

## CAPACIDAD DE PREDICCIÓN DE MODELOS APLICADOS A LA ZNS: HERRAMIENTA INFORMÁTICA PARA LA ADECUADA EVALUACIÓN DE LA BONDAD-DE-AJUSTE CON SIGNIFICACIÓN ESTADÍSTICA

A. Ritter<sup>1</sup>, R. Muñoz-Carpena<sup>2</sup>, C. M. Regalado<sup>3</sup>

<sup>1</sup>Dep. Ing., Producc. y Econom. Agraria. Universidad de La Laguna. Ctra. Geneto, 2; 38200 La Laguna. aritter@ull.es.

<sup>2</sup>Agric. and Bio. Engineering Dept. University of Florida. 287 Frazier Rogers Hall, Gainesville, FL 32611 (EEUU).

<sup>3</sup>Dep. Suelos y Riegos. Instituto Canario Investigaciones Agrarias (ICIA). Apdo. 60 La Laguna, 38200 Tenerife.

**RESUMEN.** La utilización de modelos para la predicción de variables o simulación de procesos en la ZNS es una práctica usual y cada vez más extendida, pero previamente requiere la calibración y verificación de dichos modelos. Tradicionalmente el coeficiente de determinación se ha usado para cuantificar la bondad de ajuste entre los valores observados y los calculados por el modelo, pero las limitaciones de este indicador ha motivado a diversos autores a proponer criterios alternativos. Así, actualmente existen diversos indicadores para la cuantificación de la bondad de ajuste, como por ejemplo: coeficiente de Nash-Sutcliffe, índice de ajuste, raíz del error cuadrático medio, error medio absoluto, pendiente e intersección de la recta de regresión, etc. Sin embargo, ninguno de ellos está libre de limitaciones y su interpretación puede no ser sencilla. Algunos autores se centran sólo en un único indicador, lo que puede resultar en una verificación errónea del modelo. Otros optan por el cálculo de múltiples indicadores, pudiendo ser su interpretación en conjunto complicada y poco práctica. Por otro lado, la significación estadística de los índices calculados es un aspecto generalmente obviado. Con este trabajo se pretende desarrollar un procedimiento práctico para la evaluación de la capacidad de predicción de modelos basado en la representación gráfica del ajuste y en el cálculo de un índice relativo y un estimador del error medio, incorporando significación estadística a los mismos, así como instrucciones para detectar la idoneidad del valor calculado según su sensibilidad a valores atípicos o a sesgo en la magnitudes de los datos. Para facilitar y generalizar la aplicación de este procedimiento se ha elaborado un código de fácil uso y disponible como aplicación informática ejecutable. Mediante una serie de ejemplos se ilustra la utilidad del código para llevar a cabo, y de forma sencilla, la correcta evaluación de la bondad de ajuste.

**ABSTRACT.** The use of models for predicting variables or simulation of processes in the ZNS is a normal practice and increasingly widespread, but first requires calibration and verification of these models. Traditionally, the coefficient of determination was used to quantify the goodness of fit between observed data and values calculated by the model, but the limitations of this indicator have prompted several authors to propose alternative approaches. So today there are several

indicators to quantify the goodness of fit: e.g., the Nash-Sutcliffe coefficient of efficiency, the root mean square error, the mean absolute error, the slope and intercept of the regression line, etc. However, none of them is free of limitations and their interpretation may not be simple. Some authors focus only on a single indicator, what can result in incorrect verification of the model. Others opt for the calculation of multiple indicators, but their together interpretation may be complicated and impractical. On the other hand, the statistical significance of the performance indices is an aspect generally ignored. This work aims to develop a practical procedure for evaluating the predictive ability of models based on the graphic representation of the adjustment and calculation of a relative index and a mean error estimator, incorporating statistical significance to them, and instructions to detect the adequacy of the value calculated according to their sensitivity to outliers or bias in the magnitudes of the data. To facilitate and generalize the application of this procedure, a code was developed and is available as a stand-alone software application. The usefulness of the code for the correct and easy evaluation of model performance is illustrated with a series of example cases.

