Models for the estimation of a "no effect concentration"

Ana M. Pires,* João A. Branco,* Ana Picado** and Elsa Mendonça**
*Department of Mathematics and Applied Mathematics Centre
Instituto Superior Técnico, Technical University of Lisbon
Av. Rovisco Pais, 1049-001 Lisboa, Portugal

**Environmental Technologies Institute
INETI, Az. Lameiros, 1699 Lisboa Codex, Portugal

SUMMARY: The use of a no effect concentration (NEC), instead of the commonly used no observed effect concentration (NOEC) has been advocated recently. In this paper models and methods for the estimation of a NEC are proposed and it is shown that the NEC overcomes many of the objections to the NOEC. The NEC is included as a threshold parameter in a non-linear model. Numerical methods are then used for point estimation and several techniques are proposed for interval estimation (based on bootstrap, profile likelihood and asymptotic normality). The adequacy of these methods is empirically confirmed by the results of a simulation study. The profile likelihood based interval has emerged as the best method. Finally the methodology is illustrated with data obtained from a 21 day *Daphnia magna* reproduction test with a reference substance, 3,4-dichloroaniline (3,4-DCA), and with a real effluent.

KEYWORDS: No effect concentration (NEC); No observed effect concentration (NOEC); Resampling methods; *Daphnia magna* reproduction test; Threshold parameter; Segmented regression.

1. INTRODUCTION

No observed effect concentrations (NOEC) have been extensively used in environmental toxicology studies, mainly in chronic toxicity tests. One such test is the 21 day reproduction test for *Daphnia magna* (OECD, 1997, CEE, 1979, Kühn *et al.*, 1989). EPA (1991) defines the NOEC as "that largest concentration of effluent, or test substance, for which survival, reproduction, or growth of the test organism is not significantly different (at the 95% confidence level) from that of the control organisms". However NOEC's have a lot of disadvantages that have been pointed out recently in Chapman *et al.* (1995), Chapman, Caldwell and Chapman (1996), OECD (1997) and OECD (1998). These disadvantages include:

- "The NOEC must be one of the concentrations used in the experiment"
- "The NOEC tends to increase as the precision of the experiment decreases.[1] Since a larger NOEC implies a safer chemical, the approach reward those who perform poor experiments"
- "Confidence intervals cannot be calculated for the NOEC"
- "A NOEC is not always obtainable"
- "The NOEC depends upon the choice of the type I error rate used in a significance test and also on the choice of the test".

Also, as Schwartz, Gennings and Chinchilli (1995) point out, when analyzing a similar concept (the "no observed effect level", NOEL), the dose response trend is largely ignored in the determination of the NOEC.

Important conclusions from Chapman *et al.* (1995) and OECD (1997) are that an EC_X (that is, the usual effective concentration which reduces x% of the mean response at control, in a dose-response model, for instance the logistic model) or a parametric no effect concentration (NEC) are to be preferred over a NOEC and that there is a strong need for statistical methodology — models and inference procedures — to deal with parametric NEC's.

---

[1]Here precision may be understood as the inverse of the variance of experimental data.

The first approach, the use of an EC_X (along with the traditional models), requires an 'x' to be specified which is both a biological and a statistical issue that has been discussed by several authors without a definite conclusion (Hoekstra and van Ewijk, 1993, Chapman *et al.*, 1996; OECD, 1997, OECD, 1998, Guilhermino *et al*., 1998).

On the other hand, the threshold models reviewed by Cox (1987), and also considered by Ulm (1989, 1991), and the models proposed by Kooijman (1996) (see also Kooijman and Bedaux, 1996a, 1996b) offer the potential to directly estimate a NEC.

In this paper we propose three simpler non-linear models where the NEC is included as a threshold parameter. These models are described in section 2. Models of this kind have been in the statistical literature for a long time under several names, such as, segmented regression, multiphase regression, regression with changepoints, regression with breakpoints (see e.g. Seber and Wild, 1989, Chapter 9). Point estimation is simple but the interval estimation of the NEC (which, as mentioned before, is very important) is not an easy issue. In this paper several methods for interval estimation are evaluated through a simulation experiment. To illustrate the methodology results obtained from the 21 day reproduction test for *Daphnia magna* are analyzed in Section 3. The main conclusions are summarized in Section 4..

