THE PROBABILISTIC VALIDATION OF COMPUTER SIMULATIONS USING THE BOOTSTRAP

by

Efraim Halfon

NWRI Contribution No. 88-35

Lakes Research Branch
National Water Research Institute
Canada Centre for Inland Waters
P.O. Box 5050
Burlington, Ontario L7R 4A6

MANAGEMENT PERSPECTIVE

This paper integrates statistical and system methods to reduce the computational requirements to validate a simulation model. The bootstrap is a statistical method that analyzes probabilistic events. A common way to validate ecological models is to run Monte Carlo simulations, that is, the model is run 100 to 500 times to study its probabilistic properties. By analyzing the simulation model with the bootstrap, the simulation model needs to be run only once.

ABSTRACT

A recently developed statistical method, the bootstrap, is used to compare data and model simulations. As a test case, I analyzed the validity of the relation, hypothesized by Neely, between the water solubility of an organic chemical and the ratio R of the acute lethal concentration to fish (LC50), at two different exposure times. The results of the analysis show that Neely's simulation model is likely (with an 88% probability) correct but it might have a systematic bias which makes the theoretical ratio (R) slightly higher than the observed ratio (R_0). The bootstrap is an interesting statistical method which could be commonly used in model validation using a probabilistic approach. When computing resources are limited, the bootstrap validation might substitute error analysis by Monte Carlo methods, commonly used in ecological modelling, as the mathematical model is run only once.

PERSPECTIVE-GESTION

Le présent article comprend des méthodes statistiques et systématiques visant à réduire le matériel informatique nécessaire pour valider un modèle de simulation. La méthode dite "bootstrap" est une méthode statistique servant à analyser la probabilité que certains événements se réalisent. On applique couramment les simulations de Monte Carlo pour valider les modèles écologiques et, pour ce faire, on applique le modèle de 100 à 500 fois pour étudier sa probabilité de réalisation. Lorsqu'on analyse un modèle de simulation à l'aide de la méthode "bootstrap", le modèle de simulation n'a qu'a être appliqué une seule fois.

RÉSUMÉ

On utilise une méthode statistique récemment mise au point, la méthode dite "bootstrap", pour comparer des données et des modèles de simulation. À titre de test, j'ai analysé la validité de la relation, suggérée par Neely, entre la solubilité d'une substance organique dans l'eau et le ratio R de la concentration létale aiguë pour un poisson (CL50), à deux durées d'exposition. Les résultats de l'analyse montrent que le modèle de simulation de Neely est probablement correct (probabilité de 88%), mais qu'il risque de présenter un biais systématique augmentant légèrement la valeur du ratio théorique (R) par rapport au ratio observé (R_o). La méthode "bootstrap" est une méthode statistique intéressante qui pourrait être couramment utilisée pour valider des modèles par une approche statistique. Lorsque les ressources informatiques sont limitées, la méthode de validation "bootstrap" peut remplacer l'analyse des erreurs par les méthodes de Monte Carlo qui sont communément utilisées dans les modèles écologiques, car le modèle mathématique n'est appliqué qu'une fois.

INTRODUCTION

The validation of simulation models (here defined as systems of ordinary or partial differential equations) can be done by comparing the simulations with data using an appropriate framework (e.g. Mankin et al., 1977; Halfon and Reggiani, 1978). Validation may be considered successful if the model can predict reasonably well data not used in the model formulation or calibration. The wording "reasonably well" should take into account that the model is uncertain, due to inaccurate assumptions or missing information (O'Neill et al., 1980; Halfon, 1983a,b; Halfon, 1985), and that the data set used for comparison, may be incomplete and not sufficient to disqualify a hypothesis (Halfon, 1979, 1984).

In this paper, I suggest the performance of a probablistic analysis to help the validation of an ecotoxiocological model; for this purpose I use the bootstrap (Efron, 1981a, b) with a test case: a model developed by Neely (1984). The advantage of using the bootstrap over standard error analysis Monte Carlo methods, commonly used in ecological modelling (O'Neill et al., 1981), is that with the bootstrap the mathematical model is run only once, a time saving procedure when working with large complex simulation models. The disadvantage is that the bootstrap provides less information on the adequacy of the model structure than what is usually available from error analysis.

The application of the bootstrap to the validation of mathematical models is advisable given the uncertainty in the data and, for example in ecotoxicological models, uncertainty about causal links between chemical properties and observed environmental behaviour of some pollutants. A data set, even if comprehensive, is never complete since thousands of chemical compounds presently pollute the environment and are taken up by living organisms.

The bootstrap allows modelers to analyze a data set and provides a theoretical framework on which to base generalizations; indeed when we couple, as I have done in this paper, a statistical method, the bootstrap, and a system method, a simulation model, we can analyze the validity of the assumptions included in the simulation model and provide a probabilistic answer to the question of model validation. With the bootstrap we take into consideration that not all possible data were available for model development and that those available have some uncertainty associated with them given the variety of sources used to assemble the data set.

