

Response Surfaces for Overdispersion in the Study of the Conditions for Fish Eggs Hatching

J. K. Lindsey

Biostatistics, Limburgs Universitair Centrum, Diepenbeek, Belgium

email: jlindsey@luc.ac.be

SUMMARY. Response surface methodology, originally developed for determining optimal conditions in industrial experiments, was early adapted to experiments in marine ecology. However, these involved studying the shape of the complete response surface, not only detecting the optimum, and often had counts or durations as the response variable. Thus, nonlinear, nonnormal response models were required. For counts, binomial and beta-binomial models have been used, the latter because of substantial overdispersion. In closely controlled experiments, overdispersion among units held under the same conditions might indicate that some mishap has occurred in conducting the study. One possible check is to model the dispersion as a second response surface. This procedure is used to show that overdispersion in fish egg hatching experiments has a biological explanation in that it occurs only under suboptimal hatching conditions.

KEY WORDS: Beta-binomial distribution; Generalized nonlinear model; Overdispersion; Response surface; Transformation.

1. Introduction

In the 1950s, Box and his colleagues developed the techniques of response surface analysis for use in industrial experiments to determine the optimum conditions of production, especially in chemical processes (Box and Wilson, 1951; Box, 1954; Box and Draper, 1959; Box and Hunter, 1965; Box and Hill, 1967). Shortly thereafter, Alderdice pioneered the introduction of these techniques into experimental marine ecology (Alderdice, 1963, 1976; Forrester and Alderdice, 1966; Alderdice and Forrester, 1968, 1971a, 1971b, 1974; Lindsey, Alderdice, and Pienaar, 1970; Alderdice and Velsen, 1971).

The early work on response surfaces presented a number of unsuitable constraints for its applications in marine ecology. For the detection of the point of optimal response conditions, as in industrial experiments, a regression that is quadratic in the explanatory variables is often sufficient. However, the ecological goal soon became much wider than, e.g., only determining under what conditions the maximum proportion of eggs hatch. The complete shape of the surface, under varying environmental conditions (but controlled in the laboratory), was of direct interest. Simple polynomials were not adequate for this, so that transformations of the explanatory variables were adopted (Box and Tidwell, 1962; Lindsey et al., 1970).

In addition, the original response surface methodology involved at least approximately normally distributed response variables, but, in marine ecology, the response was often a count or a duration, such as the number of eggs hatching in a tank or the time to hatch. These experiments were being conducted in the period when generalized linear models were only in the process of being developed. Thus, the initial approach to this problem was to transform the response (e.g.,

the proportion of eggs hatching) as well as the explanatory variables (Box and Cox, 1964; Lindsey et al., 1970).

The combination of these constraints soon led to the development of generalized nonlinear models for a variety of response distributions (Lindsey, 1971, 1974, 1975). However, such models did not take into account the overdispersion that might arise in such count data. For example, in a response surface design for fish eggs hatching, four sets of eggs were kept in separate cells of each tank corresponding to a point of the design chosen, i.e., to each combination of conditions studied. Subsequent analysis (Lindsey, 1993, pp. 160–164, 1995, pp. 232–234) showed that overdispersion was indeed present. There is more variability among cells within a tank, all under the same controlled conditions, than would be expected under a binomial distribution.

This is a disturbing conclusion for the overall quality of the experiments. Here I will perform a reanalysis to show that an adequate model, taking into account variation in the dispersion under different conditions, can lead to a biological explanation for the observed variability within cells of a tank.

2. Models for Varying Overdispersion

As suggested by Lindsey (1974), there is no reason that regression modeling should be restricted to changes in the mean. Changes in dispersion, or other shape parameters, under varying conditions might often be necessary. For example, Lindsey and Laurent (1996), using a double Poisson distribution (Efron, 1986), found that the dispersion of micronuclei counts can change over time after the exposure of a subject to a toxic substance.