### 1.- Introducción

La mayoría de los modelos matemáticos empleados para el cálculo de variables o la simulación de procesos en la zona no saturada del suelo han sido previamente evaluados mediante algún tipo de técnica que permite estudiar su capacidad de predicción. Generalmente, el grado de predicción del modelo es juzgado mediante la comparación entre los valores calculados y los correspondientes datos observados o medidos. Para ello, existen numerosos procedimientos propuestos por diversos autores. En varias publicaciones (Willmott, 1981; Loague and Green, 1991; Legates y McCabe, 1999; Santhi et al., 2001; Van Liew et al., 2003; Krause et al., 2005; Moriasi et al. 2007; Jain y Sudheer, 2008) pueden encontrarse descripciones de los diversos índices o indicadores de bondad de ajuste conocidos, así como discusiones rigurosas sobre la idoneidad, ventajas e inconvenientes de cada uno de ellos. Sin embargo, no existe un acuerdo general sobre el procedimiento estándar para la evaluación de la capacidad de predicción de los modelos. Dentro de las diversas sugerencias que se

pueden encontrar en la literatura, destacan tres aspectos en los que parece haber acuerdo: i) la necesidad de un procedimiento estándar para la evaluación de modelos matemáticos; ii) lo inadecuado que supone el cuantificar la bondad de ajuste con indicadores de correlación como es el *coeficiente de determinación* ( $R^2$ ) y iii) lo propuesto por Legates y McCabe (1999) en cuanto a que toda evaluación de la capacidad predictiva de los modelos debe incluir por lo menos un indicador del error, un índice adimensional (o indicador del error relativo) cuantificador de la bondad de ajuste y un procedimiento gráfico que ilustre la comparación entre los valores calculados y los observados.

La interpretación del valor de los diversos indicadores de bondad de ajuste es también controvertida. En primer lugar, porque la información que representa cada índice tiene distinto significado. En segundo lugar, porque no existen criterios estándar sobre los intervalos de valores que indican cuándo la capacidad predictiva es aceptable, buena, muy buena, etc. Como consecuencia de todo esto, suele ser práctica habitual que, en la evaluación de modelos, la conclusión sobre el grado en el que los valores calculados se aproximan a los datos observados se caracteriza por una falta de objetividad (Kirchner et al., 1996).

Otro aspecto importante y nada trivial es la significación estadística de los indicadores de bondad de ajuste. Generalmente, estos indicadores se obtienen a partir del conjunto (es decir, de una muestra) de desviaciones entre los valores calculados por el modelo y los correspondientes valores observados. Por lo tanto, como cualquier variable aleatoria, el valor del indicador calculado a partir de una muestra puede diferir del valor real, existiendo así una distribución de probabilidad subyacente (McCuen et al., 2006). Una de las ventajas del *coeficiente de determinación*,  $R^2$ , es que su distribución estadística es conocida y en consecuencia, es posible evaluar su significación estadística fácilmente. Sin embargo, éste no es el caso de los indicadores de bondad de ajuste más recomendados (Legates y McCabe, 1999), por lo que asignar a éstos significación estadística no es tarea fácil. En concreto, McCuen et al. (2006) proponen ecuaciones aproximadas para la distribución de probabilidad del *coeficiente o índice de eficiencia* de Nash y Sutcliffe (1970), pero no indica claramente en qué se basa para obtener tales ecuaciones. Teniendo en cuenta que los indicadores de bondad de ajuste no tienen por qué tener una distribución de probabilidad concreta, diversos autores (Willmot, 1985; Legates y McCabe, 1999) recomiendan el uso de *método bootstrap* (Efron y Tibshirani, 1993) para la asignación de significación estadística a los indicadores de bondad de ajuste.

El primer objetivo de este trabajo es el de contribuir a la evaluación de la capacidad de predicción de modelos matemáticos desarrollando un procedimiento práctico basado en la combinación de un número reducido de técnicas de evaluación posibles, que incorpore a los indicadores seleccionados, significación estadística utilizando *método bootstrap* y criterios para la caracterización de la bondad de ajuste. El segundo

objetivo busca la generalización de este procedimiento facilitando su aplicación mediante la elaboración de un código fuente y su difusión mediante aplicación informática orientada a usuarios.

## 2.- Materiales y métodos

### 2.1.- Selección de técnicas de evaluación de modelos

Siguiendo la recomendación de Legates y McCabe (1999), para evaluar la capacidad predictiva de los modelos se representará gráficamente los valores observados frente a los calculados ilustrando así el grado con el que los puntos se distribuyen sobre la recta 1:1 o de ajuste perfecto. El error en las predicciones del modelo se cuantificará en términos de las unidades de la variable calculada mediante la raíz del error cuadrático medio (*RMSE*). Este indicador es usado frecuentemente y su definición viene dada por

$$RMSE = \sqrt{\frac{\sum_{i=1}^N [O_i - P_i]^2}{N}} \quad (1)$$

donde  $O_i$  y  $P_i$  representan la muestra (de tamaño  $N$ ) de los valores observados y los calculados por el modelo, respectivamente. Un valor de  $RMSE=0$  indica un ajuste perfecto. Como indicador adimensional de la bondad de ajuste se ha seleccionado el coeficiente de eficiencia de Nash y Sutcliffe (1970). Se trata de un indicador ampliamente utilizado en la evaluación de modelos hidrológicos, caracterizado por su flexibilidad para ser aplicado a diversos tipos de modelos matemáticos (McCuen et al., 2006). Se calcula de la siguiente forma

$$Ceff = 1 - \frac{\sum_{i=1}^N [O_i - P_i]^2}{\sum_{i=1}^N [O_i - \bar{O}]^2} \quad (2)$$

