercepto estimado, b es la pendiente estimada y $S_{X,Y}^2$ se calcula con la ecuación Ec. (15), siendo \hat{Y}_i determinado por la Ec. (16).

$$S_{X,Y}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} \quad (15)$$

$$\hat{Y}_i = \bar{Y} + b \times (X_i - \bar{X}) \quad (16)$$

Este estadístico sigue una distribución F con 2 y $n-2$ grados de libertad. Generalmente, si el modelo original es bueno, la prueba falla el rechazo de la hipótesis nula de que la pendiente es 1 y el intercepto es 0 [ver (Haefner, 2005, p. 155) para casos especiales donde la prueba no cumple lo establecido teóricamente]. Por ello, valores pequeños de F indican que un modelo tiene buen ajuste (Haefner, 2005). Cuenta con un valor p asociado, probabilidad que define si se acepta o se rechaza la hipótesis nula.

3.9. Prueba t pareada

Una alternativa a las pruebas de regresión 1:1 (como la prueba F de la sección 3.8), es tratar el modelo y las observaciones como muestras pareadas y utilizar una prueba t pareada para probar la hipótesis nula $H_0 : \mu_x - \mu_y = 0$. En la Ec. (17) se refleja el cálculo del estadístico t .

$$t = \frac{\sum_{i=1}^n (X_i - Y_i)}{\sqrt{\frac{S^2}{n}}} \quad (17)$$

El estadístico t cuenta con un valor p asociado, probabilidad que define si se acepta o se rechaza la hipótesis nula. Igualmente, a partir del valor de t , pueden calcularse intervalos de confianza para la variable en estudio.

En análisis comparativos, (Mayer y Butler, 1993) encontraron que la regresión 1:1 tiene un carácter discriminante más fuerte que una prueba t pareada, esto es, los modelos fueron rechazados utilizando la regresión 1:1, pero aceptados con la prueba t (Haefner, 2005).

3.10. Análisis de perfil

El análisis de perfil es un método multivariante que prueba la hipótesis de que las trayectorias de las observaciones y la salida del modelo son paralelas. Hay dos grandes ventajas de este método sobre otros enfoques: primero, no se hacen supuestos acerca de la naturaleza de las relaciones de varianza o covarianza de las variables, lo que lo hace más general para problemas repetidos; segundo, permite examinar la relación entre las observaciones y el modelo para varias variables de salida simultáneamente (Haefner, 2005).

La hipótesis nula que se prueba es que la diferencia entre las observaciones y el modelo es 0 para todos los instantes de tiempo en comparación. El análisis de perfil utiliza el estadístico Hotelling T^2 , para el cual se dispone de tablas de probabilidad (Timm, 2002).

Para un ejemplo práctico, se recomienda revisar (Haefner, 2005, p. 160-163). En esta prueba se cuenta con n observaciones (réplicas) de q variables en t instantes de tiempo, así como los resultados del modelo para cada una de las q variables en los t instantes de tiempo. Sea una observación Y_{ijk} , donde i es el número de réplica, j es el número de variable y k es el instante de tiempo; y la predicción del modelo X_{jk} , donde j es el número de variable y k es el instante de tiempo. Se aplica la Ec. (18) para $i = 1, 2, \dots, n$; $j = 1, 2, \dots, q$; $k = 1, 2, \dots, t$. Con ello se obtienen q matrices de n filas y t columnas cada una.

$$\delta_{ijk} = Y_{ijk} - X_{jk} \quad (18)$$

Ahora, recorriendo cada una de las q matrices, a partir de la primera columna de cada matriz, con l definiendo el movimiento entre las columnas, que va desde 1 hasta $t - 1$, se hace la resta de cada columna l con la columna siguiente $l + 1$. La Ec. (19) define el proceso.

