NAUPLIUS, Rio Grande, 4: 139-145, 1996.

EVALUATION OF A REGRESSION MODEL EMPLOYED IN CRUSTACEAN SURVIVAL STUDIES TO COMBINATIONS OF ABIOTIC FACTORS

J. M. MONSERRAT; R. B. MIRANDA; A. BIANCHINI & E. A. SANTOS

Laboratório de Zoofisiologia; Departamento de Ciências Fisiológicas; Fundação Universidade do Rio Grande; C. Postal 474, 96201-900 Rio Grande - RS, Brasil.

ABSTRACT

The goal of many experiments is to establish crustacean tolerance zones to different combinations of abiotic factors, tipically salinity (S) and temperature (T). In this kind of assay, data analysis frequently envolve multiple regression, employing the following model: $y_{ij} = \beta_0 + \beta_1 S_i + \beta_2 T_j + \beta_3 S_i^2 + \beta_4 T_j^2 + \beta_5 S_i{*}T_j + \varepsilon_{ij}$, proposed by Alderdice (1972). However, in these experiments no mentions about verification of the regression model assumptions or any analysis in order to verify the adequacy of the fitted model are made. We employ tolerance data of male adult crabs (*Chasmagnathus granulata*, Decapoda, Grapsidae) to different combinations of salinity and temperature, in winter and summer. Data obtained for each season were submitted to the following analysis: (1) normal distributions of the residuals; (2) plot of the residuals against the predicted values for the adjusted model and (3) lack of positive autocorrelation of the error values. Also, the prediction capacity of the model was studied, employing the prediction error sum of square (PRESS) statistic. The results obtained, showed a poor prediction capacity of the model estimated. So, there is no obvious reasons to apply the model proposed by Alderdice (1972) without an objetive analysis of the model quality.

Keywords: Regression Model, Crustacea, Survival, Crab, Abiotic Factors

INTRODUCTION

The goal of many experiments is to establish crustacean tolerance zones to different combinations of abiotic factors, tipically salinity and temperature. In this kind of assay, data analysis frequently envolve multiple regression, employing the following model:

$$y_{ij} = \beta_0 + \beta_1 S_i + \beta_2 T_j + \beta_3 S_i^{2} + \beta_4 T_j^2 + \beta_5 S_i{*}T_j + \varepsilon_{ij}$$

where β_0 represents the intercept, β_1 the coefficient of the linear term of salinity (S) and β_2 the coefficient of the linear term of temperature (T). On the other hand, β_3 and β_4 are the coefficients of the quadratic terms of salinity (S^2) and temperature (T^2), respectively. Finally, β_5 is the coefficient of the interaction term of salinity and temperature (S*T) and $_{ij}$ is the error term of the model.

If *s* salinities and *t* temperatures are assayed and only one replica is run for each salinity-temperature combination then *i* = 1,2,..., *s* and *j* = 1,2,..., *t*. The experimental design consists of a random assignement of the treatments (salinity-temperature combinations) to each experimental units. Since in this kind of assay the common registered 'response' is death, the transformation $y' = \sin^{-1}(y)^{1/2}$ is usually used. In such a case, y' is the proportion of dead (or alive) organisms after a certain period of time, usually 96 hr. This transformation is applied in order to accomplish the basic assumptions required by the regression analysis. One of them, assumes that $\varepsilon_{ij} \sim N(0, \sigma^2)$, where σ^2 is the variance, equal for the different treatments. Another asumption is the lack of autocorrelation between the error terms. When considering the y variable, it must be pointed out that the experimental unit is represented by the aquarium if, for example, the experiments are performed on an aquatic species. The observation registered over each experimental unit is the proportion of dead (or alive) animals.

Authors like Yamashita (1977) and Young (1991) refer to Alderdice (1972) as reference, in order to justify the employement of the model mentioned above. Further, Laughlin and Neff (1980) used this model with four independent variables: salinity, temperature, phenanthrene (a polycyclic aromatic hydrocarbon) and osmotic shock. In every case, after the model estimation, a surface response was obtained. This is an useful graphic tool, because it permits the determination of salinity-temperature zones, in which the mortality attains to a minimum. Further, response surfaces of different populations, or of different seasons, can also be compared. Authors like Moreira *et al.* (1979) refer to Box and Youle (1955) on the use of response surface curves.

However, it must be noticed that none of the works cited above has mentioned neither a verification of the regression model assumptions nor any analysis in order to verify the adequacy of the fitted model to the data obtained. Thus, the objetive of the present work is to analyze the basic assumptions required by the model proposed by Alderdice (1972). If any violation of one or more required statistical assumptions mentioned occurs, then the model validity would be dubious (Montgomery and Peck, 1982).