El *Ceff* representa el complementario a la unidad de la relación entre el error cuadrático medio y la varianza de los valores observados. Toma valores  $-\infty \leq Ceff \leq 1$ . Un  $Ceff = 1$  indica un ajuste perfecto, mientras que un  $Ceff \leq 0$  sugiere que el promedio de los valores observados ( $\bar{O}$ ) es mejor predictor que el propio modelo utilizado (y que se pretende evaluar). Según McCuen et al. (2006), el *Ceff* puede ser un indicador de bondad-de-ajuste útil, pero teniendo en cuenta sus limitaciones, es recomendable acompañar su cálculo con un análisis del sesgo en las predicciones del modelo y del posible efecto de valores atípicos en las series de valores  $\{O_i, P_i\}$ .

### 2.2.- Caracterización cualitativa de la bondad de ajuste

No existen criterios globalmente aceptados sobre los intervalos que deben utilizarse para la interpretación cualitativa de los valores de *Ceff*. En este sentido, algunos autores (como indican Legates y McCabe, 1999) ya advierten de que existe una equivocada percepción a la hora de interpretar el *Ceff* como consecuencia de la interpretación fácil que tiene el coeficiente de determinación: Así por ejemplo, mientras que  $R^2=0.5$  sí

indica que el modelo explica el 50% de la varianza en los datos observados, un  $Ceff = 0.5$  tiene un significado totalmente diferente.

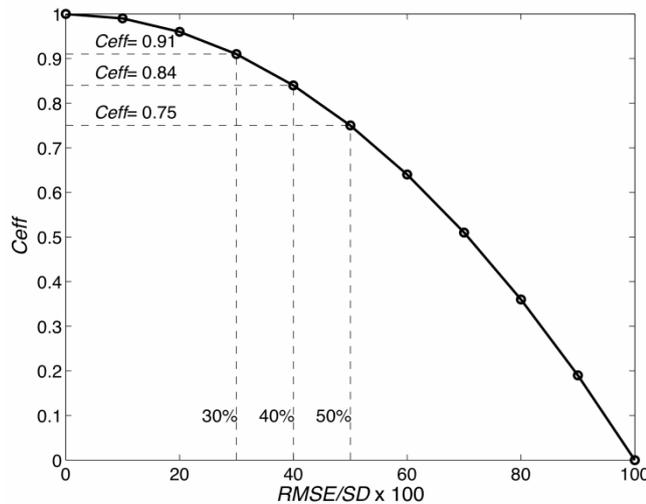
Para interpretar los valores de  $Ceff$  en este trabajo, se parte de la sugerencia de Sign et al. (2004), en la que valores de  $RMSE$  inferiores a la mitad de la desviación estándar de los valores observados ( $SD$ ) pueden considerarse bajos ( $RMSE \leq 0.5SD$ ). De acuerdo con esto, podemos interpretar este criterio como que valores “bajos” son satisfactorios. Teniendo en cuenta que el  $Ceff$  puede definirse con las variables anteriores tal que,

$$Ceff = 1 - \left( \frac{RMSE}{SD} \right)^2 \quad (3)$$

se deduce que valores de  $Ceff \geq 0.75$  son satisfactorios o aceptables. Esta deducción se aplica igualmente para caracterizar la bondad de ajuste dentro del intervalo  $0.75 \leq Ceff < 1$ , quedando así establecido un total de cuatro clases (Tabla 1 y Fig. 1).

**Tabla 1.** Criterios para la evaluación cualitativa de la bondad de ajuste.

Evaluación	$RMSE/SD \cdot 100$	$Ceff$
Muy buena	$\leq 30\%$	$\geq 0.91$
Buena	30 - 40%	0.84 - 0.91
Aceptable	40 - 50%	0.75 - 0.84
No aceptable	$> 50\%$	$< 0.75$



**Fig. 1.** Relación entre el  $Ceff$  y  $RMSE/SD$ .

### 2.3.- Significación estadística de los indicadores seleccionados

Conocida la distribución de probabilidad de los indicadores seleccionados ( $RMSE$  y  $Ceff$ ), es posible asignarle a éstos su significación estadística. Como estas distribuciones son desconocidas, se recurre a técnicas de *bootstrapping* para generarlas. La metodología *bootstrap* debe su nombre y su formulación original a Efron (1979), y, al igual que otros métodos de re-muestreo, básicamente consiste en reutilizar los datos disponibles como si éstos constituyeran una población de la cual extraer repetidas muestras. Es decir, se asume que la matriz de valores