Thus, it seems worthwhile to investigate whether such a phenomenon was present in the marine ecology experiments

described in Section 1. What we require will be two response surfaces for a given experiment: the usual one describing changes in hatching success under the different assigned conditions and another describing changes in the heterogeneity among cells within tanks (or, what is the same thing, correlation among eggs within each cell) under these same conditions.

The usual model for overdispersed binomial data is the beta-binomial distribution (Skellam, 1948). One way that this can be obtained is by assuming that the binomial probability varies in a heterogeneous population (here, among cells under identical conditions) according to a beta distribution; this is then integrated to obtain the marginal beta-binomial distribution of the counts. Here, I will use the following parametrization for this distribution:

$$f(y; \pi, \psi) = \binom{n}{y} \frac{B(\pi e^\psi + y, (1 - \pi)e^\psi + n - y)}{B(\pi e^\psi, (1 - \pi)e^\psi)} \quad (1)$$

where, for our data, n is the number of eggs in a cell of tank, y the number hatching, π the probability of an egg hatching, ψ a dispersion parameter, and $B(\cdot)$ the beta function. The mean and variance are, respectively, $\mu = n\pi$ and

$$\sigma^2 = \pi(1 - \pi) \frac{e^\psi + n}{e^\psi + 1}.$$

The correlation among eggs in a cell is given by $\rho = 1/(\exp(\psi) + 1)$, with positive values (i.e., $\psi < \infty$) indicating overdispersion (Lindsey, 1993, pp. 159–160). When $\psi \rightarrow \infty$, it becomes the binomial distribution.

I will use logistic regression for modeling variation in the probability of hatch:

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \eta_1(\mathbf{x}_i, \boldsymbol{\theta}), \quad (2)$$

where \mathbf{x}_i is a vector of (usually two or three) explanatory variables defining the conditions at design point i , $\boldsymbol{\theta}$ is the unknown parameter vector, and $\eta_1(\cdot)$ is some function describing the response surface for the probability of hatch, in the simplest cases, a polynomial in \mathbf{x}_i .

To this, I add a regression equation for the dispersion:

$$\psi_i = \eta_2(\mathbf{x}_i, \boldsymbol{\zeta}), \quad (3)$$

where $\boldsymbol{\zeta}$ is another unknown parameter vector and $\eta_2(\cdot)$ is some function describing the response surface for the heterogeneity among cells of a tank under the same conditions, again, in the simplest cases, a polynomial in \mathbf{x}_i . Note that this regression model induces a logistic regression for the correlation, ρ_i , among the eggs in a cell and that the way in which the variance changes will be a rather complex combination of the two regression equations.

All model parameters were estimated by maximum likelihood. The models were fitted by substituting the appropriate regression equations based on (2) and (3) directly into the log likelihood derived from equation (1). This function was then minimized by a Newton–Raphson procedure that calculates numerical second derivatives.

All the models to be considered will be compared using a direct likelihood approach based on the full log likelihood so obtained. In other words, models that predict the observed data better (i.e., make them more probable) will be considered to be more suitable for describing the conditions for fish

eggs hatching. When the numbers of parameters estimated differ among models, the log likelihood will be penalized by adding to it the number of estimated parameters, as in the Akaike information criterion (AIC; see Akaike, 1973). Smaller values indicate a preferable model, but the absolute size has no meaning because it depends on which proportionality constants have been included. (Here, all constants of the probability distribution are included.) For example, one model with an AIC that is 1 point lower than another model could have a completely redundant parameter added and still fit as well as the second model. Notice that the usual problems of inference in the presence of overdispersion do not arise with this approach; the need for each regression coefficient can be checked by removing it from the model and comparing AICs.