MATERIAL AND METHODS

In this work the data obtained by Miranda (1994) is analyzed. This author studied the tolerance of male adult crabs (*Chasmagnathus granulata*, Decapoda, Grapsidae) to different combinations of salinity (0, 5, 10, 30, 40 or 45) and temperature (5, 10, 20, 30 or 35°C), in winter and summer. Ten crabs were exposed to each treatment during 96 hr and the death criteria employed was the absence of locomotory and respiratory activities. The variable registered was the proportion of death animals in each treatment after 96 h (y). The winter and summer data were analyzed separately. The transformation $y' = \sin^{-1}(y)^{1/2}$ was applied to the original data. The transformed data were adjusted by minimum squares to the multiple regression model mentioned in the Introduction section. Data obtained for each season were submitted to the following analysis in order to verify the model assumptions: (1) normal distributions of the residuals (which are the estimators of the ε_{ij} values), employing the Kolmogorov-Smirnov test (Montgomery, 1984); (2) plot of the residuals against the predicted values for the adjusted model and (3) lack of positive autocorrelation of the error values, by means of the Durbin-Watson statistic (Montgomery and Peck, 1982). Also, the prediction capacity of the model was studied employing the prediction error sum of square (PRESS) statistic (Montgomery and Peck, 1982). This statistic is calculated by eliminating one of the *n* values employed to estimate the model and re-estimating the model using the other *n* -1 values. With the new model, the value of the eliminated i-th data point ($y_{hat(i)}$) is estimated and the error between the estimated and the observed values ($e_i = y_{hat(i)} - y_i$) is calculated. This procedure is repeated over all the *n* values, and the PRESS statistic then calculated according to the expression: $PRESS = \Sigma^{n}_{i=1} e^2_i$. The PRESS value is an useful measure of the prediction quality of a model. Further, it can be used when two or more alternative models are compared, since the model presenting the lowest PRESS value will be that with greatest prediction capacity.

RESULTS

For the winter data (Table 1.A), only the S*T coefficient was not statistically different from zero. The R^2 adjusted to the number of parameters estimated was 0.6762. The residuals distribution was not significantly different from a normal distribution. Figure 1.A shows the residuals as function of the predicted values. A clear pattern of parallel lines could be observed. The Durbin-Watson statistic gave a value of d = 1.297. The test used to detect positive autocorrelation uses d_L and d_U as critical values (Montgomery and Peck, 1982). If $d<d_L$, the null hipothesis of absence of autocorrelation is rejected and if $d>d_U$ the null hipothesis is not rejected. When $d_U<d<d_L$, the test is inconclusive. Using an α= 0.01 and n =30, d_U = 0.88 and d_L= 1.61. So, for the winter data, the test to detect presence of autocorrelation is inconclusive.

Now considering the summer data, it can be observed in Table 1.B that all the coefficientes were statistically different from zero. The R^2 adjusted to the numbers of parameters estimated was 0.7010. As in the winter data, the residuals distribution was not significatively different from a normal distribution. When the residuals are plotted against the predicted values, the same pattern of parallel lines mentioned for the winter data is observed (Fig. 1.B). The value of Durbin-Watson statistic was d = 1.664. So, with an α= 0.01 and n = 30, the autocorrelation test is again inconclusive.

Table 1. Regression coefficients of the models adjusted by minimum squares.

A. Winter

Coefficient	Estimate	Std. Error	t value	p
β_0	2.3617	0.3294	7.17	<0.01
β_1	-0.0644	0.0194	-3.32	<0.01
β_2	-0.2166	0.0343	-6.31	<0.01
β_3	0.0011	0.0004	2.86	<0.01
β_4	0.0057	0.0008	6.90	<0.01
β_5	0.0002	0.0004	0.68	>0.05

B. Summer

Coefficient	Estimate	Std. Error	t value	p
β_0	2.4379	0.3317	7.35	<0.01
β_1	-0.1040	0.0193	-5.38	<0.01
β_2	-0.1758	0.0347	-5.07	<0.01
β_3	0.0015	0.0004	4.01	<0.01
β_4	0.0044	0.0008	5.30	<0.01
β_5	0.0011	0.0004	2.92	<0.01