observados y calculados  $\{O_i, P_i\}$  constituyen los valores de toda la población, y se van extrayendo con reemplazo  $M$  muestras (de tamaño  $N$ ), de conjuntos de valores observados y calculados. Para cada re-muestreo se calcula el  $RMSE$  y el  $Ceff$ . De esta forma se construye empíricamente la distribución de probabilidad de ambos indicadores. Este procedimiento requiere gran potencia computacional, dependiendo del número de re-muestréos a realizar (en este trabajo se ha elegido un total de re-muestréos  $M = 2000$ ). A partir de las funciones de distribución de probabilidad, se obtiene la significación estadística de los indicadores mediante el intervalo de confianza de 95%, que se calcula por el método corregido para el sesgo y acelerado, *Bca*, (DiCiccio y Efron, 1996). La fiabilidad de la bondad de ajuste puede abordarse también, y con más detalle, mediante el valor de la probabilidad de que el  $Ceff$  se encuentre dentro de los valores límites que se indican en la Tabla 1.

### 2.4.- Efecto de predicciones sesgadas sobre el $Ceff$

Debido a errores sistemáticos, los cálculos de modelos matemáticos pueden estar sesgados. Tal y como ilustran McCuen et al. (2006) el valor de  $Ceff$  puede verse afectado por predicciones del modelo sesgadas. De acuerdo con estos autores, el  $Ceff$  debe ir acompañado de la cuantificación del sesgo relativo (definido como el ratio entre el promedio de las desviaciones entre valores observados y calculados, y el promedio de los observados ( $\bar{O}$ ). Un sesgo relativo (en valor absoluto) superior al 5%, puede considerarse relevante de tal forma que las predicciones del modelo pueden estar sobre- o infracalculadas. La presencia de valores negativos en las series de observados y calculados puede afectar a la correcta estimación del sesgo. Por tanto, éste se calcula con los valores absolutos de ambas series, es decir con  $\{O_i^*, P_i^*\} = \{\text{abs}(O_i), \text{abs}(P_i)\}$ .

$$Sesgo\ relativo = \frac{\frac{1}{N} \sum_{i=1}^N (P_i^* - O_i^*)}{\bar{O}^*} \cdot 100 \quad (4)$$

### 2.5.- Efecto de valores atípicos presentes en los datos sobre los indicadores de bondad de ajuste

Tanto el  $Ceff$  como el  $RMSE$  se calculan elevando al cuadrado las desviaciones entre valores observados y calculados por el modelo. Esto implica que ambos indicadores pueden verse afectados, es decir presentar un ligero sesgo, cuando hay valores atípicos grandes de las variables observadas y calculadas. El grado en el que el  $RMSE$  es superior al error absoluto medio ( $MAE$ ) es un indicador de la presencia de valores atípicos (Legates y McCabe, 1999). Para detectar estos casos, tanto el  $RMSE$  como el  $MAE$  se expresan como porcentaje del intervalo que abarcan los valores observados  $[\min(O_i) - \max(O_i)]$  (lo cual se indica mediante \*). A partir del análisis de numerosos ejemplos, en este trabajo se propone que una diferencia entre  $RMSE^*$  y  $MAE^*$  de más de 3 puntos porcentuales se considera como advertencia de que tanto el  $RMSE$  y  $Ceff$  pueden estar sensiblemente afectados por la presencia de valores atípicos grandes en la series  $\{O_i,$

$P_{ij}$ .

### 2.6.- Descripción de la aplicación informática

El procedimiento descrito anteriormente para la evaluación de la bondad de ajuste de modelos matemáticos se ha implementado en un código desarrollado en MatLab® 2007b. El código, denominado FITEVAL, está disponible como aplicación informática en <http://webpages.ull.es/users/aritter/software.html> y requiere de un archivo en formato ASCII que debe contener dos vectores o columnas de valores (que deberá preparar el usuario): la primera con los valores observados y la segunda con aquellos calculados por el modelo a evaluar. Nótese que el archivo sólo debe contener dos vectores de valores usando el punto como separador de decimales. Una vez ejecutada la aplicación informática, ésta realiza la evaluación de la bondad de ajuste proporcionando la gráfica de ajuste frente a la línea 1:1, el cálculo del  $Ceff$  y  $RMSE$  con sus correspondientes intervalos de confianza de 95%, la evaluación cualitativa del ajuste, la verificación de la presencia de sesgos o la posible presencia de valores atípicos. Los resultados se almacenan tanto en un archivo pdf como eps, jpg y txt.