$$\Delta_{ijl} = \delta_{ijl} - \delta_{ij(l+1)} \quad (19)$$

Ahora se cuenta con q matrices de n filas y $t - 1$ columnas. Estas matrices se unen, para formar una sola, con lo que se tendrá una matriz Δ con $v = q \times (t - 1)$ columnas. Sea m quien define la columna actual $m = 1, 2, \dots, v$. El valor promedio de cada columna, definido por la Ec. (20), se agrupa en un vector fila denotado como $\overline{\Delta}$.

$$\overline{\Delta}_m = \frac{\sum_{i=1}^n \Delta_{im}}{n} \quad (20)$$

Sea S la matriz de covarianza de Δ , el estadístico Hotelling T^2 (Timm, 2002; Haefner, 2005) se define con la Ec. (21), donde $\overline{\Delta}^T$ es la matriz traspuesta de $\overline{\Delta}$.

$$T^2 = n \times \overline{\Delta} \times S \times \overline{\Delta}^T \quad (21)$$

Para determinar el nivel de significancia del estadístico Hotelling T^2 , se utiliza el valor tabulado con los grados de libertad definidos por la Ec. (22).

$$T^\alpha(p, v) = T^\alpha(q \times (t - 1), n - 1) \quad (22)$$

Además del análisis de perfil existen otros métodos multivariantes que pueden ser utilizados para validar modelos de simulación, tales como la prueba de igualdad de matrices de covarianza, donde se evalúa la hipótesis de que las matrices de covarianza sean iguales, haciendo uso del estadístico M de Box (verosimilitud modificado), que a su vez puede ser aproximado por una prueba Chi cuadrado o una prueba F (Timm, 2002; Hair et al., 2009); el análisis de correlación canónico, que hace uso de una prueba Chi cuadrado de Bartlett o una prueba F aproximada, para probar la hipótesis de que las correlaciones canónicas son cero (Afifi y Clark, 1996); y el análisis de varianza multivariante que hace uso de estimadores de verosimilitud para probar igualdad de medias (Peña, 2002) o de estadísticos como la raíz característica más grande de Roy (*Roy's greatest characteristic root*) o el lambda de Wilks, para probar la hipótesis de que no hay diferencia entre grupos (Hair et al., 2009).

3.11. Coeficiente de información de Akaike

Akaike propuso una aproximación no sesgada que puede ser aplicada a datos empíricos basados en la función del logaritmo de máxima verosimilitud. Esta aproximación se conoce como el criterio de información de Akaike (AIC, por sus siglas en inglés, Akaike information criterion) que permite discriminar entre distintos modelos. La Ec. (23) define el cálculo de AIC para cada modelo j (Burnham y Anderson, 2002).

$$AIC_j = -2 \times \ln L_j + 2 \times K \quad (23)$$

Donde L es el estimador de máxima verosimilitud de los parámetros del modelo dadas las observaciones Y , los resultados del modelo X y la ecuación del modelo, donde K es el número de parámetros estimados en el ajuste del modelo a los datos. El valor K incorpora los parámetros conocidos en el modelo más los parámetros de la distribución del error que deben ser especificados (Burnham y Anderson, 2002).

El estimador de máxima verosimilitud utilizado por (Haefner, 2005) se muestra en la Ec. (24).

$$\ln L_j = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{n}{2} \ln(2 \times \pi) - \frac{n}{2} \quad (24)$$

Donde $\hat{\sigma}^2$ se define con la Ec. (25).

$$\hat{\sigma}^2 = \frac{(Y_i - X_i)^2}{n} \quad (25)$$

También existe una variante del AIC, llamada AIC corregido (AICc) que se define con la Ec. (26).

$$AICc_j = AIC_j + \frac{2 \times K \times (K + 1)}{n - K - 1} \quad (26)$$

El AIC es una herramienta útil para seleccionar el mejor modelo de un conjunto particular; sin embargo, si todos los modelos son deficientes, el AIC seguiría seleccionando el mejor entre ellos, pero incluso es modelo puede no ser bueno a nivel general. Entonces, al momento de utilizar el AIC, debe hacerse un esfuerzo para asegurar un conjunto de modelos de utilidad. Para un mayor entendimiento del AIC y su uso, se recomienda revisar (Burnham y Anderson, 2002) y (Haefner, 2005).

