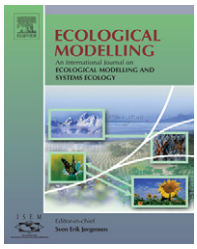
available at www.sciencedirect.com

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.

© 2008 Elsevier B.V. All rights reserved.

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 © 2008 Elsevier B.V. All rights reserved.
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.

$$X = \{1, 2, 3, \ldots 60\} \quad (1)$$

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

$$Y_{\text{Lin}} = X + \varepsilon \quad (2)$$

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

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

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):

$$\hat{Y}_{\text{Lin}} = aX + b \quad (5)$$

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

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

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

Download English Version:

https://daneshyari.com/en/article/4378010

Download Persian Version:

https://daneshyari.com/article/4378010

Daneshyari.com