DISCUSSION

There is no evident biological meaning in the model proposed by Alderdice (1972), although, as previously mentioned, it has been widely used. It seems that the goal is to estimate this model in order to obtain response surface curves. Then, different experimental or ambiental conditions can be compared. The comparisons are made only by visual inspection of the model. It must be pointed out that this kind of comparison or even more formal ones, rely on the quality of the model estimation. Also, in the original article of Alderdice (1972), he proposed other kinds of models. There are no reasons *a priori* to select one or another model. So, the use of a particular model without any statistical evaluation is objectable. For example, two different kind of models can be considered:

$$(1)\ y_{ij} = \beta_0 + \beta_1 S_i + \beta_2 T_j + \beta_3 S_i^2 + \beta_4 T_j^2 + \beta_5 S_i{*}T_j + \varepsilon_{ij},$$
$$(2)\ y_{ij} = \beta_0 + \beta_1 S_i + \beta_2 T_j + \beta_3 S_i^3 + \beta_4 T_j^3 + \beta_5 S_i{*}T_j + \varepsilon_{ij},$$

Model (1) was analyzed above. If model (2), with cubic terms, is applied to the winter data, an R^2 (adjusted to the number of parameters) of 0.6823 is obtained, which is a little better that the obtained with the model (1). Residual analysis showed no statistical differences with a normal distribution and lack of

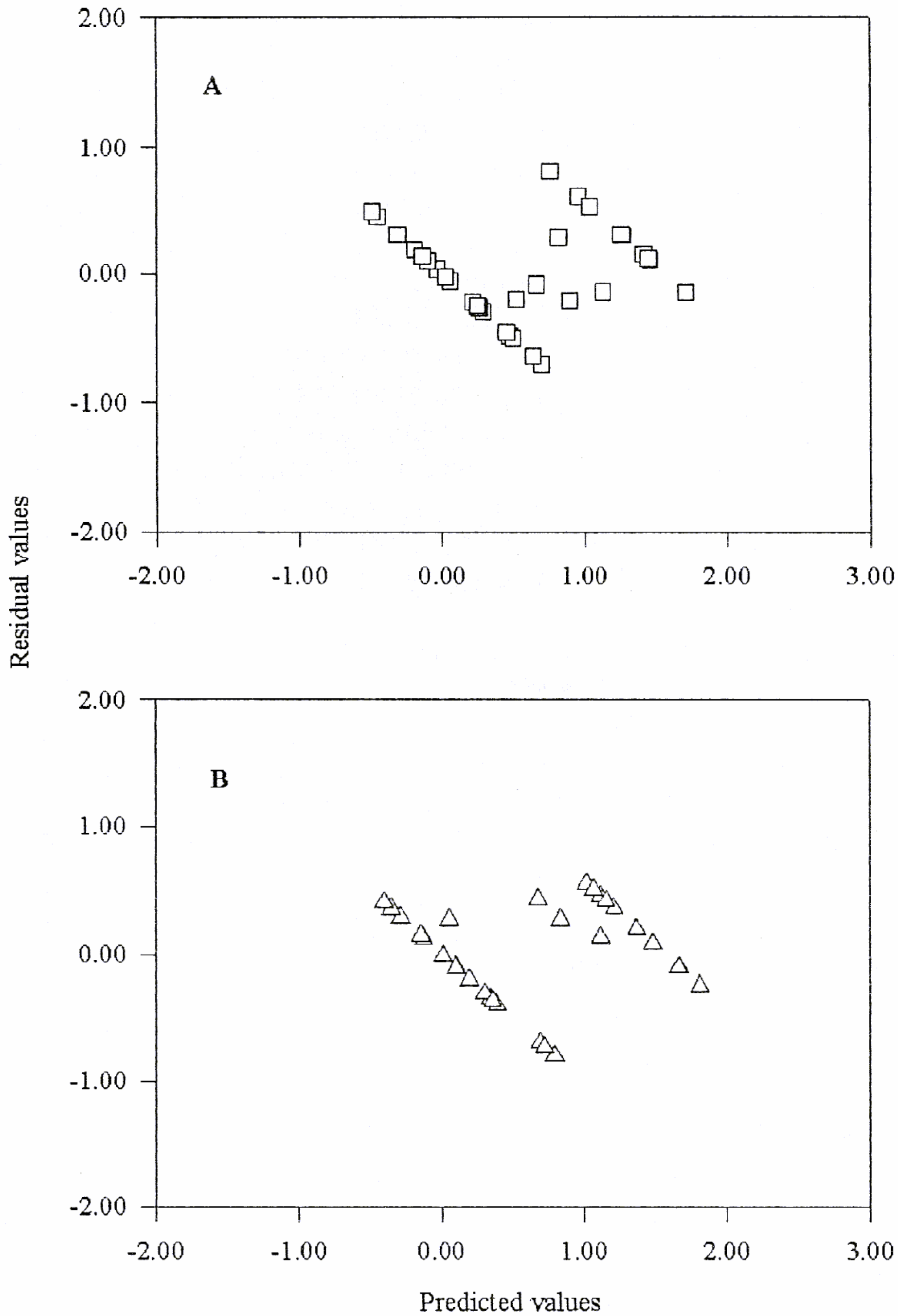

Figure 1. Residuals plots against predicted values by the models ajusted to the winter (A) and summer data (B).

positive autocorrelation. However, the PRESS value obtained with the model (2) was of 4.2234 whereas the obtained with model (1) was of 6.2924. It seems that model (2) have better predictive properties that the commonly employed model (1).