3.12. Inferencia bayesiana

El método de máxima verosimilitud ordena cuantitativamente la pertinencia de un conjunto de modelos por su habilidad para ajustarse a los datos, pero no calcula la probabilidad de que esos modelos sean correctos. Un método para calcular esta probabilidad es el Teorema de Bayes (Chick, 2000).

Las estadísticas bayesianas permiten la inclusión de conocimiento previo sobre los modelos en estudio. Para calcular la probabilidad de que un modelo i sea verdadero, a partir de las observaciones y de un conjunto finito de m modelos, puede utilizarse el Teorema de Bayes (Chick, 2000).

La probabilidad de que un modelo i sea verdadero, dadas las observaciones, se define con la Ec. (27) (Chick, 2000).

$$P(M_i/Y) = \frac{P(M_i) P(Y/M_i)}{\sum_{j=1}^m P(M_j) P(Y/M_j)} \quad (27)$$

Donde $P(M_i)$ es la probabilidad a priori que de que el modelo i sea verdadero y la $P(Y/M_i)$ es la probabilidad de que se tengan los valores de las observaciones dado que el modelo M_i es verdadero. Esta última cantidad puede calcularse con el estimador de máxima verosimilitud de Y . Con ello, la Ec. (27) puede reescribirse como la Ec. (28).

$$P(M_i/Y) = \frac{P(M_i) L_i}{\sum_{j=1}^m P(M_j) L_j} \quad (28)$$

Cuando no se tenga la información a priori sobre $P(M)$ se puede dar la misma probabilidad previa a cada uno de los modelos. Para un mayor entendimiento de la inferencia bayesiana y su uso, se recomienda revisar (Chick, 2000).

4. Descripción de la librería

Haciendo uso de las técnicas descritas en la sección 3 se programó la librería **Validacion** en el programa estadístico R.

Esta librería cuenta con 12 funciones y 3 conjuntos de datos. La mayor parte de las funciones, tienen características similares en cuanto al formato de entrada, eliminación de datos faltantes y presentación de los resultados del uso de la técnica correspondiente. A continuación se da el caso general y, posteriormente, al describir cada función, se particulariza en caso de ser necesario.

4.1. Formato general de las funciones

A continuación se describen las características comunes que tienen las funciones de la librería, en cuanto a la entrada, el procesamiento de datos faltantes y la salida.

4.1.1. Formato de entrada

Todas las funciones reciben tanto los datos observados como los resultados del modelo con el tipo de dato siendo vector, matriz o marco de datos (*data frame*).

La aplicación de cada técnica debe hacerse para una variable en particular cada vez. Se recibe primero el conjunto de datos correspondiente a los resultados del modelo para una variable y luego el conjunto de datos observados para la misma variable. Tanto para los resultados del modelo como para las observaciones, puede disponerse bien sea, de varios resultados del modelo para la misma variable, o de varias réplicas para las observaciones de una misma variable; si este es el caso, la función se encarga de realizar el cálculo de los estadísticos de acuerdo con las posibles comparaciones.

Por ejemplo, la entrada de una función puede estar conformada por una matriz $X_{10 \times 2}$ para las observaciones (se tendrían dos réplicas de la misma variable, con 10 datos en distintos instantes de tiempo para cada réplica) y una matriz $Y_{10 \times 3}$ para los resultados del modelo (se tendría la salida de tres modelos para la misma variable, con 10 valores en distintos instantes de tiempo para cada uno), con lo cual debe realizarse el cálculo de los estadísticos de la función en 6 ocasiones.

4.1.2. Eliminación de datos faltantes

En el caso de que los datos ingresados por el usuario estén incompletos (cuando falten algunas observaciones) las posiciones en las cuales deberían estar los elementos faltantes deben catalogarse como *NA* (estructura utilizada por el programa estadístico R para manejar datos ausentes). Las filas en las cuales existan valores *NA* son eliminadas al momento de realizar la comparación entre datos observados en una variable particular y el resultado del modelo correspondiente. Siguiendo con el ejemplo, si en una réplica falta alguno de los 10 datos, la posición que debería ocupar dicho dato debe catalogarse como *NA* y con ello el cálculo del estadístico se realizará con 9 datos, eliminando la pareja de este dato en el modelo.

