ECOLOGICAL MODELLING 216 (2008) 316–322

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel

How to evaluate models: Observed vs. predicted or predicted vs. observed?

Gervasio Piñeiro[a,*], *Susana Perelman*[b], *Juan P. Guerschman*[b,1], *José M. Paruelo*[a]

[a] *IFEVA, Cátedra de Ecología, Laboratorio de Análisis Regional y Teledetección, Facultad de Agronomía, Universidad de Buenos Aires/CONICET, San Martin 4453, C1417DSE Capital Federal, Argentina*
[b] *IFEVA, Cátedra de Métodos Cuantitativos Aplicados, Facultad de Agronomía, Universidad de Buenos Aires/CONICET, Argentina*

ARTICLE INFO

Article history:
Received 2 July 2007
Received in revised form
24 April 2008
Accepted 19 May 2008
Published on line 2 July 2008

Keywords:
Measured values
Simulated values
Regression
Slope
Intercept
Linear models
Regression coefficient
Goodness-of-fit
1:1 line

ABSTRACT

A common and simple approach to evaluate models is to regress predicted vs. observed values (or vice versa) and compare slope and intercept parameters against the 1:1 line. However, based on a review of the literature it seems to be no consensus on which variable (predicted or observed) should be placed in each axis. Although some researchers think that it is identical, probably because r^2 is the same for both regressions, the intercept and the slope of each regression differ and, in turn, may change the result of the model evaluation. We present mathematical evidence showing that the regression of predicted (in the *y*-axis) vs. observed data (in the *x*-axis) (PO) to evaluate models is incorrect and should lead to an erroneous estimate of the slope and intercept. In other words, a spurious effect is added to the regression parameters when regressing PO values and comparing them against the 1:1 line. Observed (in the *y*-axis) vs. predicted (in the *x*-axis) (OP) regressions should be used instead. We also show in an example from the literature that both approaches produce significantly different results that may change the conclusions of the model evaluation.

1. Introduction

Testing model predictions is a critical step in science. Scatter plots of predicted vs. observed (or vice versa) values is one of the most common alternatives to evaluate model predictions (i.e. see articles starting on pages 1081, 1124 and 1346 in *Ecology* vol. 86, No. 5, 2005). However, it is unclear if models should be evaluated by regressing predicted values in the ordinates (*y*-axis) vs. observed values in the abscissas (*x*-axis) (PO), or by regressing observed values in the ordinates vs. predicted values in the abscissas (OP). Although the r^2 of both regressions is the same, it can be easily shown that the slope and the intercept of these two regressions (PO and OP) differ. The analysis of the coefficient of determination (r^2), the slope and the intercept of the line fitted to the data provides elements for judging and building confidence on model performance. While r^2 shows the proportion of the total variance explained by the regression model (and also how much of the linear variation in the observed values is explained by the variation in the predicted values), the slope and intercept describe the consistency and the model bias, respectively (Smith and Rose, 1995; Mesple et al., 1996). It is interesting to note that even in widely

* *Corresponding author.*
E-mail address: pineiro@ifeva.edu.ar (G. Piñeiro).
[1] Current address: CSIRO Land and Water-GPO Box 1666, Canberra, ACT 2601, Australia.
0304-3800/$ – see front matter
doi:10.1016/j.ecolmodel.2008.05.006

Table 1 – Number of papers published in *Ecological Modelling* in 2000 using different types of model evaluation

	Total papers	Papers that evaluate models	Papers plotting predicted and observed data	Using visual graph interpretation, r^2 or other method	Estimating intercept or slope
Predicted vs. observed (PO)			11	6	5
Observed vs. predicted (OP)			6	2	4
Both regressions			2	1	1
Total	204	61	19	9	10

used software packages (like Statistica or Math Lab), default scatter plots available to evaluate models differ in the variable plotted in the x-axis. Is it important to care on what to put in each axis? Do scientists care?