2. METHODS

2.1. Models

A reasonable model for describing the type of data where it is necessary to estimate a NEC, is to relate the response variable, *y*, to the explanatory variable, *x*, (always a concentration) by a function which is constant near $x = 0$ (control) up to a point c (playing the role of the NEC), followed by a decay. Two types of decay are considered. The simplest one is a linear decay (in the range of concentrations analyzed), giving Model I:

$$y_j = l - m(x_j - c)I(x_j - c) + \varepsilon_j \qquad (1)$$

where

- (x_j, y_j), $j = 1,\ldots,n$, are the pairs of observations;
- l, m and c are parameters (l is the constant level near $x = 0$, and m is the slope of the decay;
- $I(x)$ is an indicator function

$$I(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

- ε_j are independent random errors with zero mean. Although it is not absolutely necessary it is convenient that these errors can be assumed to follow a normal distribution with constant variance, σ^2.

The mean value of the response variable, $E(y|x) = l - m(x-c)I(x-c)$, according to Model I, is represented in Figure 1.

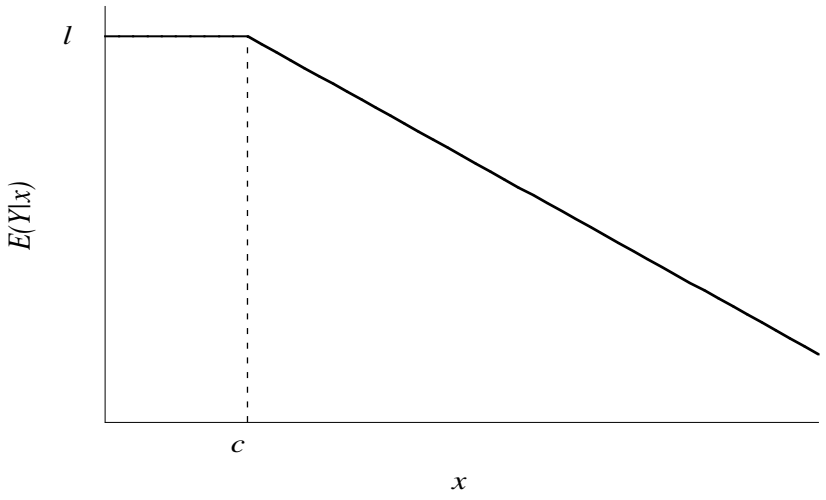

Figure 1: Mean response function with a linear decay

It can be argued that Model I is unreasonable in many situations where the response variable is intrinsically positive and very high concentrations are tested. In that case an exponential decay and multiplicative log-normal errors may be more adequate. This is Model II:

$$y_j = l' \exp\left[-m'(x_j - c')I(x_j - c')\right] \times \varepsilon'_j \qquad (2)$$

where l' and c' have the same meaning of l and c of (1) but m' is now a rate of decay

and ε_j' are random errors approximately following a log-normal distribution. Assuming that all quantities involved are positive Model II can be written in the form of Model I by applying a log-transformation, yielding

$$\log y_j = \log l' - m'\left(x_j - c'\right)I\left(x_j - c'\right) + \log \varepsilon_j' \tag{3}$$

Therefore for estimation purposes the two models are similar. In order to decide which model is more appropriate for a particular data set, it may be helpful to plot y_j versus x_j and $\log y_j$ versus x_j.

Unfortunately there are still some practical situations where it is impossible to use either of the above models, namely when a linear decay is not indicated from the plots, and when for some *j*, $y_j = 0$. If the response variable *y* is measured on a continuous scale the zero values may be replaced by half the measurement unit and we can still attempt to fit Model II. However, in many situations, the response is a count (e.g. the example in Section 3) and that is no longer reasonable. Therefore we need another model that can accommodate an exponential decay and response counts. This is achieved by our proposed Model III, where we consider that each response, y_j, has a Poisson distribution with mean λ_j, and

$$\lambda_j = l'' \exp\left[-m''\left(x_j - c''\right)I\left(x_j - c''\right)\right] \tag{4}$$

We note that Models II and III have similar mean response functions (see Figure 2) but the parameter estimates may be different.

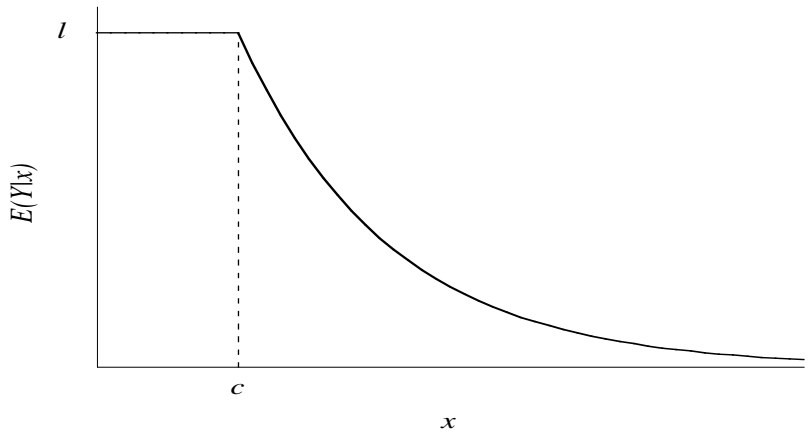

Figure 2: Mean response function with an exponential decay

As one referee pointed out, a more general setup of the models proposed is through the Generalized Linear Models. In this setup we could consider a family of models from which the three models above are special cases.