4.1.3. Presentación de resultados

Cada función calcula uno o más estadísticos al ser utilizada. La presentación de resultados se hace, por pantalla, para cada una de las funciones, mostrando el cálculo de los estadísticos para todas las posibles comparaciones entre las réplicas y los distintos resultados del modelo que se tengan, en forma de matriz, donde las filas representan cada una de las réplicas y las columnas cada uno de los resultados del modelo con que se cuenta. Siguiendo el ejemplo, para cada estadístico, se mostrará una matriz de dimensión 2×3 .

4.2. Descripción de las funciones de la librería

A continuación se describe brevemente cada una de las funciones de la librería, en el mismo orden en que se describieron las técnicas al inicio.

ValMSE Función que se encarga de calcular las medidas del error cuadrático medio. Retorna tres estadísticos, que son el *MSE*, el *RMSE* y el *RMSPE*.

ValTheil Función que se encarga de calcular las medidas de descomposición del error del estadístico de Theil. Retorna tres estadísticos, que son el U^M , el U^S y el U^C .

ValMAE Función que se encarga de calcular las medidas del error absoluto medio. Retorna dos estadísticos, que son el *MAE* y el *MAPE*.

ValWillmott Función que se encarga de calcular el índice de acuerdo de Willmott. Retorna el estadístico *d*.

ValEF Función que se encarga de calcular la eficiencia del modelo de Nash Sutcliffe. Retorna el estadístico *EF*.

ValCorrela Función que se encarga de calcular el coeficiente de correlación de Pearson. Retorna el estadístico *r*.

ValR2 Función que se encarga de calcular el coeficiente de determinación. Retorna el estadístico *R2*.

ValPruebaF Función que se encarga de calcular las medidas del estadístico F. Retorna dos valores que son *F* y *p*.

ValPruebat Función que se encarga de calcular las medidas del estadístico t. Además de *X* y *Y*, también recibe el nivel de significancia del intervalo de confianza, por defecto en 0,95. Retorna cuatro valores que son *t*, *p*, el límite inferior del intervalo de confianza *LimInfIC* y el superior *LimSupIC*.

ValProfile Función que se encarga de realizar un análisis de perfil y que calcula las medidas del estadístico Hotelling T^2 . La estructura de los conjuntos de datos *X* y *Y* varía con respecto a las otras técnicas, haciendo uso de la señalada en la sección 3.10, que incluye en el caso del modelo, *q* variables en *t* instantes de tiempo, y en las observaciones, *q* variables en *t* instantes de tiempo con *n* réplicas. Además, como parámetro de entrada, recibe el valor *t*. Retorna dos valores que son el estadístico Hotelling T^2 y su valor *p*.

ValAIC Función que se encarga de calcular el coeficiente de información de Akaike. Además de *X* y *Y*, también recibe el número de parámetros del modelo, definido como *K*. Retorna dos valores que son el *AIC* y el *AICc*.

ValBayes Función que se encarga de calcular la probabilidad a posteriori de que un modelo sea verdadero, dadas las observaciones. *X* es una estructura que contiene los *m* modelos a evaluar, *Y* tiene el mismo formato de las demás funciones y también recibe las probabilidades a priori de que el modelo sea verdadero, definidas en el parámetro *apriori*. Retorna la probabilidad a posteriori para cada uno de los *m* modelos.

4.3. Descripción de los datos de la librería

Como se dijo, la librería cuenta con tres conjuntos de datos. A continuación se describe brevemente cada uno de ellos.