Quantitative models are a common tool in ecology as shown by (Lauenroth et al., 2003), who found that 15% of the papers published in *Ecology* and 23% of the ones published in *Ecological Application* contained some dynamic quantitative modeling. In order to analyze how ecologists evaluate their quantitative models we reviewed all articles published in the journal that more focuses on quantitative modeling (*Ecological Modelling*): For year 2000 we selected the papers that used either PO or OP regressions to evaluate their models. The papers were considered in the analysis if a model was evaluated. Articles that evaluated a model using the regression of predicted vs. observed (or vice versa), were separated in two categories: those that considered slope or intercept in the analysis and those that used only visual interpretation of the data or r^2. We found 61 papers out of 204 published during 2000 in *Ecological Modelling* that evaluated models and 19 of them did it by regressing either PO or OP data (Table 1). Papers that did not use regression techniques evaluated model predictions mostly based on plotting observed and predicted values both in the y-axis, and time (or some other variable) in the x-axis. Thus, most papers did not present a formal evaluation of their models at the level of the prediction although they have data to do so. Almost half of the 19 papers that evaluated a model using regression techniques performed just a visual interpretation of the data or used only the r^2. The other half estimated the regression coefficients and compared them to the 1:1 line. Of these 19 papers, 58% regressed PO data, 32% regressed OP values and 10% did both analyses. The survey showed that regression of simulated and measured data is a frequently used technique to evaluate models, but there is no consensus on which variable should be placed in each axis.

Several methods have been suggested for evaluating model predictions, aimed in general to quantify the relative contribution of different error sources to the unexplained variance (Wallach and Goffinet, 1989; Smith and Rose, 1995; van Tongeren, 1995; Mesple et al., 1996; Monte et al., 1996; Loehle, 1997; Mitchell, 1997; Kobayashi and Salam, 2000; Gauch et al., 2003; Knightes and Cyterski, 2005). The use of regressions techniques for model evaluation has been questioned by some authors (Mitchell, 1997; Kobayashi and Salam, 2000). However, the scatter plot of predicted and observed values or vice versa is still the most frequently used approach (as shown in our survey). Thus, it seems that plotting the data and showing the dispersion of the values is important for scientists (an often undervalued issue), that probably promote authors to use graphic plots of predicted and observed data. However, we think that this approach should be complemented (not substituted) by other statistics that add important information for model evaluation as suggested further on.

In this article we show that there are conceptual and practical differences between regressing predicted in the y-axis vs. observed in the x-axis (PO) or, conversely, observed vs. predicted (OP) values to evaluate models. We argue that the latter (OP) is the correct procedure to formulate the comparison. Our approach includes both an empirical and algebraic demonstration. We also use a real example taken from the literature to further show that using a PO regression can lead to incorrect conclusions about the performance of the model being analyzed, and suggest other statistics to complement model evaluation.

2. Materials and methods

Since the slope and intercept derived from regressing PO or OP values differ, we investigated which of the two regressions should be used to evaluate model predictions. We constructed a X vector with continuous values ranging from 1 to 60.

$$X = \{1, 2, 3, \ldots 60\} \tag{1}$$

Y vectors were constructed to have either a linear, quadratic or logarithmic relationship with the X vector

$$Y_{\text{Lin}} = X + \varepsilon \tag{2}$$

$$Y_{\text{Quad}} = -0.05X^2 + 3X + \varepsilon \tag{3}$$

$$Y_{\text{Ln}} = 30\ \text{Ln}(X) + \varepsilon \tag{4}$$

where ε is a random error with normal distribution (mean = 0, Stdev = 15). Both vectors X and Y are named as observed X and observed Y, since they mimic data normally observed or measured in the experiments. Using regression analyses we adjusted a linear, quadratic or logarithmic model for each Y vector (see examples in Fig. 1a–c, respectively):

$$\hat{Y}_{\text{Lin}} = aX + b \tag{5}$$

$$\hat{Y}_{\text{Quad}} = aX^2 + bX + c \tag{6}$$

$$\hat{Y}_{\text{Ln}} = a\ \text{Ln}(X) + b \tag{7}$$

Eqs. (5)–(7) allowed us to generate a vector of predicted values $\hat{Y}$. Each $\hat{Y}$ vector contains 60 $\hat{y}_i$ predicted values for each x_i