### 3.- Resultados y discusión

#### 3.1.- Ejemplos ilustrativos

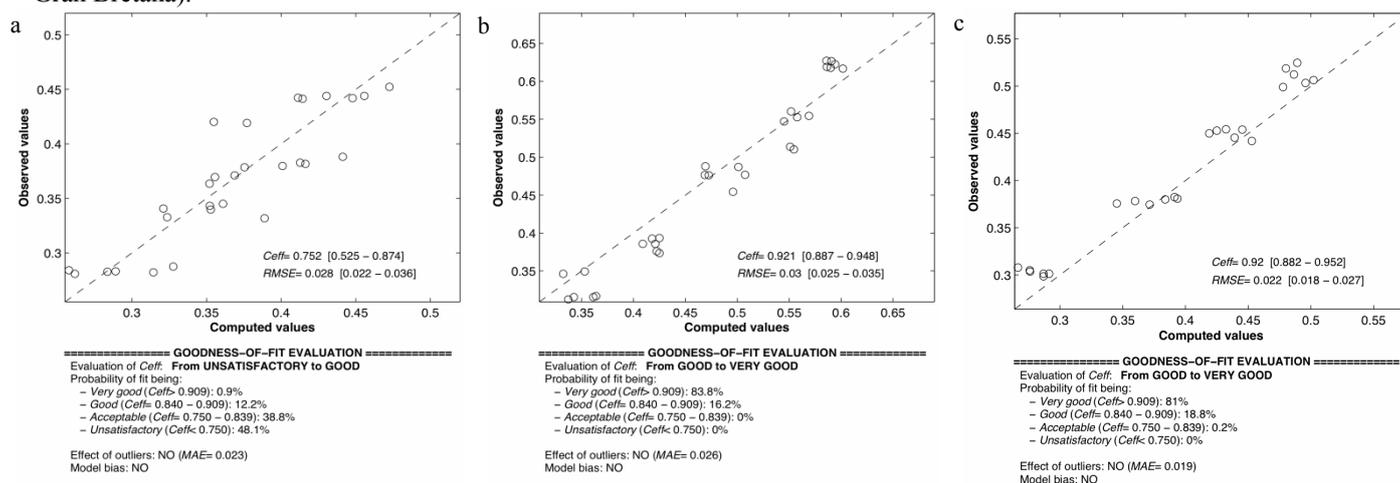
Para ilustrar la utilidad del procedimiento y los resultados que proporciona el programa informático, se han usado los datos presentados por Ritter y Regalado (2007) sobre la calibración de modelos matemáticos que permiten determinar el contenido de agua del suelo ( $\theta$ ) y la conductividad eléctrica de la solución del mismo ( $\sigma_w$ ) a partir de las lecturas que proporciona el sensor dieléctrico comercial WET Sensor (Delta-T Devices Ltd., Burwell, Gran Bretaña).

En concreto, en ese trabajo se observa que las lecturas del sensor en tres suelos forestales de origen volcánico y en una arena conducen a estimaciones erróneas de  $\theta$  y  $\sigma_w$  como consecuencia de la baja frecuencia de trabajo que utiliza el sensor y de la inadecuación de los modelos que proporciona el fabricante. Tras el calibrado de las lecturas del sensor usando otros modelos alternativos, propuestos por Evett et al. (2005) y por Muñoz-Carpena et al. (2005), la determinación de  $\theta$  y  $\sigma_w$ , mejora considerablemente.