Lacteos Datos provenientes del estudio realizado por (Márquez y Ramírez, 2009), referentes a la producción de quesos madurados. Contiene las observaciones y los resultados de un modelo para 12 variables. El modelo de simulación fue elaborado utilizando la metodología de Dinámica

de Sistemas (Sterman, 2000) y el software *Vensim PLE* (Ventana, 2010).

Plankton Datos hipotéticos y respuesta del modelo para 6 réplicas en tres instantes de tiempo para la biomasa de fitoplancton y zooplancton, como se muestran en (Haefner, 2005).

Mixteque Datos de un estudio de la dinámica de agua en el suelo durante el ciclo del cultivo de papa en la población de Mixteque en el páramo andino. Se cuenta con tres series de datos para tres posiciones topográficas diferentes (ladera baja, media y alta). Estos datos fueron recolectados por (Díaz, 2007). En cada una de las laderas se ejecutaron tres modelos distintos que corresponden a diferentes maneras de modelizar el proceso de la evapotranspiración, desde un modelo simple (Modelo 1) hasta un modelo de mayor realismo (Modelo 3) (Quintero et al., 2010).

5. Uso de la librería

En esta sección se muestran dos ejemplos, basados en los conjuntos de datos que tiene la librería, de cómo hacer uso de las funciones de la misma.

5.1. Ejemplo 1: *Lacteos*

Como se dijo en la sección 4.3, el marco de datos *Lacteos* tiene 12 variables, cada una con los resultados de la simulación y con los datos observados. En este caso, utilizaremos la variable *Inventario en Gruta Bacteriana*, para ilustrar el uso de la librería. La figura 1 muestra el comportamiento de dicha variable en las observaciones y en el modelo.

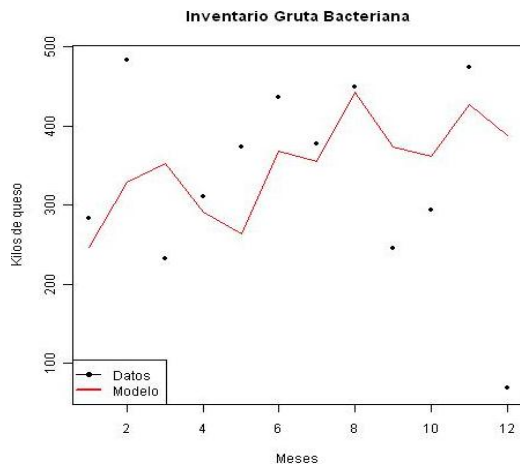


Figura 1. Resultados del modelo y datos observados para la variable *Inventario en Gruta Bacteriana*.

A continuación se muestra cómo hacer uso de una de las funciones de la librería, en particular, *ValMSE*. Se carga la librería y luego se cargan los datos.

```
> library(Validacion)
Loading required package: ICSNP
Loading required package: mvtnorm
Loading required package:
ICS Loading required package: survey

> data(Lacteos)
```

Dentro de los datos debe accederse a las variables que corresponden, haciendo uso de las herramientas que tiene R. Se utiliza la función *ValMSE* para calcular los estadísticos correspondientes.

```
> ValMSE(
  Lacteos$InventarioGrutaBacterianaSimula,
  Lacteos$InventarioGrutaBacterianaDatos)
```

```
$MSE
      Modelo 1
Observación 1 15164.23
```

```
$RMSE
      Modelo 1
Observación 1 123.1431
```

```
$RMSPE
      Modelo 1
Observación 1 36.71339
```

Para el caso del marco de datos *Lacteos* se cuenta con un sólo modelo, por lo cual las funciones *ValProfile*, *ValAIC* y *ValBayes* no pueden ser utilizadas.

La tabla 1 muestra algunos de los estadísticos que pueden calcularse y su interpretación. Como se ha dicho anteriormente, los resultados a veces pueden ser contradictorios, con buenas medidas de ajuste para algunos estadísticos, pero con medidas deficientes para otros.

Tabla 1. Valores de los estadísticos obtenidos para la variable *Inventario en Gruta Bacteriana* y su interpretación.