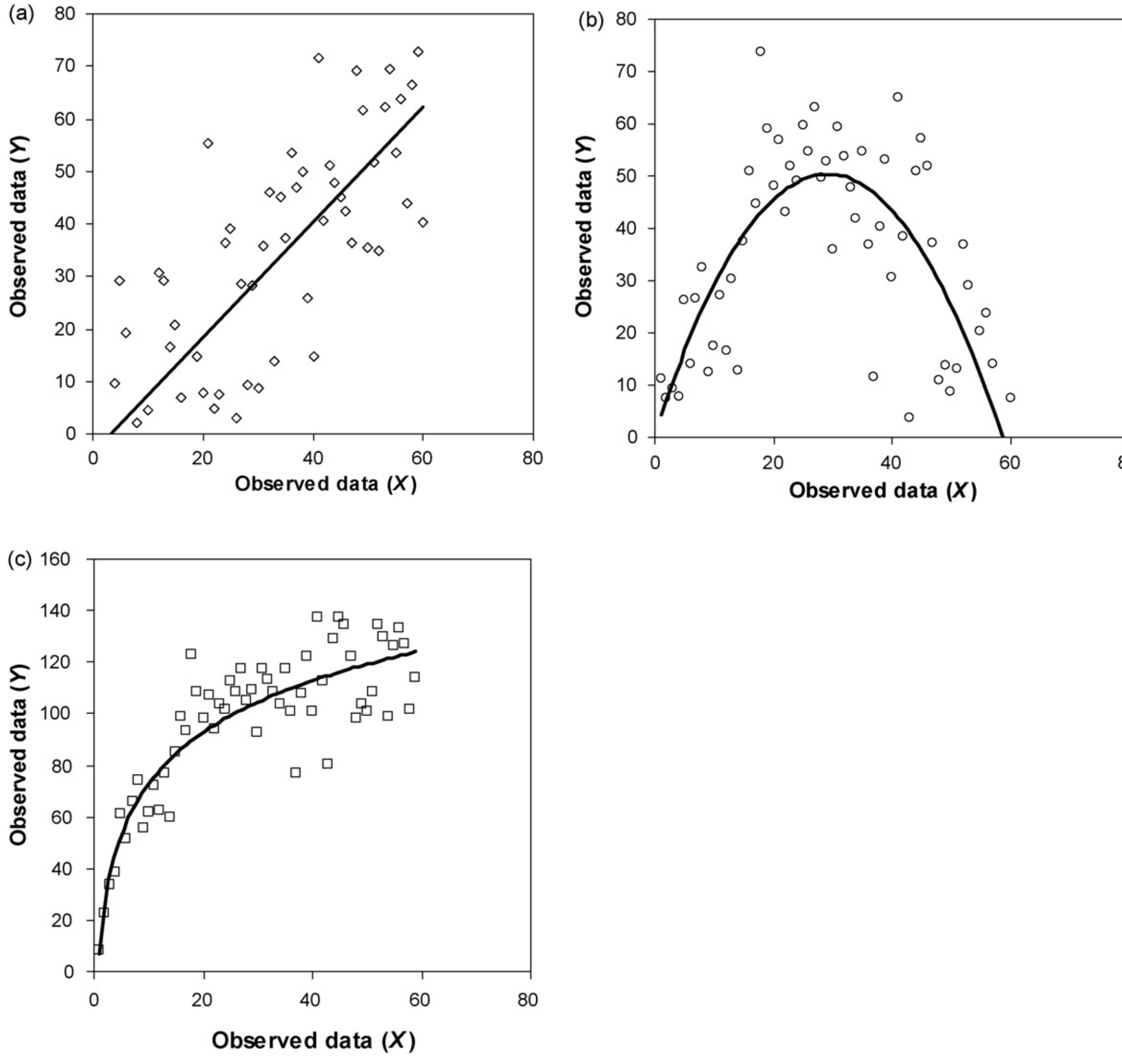

Fig. 1 – Examples of regressions generated using X and Y vectors. (a) Linear $Y_{Lin} = X + \varepsilon$, (b) quadratic $Y_{Quad} = -0.05X^2 + 3X + \varepsilon$, and (c) logarithmic $Y_{Ln} = 30\,\mathrm{Ln}(X) + \varepsilon$. Y vectors have a random error with normal distribution, mean = 0 and Std = 15.

value of the X vectors. We repeated this procedure 100 times for each type of model obtaining 300 pairs of Y and $\hat{Y}$ vectors, each one with 60 elements. We evaluated model predictions ($\hat{Y}$) by plotting and calculating the linear regression equations of each paired Y (observed values) and $\hat{Y}$ (predicted values) vectors, for either PO ($\hat{y} = b_1 y + a_1$) and OP ($y = b_2\hat{y} + a_2$) values. We then plotted the distribution of slope and intercept parameters achieved in the 100 simulations for the linear models. Since the same data were used to construct the model and to evaluate model predictions, we expect no bias in the slope nor the intercept of the regression between Y and $\hat{Y}$. Thus, b_1 and b_2 should be 1, and a_1 and a_2 should be 0.