2.2. Point estimation of the parameters

For the models previously described it is necessary to estimate its three parameters l, m and c, usually c being the one of most interest as it represents the NEC. We propose to use the method of maximum likelihood (which coincides with the method of least squares for Model I and for Model II after the log-transformation, under the distributional assumptions made). Because c is a threshold the usual method of differentiation can not be used directly. But it is still possible to obtain explicit solutions of the normal equations for the parameters l and m of Models I and II and for the parameter l of Model III and use a grid search for the parameter c. (A description of the grid search procedure is given in the Appendix).

Alternatively the exact algorithm described in Küchenhoff (1997) could be adapted to this case, resulting in a significant reduction in the number of evaluations of the likelihood function. However, this algorithm is not used in this paper because, as will be seen in the next Subsection, computing the likelihood on a large number of points will still be necessary for interval estimation purposes.

2.3. Interval estimation for the parameter *c*

Given the type of model and parameter under consideration a theoretical derivation of the sampling distribution of $\hat{c}$ does not seem feasible. Two approaches are possible: to use a (non-parametric) bootstrap method, or to rely on the asymptotic distribution of $\hat{c}$ as a maximum likelihood estimator.

There are several possibilities for resampling as well as several bootstrap confidence intervals (see e. g. Shao and Tu, 1995). The resampling schemes considered were:

(1) paired bootstrap (Shao and Tu, 1995, p. 291);

(2) bootstrap separately on each level of x (if the experiment has such a design);

(3) bootstrap based on residuals, which is model dependent, allowing for possible heterocedasticity, even if not considered in the model (Shao and Tu, 1995, p. 289).

When applied to the data sets described in Section 3 and also on some simulated data sets we verified that the three schemes give similar distributions for $\hat{c}$.

In the simulation study described next only scheme (3) is used because it keeps the design of the actual experiment (as does scheme (2)) but is has a much larger number of possible resamples.

Three types of bootstrap confidence intervals are considered (Shao and Tu, 1995, Chapter 4):

(BP) the bootstrap percentile;

(BC) the bootstrap bias corrected percentile, with the transformation $\Psi = \Phi$, the standard normal distribution;

(BT) the bootstrap-t, with the standard deviation of $\hat{c}$ estimated by the standard jackknife method ($\hat{\sigma}_{\{1\}}$);

and also three types of asymptotic intervals:

(PL) the profile likelihood based interval which results from inverting the asymptotic likelihood ratio test (Cox and Hinkley, 1974, Chapter 9, page 343), that is the solution of $2\log L\left(c,\hat{l}(c),\hat{m}(c)\right) \geq 2\log L\left(\hat{c},\hat{l},\hat{m}\right) - \chi^2_{1;0.95}$ (for 95% confidence).

(NJ1) assuming a normal distribution for $\hat{c}$ and standard deviation estimated as in the (BT), that is $\hat{c} \pm 1.96\hat{\sigma}_{\{1\}}$ (for 95% confidence);

(NJ2) same as (NJ1) but with the standard deviation estimated by the delete-2 jackknife (Shao and Tu, 1995, Section 2.3), that is $\hat{c} \pm 1.96\hat{\sigma}_{\{2\}}$ (for 95% confidence);

To compare the confidence intervals produced by the various methods a simulation study was performed. This simulation is based on a 1000 samples generated from Models I and III, with the parameters given in Table 1.

Table 1
Parameters used in the simulation

Model	c	l	m	σ
I	0.005 ; 0.01	32.5	670	4
III	0.005 ; 0.01	32.5	95	--

Each sample consists of 50 observations with 10 observations on each of five possible values of the concentration (x = 0; 0.005; 0.01; 0.02 and 0.05). These conditions were chosen in order to reflect a real experimental situation, including the magnitude of the sample size. In fact they are similar to those found in one of the DCA data sets analyzed in the next section. The number of bootstrap replicates was set on 500.