5.2. Ejemplo 2: *Mixteque*

Para ilustrar el uso del coeficiente de información de Akaike (AIC) usaremos el marco de datos *Mixteque* incluido en la librería y descrito en la sección 4.3. La figura 2 ilustra los resultados de los diferentes modelos y los datos asociados para los tres sitios de este proyecto. Note la dificultad de discriminar el mejor modelo en cada sitio utilizando sólo análisis visual. Con referencia a este trabajo, las preguntas relevantes son:

Medida	Interpretación
$MSE = 15164,23$	El error es muy alto, en el orden del 36 %
$RMSE = 123,14$	
$RMSPe = 36,71$	
$U^M = 0,014$	La media influye poco en el error.
$U^S = 0,230$	La varianza influye poco en el error.
$U^C = 0,756$	La fuente del error puede ser la aleatoriedad
$d = 0,432$	El ajuste del modelo es regular.
$R^2 = 0,018$	El modelo explica en forma deficiente la varianza en los datos.
$F = 0,73$	Se acepta la hipótesis nula de que la pendiente es 1 y el intercepto 0
$p = 0,51$	Se acepta la hipótesis nula de que las medias son iguales.
$t = 0,39$	
$p = 0,70$	

¿Es el modelo más sofisticado (el modelo 3) mejor que el modelo más simple (el modelo 1), tal como se esperaba? ¿Para todos los sitios el mejor modelo es el mismo?

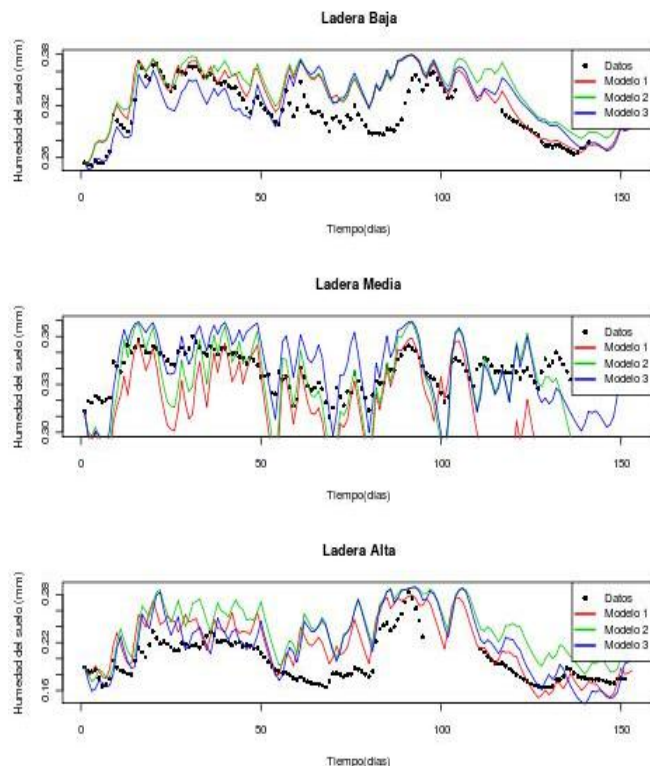


Figura 2. Datos observados, junto con los tres modelos, para cada una de las laderas.

Utilizaremos la librería **Validacion** para responder a estas preguntas. Primero, se carga la librería, tal como en el caso previo:

```
> library(Validacion)
Loading required package: ICSNP
Loading required package: mvtnorm
```

```
Loading required package: ICS
Loading required package: survey
```

Luego, inicializamos la matriz donde se van a almacenar los resultados del modelo *AICmat* y el vector *K* que contiene el número de parámetros de cada modelo:

```
> AICmat<-matrix(0,3,3)
> K<-c(7,12,12)
```

Ahora estamos listos para comenzar los cálculos para la primera posición: ladera baja. Primero, se leen y asignan los resultados de los tres modelos de ladera baja en *ModelosBaja* y los datos correspondientes en *DatosBaja*:

```
>ModelosBaja<-data.frame(
Mixteque$Mod1SimulaBaja,
Mixteque$Mod2SimulaBaja,
Mixteque$Mod3SimulaBaja)
DatosBaja<-data.frame(
Mixteque$Mod1DatosBaja)
```