3. Modeling Fish Eggs Hatching

In the context of fish eggs hatching, there are (as yet) no theoretical equations for $\eta_1(\cdot)$ and $\eta_2(\cdot)$ based on the underlying biological mechanisms. Thus, in the models to follow, I will take polynomial regression equations for the probability ($k = 1$) and dispersion ($k = 2$) parameters of the beta binomial distribution,

$$\eta_k(\mathbf{x}_i, \boldsymbol{\theta}) = \beta_{0k} + \beta_{1k}g_1(x_{1i}) + \beta_{2k}g_1(x_{1i})^2 + \beta_{3k}g_2(x_{2i}) + \beta_{4k}g_2(x_{2i})^2 + \beta_{5k}g_1(x_{1i})g_2(x_{2i}),$$

for two explanatory variables, with the obvious extension for three, where $g_j(x_{ji}) = x_{ji}$ without transformation and $g_j(x_{ji}) = x_{ji}^{\alpha_j}$ when power transformed. Because these are not theoretical but empirical functions, we are not interested in estimating an exact value of the transformations. Instead, the power transformations are usually rounded to interpretable values such as the square root, logarithm, or reciprocal; this will be done here. They will, nevertheless, be counted in the penalty for the AIC because they have been obtained from the data.

3.1 English Sole

The English (or lemon) sole (*Parophrys vetulus*) is found in the Pacific Ocean on the west coast of North America. The experiment considered here was conducted at the Fisheries Research Board of Canada Biological Station in Nanaimo, British Columbia, in 1966 to study total and viable hatching of the eggs (as well as incubation period and size of larvae) under an experimental design assigning various levels of salinity and temperature. All eggs were from the same source. There were 17 combinations of conditions, with one tank containing four separate cells, for each condition except one point, which had two tanks (eight cells). Thus, the four cells per tank represent a pseudoreplication. Alderdice and Forrester (1968) provide the data on total and viable hatch. The former, which are used here, are the total numbers hatching out of those in the cell, whether viable or not; they are also reproduced in Lindsey (1993, p. 162; 1995, p. 252).

Salinity ranged from 10 to 40 parts per thousand, with maximum standard error of 0.02 over the course of the experiment. Temperature ranged from 4°C to 12°C, with a maximum standard error of 0.1. Thus, the experimental conditions were very tightly controlled.

Standard normal theory polynomial regression was originally used to estimate response surfaces for all these variables, as described by Alderdice and Forrester (1968). Lindsey et al. (1970) reanalyzed the data for total hatch using normal (quadratic) regression, but with estimated power transformations for both response and explanatory variables. When a binomial distribution is used (nonlinear logistic regression), the transformations are estimated to be about 0.2 for both salinity and temperature (Lindsey, 1993, p. 164). The same transformations are also reasonable for the beta-binomial distribution and will be used here; the nonlinear optimization problem proved too difficult to estimate them independently for this distribution.

Overdispersion can be detected because four cells were used in each tank at each combination of conditions. The AIC for a model using a factor variable so that there is a different probability of hatch for each of the 72 cells is 292.4, whereas that with a different probability for each design point (pair of conditions) but the same for the set of four cells at that point is 568.6, indicating substantial overdispersion. The AICs for quadratic response surface models using the binomial distribution (logistic regression) without and with the transformations of the two explanatory variables are given in the first line of Table 1. We see that, even with the transformed explanatory variables, the binomial model fits substantially less well than the model with a different probability at each design point.

When we fit a standard beta-binomial regression model, we obtain a substantial improvement over the binomial (which is a special case of it), as seen in the second line of the table. (This line also confirms that the transformation parameters taken from the binomial model improve the beta-binomial model even though they have not been reestimated.) Finally,

Table 1

AICs for various models fitted to the data on hatching of English sole eggs. The transformed model has both salinity and temperature raised to the power 0.2.

	Untransformed	Transformation
Binomial	1383.8	868.3
Beta binomial, constant dispersion	392.1	371.1
Beta binomial, dispersion regression	383.4	365.8

we can allow the dispersion, or correlation among eggs in a cell, to vary with the experimental conditions; we find yet more improvement in the AIC, as shown in the last line of the table. Using transformations for the covariates in the dispersion regression does not improve the model, so they are left untransformed. This regression for the dispersion can be further simplified by eliminating two terms, yielding a final AIC of 363.9 and the equations

$$\begin{aligned} \eta_1 &= -264.34 + 108.62x_{1i}^{0.2} - 35.63x_{1i}^{0.4} + 217.26x_{2i}^{0.2} \\ &\quad - 80.70x_{2i}^{0.4} + 15.73x_{1i}^{0.2}x_{2i}^{0.2} \\ \eta_2 &= 5.02 - 0.0045x_{1i} - 0.029x_{2i} + 0.020x_{1i}^2, \end{aligned}$$

where x_1 is salinity and x_2 temperature.