Table 2
Results of the confidence interval simulation

		c=0.005			c=0.01		
Model	Method	CP	AL×10^2	SL×10^2	CP	AL×10^2	SL×10^2
I	BP	0.999	1.487	0.329	0.998	1.669	0.309
	BC	0.994	1.502	0.334	0.994	1.616	0.344
	BT	0.955	1.868	1.954	0.980	2.747	2.364
	PL	**0.949**	**0.807**	**0.204**	**0.946**	**0.922**	**0.262**
	NJ1	0.961	0.879	0.604	0.959	1.042	0.671
	NJ2	0.952	0.863	0.537	0.952	1.021	0.611
III	BP	1.000	1.153	0.156	1.000	1.738	0.078
	BC	1.000	1.183	0.186	1.000	1.477	0.204
	BT	1.000	1.130	0.354	1.000	1.910	0.695
	PL	**0.945**	**0.350**	**0.074**	**0.948**	**0.450**	**0.100**
	NJ1	0.871	0.539	0.585	0.898	0.717	0.595
	NJ2	0.942	0.529	0.360	0.941	0.696	0.447

The results of the simulation study are presented in Table 2. CP is an estimate of the coverage probability of the given interval, that is the proportion of the confidence intervals which contained the true value of the parameter. AL is the average length of the intervals and SL the corresponding standard deviation. The nominal confidence level is 95% and we have used only two-sided intervals. In (BT), (NJ1) and (NJ2) intervals the lower

confidence bound was set to zero whenever the calculations led to a negative number. This affects the lengths but it reflects the nature of the parameter.

From the results in Table 2 it is clear that the best interval is (PL): it meets the specified coverage probability with the shortest AL and shortest SL in all cases. (NJ2) performs almost as well in terms of CP but has worse AL and SL. (PL) is also the simplest interval to compute (if the profile likelihood has already been computed to obtain the point estimates) and it does not involve resampling . In terms of computational effort the worst method is (BT), requiring nb estimation cycles, where b is the number of bootstrap samples, while (NJ1) requires n and (NJ2) $n(n\text{-}1)/2$.

In similar simulation exercises, not reported here, for smaller n (n=25) and different combinations of the parameters the conclusions were the same.

A simple explanation for the failure of the normal based intervals is the non-differentiability of the profile log likelihood at some points ($c = x_i$, $i = 1,\ldots,n$). As a consequence a second order polynomial may be a very poor approximation (to the profile log likelihood). We can also infer that it is not worth trying other Wald-type confidence intervals.

The same fact may also explain the poor behavior of the bootstrap intervals because the consistency of the bootstrap distribution estimator requires some smoothness conditions that are almost the same as those required for the asymptotic normality (Shao and Tu, 1995, Chapter 3, page 128).

Finally it is worth noting that the (PL) confidence interval is likely to be conservative if the true c is very close to zero, because in that case $-2\log L$ is not asymptotically χ^2_1 (Feder, 1975; Self and Liang, 1987; Ulm, 1989, 1991; Silvapulle, 1991).

2.4. Model selection and diagnosis

If in a given application it is possible to apply more than one of these models it is important to have a selection criterion. We propose the use of Akaike Information

Criterion (AIC):

$$\text{AIC(Model)} = -2\log\hat{L} + 2k$$

where $\hat{L}$ is the maximum of the likelihood function under the model and k is the number of parameters in the model. Our Models I and II have $k = 4$ (c, l, m and σ) while Model III has $k = 3$ (c, l, and m). Under the distributional assumptions made $\log\hat{L}$ has the following expressions:

Model I: $\log\hat{L} = -\frac{n}{2}\left(\log 2\pi + 1 + \log\frac{\hat{Q}}{n}\right)$, where $\hat{Q}$ is given by formula (5) in the Appendix with l, m and c replaced by $\hat{l}$, $\hat{m}$ and $\hat{c}$;

Model II: $\log\hat{L} = -\frac{n}{2}\left(\log 2\pi + 1 + \log\frac{\hat{Q}}{n}\right) - \sum_{j=1}^{n}\log y_j$, where $\hat{Q}$ is as in Model I but with y_j replaced by $\log y_j$ and $\hat{l}$ by $\log\hat{l}$;

Model III: $\log\hat{L} = \sum_{j=1}^{n} y_j \log\hat{\lambda}_j - \sum_{j=1}^{n}\hat{\lambda}_j - \sum_{j=1}^{n}\log(y_j!)$, where $\hat{\lambda}_j$ is given by formula (4) with l, m and c replaced by $\hat{l}$, $\hat{m}$ and $\hat{c}$.

In addition, the usual plots of the residuals must be inspected in order to (informally) check the assumptions underlying the models.

An asymptotic lack of fit test based on the ratio between the likelihood of the model under consideration and the likelihood of a corresponding saturated model is still under investigation. A preliminary simulation, not reported here, suggests that the number of degrees of freedom of such a test may not be the conventional.