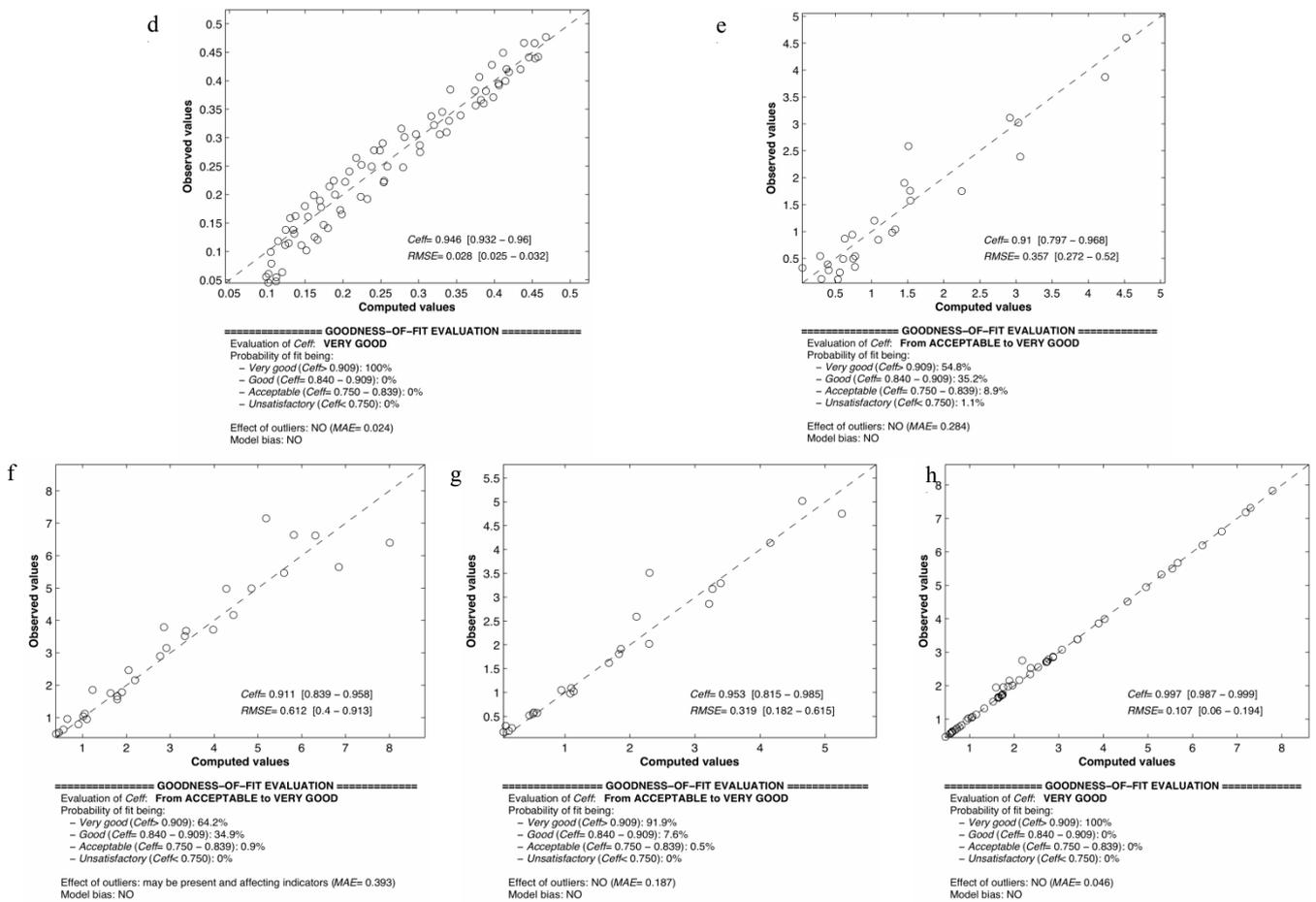
La Fig. 2 recoge los resultados que proporciona la aplicación informática para cada caso evaluado. Estos ejemplos ilustran diversas situaciones en las que teniendo en cuenta la significación estadística del  $Ceff$ , la evaluación de la bondad de ajuste no puede englobarse en una única clase (*No acceptable, Aceptable, Buena o Muy buena*).

Nótese como la indicación del valor de la probabilidad de que el ajuste pueda considerarse dentro de una clase determinada, proporciona información interesante que enriquece la evaluación de la capacidad predictiva de los modelos evaluados. En estos ejemplos presentados las predicciones de las variables no se encuentran sesgadas y sólo se advierte de un posible efecto en el  $RMSE$  y  $Ceff$  por la presencia de valores atípicos en el caso de la Fig. 2f.

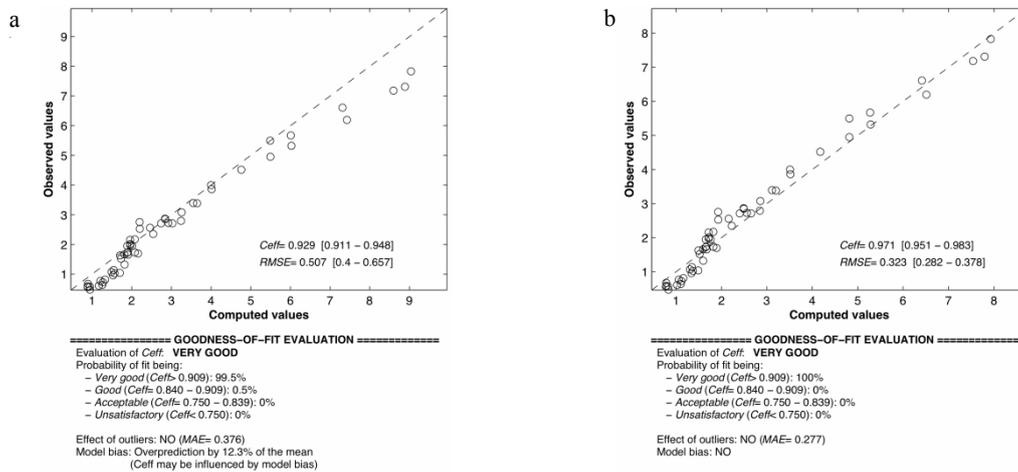
Para ilustrar el efecto del sesgo, se incluye un ejemplo en el cuál se obtiene la  $\sigma_w$  de la arena del caso anterior (Ritter y Regalado, 2007) mediante las lecturas que proporciona el WET Sensor, pero usando esta vez la ecuación de calibración que proporciona el fabricante. Así, de acuerdo con la Fig. 3a, aunque en este caso, la predicción de  $\sigma_w$  es *Muy buena* ( $Ceff=0.929$ ), los valores muestran un sesgo relativo del 12.3%. Si los valores calculados de  $\sigma_w$  se reducen en un 12.3% y se comparan con los observados, entonces puede comprobarse (Fig. 3b) como la bondad de ajuste mejora ( $Ceff=0.971$ ).