The response contours for the two parameters of the beta-binomial distribution described by these two equations are plotted in Figure 1. Although the distributional assumptions and the transformation of temperature are quite different, the surface for the probability of hatch resembles that obtained

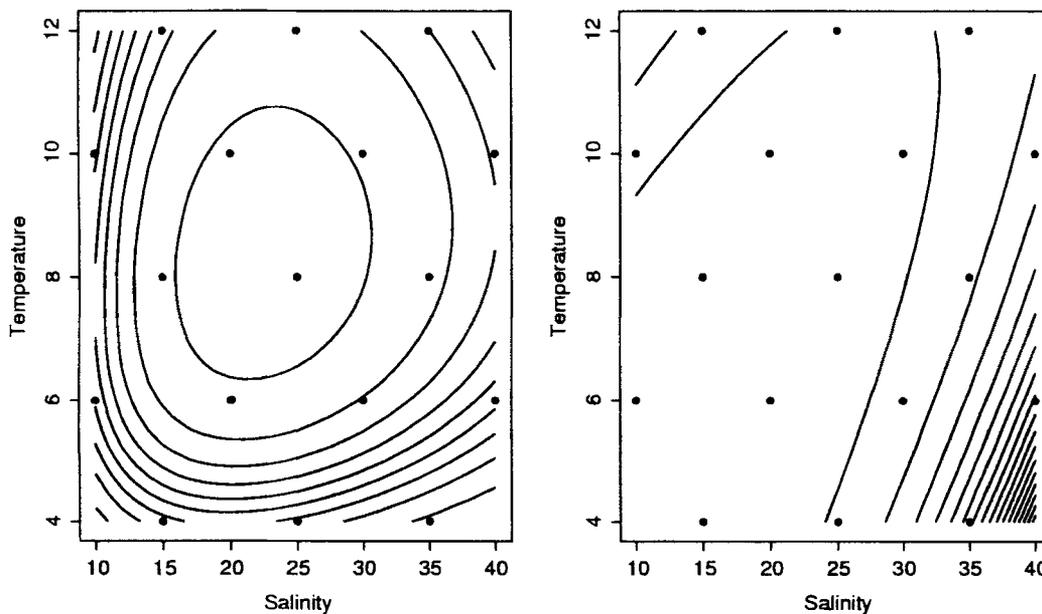


Figure 1. Contours for the response surfaces for the probability of sole eggs hatching (left) and for the correlation among the eggs (right) along with the design points where observations were made. Probability contours range from 0.1 to 0.9 in steps of 0.1; correlation contours range from 0.04 to 0.32 in steps of 0.02.

by Lindsey et al. (1970). As we have seen, the transformations substantially reduce the AIC; without them, the contour plot for the probability of eggs hatching would have an elliptical form. The contours for the correlation are indistinguishable with and without the transformation.

The maximum probability of hatch (94%) occurs at about (22.3, 8.4). However, the interesting result is that the correlation response surface is very flat at about zero in the whole region of good hatching conditions but increases steeply away from them (in the lower-right corner). Thus, we find that the overdispersion is present only when the conditions for the eggs are very poor; it is especially dependent on salinity. According to this model, under good hatching conditions, the results from a set of four cells are very similar, not exceeding binomial variation.

Note, however, that the fit for this final model is considerably worse than that for the completely saturated model with a different probability for each cell. This appears to be due primarily to a lack of adequate flexibility of the quadratic regression, even with the transformed explanatory variables. (Only slight further improvement of our model, according to the AIC, can be obtained by using the exact maximum likelihood estimates of the two transformations.)