Ahora, llamamos a la función *ValAIC*:

```
> AICmat[1,]<-ValAIC(ModelosBaja,
DatosBaja,k)$AIC
> AICmat
```

	Modelo 1	Modelo 2	Modelo 3
Ladera 1	-1134.593	-1076.803	-1092.405
Ladera 2	0.000	0.000	0.000
Ladera 3	0.000	0.000	0.000

Para la ladera baja, el menor AIC (mejor modelo) se obtiene para el modelo 1. Ahora, repetimos el mismo proceso para las laderas restantes completando *AICmat*:

```
> AICmat
```

	Modelo 1	Modelo 2	Modelo 3
Ladera 1	-1134.593	-1076.8033	-1092.405
Ladera 2	-996.774	-1199.2870	-1244.462
Ladera 3	-1124.335	-967.3274	-1044.859

Para facilitar la interpretación de los resultados se substrahe el menor AIC de los modelos de una ladera dada a todos los demás. Así, se reescala el resultado y con un AIC de 0 se señala el mejor modelo:

```
> minimo<-apply(AICmat,1,min)
> AICmatG<-sweep(AICmat,1,minimo)
> AICmatG
```

	Modelo 1	Modelo 2	Modelo 3
Ladera 1	0.0000	57.78995	42.18788
Ladera 2	247.6883	45.17514	0.00000
Ladera 3	0.0000	157.00741	79.47540

El modelo 1 es el mejor modelo para ladera 1 (baja) y ladera 3 (alta). Para la ladera 2 (media) el mejor modelo resulta ser el modelo 3.

6. Conclusiones y recomendaciones

La validación es una etapa de gran importancia al momento de elaborar un modelo de simulación. En este trabajo se han discutido numerosas técnicas para validar modelos de simulación continua, desde técnicas sin réplicas, con réplicas, hasta la discriminación de modelos.

La librería en R **Validacion** provee una herramienta de gran utilidad al momento de evaluar qué tan bueno es un modelo de simulación o qué tan preferible es un modelo de simulación en comparación con otro. El hecho de estar implantada en software libre le da un valor agregado.

Esta es la primera versión de la librería, por lo cual hay aún mucho por hacer. Pueden incluirse nuevas técnicas como el estadístico de desigualdad de Theil, Janus, Turing, entre otras. Igualmente debe evaluarse si tanto la forma de la entrada en las funciones como la salida de las mismas es la más apropiada, así como también si es necesario incluir una funcionalidad que permita realizar la comparación de más de una variable a la vez.

