Ecological Modelling, 49 (1989) 1–6
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

APPLICATION OF A GUARANTEED REGRESSION MODEL TO TROPHIC INTERACTION IN AN AQUATIC SYSTEM

VLASTIMIL KŘIVAN and JAROMÍR SEĎA

South Bohemian Biological Research Centre, Czechoslovak Academy of Sciences, 370 05 České Budějovice (Czechoslovakia)

(Accepted 24 May 1989)

ABSTRACT

Křivan, V. and Sed'a, J., 1989. Application of a guaranteed regression model to trophic interaction in an aquatic system. *Ecol. Modelling*, 49: 1–6.

In this paper the method of guaranteed estimation is used to estimate the unknown parameters of two nonlinear regression models. The first describes the relations between the body size structure of zooplankton and the biomass of planktivorous fish. The second estimates time-lag between the dynamics of cladoceran egg production and dynamics of phytoplankton.

INTRODUCTION

In ecological studies, regression models are widely used to describe the various relations among important variables. In general, such a model has the following form:

$$y = f(x, a_1, \ldots, a_p)$$

where x denotes the independent variable, $a_1, \ldots, a_p$ are the parameters, and y denotes the dependent variable. The main task is to estimate the unknown parameters from the experimental data (x_i, y_i), $i = 1, \ldots, N$. Usually some standard statistical method is used. It is assumed that the model can be written in the form:

$$z_i = f(x_i, a_1, \ldots, a_p) + \varepsilon_i'$$

$$y_i = z_i + \varepsilon_i'' \qquad i = 1, \ldots, N$$

or, in more compact form:

$$y_i = f(x_i, a_1, \ldots, a_p) + \varepsilon_i \qquad \begin{array}{l} \varepsilon_i = \varepsilon_i' + \varepsilon_i'' \\ i = 1, \ldots, N \end{array} \tag{1}$$

0304-3800/89/$03.50

where the noise ε_i' and ε_i'' is governed by a random mechanism. For example, if the random variable ε_i is supposed to have normal distribution and all ε_i are jointly independent, the standard method for parameter estimation is the least squares method. For the case of f linear in parameters, fairly complete statistical theory has been developed. In the case of biological data it may be difficult or impossible to verify the above-mentioned assumptions, especially because of the lack of data.

An alternative method to statistical parameter estimation based on 'guaranteed estimation' (Kurzhanski, 1988) was developed. Contrary to conventional statistical approach it is assumed that there are no statistical data on the noise ε_i', ε_i'' available. It is only assumed that the noise is bounded, i.e.

$$|\varepsilon_i'| \le P_1, |\varepsilon_i''| \le P_2$$

Then we are looking for all those values of the unknown parameters $a_1, \ldots, a_p$ that are consistent with given experimental data in the following sense:

$$y_i = f(x_i, a_1, \ldots, a_p) + \varepsilon_i \qquad \begin{array}{l} |\varepsilon_i| \le P_1 + P_2 = P \\ i = 1, \ldots, N \end{array} \tag{2}$$

We define *admissible set A(N)* that depends on the number of data N:

$$A(N) := \{(a_1, \ldots, a_p) \| y_i - f(x_i, a_1, \ldots, a_p)| \le P \qquad i = 1, \ldots, N\}$$

The set $A(N)$ contains all values of parameters $a_1, \ldots, a_p$ that are consistent with the measurements (x_i, y_i), $i = 1, \ldots, N$. It is easy to see that $A(N+1) \subset A(N)$.

We tested this method on two data sets. The first consists of the measurements of the size structure of zooplankton and the density of planktivorous fish in Římov Reservoir (see Sed'a et al., 1989). The second evaluates time lag between the time course of cladoceran egg production and the time course of phytoplankton (Sed'a, 1989).

METHODS

In the first case, where the biomass of the planktivorous fish is taken as the independent variable (x) and the proportion of the biomass of large cladocerans to the total biomass of zooplankton is dependent variable (y), the following regression function:

$$y = a_1 e^{a_2 x} \tag{3}$$

was used (Sed'a et al., 1989). Since, for the computation, only annual averages for both variables can be used, only data for a 10-year period were available. This is of course not enough to validate the assumptions about the statistical properties of the noise. It is only assumed that we know the upper bound for the noise P, i.e., $|\varepsilon_i| \leq P$, $i = 1, \ldots, 10$. It follows:

$$y_i \in \left[a_1\, e^{a_2 x_i} - P,\ a_1\, e^{a_2 x_i} + P\right] \qquad i = 1, \ldots, 10$$

and consequently

$$f(a_2) := \max_{i=1,\ldots,10} \left((y_i - P)\, e^{-a_2 x_i}\right) \leq a_1$$