2.5. Comparison between the model based NEC and the NOEC

It is obvious from the description in the previous subsections that the estimate of the model based NEC does not have any of the disadvantages of the NOEC mentioned in Section 1:

- The estimated NEC does not need to be one of the concentrations used in the experiment;
- The estimated NEC does not tend to increase as the precision of the experiment decreases;

- Confidence intervals can be calculated for the NEC and their length reflects the precision of the experiment;[2]
- An estimate is always obtainable;
- The estimated NEC does not depend upon the choice of a type I error rate, nor on the choice of a test, although it depends upon the choice of a model.

The main problems with the NOEC were thus overcome but we can still ask the question how do the estimated NEC and the NOEC compare with the same data. To answer this question we conducted a small simulation study. 1000 samples were generated with the parameters *l*, *m* and σ given in Table 1 and parameter *c* given in Table 3. Each sample has the design described in Subsection 2.3. The NOEC was calculated using an ANOVA model and one-sided Dunnett's multiple comparison procedure (Dunnett, 1964) with $\alpha = 0.05$ (another possibility is Williams' multiple comparison procedure, Williams, 1972). The results of the simulation are also presented in Table 3.

Table 3
Comparison between NEC and NOEC, mean×10^3 (and standard deviation×10^3) over 1000 simulations.

	c=0.005		*c*=0.0075		*c*=0.01	
Model	NEC	NOEC	NEC	NOEC	NEC	NOEC
I	4.96	8.07	7.46	9.40	10.07	10.56
	(2.24)	(2.57)	(2.76)	(1.88)	(2.44)	(2.58)
III	5.13	4.71	7.46	5.57	10.44	9.84
	(0.91)	(1.16)	(1.02)	(2.17)	(1.27)	(1.00)

It is clear from this study that the NEC estimator is not biased but that the NOEC is biased if the true NEC is not one of the concentrations used in the experiment (case *c* = 0.0075) or when the concentrations are closer and the decrease in the response variable is not very steep (case *c* = 0.005 for Model I). In this situation, slow decay of the response,

[2]In this case a theoretical definition of the precision of the experiment can be given in terms of the inverse of the variance of the random component of the model (but it can also be related to the idea of empirical precision used in the Introduction).

the bias of the NOEC in both cases (c = 0.005 and c = 0.0075) is positive, that is in the wrong direction in what concerns environmental safety.

3. APPLICATION

The data here described were obtained at the Environmental Technologies Institute of INETI with the 21 day reproduction test for *Daphnia magna* (a water flea species used in toxicity tests). This test is recommended to detect chronic effects of chemicals, complementing the results of acute tests (CEE, 1979).

3.1. Data

The first experiment was a controlled experiment with fixed concentrations of the reference substance 3,4-dichloroaniline (3,4 DCA). It has been established (Kühn *et al.*, 1989) that for this substance the most sensitive response variable is the reproduction rate (total number of offsprings per parent). The fixed concentrations (in $mg.l^{-1}$) used were x=0, 0.005, 0.01, 0.02 and 0.05. For each concentration 10 parent animals were observed during 14 days and 5 of them during 21 days. The results are shown in Table 4.

For the second experiment a solid waste from a chemical industry was sampled. To run the test an elutriate was prepared according to DIN 38414-S4. The fixed concentrations (in %) used were x = 0, 0.6, 0.9, 1.3, 2.0 and 3.0. For each concentration 10 parent animals were observed at 14 days and at 21 days. The results are given in Table 5.

In both experiments the symbol † means that the parent animal died. According to the recommendations in OECD (1997, pp. 44-59) these results were eliminated, that is, only the data for parents that survived until the observation time was used in the model building and estimation procedures.

Table 4
DCA data

x	Days	Total number of offsprings observed									
0	14	28	29	33	31	38	34	35	33	31	33
	21						95	84	93	89	93
0.005	14	35	36	33	27	36	34	32	29	32	33
	21						92	93	95	94	87
0.01	14	25	22	33	32	31	27	22	21	18	30
	21						73	67	66	61	81
0.02	14	28	27	27	29	29	22	22	23	15	28
	21						55	52	53	46	†
0.05	14	0	0	12	†	†	0	1	1	†	†
	21						4	2	1	†	†

Table 5
Chemical effluent data

x	Days	Total number of offsprings observed									
0	14	46	41	58	40	58	42	56	49	61	42
	21	99	83	107	96	108	95	102	102	111	97
0.6	14	42	†	†	43	42	†	50	51	†	†
	21	92			105	104		105	106		
0.9	14	43	38	42	40	†	44	39	39	38	39
	21	98	92	105	103		89	95	91	91	99
1.3	14	†	38	39	†	29	41	†	42	39	46
	21		97	93		84	105		96	91	106
2.0	14	†	†	6	12	23	23	†	†	15	23
	21			55	65	71	73			60	72
3.0	14	18	21	17	22	10	†	21	†	20	22
	21	56	60	61	47	27		53		48	55

At this stage we have decided not to use the data for the last concentration as these are clearly leverage points, which can have a dangerous influence both on the estimates and on the model selection, and a good fit is clearly more important in the lower concentration range.