In a second step, we further demonstrate analytically our empirical findings using basic algebra. In this mathematical approach we illustrate the relationship between a_1 and a_2, and between b_1 and b_2. We also relate both slopes to r^2.

Finally, we took an example from the literature and analyzed the effects of evaluating model predictions by regressing either PO or OP values. The paper by (White et al., 2000), presented regressions of predicted (in the ordinates) vs. observed (in the abscissas) (PO) values and had a table with the data used, so it was easy to generate the opposite regressions of OP values. We compared the regression parameters of both approaches and tested the hypothesis of slope = 1 and intercept = 0 to assess statistically the significance of regression parameters. This test can be performed easily with statistical computer packages with the models:

$$y_i - \hat{y}_i = a_1 + b_1 y_i + \varepsilon_i \quad (8)$$

$$\hat{y}_i - y_i = a_2 + b_2 \hat{y}_i + \varepsilon_i \quad (9)$$

The significance of the regression parameters of these models corresponds to the tests: b_1, $b_2 = 1$ and a_1, $a_2 = 0$, for either regression of PO (Eq. (8)) or OP values (Eq. (9)). If the null hypothesis for the slope is rejected the conclusion is that model predictions have no consistency with observed values. If this hypothesis is not rejected but the hypothesis for the intercept is, then the model is biased. If both null hypotheses are not rejected, then disagreement between model predictions and observed data is due entirely to the unexplained variance.

We also calculated for Whites et al.'s, data, Theil's partial inequality coefficients (U_{bias}, U_{slope} and U_{error}), which separate total error of the predictions (the squared sum of the predictive error), into different components and complement the assessment of model performance made with the regression (Smith and Rose, 1995; Paruelo et al., 1998). Theil's coefficients partition the variance of observed values not explained by the predicted values (called the squared sum of the predictive error), being: $U_{bias,}$ the proportion associated with mean differences between observed and predicted values, U_{slope} the proportion associated with the slope of the fitted model and the 1:1 line, and U_{error} the proportion associated with the unexplained variance (see Paruelo et al., 1998, for a simple formula to calculate Theil's coefficients). Additionally, we estimated for White et al.'s data the root mean squared deviation (RMSD) as

$$\mathrm{RMSD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{10}$$

which represents the mean deviation of predicted values with respect to the observed ones, in the same units as the model variable under evaluation (Kobayashi and Salam, 2000; Gauch et al., 2003).

3. Results and discussion

Since model predictions were tested using the same data used in their construction (the same Y vector), commonly called an evaluation of the calibration procedure, the regression of PO values is expected to have no bias from the 1:1 line. As a consequence, we expected that the parameters of the regression $\hat{y} = b_1 y + a_1$, be: $b_1 = 1$ and $a_1 = 0$. The dispersion of the data is a consequence of the random error introduced in the process of model generation. However, as shown in Fig. 2a, when regressing PO data the slope b_1 was always lower than 1 (and the most frequent value was similar to r^2) and the intercept a_1 was always higher than 0. Only when the regression was performed with OP data $y = b_2\hat{y} + a_2$, then $b_2 = 1$ and $a_2 = 0$ (Fig. 2b). This empirical analysis suggests that regressions to evaluate models should be performed placing observed values in the ordinates and predicted values in the abscissas (OP). The same results were obtained for the quadratic and logarithmic models (data not shown).

These results can be also demonstrated algebraically. The slope of the regression of PO values (b_1) can be calculated as

$$b_1 = \frac{S\hat{y}y}{Syy} \tag{11}$$

where $S\hat{y}y$ is the sum of the cross products of centered predicted and observed values and Syy is the sum of squares of centered observed values. The slope of the regression of OP values (b_2) can be calculated as

$$b_2 = \frac{S\hat{y}y}{S\hat{y}\hat{y}} \tag{12}$$

where $S\hat{y}\hat{y}$ is the sum of squares of centered predicted values. The coefficient of determination of the regression of PO values (r_1^2) is then:

$$r_1^2 = \frac{(S\hat{y}y)^2}{Syy\ S\hat{y}\hat{y}} \tag{13}$$

and the coefficient of determination of the regression of OP values (r_2^2) is

$$r_2^2 = \frac{(S\hat{y}y)^2}{S\hat{y}\hat{y}\ Syy} \tag{14}$$

thus, the two coefficients of determination are equal, and also related to b_1 and b_2 as

$$r_1^2 = r_2^2 = b_1 b_2 \tag{15}$$

Considering once more that our vector $\hat{Y}$ was estimated from the vector X and Eqs. (2), (3) or (4), and that because of that the relation between Y and $\hat{Y}$ is exact, with no distortion or bias, then each observed value can be defined as prediction plus a random error ($y_i = \hat{y}_i + \varepsilon_i$). Consequently:

$$\begin{aligned} S\hat{y}y &= \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y}) = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})y_i = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(\hat{y}_i + \varepsilon_i) \\ &= \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})\hat{y}_i + \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})\varepsilon_i = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 = S\hat{y}\hat{y} \end{aligned} \tag{16}$$