**Fig. 2.** Evaluación de la bondad de ajuste del modelo para estimar  $\theta$  ( $m^3/m^3$ ) en tres tipos de suelos volcánicos (a, b y c) y en una arena gruesa (d). Evaluación de la bondad de ajuste del modelo para estimar  $\sigma_w$  (DS/m) en tres tipos de suelos volcánicos (e, f y g) y en una arena gruesa (h).



**Fig. 2 (continuación).** Evaluación de la bondad de ajuste del modelo para estimar  $\theta$  ( $\text{m}^3/\text{m}^3$ ) en tres tipos de suelos volcánicos (a, b y c) y en una arena gruesa (d). Evaluación de la bondad de ajuste del modelo para estimar  $\sigma_w$  (dS/m) en tres tipos de suelos volcánicos (e, f y g) y en una arena gruesa (h).



**Fig. 3.** Efecto del sesgo en la variable calculada sobre la evaluación de la bondad de ajuste del modelo para estimar  $\sigma_w$  (dS/m) en una arena gruesa.

El efecto que un valor atípico alto puede tener sobre el  $RMSE$  y el  $Ceff$  puede observarse con el caso de la Fig. 2f. Eliminando el punto (5.18, 7.15) que puede ser un valor atípico, tanto el  $Ceff$  como el  $RMSE$  mejoran ( $Ceff=0.933$ ;  $RMSE=0.504$  frente a  $Ceff=0.911$ ;  $RMSE=0.612$ ) y la probabilidad de que el ajuste sea *Muy bueno*, se incrementa desde un 64.2% a un 87.5% (Fig. 4a).

Este efecto de la presencia de valores atípicos se ilustra también con el caso de la Fig. 4b y 4c donde se observa

cómo al eliminar de las series de datos correspondiente a la Fig. 4b el valor atípico (0.96, 2.73), la mejora en los indicadores es mayor (Fig. 4c). Nótese como la probabilidad de que el ajuste sea *No aceptable* se reduce desde un 37.4% hasta un 14.3%. Igualmente las probabilidades de que el ajuste sea *Bueno* o *Muy bueno* aumentan (desde 19.8% a 40.3% y desde 7.7% a 20.9%, respectivamente).

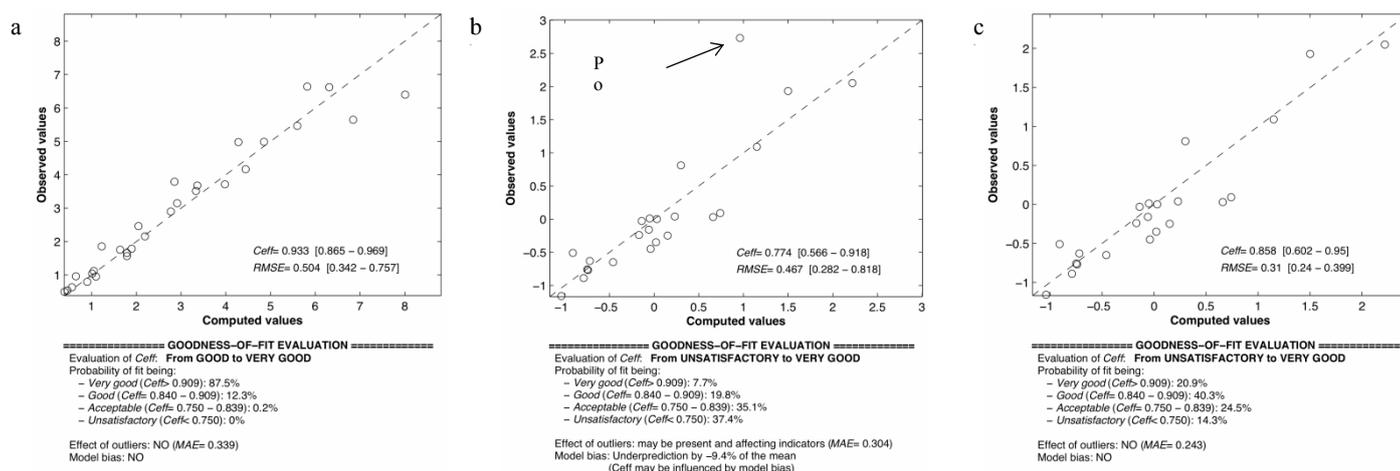


Fig. 4. Efecto de la presencia de valores atípicos: a) estimación de  $\sigma_w$  (dS/m) en el suelo volcánico c) eliminando un posible valor atípico; Ejemplo adicional de series de datos con y sin valor atípico (b y c, respectivamente).