3.2. Results

In standard use of the 21 day reproduction test only the data for 21 days is analyzed (as the name suggests) but here we have also considered the 14 days data, because the evolution of the parameter c in time is important for the ecotoxicologists. We have applied the model selection and estimation procedure for each time separately. The results obtained are summarized in Table 6. Figure 3 shows, as an illustration, the data and adjusted model for the DCA data set (21 days) along with the (relevant part of the) profile log-likelihood, that is $\log L\left(c, \hat{m}(c), \hat{l}(c)\right)$ versus c.

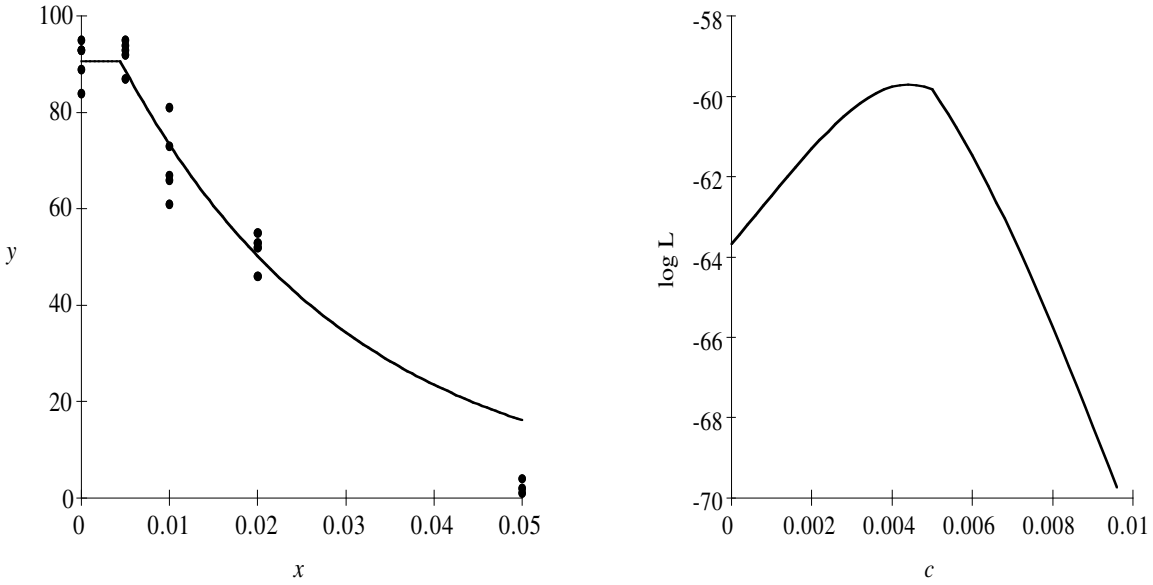

Figure 3: DCA data (21 days), adjusted model and profile log-likelihood.

The results in Table 6 show that the estimated NEC is always smaller than the NOEC, however the differences are not statistically significant (according to the confidence intervals obtained), and were also not considered significant in practical terms, by the ecotoxicologists.

The relevant contribution of using a parametric NEC is the possibility to assess the precision of the estimates via the confidence intervals.

Table 6
Results for the data sets in Tables 4 and 5 (without the last concentration)

Data	NOEC	Model	$\hat{c}$	$\hat{m}$	$\hat{l}$	AIC	95% Conf. Interval (PL)
DCA 14 days	0.005	**I**	**0.0016**	**454.9**	**32.5**	**235.2**	**(0;0.0073)**
		II	0.0016	16.7	32.4	244.2	(0;0.0075)
		III	0.0030	18.0	32.5	239.4	(0;0.0085)
DCA 21 days	0.005	I	0.0040	2610.2	98.9	129.1	(0.0018;0.0059)
		II	**0.0044**	**37.97**	**90.7**	**127.4**	**(0.0017;0.0061)**
		III	0.0050	42.0	90.8	130.6	(0.0007;0.0071)
Chemical 14 days	0.6	**I**	**0.52**	**19.9**	**49.2**	**248.5**	**(0.17;1.23)**
		II	1.2	1.33	44.6	272.2	(0.97;1.99)
		III	1.18	1.20	45.2	252.5	(1.01;1.29)
Chemical 21 days	1.3	**I**	**1.24**	**43.3**	**98.9**	**255.9**	**(1.05;1.99)**
		II	1.24	0.54	98.7	258.6	(1.09;1.99)
		III	1.24	0.54	99.1	260.8	(1.03;1.95)

An interesting aspect of the results obtained is that the estimated NEC increases from 14 days to 21 days in both experiments (a trend also observed for the NOEC in the second experiment). This can be explained as an adaptation of the organism to adverse conditions. On the other hand the confidence intervals are narrower, meaning more reliable estimates, for 21 days than for 14 days, even when the number of observations decreases. A similar behavior was observed in Guilhermino *et al*. (1998) for EC_X estimates.