We demonstrated that $S\hat{y}y = S\hat{y}\hat{y}$ if $y_i = \hat{y}_i + \varepsilon_i$. Thus, we confirm algebraically that for our experiment $b_2 = 1$ and that $b_1 = r^2$, founded on Eqs. (11), (12) and (15). Consequently, b_1 will be always smaller than 1 when any $\varepsilon_i \neq 0$. Additionally, since:

$$a_1 = \bar{\hat{y}} - b_1\bar{y}, \qquad a_2 = \bar{y} - b_2\bar{\hat{y}} \tag{16}$$

and because $b_1 = r^2$, then $a_1 = 1 - r^2$ (always >0) when observed and predicted values have the same mean (model predictions are not biased). In identical conditions $a_2 = 0$, because $b_2 = 1$. However, in real comparisons between observed and predicted values, b_1 will approximate r^2 when b_2 approximates to 1.

The theoretical evidence presented before shows that the proper slope and y-intercept to compare observed and predicted values must be calculated only by regressing OP data. A spurious estimate will be obtained by regressing PO values. Wrong conclusions on model performance will be drawn in the latter case. Eq. (15) also revealed that the differences between the two slopes calculated will increase as r^2 decreases. In addition, in Eq. (8) the error term represents the variation in the predicted values and residuals are independent of the observed values. In the second Eq. (9) the residuals are independent of the predicted values which are what we want to evaluate. This line of reasoning adds additional theoretical basis for using Eq. (9) of OP values instead of Eq. (8) of PO values in model evaluation.

The reanalysis of the data presented by (White et al., 2000), showed with real data that slope and intercept vary when regressing OP values instead of PO values, changing the results