THE TEST MODEL

Neely (1984) compared measured 96h and 24h acute LC50 fish data with aquatic toxicity data estimated by a theoretical model based on water solubility of 24 chemicals. His results seem to prove the validity of the hypothesis that prediction of acute LC50 fish data is possible using his theoretical model.

The Mathematical Model

Neely established a theoretical relation between the water solubility of an organic chemical and the ratio R of the acute fish LC50 at two different time periods. The acute fish LC50 is the water concentration of a contaminant lethal to 50% of the fish in the experimental set up during a given time t, e.g., 24 hours. To prove this theoretical relation, Neely first analyzed a model of contaminant uptake and release in fish. If uptake and release are treated as first order processes, then the model is

$$dC_f/dt = k_1 C_w - k_2 C_f \qquad (1)$$

where k_1 is the rate constant for upake [litres of water flowing through the fish (kg fish hour)$^{-1}$], k_2 is the rate constant for clearance [hour^{-1}], C_f is the concentration of the contaminant in the fish [ng g^{-1}] and C_w is the concentration of the contaminant in water [ng L^{-1}].

Neely (1979) showed that

$$\log_{10} k_2 = 0.4 \log_{10} S - 2.19 \qquad (2)$$

where S is the solubility of the contaminant in mg L^{-1}.

The analytical solution of the fish model (Eq. 1) is

$$C_f = C_w \, (1-\exp\,(-k_2 t)\, k_1/k_2 \qquad (3)$$

If the water has a critical concentration, C_w*, of the contaminant capable of causing death in 50% of the fish in the experimental aquarium, then C_w* is the fish LC50 within a specified time interval, e.g. 24 or 96 hours. The lethal concentration in fish is C_f* at time t_a. In mathematical terms

$$LC50 = (k_2/k_1)\; C_f^*/(1 - \exp\,(-k_2 t_a)). \qquad (4)$$

The ratio R of the LC50 at two different times t_a and t_b is therefore

$$R = \frac{LC50\ (t_a)}{LC50\ (t_b)} = \frac{(1 - \exp\,(-k_2 t_b))}{(1 - \exp\,(-k_2 t_a))} \qquad (5)$$

Note that for large values of k_2, or high fish clearance rates, the exponents approach zero and R (Eq. 5) approaches 1; in this case the lethal concentrations are similar at times t_a and t_b. For small values of k_2, R approaches t_b/t_a since

$$\text{Lim}\ (1 - \exp\,(-x)) \rightarrow x \text{ as } x \rightarrow 0.$$

Neely (1984) showed that the ratio (Eq. 5) could be predicted by knowing the solubility of different contaminants and inserting Eq. 2 in Eq. 5. The correlation between predicted R and observed ratios R_o has a value of 0.847.

MODEL VALIDATION

One of the methods to compare predicted and observed data is to compute a goodness of fit index, another comparison method is to use a statistical linear regression model to predict the observed data using the predicted ones. If the predicted values agree with the observed values exactly, then a one to one correspondence would yield a slope of 1.0 and an intercept of zero between the two ratios, i.e.

$$R_o = 0.0 + 1.0\ R \tag{6}$$

Neely (1984) also used regression analysis to compare the observed (96h LC50/24h LC50) ratio (R_o) with a theoretical ratio (R), which was predicted by a mathematical model of contaminant dynamics in fish (Eq. 5).

The slope computed from Neely's published data using standard regression analysis is 0.845 with an intercept of 0.033; the correlation factor is 0.847. Since the slope is different from 1.0, is Neely's hypothesis wrong or is there a bias in the model?

Linear regression with errors in both variables

A fact that Neely did not take into consideration while computing the coefficients of the linear regression model (Eq. 6) is that both sets of data contained errors. The observed values have errors of observations and natural variability, the theoretical values are also uncertain since they were derived from a mathematical model, uncertain by definition. One procedure to compare uncertain variables and to compute the parameters of a linear regression model is the geometric mean (GM) linear regression method (Teisser, 1948; Quenouille, 1949; Halfon, 1985b); the method takes into account that both the X's and the Y's are measured or estimated with error. The slope b is computed from the formula,

$$b = \text{sign}(r)\ \sqrt{(Sy^2/Sx^2)} \qquad (7)$$

where

$$Sx^2 = \sum X^2 - (\sum X)^2/N, \qquad (8)$$

and

$$Sy^2 = Y^2 - (\sum Y)^2/N, \qquad (9)$$

and sign (r) is the sign, + or -, of the correlation coefficient r. N is the number of paired observations.

The intercept a is estimated as usual as

$$a = \overline{Y} - b\,\overline{X}, \qquad (10)$$

where $\overline{Y}$ and $\overline{X}$ are the averages of the Y's and X's, respectively.