$$\leq \min_{i=1,\ldots,10} \left((y_i + P)\, e^{-a_2 x_i}\right) =: g(a_2)$$

The admissible set after ten measurements is:

$$A(10) := \{(a_1, a_2) \mid f(a_2) \leq a_1 \leq g(a_2)\}$$

In the second case we tried to estimate the time-lag between two time-series. The standard approach to handle such a kind of problem is based on computation of a cross-correlation coefficient for different time-lags and then test the significance of the highest value. In this case we can use guaranteed approach too. As an example we used data for the dynamics of cladoceran egg production and the dynamics of phytoplankton as potential food source for cladocerans. We assume that these two time series are linearly dependent with added 'noise', i.e.

$$y_i = k x_{i-\tau} + \varepsilon_i$$

Moreover we assume that the upper bound for ε_i is given, i.e.

$$|\varepsilon_i| \leq P \qquad i = 1, \ldots$$

It follows that the measurements must satisfy:

$$y_i \in \left[k x_{i-\tau} - P,\ k x_{i-\tau} + P\right] \qquad i = 1, \ldots, N$$

The admissible set $A(N)$ is then:

$$A(N) := \left\{(k, \tau) \mid \max_{i=1,\ldots,N} \frac{y_i - P}{x_{i-\tau}} \leq k \leq \min_{i=1,\ldots,N} \frac{y_i + P}{x_{i-\tau}}\right\}$$

RESULTS AND DISCUSSION

The data on body size structure for which calculations were performed are plotted together with the corresponding regression function in Fig. 1. The values of the parameters were calculated using the nonlinear least-squares method. The admissible sets A for different upper bounds of the noise (different values of P) are plotted in Fig. 2. The main difference between the method of guaranteed estimation and the other statistical methods for parameter estimation lies in the fact that in the first case we do not assume

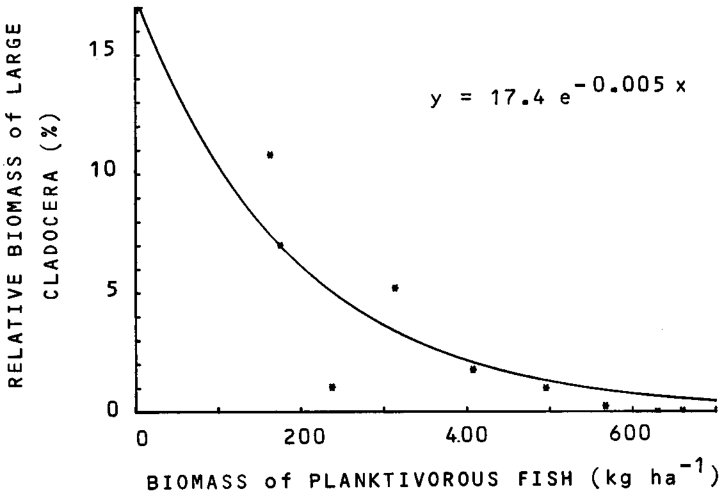

Fig. 1. Relationship between the body size structure of zooplankton (y) and the biomass of planktivorous fish (x) (Sed'a et al., 1989).

anything about statistical properties of the noise. What we only assume is that the lower and upper bounds for the noise are given. Sometimes it might be even quite difficult to give these bounds. In general it is not difficult to

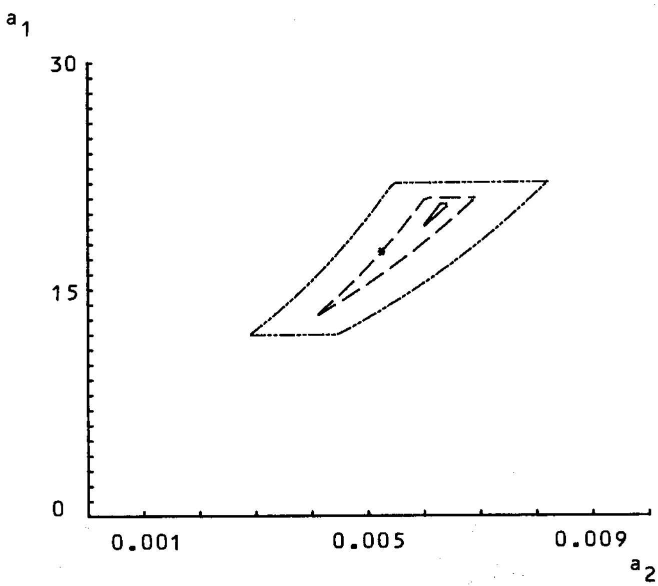

Fig. 2. Admissible set A for different values of P: —‑‑—, 5; — — —, 4; ——— 3.6. The point denotes the estimated values of parameters via the least squares method.