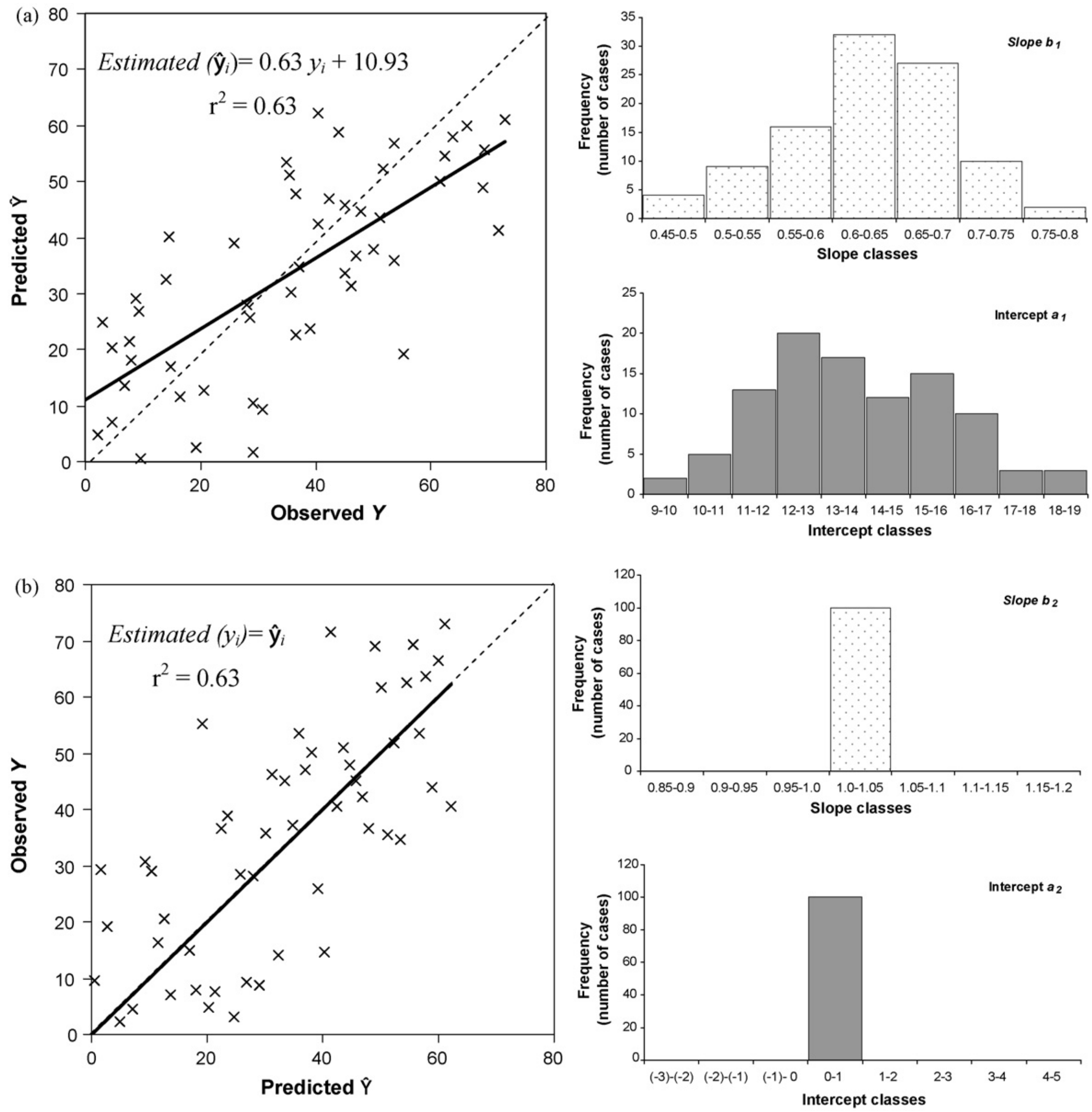

Fig. 2 – Predicted vs. observed (a) (PO) and observed vs. predicted (b) (OP) regression scatter plots derived from the linear model presented in Fig. 1a. Regression equations are shown in the graphs. Small graphs show the distribution of slope and intercept estimates obtained from regressing 100 paired Y and Ŷ vectors either as PO (a) and OP (b).

of the analysis. In their paper, White and collaborators used a simple physiological model for estimating biomass accumulation in New Zealand vegetation. Model predictions were compared with observed values collected in several studies. The slope of the regression of PO values of total biomass presented by the authors in their Fig. 3, differed by 0.40 units from 1, while the slope of regressing OP values differed by only 0.27 units (almost half) (Fig. 3a and b). Looking at the graphs we can state that the authors probably thought that their model overestimated observed data at low values and underestimated it at high values, thus the slope of the regression was significantly different from 1.

Opposite results are obtained when testing the significance of the intercept and the slope for both regressions. For total biomass records in White et al. (2000), the intercept and slope were significantly different from 1 and 0 when regressing PO data as the authors did ($p = 0.024$ and $p = 0.0059$, respectively), but they were both not significant with the correct regression of OP values (Table 2). The conclusions of model evaluation changed completely when exchanging the variables plotted in each axis. The regression of OP values (Fig. 3b) showed that the model had a similar bias throughout all the range of values and that the slope did not differ significantly from 1 (Table 2). Theil's coefficients also showed that most of the errors in model predictions were due to unexplained variance (77%), and not to bias or slope misleading (Table 2).

The lack of symmetry in the computation of several parameters when regressing OP or PO, has been noted by several authors, but not thoroughly examined (Kobayashi and Salam, 2000; Gauch et al., 2003). Mitchell (1997) writes in page 315: "*Prediction and observation are plotted on a scatter graph. For the*