#### 4.- Conclusiones

Se ha desarrollado con éxito un procedimiento práctico para la evaluación de la capacidad predictiva de modelos matemáticos empleados para el cálculo de variables o la simulación de procesos en la zona no saturada del suelo. Éste emplea técnicas gráficas y el cálculo de indicadores de bondad de ajuste (*Ceff* y *RMSE*). Hay que destacar que se ha abordado la asignación de significación estadística al *Ceff* y *RMSE* obteniendo empíricamente su distribución de probabilidad mediante *bootstrapping*, así como la caracterización cualitativa de la capacidad de predicción del modelo. La implementación de este procedimiento en un programa informático permite aplicarlo de forma rápida y sencilla, proporcionando información que facilita la interpretación de los resultados. En futuros trabajos se plantea implementar el código desarrollado como complemento (add-in) de Microsoft Excel.

*Agradecimientos:* Este trabajo se ha realizado en el marco del proyecto de investigación RTA2009-161 financiado por el Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Ministerio de Ciencia e Innovación.

#### 5.- Referencias

DiCiccio, T.J. y Efron, B. 1996. Bootstrap confidence intervals. *Statistical Science* 11, 189-228.  
 Efron, B. 1979. Bootstrap methods. *Annals of statistics* 7, 1-26.  
 Efron, B. y Tibshirani, R.J. 1993. An introduction to the Bootstrap. N.Y.: Chapman & Hall.  
 Evett, S.R., Tolk, J.A. y Howell, T.A. 2005. TDR laboratory calibration in travel time, bulk electrical conductivity, and effective frequency. *Vadose Zone J.* 4, 1020-1029.  
 Jain, S.K. y Sudheer, K.P. 2008. Fitting of hydrologic models: A close look at the Nash-Sutcliffe Index. *J. Hydrol. Eng.* 13, 981-986.  
 Kirchner, J.W., Hooper, R.P., Kendall, C., Neal, C. y Leavesley, G. 1996. Testing and validating environmental models. *Science of the Total Environment* 183, 33-47.

Krause P., Boyle, D.P. y Bäse, F. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89-97.  
 Legates, D.R. y McCabe, Jr., G.J. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Res.* 35, 233-241.  
 Loague, K. y Green, R.E. 1991. Statistical and graphical methods for evaluating solute transport models: Overview and application. *J. Contaminant Hydrol.* 7, 51-73.  
 McCuen, R.H., Knight, Z. y Cutter, A.G. 2006. Evaluation of the Nash-Sutcliffe Efficiency Index. *J. Hydrol. Eng.* 11, 597-602.  
 Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. y Veith, L. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50, 885-900.  
 Muñoz-Carpena, R., Regalado, C.M., Ritter, A., Álvarez-Benedí, J. y Socorro, A.R. 2005. TDR estimation of saline solutes concentration in a volcanic soil. *Geoderma* 124, 399-413.  
 Nash, J.E. y Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, part I: A discussion of principles. *J. Hydrol.* 10, 282-290.  
 Ritter, A. y Regalado, C.M. 2007. Eficacia de un sensor de capacitancia para medir simultáneamente salinidad y contenido hídrico. pp 145-151. En: J.V. Giráldez Cervera y F.J. Jiménez Hornero (eds.). *Estudios de la Zona No Saturada del Suelo*, Vol. VIII. Universidad de Córdoba.  
 Santhi, C., Arnold, J.G., Williams, J.R., Dugas, W.A., Srinivasan, R. y Hauck, L.M. 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. Amer. Water Resources Assoc.* 37, 1169-1188.  
 Singh, J., Knapp, H. V. y Demissie, M. 2004. Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. *ISWS CR 2004-08*. Champaign, Ill.: Illinois State Water Survey. Available at: www.sws.uiuc.edu/pubdoc/CR/ISWSCR2004-08.pdf. Consultado el 15 de marzo 2011.  
 Van Liew, M.W., Arnold, J.G. y Garbrecht, J.D. 2003. Hydrologic simulation on agricultural watersheds: Choosing between two models. *Trans. ASAE* 46, 1539-1551.  
 Willmott, C.J. 1981. On the validation of models. *Physical Geographer* 2, 184-194.  
 Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddesma, J.J., Klink, K.M., Legates, D.R., O'Donnell J. y Rowe, C.M. 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research* 90, 8995-9005.