Referencias

- Affi AA y Clark V, 1996, *Computer Aided Multivariate Analysis*, Chapman & Hall/CRC, Florida.
- Alvarez Y, López B y Begoña M, 2007, Análisis dinámico de la cadena de franquicia, *Dinámica de Sistemas*, Vol. 3, No. 1, pp. 48-78.
- Barlas Y, 1996, Formal aspects of model validity and validation in system dynamics, *System Dynamics Review*, Vol. 12, No. 3, pp. 183-210.
- Burnham K y Anderson D, 2002, *Model Selection and Multimodel Inference: a practical information-theoretic approach*, Springer, New York.
- Chick S, 2000, *Bayesian Methods For Simulation*, Proceedings of the 2000 Winter Simulation Conference.
- Diaz C, 2007, Balance hídrico y de nutrientes y procesos erosivos en un agroecosistema de papa bajo diferentes posiciones topográficas en los Andes Venezolanos, Avance de tesis Doctoral, Postgrado en Ecología Tropical, Universidad de los Andes, Facultad de Ciencias.
- Ferrer F, Villar JM y Stockle CO, 2000, Evaluación del modelo de simulación CropSyst para maíz de regadío en el valle del ebro, *Investigación Agraria: Serie Producción y Protección Vegetales*, Vol. 15, No. 3, pp. 237-252.
- Haefner J, 2005, *Modeling Biological Systems: Principles and Applications*, Springer, New York.
- Hair J, Black W, Babin B y Anderson R, *Multivariate Data Analysis*, Prentice Hall, Upper Saddle River, N.J.
- Kaur R, Srinivasan R, Mishra K, Dutta D, Prasad D y Bansal G, 2003, Assessment of a SWAT model for soil and water management in India, *Land Use and Water Resources Research*, Vol. 3, No. 4, pp. 1-7.
- Márquez R y Ramírez V, 2009, Un modelo de simulación de la producción de quesos madurados, *Revista Agroalimentaria*, Vol. 15, No. 28, pp. 107-122.
- Martis MS, 2006, Validation of Simulation Based Models: A Theoretical Outlook, *The Electronic Journal of Business Research Methods*, Vol. 4, No. 1, pp. 39-46.
- Mayer DG, y Butler DG, 1993, Statistical validation, *Ecological Modelling*, Vol. 68, No. 1-2, pp. 21-32.
- Orhan H, Eydurán E y Akba Y, 2010, Defining the best covariance structure for sequential variation on live weights of anatolian merinos male lambs, *The journal of animal and plant sciences*, Vol. 20, No. 3, pp. 158-163.
- Pansu M, Sarmiento L, Rujano MA, Ablan M, Acevedo D y Bottner P, 2010, Modeling organic transformations by microorganisms of soils in six contrasting ecosystems: Validation of the MOMOS model, *Global Biogeochem Cycles*, Vol 24, GB1008, doi:10.1029/2009GB003527.
- Peña D, 2002, *Análisis de datos multivariantes*, McGraw-Hill, Madrid.
- Power M, 1993, The predictive validation of ecological and environmental models, *Ecological Modelling*, Vol. 68, No. 1-2, pp. 33-50.
- Quintero C, Ablan M y Sarmiento L, 2010, Ensamblaje de modelos de balance hídrico de agroecosistemas en suelos bien drenados: el rol de la evapotranspiración, *Jornadas de Modelado y Simulación*, 24 al 26 Noviembre del 2010, Mérida, Venezuela.
- Sourisseau S, Bassères A, Périe F y Caquet T, 2008, Calibration, validation and sensitivity analysis of an ecosystem model applied to artificial streams, *Water Research*, Vol. 42, No. 4-5, pp. 1167 - 1181.
- Sterman JD, 1984, Appropriate Summary Statistics for Evaluating the Historic Fit of the System Dynamics Models, *Dynamica*, Vol. 10, Parte II, Winter, pp. 51-66.
- Sterman JD, 2000, *Business dynamics : systems thinking and modeling for a complex world*, McGraw Hill, New York.
- Tedeschi L, 2006, Assessment of the adequacy of mathematical models, *Agricultural Systems Review*, Vol. 89, No. 2-3, pp. 225-247.
- Timm N, 2002, *Applied Multivariate Analysis*, Springer, New York.
- Ventana, 2007, Vensim PLE (personal learning edition), se encuentra en <http://www.vensim.com>, Fecha de consulta: 08 Octubre 2010.
- Wainwright J y Mulligan M, 2002, *Environmental Modelling: Finding Simplicity in Complexity*, John Wiley and Sons, Chichester.

Waller L, Smith D, Childs J y Real L, 2003, Monte Carlo assessments of goodness-of-fit for ecological simulation models, *Ecological Modelling*, Vol. 164, No. 1, pp. 49-63.

Warnock R, Madriz P y Trujillo A, 2005, Determinación de los coeficientes genéticos del modelo cropgro para siete cultivares de caraota en Venezuela, *INCI*, Vol. 30, No. 4, pp. 205-212.

Willmott CJ, 1981, On the validation of models, *Physical Geography*, Vol. 2, pp. 184-194.

Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O'Donnell J y Rowe CM, 1985, Statistics for the evaluation and comparison of models, *Journal of Geophysical Research*, Vol. 90, No. C5, pp. 8995-9005.