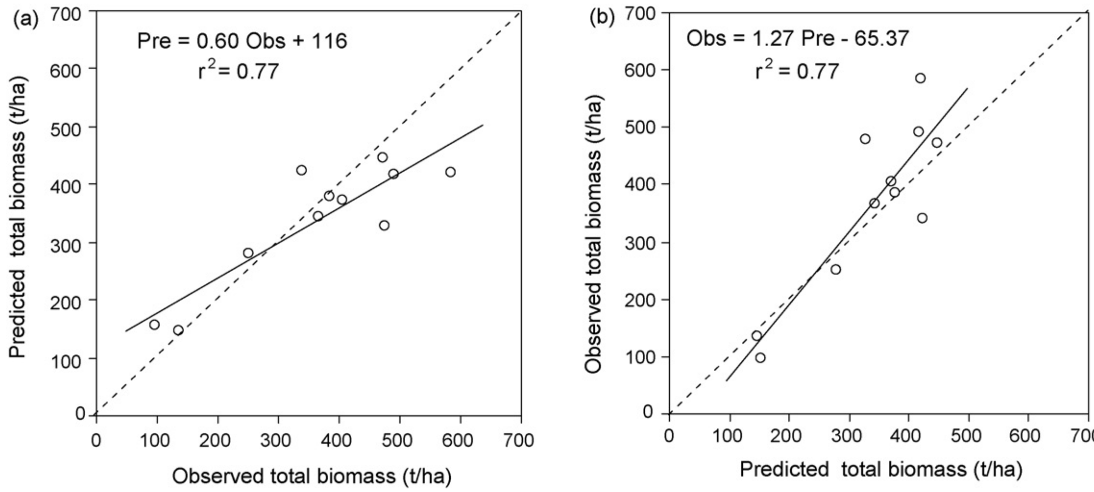

Fig. 3 – Predicted vs. observed (a) (PO) and observed vs. predicted (b) (OP) regression scatter plots of data from White et al., 2000. (a) is Fig. 3 presented in White et al. paper's and (b) is the regression obtained with the same data but changing the variables from one axis to the other. Note that although r^2 is the same, regression coefficients (that describe the similarity of the regression line with the 1:1 line) change notably.

purpose of the arguments set out below it makes little difference whether predictions or observations are the independent variable on the x-axis". Smith and Rose (1995) suggested in page 53 that Theil's coefficients and goodness of fit analysis are easy to perform when regressing OP values, and "*not as straightforward*" to calculate when regressing PO values. Here we have shown that this last approach is, simply, incorrect.

The validity of r^2 in regressions of predicted and observed values has been questioned, because it characterizes the mean deviation of observed values (placed in the y-axis) from the regression line (the regression sum of squares divided by the total sum of squares). It may have little importance to evaluate how much the observed values differ from the regression line of OP values (Kobayashi and Salam, 2000; Gauch et al., 2003). However, although the r^2 can be estimated by dividing the regression sum of squares by the total sum of squares, it can be also calculated from Eq. (14). This equation shows that r^2 also represents the proportion of the linear covariance of y and $\hat{y}$, with respect to the total variance of y and $\hat{y}$. In this sense, the r^2 indicates how much of the linear variation of observed values (y) is explained by the variation of predicted values ($\hat{y}$). Linearity between observed and predicted values can be tested following (Smith and Rose, 1995). Thus, the r^2 of OP values is a valid parameter that gives important information of model performance.

Conversely, the root mean squared error (RMSE) a commonly used statistic to show model performance (Weiss and Hays, 2004; Doraiswamy et al., 2005; Lobell et al., 2005), should not be applied for the regression of OP data, instead the root mean squared deviation (RMSD) (see Eq. (10)) should be reported (Wallach and Goffinet, 1989; Kobayashi and Salam, 2000; Gauch et al., 2003). The RMSE is a proxy of the mean deviation (not exactly the mean because it is squared and divided by $n-1$) of values in the y-axis against the regression line. When reporting the RMSE for the OP or PO regression, we are not estimating the mean deviation between estimated and predicted data. Instead, we are estimating the root mean squared error of the observed values against the regression line of observed vs. predicted values (in the case of regressing OP) and the root mean squared error of the predicted values against the regression line of predicted vs. observed values (in the case of PO). The correct comparison is to calculate the deviation of each predicted values against the 1:1 line and not against the regression line of either OP or PO. RMSE will be always smaller than RMSD and thus represents an underestimation of the real error between observed and

Table 2 – Regression parameters and hypothesis testing for PO or OP regressions, from data presented in White et al. (2000)

	Predicted vs. observed (PO)	Observed vs. predicted (OP)
a	116.9	−65.37
Significance of Test $a = 0$	0.024	0.44
b	0.60	1.27
Significance of Test $b = 1$	0.0059	0.27
U_{bias} (%)[a]	–	0.11
U_{slope} (%)[a]	–	0.12
U_{error} (%)[a]	–	0.77
RMSD (tons/ha)	–	82.6

Theil's partial inequality coefficients and the root mean squared deviation (RMSD) are shown when applicable. RMSD estimates the mean deviation of predicted values respect to the observed ones, in the same units as the model variable under evaluation.