3.2 Pacific Cod

Let us now further investigate our hypothesis about the source of overdispersion in fish hatching experiments by considering a second study performed in the same laboratory. The one that I will look at has the added attraction of involving the simultaneous variation of three conditions: salinity, temperature, and oxygen content of the water. The full quadratic model will have 10 regression coefficients for each of the two parameters plus the three power transformations for a total of 23 parameters.

This second study involves the Pacific cod (*Gadus macrocephalus*), a fish that occurs around the rim of the north Pacific Ocean. A first experiment (Forrester and Alderdice, 1966), similar to that described previously for the English sole, was conducted in 1965 using only salinity and temperature. These preliminary results allowed a more complex experiment to be designed in 1966 whereby oxygen content was also introduced, as described by Alderdice and Forrester (1971b). For more precision near the optimum, additional design points were added in a complementary experiment in 1968. Here, over the two years, there were a total of 19 design points, all distinct, with four cells at each. Again, we look at total eggs hatched out of those in each cell; the data are provided in Alderdice and Forrester (1971b) and reproduced in Lindsey (1995, p. 233).

For this experiment, the original publication used normal regression with transformations of all variables, including the response. Reanalysis using a binomial or beta-binomial distribution (Lindsey, 1995, pp. 231–234) yields, respectively, approximately -1 for salinity, -0.5 for temperature, and 2 for oxygen content. These transformations will be used in what follows.

The results are remarkably similar to those obtained for English sole, which is surprising given the very different nature of the two species. The AIC for a model with a different probability of hatch for each of the 76 cells is 287.9, whereas

that with a different probability for each design point but the same for the set of four cells at that point is 681.0, again indicating substantial overdispersion. The AICs for quadratic response surface models using the binomial distribution (logistic regression) without and with the transformations of the three explanatory variables are given in the first line of Table 2. As expected, these models do not fit well. The results in the lower part of the table are also similar to those obtained previously for the English sole. The apparent poorer fit of the transformed regression for the dispersion parameter is due to the fact that 5 of the 10 terms can be eliminated from that equation, giving an AIC of 360.2 and the equations

$$\begin{aligned}\eta_1 &= -26.36 + 396.58/x_{1i} - 2992.5/x_{1i}^2 + 60.65/\sqrt{x_{2i}} \\ &\quad - 73.88/x_{2i} + 0.0550x_{3i}^2 - 0.000177x_{3i}^4 \\ &\quad + 98.39/(x_{1i}\sqrt{x_{2i}}) - 1.31x_{3i}^2/x_{1i} + 0.0986x_{3i}^2/\sqrt{x_{2i}} \\ \eta_2 &= 2.34 + 913.5/x_{1i}^2 - 74.86/(x_{1i}\sqrt{x_{2i}}) \\ &\quad - 1.34x_{3i}^2/x_{1i} + 0.162x_{3i}^2/\sqrt{x_{2i}},\end{aligned}$$

where x_1 is salinity, x_2 temperature, and x_3 oxygen content. In contrast to the model for the English sole, here the transformations have been retained in the dispersion regression.

Because we have a four-dimensional model, no simple representation of the complete response surface is possible. Contour plots for cross-sections, approximately through the maximum hatch of 0.92, at about (22.0, 3.5, 11.6), are shown in Figure 2. This maximum is considerably different than that found in the original publication, using a Box–Cox transformation: maximum probability of hatch, 1.024 (!) at about (15.5, 3.7, 9). The new results correspond much more closely to the empirical observations.

Because of the small number of points in a three-dimensional design space, much less information is available about changing dispersion than for the English sole. However, it is worth considering the results here because of the agreement between the two experiments. For the Pacific cod, the correlation response surface is also flat at zero under optimal hatching conditions but rises rapidly under poor conditions: The upper left for salinity–temperature, the upper left and lower right for salinity–oxygen, and the upper right and lower left for temperature–oxygen in the lower panel of Figure 2. For both these and the English sole data, inspection of the raw data confirms the much higher variability of response among a set of four cells when under suboptimal conditions.