Considering the parallel lines pattern observed when the residuals are plotted against the predicted values, Searle (1988) pointed out that this situation arised when there is little data variability. Considering the original winter data (Miranda, 1994), 58% of them is zero and 28% is 1. In the summer data, 58% is zero and 30% is 1. The variable studied (proportion of death animals) can be considered to be almost dichotomic. Probably, an experimental design with a better selection of salinity-temperature combinations could produce a greater variability. It must be pointed out that regression analysis assumes that the variable studied is continuous.

Considering the prediction quality of the models adjusted for the winter and summer data, Table 2 show predicted values for some salinity-temperature combinations for the winter and summer data, respectively. For the summer data, with a salinity of 30‰ and a temperature of 10°C, the predicted value is highly negative (-0.2866), which is inconsistent with the transformation applied to the data. If the transformation $y' = \sin^{-1}(y)^{1/2}$ is applied, the original values for the variable y (proportion of alive or dead animals) lie between 0 and 1, whereas the y' values lie between 0 and 1.5708. So, a negative value predicted by the model is inconsistent with the transformation applied.

Table 2. Mortality values observed and predicted by the models adjusted to the winter and summer data for some salinity-temperature combinations. $y' = \sin^{-1}(y)^{1/2}$. y = proportion of death animals

A. Winter

Salinity (‰)	Temperature (°C)	Observed value (y_i)	Predicted value ($\hat{y}_i$)
0	10	1.5708	0.7619
5	10	0	0.4799
30	10	0	-0.0989
45	10	0	0.2189
40	30	0	0.4542
45	30	0	0.6403

B. Summer

Salinity (‰)	Temperature (°C)	Observed value (y_i)	Predicted value ($\hat{y}_i$)
0	10	1.2490	1.1190
5	10	0	0.6918
30	10	0	-0.2866
45	10	0.3218	0.0526
40	30	0	0.7235
45	30	1.5708	1.0218

To sum up, according to the example analyzed here, there is no obvious reasons to apply the model proposed by Alderdice (1972) without an objective analysis of the model quality.

REFERENCES

ALDERDICE, D.F. 1972. Factor combinations-responses of marine poikilotherms to environmental factors acting in concert. In : Marine Ecology (O. Kinne, Ed.), Vol. I, Part 3, Wiley Interscience, New York. pp. 1659-1722

BOX, G. E. & P.V. YOULE. 1955. The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanisms of the system. Biometrics 11: 87-323.

LAUGHLIN JR., R.B. & J.M. NEFF. 1980. Influence of temperature, salinity and phenanthrene (a petroleum derived polycyclic aromatic hydrocarbon) on the respiration of larval mud crabs, *Rhithropanopeus harrisii.* Estuar. Coast. Mar. Science 10:655-669.

MIRANDA, R.B. 1994. Efeitos da temperatura e da salinidade sobre a tolerância e a ionorregulação de *Chasmagnathus granulata* (Decapoda, Grapsidae). Master Thesis, Fundação Universidade de Rio Grande, Rio Grande, Brazil, 176 pp.

MONTGOMERY, D.C. & E.A. PECK. 1982. Introduction to Linear Regression Analysis. - John Wiley & Sons, New York. 503 pp.

MONTGOMERY, D.C. 1984. Design and Analysis of Experiments. second edition, John Wiley & Sons, New York. 536 pp.

MOREIRA, G.S., J.C. MCNAMARA & P.S. MOREIRA. 1979. The combined effects of temperature and salinity on the survival and moulting of early zoeae of *Macrobrachium holthuisi* (Decapoda: Palaemonidae). Bol. Fisiol. Animal 3:81-93.

SEARLE, S.R. 1988. Parallel lines in residual plots. The American Statistician 42:211.

YAMASHITA, C. 1977. Efeitos combinados de temperatura e salinidade na sobrevivência do copépode *Euterpina acutifrons* (Dana, 1847), da região de Santos e do canal de São Sebastião. Bol. Fisiol. Animal 1:91-120.

YOUNG, A.M. 1991. Temperature-salinity tolerance of two latitudinally separated populations of the longirostris hermit crab *Pagurus longicarpus* Say (Crustacea, Decapoda, Paguridae). Ophelia 34:29-39.