4. CONCLUSIONS

Following the increasing concern about the use of No Observed Effect Concentrations (NOEC) in toxicology studies we proposed three simple non-linear models where a parametric No Effect Concentration (NEC) is included as a threshold parameter.

This approach does not have the disadvantages of the NOEC and has proved to work well both in simulated and in real data, as shown in Subsections 2.5 and 3.2.

The models proposed are simpler than those considered in Cox (1987) (which in any case need some modification before being used on the kind of data under study) and in Kooijman and Bedaux (1996a, 1996b).

An important conclusion from this work is the failure of the resampling methods, particularly of the bootstrap. This can be explained by a failure on some regularity condition necessary for the consistency of the bootstrap distribution of the estimator under consideration, and related to the non differentiability of the log likelihood. This case constitutes thus a warning against the blind use of resampling methods.

If these models are to be applied in practice a careful look shall be given to the design of the experiment. This is an issue requiring further investigation. For the moment we recommend the use of equally spaced concentrations located between the origin and three to four times the expected location of the NEC (usually known from previous or preliminary experiments).

Another advantage of these models, not mentioned previously, is the possibility to extend them in order to include relevant covariates, such as body size of the test animal. Time can also be included in the model (if the number of observation times is enough) but not as an ordinary covariate because observations at different times are usually longitudinal.

A problem deserving further study is the robustness properties of the proposed estimation procedures against small deviations from the underlying assumptions. An alternative is iteratively reweighted least squares with small weights for possible outliers and leverage points, but at the cost of additional computational effort.

All the computations described in this paper were run on a DEC 7620 (under Open VMS AXP). A Fortran code for point and interval estimation using the methods discussed is available by e-mail from the first author.

Acknowledgments

This work was partially supported by the project Praxis/PCEX/P/MAT/41/96. We would like to thank Trevor Cox for many helpful suggestions. We also thank two anonymous referees for their suggestions regarding the contents of the paper and possible extensions of this work.

Appendix

For Model I (and Model II with y_j replaced by $\log y_j$ and l by $\log l'$) we have to minimize:

$$Q = \sum_{j=1}^{n} \left[y_j - l + m(x_j - c) I(x_j - c) \right]^2 \tag{5}$$

The normal equations give

$$\hat{l} = \bar{y} + \hat{m} \frac{\sum_{j=1}^{n} (x_j - c) I(x_j - c)}{n} \tag{6}$$

$$\hat{m} = \frac{\bar{y} \sum_{j=1}^{n} (x_j - c) I(x_j - c) - \sum_{j=1}^{n} y_j (x_j - c) I(x_j - c)}{\sum_{j=1}^{n} (x_j - c)^2 I(x_j - c) - \left[\sum_{j=1}^{n} (x_j - c) I(x_j - c) \right]^2 / n} \tag{7}$$

(These are easy to check because, given *c*, the model is just a simple linear regression with independent variable $(x - c)I(x - c)$.)

In order to find $\hat{c}$, the value of *c* is chosen in turn as each point of a grid covering the range of possible values of c (noting that $0 \le c \le \max\{x_1, \ldots, x_n\}$). For each *c*, the corresponding solutions for $\hat{l}$ and $\hat{m}$, using equations (6) and (7), are found and substituted into equation (5), obtaining $Q(c)$. A plot of $Q(c)$ versus *c* will yield the desired estimate, $\hat{c}$.