Table 2
AICs for various models fitted to the data on hatching of Pacific cod eggs

	Untransformed	Transformation
Binomial	1109.4	804.3
Beta binomial, constant dispersion	382.6	362.5
Beta binomial, dispersion regression	383.6	362.8

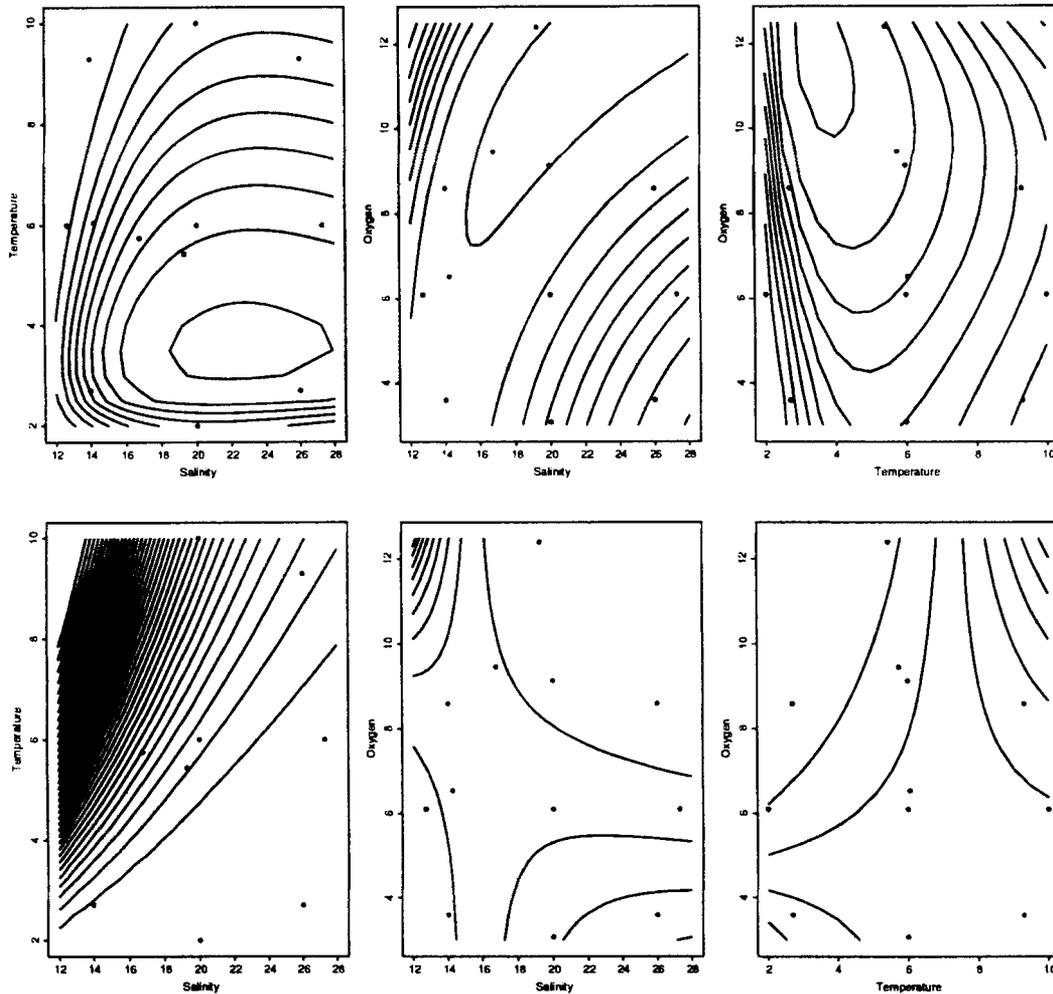


Figure 2. Three cross-sections of contours for the response surfaces for the probability of cod eggs hatching (top panel) and for the correlation among the eggs (bottom panel) along with the design points where observations were made. Probability contours range from 0.1 to 0.9 in steps of 0.1; correlation contours range from 0.02 to 1.0 in steps of 0.02.