[a] Theil's coefficients partition the variance of observed values not explained by the predicted values (called the squared sum of the predictive error), being: $U_{bias,}$ the proportion associated with mean differences between observed and predicted values, U_{slope} the proportion associated with the slope of the fitted model and the 1:1 line, and U_{error} the proportion associated with the unexplained variance.

simulated values. For example, in White's and collaborators paper the RMSD was 82.6 tons/ha (Table 2), while the RMSE changed between the regression of PO and OP values (52.7 and 76.2 tons/ha, respectively), and is always smaller than RMSD.

4. Conclusions

We showed empirically and demonstrated analytically that model evaluation based on linear regressions should be done placing the observed values in the *y*-axis and the predicted values in the *x*-axis (OP). Model evaluation based on the opposite regression leads to incorrect estimates of both the slope and the *y*-intercept. Underestimation of the slope and overestimation of the *y*-intercept increases as r^2 values decrease.

We strongly recommend scientists to evaluate their models by regressing OP values and to test the significance of slope = 1 and intercept = 0. This analysis can be complemented by decomposing the variation of observed values not explained with the predictions (the squared sum of the predictive error), through calculating Theil's partial inequality coefficients (U). The coefficient of determination r^2 can be used as a measure of the proportion of the variance in observed values that is explained by the predicted values. If replicates of observed values are available then a goodness of fit test can be performed following (Smith and Rose, 1995). RMSE should not be reported for the OP regression, but the RMSD adds important information to model evaluation.

Acknowledgments

We thank the students of the "Estadistica Aplicada a la Investigación Biológica" class, held at the EPG-FA, University of Buenos Aires in year 2001, for encouraging discussions on the topic of the paper. Fernando Tomasel gave helpful advises for starting this work. We thank Gonzalo Grigera and two anonymous reviewers that made lots of insightful comments which improved the contents of this manuscript. This work was supported by the University of Buenos Aires by the "Proyecto Estrategico" Res. (CS) No. 5988/01, the IAI-CRN II 2031, FONCYT PICT 06-1764, by UBACYT G-071 and UBACYT G-078. Gervasio Piñeiro was a PhD student funded by CONICET, Argentina.

REFERENCES

Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. Remote Sens. Environ. 97, 192–202.

Gauch Jr., H.G., Hwang, J.T.G., Fick, G.W., 2003. Model evaluation by comparison of model-based predictions and measured values. Agron. J. 95, 1442–1446.

Knightes, C.D., Cyterski, M., 2005. Evaluating predictive errors of a complex environmental model using a general linear model and least square means. Ecol. Model. 186, 366–374.

Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. Agron. J. 92, 345–352.

Lauenroth, W.K., Burke, I.C., Berry, J.K., 2003. The status of dynamic quantitative modeling in ecology. In: Canham, C.D., Cole, J.C., Lauenroth, W.K. (Eds.), Models in Ecosystem Science. Princeton University Press, New Jersey.

Lobell, D.B., Ortiz-Monasterio, J.I., Asner, G.P., Naylor, R.L., Falcon, W.P., 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agron. J. 97, 241–249.

Loehle, C., 1997. A hypothesis testing framework for evaluating ecosystem model performance. Ecol. Model. 97, 153–165.

Mesple, F., Troussellier, M., Casellas, C., Legendre, P., 1996. Evaluation of simple statistical criteria to qualify a simulation. Ecol. Model. 88, 9–18.

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

Monte, L., Hakanson, L., Bergstrom, U., Brittain, J., Heling, R., 1996. Uncertainty analysis and validation of environmental models: the empirically based uncertainty analysis. Ecol. Model. 91, 139–152.

Paruelo, J.M., Jobbágy, E.G., Sala, O.E., Lauenroth, W.K., Burke, I.c., 1998. Functional and structural convergence of temperate grassland and shrubland ecosystems. Ecol. Appl. 8, 194–206.

Smith, E.P., Rose, K.A., 1995. Model goodness-of-fit analysis using regression and related techniques. Ecol. Model. 77, 49–64.

van Tongeren, O.F.R., 1995. Data analysis or simulation model: a critical evaluation of some methods. Ecol. Model. 78, 51–60.

Wallach, D., Goffinet, B., 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecol. Model. 44, 299–306.

Weiss, A., Hays, C.J., 2004. Simulation of daily solar irradiance. Agric. Forest Meteorol. 123, 187–199.

White, J.D., Coops, N.C., Scott, N.A., 2000. Estimates of New Zealand forest and scrub biomass from the 3-PG model. Ecol. Model. 131, 175–190.