For Model III the procedure is similar. The function to maximize is

$$\log L = -\sum_{j=1}^{n} \lambda_j + \sum_{j=1}^{n} y_j \log \lambda_j - \sum_{j=1}^{n} \log(y_j!) \tag{8}$$

with λ_j defined in equation (4). The maximum likelihood estimates are given by the equations

$$\hat{l} = \frac{\sum_{j=1}^{n} y_j}{\sum_{j=1}^{n} \exp\left[-\hat{m}(x_j - c)I(x_j - c)\right]} \tag{9}$$

$$\sum_{j=1}^{n}\left[(x_j - c)I(x_j - c) - K(c)\right]\exp\left[-\hat{m}(x_j - c)I(x_j - c)\right] = 0 \tag{10}$$

where

$$K(c) = \frac{\sum_{j=1}^{n} y_j (x_j - c)I(x_j - c)}{\sum_{j=1}^{n} y_j}.$$

For each *c*, equation (10) can be solved numerically for $\hat{m}$ (for instance using the bisection method), the result introduced into equation (9), to obtain $\hat{l}$, and both $\hat{m}$ and $\hat{l}$ substituted into equation (8) to find $\log L(c)$. Again note that given *c,* the problem of finding $\hat{l}$ and $\hat{m}$ can be seen as an ordinary Poisson regression with covariate $(x - c)I(x - c)$, and a statistical package handling Poisson regression can be used instead of the above equations.

REFERENCES

CEE. 1979. Directive du Conseil. 79/831. J.O.L. 259/10, 15-10-79.

Chapman PF, Crane M, Wiles JA., Noppert F, McIndoe, EC. Eds. 1995. *Asking the Right Questions: Ecotoxicology and Statistics. (The report of a Workshop held at Royal Halloway University of London, Egham, Surrey, United Kingdom, 26-27 April, 1995)* Society of Environmental Toxicology and Chemistry — Europe: Brussels.

Chapman PM, Caldwell RS, Chapman PF. 1996. A warning: NOECs are inappropriate for regulatory use. *Environmental Toxicology and Chemistry* **15**: 77-79.

Cox C. 1987. Threshold dose-response models in toxicology. *Biometrics* **43**: 511-523.

Cox DR, Hinkley DV. 1974. *Theoretical Statistics*. Chapman and Hall: London.

Dunnett CW. 1964. New tables for multiple comparisons with control. *Biometrics* **20**: 482-491.

EPA. 1991. *Technical support document for water quality-based toxics control.*

EPA/505/2-90-001. U. S. Environmental Protection Agency, Office of Water: Washington DC.

Feder PI. 1975. The log likelihood ratio in segmented regression. *The Annals of Statistics* **3**: 84-97.

Guilhermino L, Sobral O, Chastinet C, Ribeiro R, Gonçalves F, Silva MC, Soares AMVM. 1999. A *Daphnia magna* first-brood chronic test: an alternative to the conventional 21-day chronic bioassay? *Ecotoxicology and Environmental Safety* **42**: 67-74.

Hoekstra JA, van Ewijk. 1993. Alternatives for the no-observed-effect level. *Environmental Toxicology and Chemistry* **12**: 187-194.

Kooijman SALM. 1996. An alternative for NOEC exists but the standard model has to be abandoned first. *Oikos* **75**: 310-316.

Kooijman SALM, Bedaux JJM. 1996a. Analysis of toxicity tests on *Daphnia* survival and reproduction. *Water Research* **30**: 1711-1723.

Kooijman SALM, Bedaux JJM. 1996b. Some statistical properties of estimates of no-effects concentrations. *Water Research* **30**: 1724-1728.

Küchenhoff H. 1997. An exact algorithm for estimating breakpoints in segmented generalized linear models. *Computational Statistics* **12**: 235-247.

Kühn R, Pattard M, Pernak K-D, Winter A. 1989. Results of the harmful effects of water pollutants to *Daphnia magna* in the 21 day reproduction test. *Water Research* **23**: 501-510.

OECD. 1997. *Report of the Final Ring Test of the* Daphnia magna *Reproduction Test.* OECD Environmental Health and Safety Publications. Series on Testing and Assessment, nº 6, OECD/GD(97)19: Paris.

OECD. 1998. *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data*. OECD Environmental Health and Safety Publications. Series on Testing and Assessment, nº 10, ENV/MC/CHEM(98)18: Paris.

Schwartz PF, Gennings C, Chinchilli V. 1995. Threshold models for combination data

from reproductive and developmental experiments. *Journal of the American Statistical Association* **90**: 862-870.

Seber GAF, Wild CJ. 1989. *Nonlinear Regression*. Wiley: New York.

Self SG, Liang K-Y. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**: 605-610.

Shao J, Tu D. 1995. *The Jackknife and Bootstrap*. Springer: New York.

Silvapulle JM. 1991. On testing for a threshold value (correspondence). *Biometrics* **47**: 1628-1629.

Ulm KW. 1989. On the estimation of threshold values (correspondence) *Biometrics* **45**: 1324-1326.

Ulm KW. 1991. A statistical method for assessing a threshold in epidemiological studies. *Statistics in Medicine* **10**: 341-349.

Williams DA. 1972. The comparison of several dose levels with a zero dose control. *Biometrics* **28**: 519-531.