4. Discussion

Accounting for overdispersion in count data is usually considered to be a technique that is necessary to draw correct inferences about the mean (or probability) parameters. Here, I have demonstrated that appropriate study of overdispersion can have substantial interest in its own right. As might be expected, these more complex models, accounting for much of the overdispersion, do not noticeably modify the estimated response surface for the probability of eggs hatching. However, the second response surface, for the dispersion, heterogeneity, or correlation, contributes important new information.

In a tightly controlled laboratory experiment, heterogeneity among units (here cells), supposedly under identical conditions, can indicate some mishap in conducting the experiment. The present analysis demonstrates that, in these experiments, the overdispersion might have a biological explanation linked with the suboptimal conditions. This could perhaps be due to greater sensitivity to manipulation or to interaction (competition) among such eggs within cells subjected to the most unfavorable conditions.

If a proper replication of tanks under the same conditions had been used, as well as the pseudoreplication by means of cells within a tank, the model for overdispersion would have had to be more complex than that used here. In such a case, variability among cells within a tank would probably be less than that among tanks under the same conditions.

Diagnostic tools for models involving varying dispersion or shape parameters have not been developed. Those traditionally used for generalized linear models perform badly (Lindsey and Jones, 1997), often not even detecting obvious problems found by simply inspecting the data. Here, there are several sources of lack of fit: The beta-binomial distribution and logit link and the shapes of the two response surfaces as constrained by the transformed polynomials. Nevertheless, most of the variability in the data has been accounted for: the AIC was reduced from 1384 for the untransformed binomial model to 364, with a lower bound of 292 for the English sole and from 1109 to 360, with a lower bound of 288 for the Pacific cod. The regression equations should capture the essential shape of the response surfaces, although it would be interesting to see what

results a nonparametric procedure, such as local polynomial smoothing, yields.

Following on the tradition of normal regression models with their constant variance, statisticians are unaccustomed to considering regression equations for other parameters than the mean. The fact that the variance is a function of the mean in generalized linear models has not greatly modified this situation, despite the recommendations of Lindsey (1974). Simultaneous regression models for several parameters of a distribution can often provide enlightening information.

The models were all fitted using a general nonlinear regression function written in R (Ihaka and Gentleman, 1996) that allows twin regressions for about a dozen different distributions, including the beta binomial.

ACKNOWLEDGEMENTS

Thanks to Robert Gentleman and Ross Ihaka for developing the R software and to Marc Aerts and Gerda Claeskens for pointing out that the dispersion can vary over the design space in this type of response surface data. Two referees and an associate editor provided detailed comments that led to substantial improvements in the paper.

RÉSUMÉ

La méthode des surfaces de réponses, développée à l'origine pour déterminer les conditions optimales dans des expériences industrielles, a très tôt été adaptée aux expériences d'écologie marine. Cependant, elle était utilisée pour l'étude de la forme de la surface de réponse, ne se limitait pas à la seule recherche de l'optimum, et avait souvent des comptages ou des durées comme variables de réponse. Alors, des modèles non linéaires, non Gaussiens devenaient nécessaires. Pour les comptages, des modèles binomiaux ou beta-binomiaux ont été utilisés, ces derniers dans des cas de surdispersion. Dans des expériences très contrôlées, la surdispersion pour les unités apparue dans les mêmes conditions peut révéler la survenue d'accidents durant l'étude. Une des vérifications possibles est de modéliser la dispersion par une deuxième surface de réponse. Cette procédure est utilisée pour montrer que la surdispersion observée dans des expériences de couvainon d'œufs de poissons a une explication biologique car elle apparaît seulement dans des conditions de couvainon suboptimales.