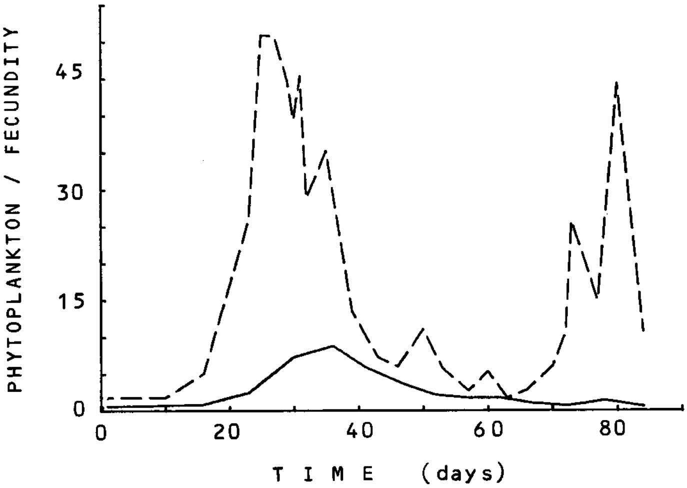

Fig. 3. Time-course of the cladoceran fecundity ([eggs per adult female], full line) and the density of phytoplankton ([μg chlorophyll *a* l^{-1}], dashed line) (Sed'a, 1989).

give a-priori the bound P_2 for the observation (or data measurement) noise ε_i''. On the other hand the estimation of ε_i' is much more difficult, or even impossible. It makes sense to start the guaranteed parameter estimation with $P = P_2$ (i.e. the optimal situation when $P_1 = 0$). Of course it is highly

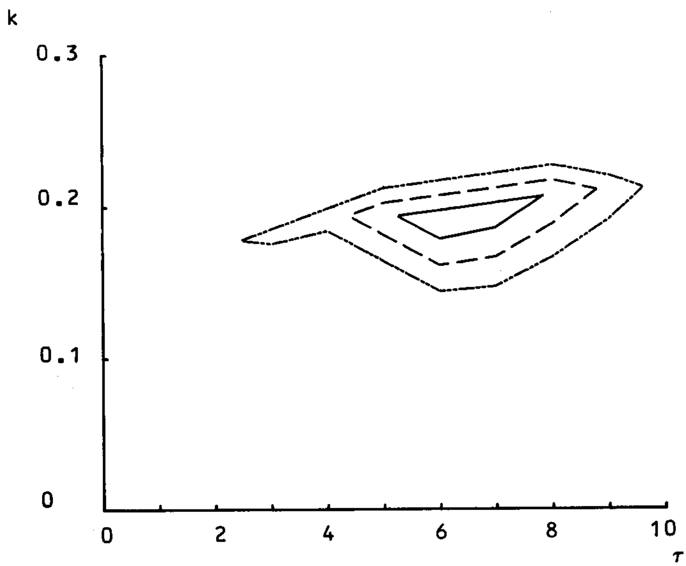

Fig. 4. Admissible set *A* for different values of *P*: ·–·–·, 3.5; — — —, 3; ———, 2.5.

probable that in this way we underestimate the upper bound P and the admissible set will be empty. Then we can increase, step by step, the upper bound P until we get a nonempty admissible set A. On the other hand, if we overestimate this bound then the admissible set may be large. We can try to find the minimal value of P such that the admissible set A consists of one point only.

In Fig. 3 we plotted data describing the dynamics of cladoceran egg production and the dynamics of phytoplankton. Using the guaranteed method we estimated both time-lag τ and proportionality coefficient k. The admissible set A is plotted in Fig. 4 for different values of P.

In the case where we are estimating three and more parameters the method of guaranteed estimation is more difficult to use since we cannot visualize the admissible set $A(N)$. More advanced mathematical methods must be used.

REFERENCES

Kurzhanski, A.B., 1988. Identification–a theory of guaranteed estimates. IIASA WP-88-55, International Institute for Applied Systems Analysis, Laxenburg, Austria, 78 pp.

Sed'a, J., 1989. Main factors affecting spring development of herbivorous Cladocera in the Římov reservoir (Czechoslovakia). Arch. Hydrobiol. Beih. Ergeb. Limnol. (in press).

Sed'a, J., Kubečka, J. and Brandl, Z., 1989. Zooplankton structure and fish population development in the Římov reservoir. Arch. Hydrobiol. Beih. Ergeb. Limnol. (in press).