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Inference Theory*, B. N. Petrov and F. Csàki (eds), pp. 267–281. Budapest: Akadémiai Kiadó.
- Alderdice, D. F. (1963). Some effects of simultaneous variation in salinity, temperature and dissolved oxygen on the resistance of young Coho Salmon to a toxic substance. *Journal of the Fisheries Research Board of Canada* **20**, 525–550.
- Alderdice, D. F. (1976). Some concepts and descriptions of physiological tolerance: rate-temperature curves of poikilotherms as transects of response surfaces. *Journal of the Fisheries Research Board of Canada* **33**, 299–307.
- Alderdice, D. F. and Forrester, C. R. (1968). Some effects of salinity and temperature on early development and survival of the English Sole (*Parophrys vetulus*). *Journal of the Fisheries Research Board of Canada* **25**, 495–521.
- Alderdice, D. F. and Forrester, C. R. (1971a). Effects of salinity and temperature on embryonic development of the Petrale Sole (*Eopsetta jordani*). *Journal of the Fisheries Research Board of Canada* **28**, 727–744.
- Alderdice, D. F. and Forrester, C. R. (1971b). Effects of salinity, temperature, and dissolved oxygen on early development of the Pacific Cod (*Gadus macrocephalis*). *Journal of the Fisheries Research Board of Canada* **28**, 883–902.
- Alderdice, D. F. and Forrester, C. R. (1974). Early development and distribution of the Flathead Sole (*Hippoglossoides elassodon*). *Journal of the Fisheries Research Board of Canada* **31**, 1899–1918.
- Alderdice, D. F. and Velsen, F. P. J. (1971). Some effects of salinity and temperature on early development of Pacific Herring (*Clupea pallasii*). *Journal of the Fisheries Research Board of Canada* **28**, 1545–1562.
- Box, G. E. P. (1954). The exploration and exploitation of response surfaces: Some general considerations and examples. *Biometrics* **10**, 16–60.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26**, 211–252.
- Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of response surface designs. *Journal of the American Statistical Association* **54**, 622–654.
- Box, G. E. P. and Hill, W. J. (1967). Discrimination among mechanistic models. *Technometrics* **9**, 57–71.
- Box, G. E. P. and Hunter, W. G. (1965). The experimental study of physical mechanisms. *Technometrics* **7**, 23–42.
- Box, G. E. P. and Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics* **4**, 531–550.
- Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimal conditions. *Journal of the Royal Statistical Society, Series B* **13**, 1–45.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**, 709–721.
- Forrester, C. R. and Alderdice, D. F. (1966). Effects of salinity and temperature on embryonic development of the Pacific Cod (*Gadus macrocephalus*). *Journal of the Fisheries Research Board of Canada* **23**, 319–340.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational Graphics and Statistics* **5**, 299–314.
- Lindsey, J. K. (1971). Programs for the analysis of nonnormal data. *Fisheries Research Board of Canada Technical Report* **275**, Fisheries Research Board of Canada. Ottawa.
- Lindsey, J. K. (1974). Construction and comparison of statistical models. *Journal of the Royal Statistical Society, Series B* **36**, 418–425.
- Lindsey, J. K. (1975). Likelihood analyses and tests for binary data. *Applied Statistics* **24**, 1–16.
- Lindsey, J. K. (1993). *Models for Repeated Measurements*. Oxford: Oxford University Press.
- Lindsey, J. K. (1995). *Modelling Frequency and Count Data*. Oxford: Oxford University Press.
- Lindsey, J. K., Alderdice, D. F., and Pienaar, L. V. (1970). Analysis of nonlinear models—the nonlinear response surface. *Journal of the Fisheries Research Board of Canada* **27**, 765–791.

- Lindsey, J. K. and Jones, B. (1997). Treatment–patient interactions for diagnostics of cross-over trials. *Statistics in Medicine* **16**, 1955–1964.
- Lindsey, J. K. and Laurent, C. (1996). Estimating the proportion of epithelial cells affected by exposure to ethylene oxide through micronuclei counts. *The Statistician* **45**, 223–229.
- Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. *Journal of the Royal Statistical Society, Series B* **10**, 257–261.

*Received January 1997. Revised November 1997.
Accepted December 1997.*