Using the GM linear regression model with the data published by Neely produces a slope of 0.997 with an intercept of -0.092; the computed slope is very close to the one to one correspondence hypothesized by Neely, or

$$R_o = -0.092 + 0.997\,R \qquad r = 0.847,\ n = 24. \qquad (11)$$

while a negative intercept indicates that the theoretical values are slightly higher than the data.

THE BOOTSTRAP AND MODEL VALIDATION

The fact that the GM model has a slope very similar to the theoretical 1.0 does not imply that the model is generally valid for all chemical classes represented by the 24 chemicals used by Neely. The question is what is the probability of Neely's model being correct. The bootstrap can be used to seek this generalization, which is very useful since chemicals with different structures and chemical properties are used in Neely's model. With the bootstrap, we can infer from the observed data the validity of Neely's hypothesis for all other chemicals with similar properties without having to perform more experiments on fish toxicity.

To perform the bootstrap test, each of the 24 data points is sampled with replacement 1000 times, the bootstrap samples; from a practical point of view a random number generator is used. The statistics of interest, in this case the standard errors and the confidence limits of the slope of the intercept and of the correlation coefficient, are computed for each such bootstrap sample. Given the fact that the assumption of normality has been abandoned, the confidence limits may not be symmetrical around the mean, if the probability density function is skewed.

Results from the bootstrap calculations are shown in Table 1. The bootstrap average estimate of the correlation coefficient is 0.835; the one standard deviation (34%) confidence limits of the correlation are 0.76 and 0.90 while the 95% confidence limits are skewed to the right and the limits are not symmetrical. For the GM linear model 34% confidence limits for the slope are 0.99 and 1.15 with a bootstrap average of 1.018, and for the intercept -0.170, -0.039 and -0.112, respectively (Table 1).

DISCUSSION

The validity of the relation, hypothesized by Neely (1984), between the water solubility of an organic chemical and the ratio of the acute fish LC50 at two different time periods has been tested using the bootstrap method. The results show that the correlation between predicted and observed data is statistically significant within one standard deviation, but sometimes it may not be significant at the 95% confidence limit. However, since the observed distribution $\tilde{F}$ is skewed to the right with an average correlation of 0.835 and an upper limit of 0.98, the theory is probably correct but for a few chemicals (the lower 95% confidence interval is 0.48). For all chemicals the predicted average ratio is 1.018 with the hypothesized 1.0 falling within one standard deviation; however, the average intercept is -0.112 and the probability of the intercept being 0.0 or positive is only about 12%. Therefore the model is probably correct but it might have a systematic bias which makes the theoretical ratio somewhat higher than the observed ratio.

The GM functional regression method (Eq. 7) should be used to compute the coefficients of a linear model when measurement errors or natural uncertainty is expected in both variables X and Y. The theoretical ratios that Neely used as independent variables in the linear model were clearly uncertain; the standard linear regression method that he used produced a large underestimate of the slope, thus undermining his hypothesis whereas the correct statistical procedure showed his hypothesis right, even if the linear model might have a positive systematic bias.

The bootstrap is a computer intensive statistical method; the present analysis was performed on a CDC Cyber 171 computer and it took 27 CPU seconds for 1000 replications or bootstrap samples; Efron (1981a, 1982) suggests 128 to 512 replications since the method converges asymptotically. The method is simple enough numerically that it can be programmed on a desk microcomputer. The application to modeling problems is intriguing given the fact that large scale ecological data sets are notoriously incomplete and validation of ecological models can not be easily defended given the above mentioned uncertainty in the assumptions, model structure and data. The bootstrap is an interesting statistical method that could be coupled with system methods to establish the uncertainty of a hypothesis and to quantify a model validity in probabilistic terms.

REFERENCES

Diaconis, P. and B. Efron 1983. Computer intensive methods in statistics. Scientific American, 248: 116-130.

Efron, B. 1981a. Nonparametric standard errors and confidence intervals. Can. J. Statistics 9: 139-172.

Efron, B. 1981b. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika, 68: 589-599.

Efron, B. 1982. The jackknife, the bootstrap and other resampling methods. SIAM, 92 pp.

Halfon, E. 1979. The effects of data variability in the development and validation of ecosystem models. In B.P. Zeigler, M.S. Elzas, G.J. Klir and T.I. Oren (Eds.) Methodology in Systems Modelling and Simulation. North Holland, Amsterdam/New York/Oxford. pp. 335-343.

Halfon, E. 1983a. Is there a best model structure ? I *Ecological Modelling*, 20: 135-152.

Halfon, E. 1983b. Is there a best model structure? II *Ecological Modelling*, 20: 153-163.

Halfon, E. 1984. Predicting the environmental fate of toxic contaminants in large lakes: data requirements for mathematical models. In K.L.E. Kaiser (Ed.) *QSAR in Environmental Toxicology.* Reidel. Dordrecht/Boston/Lancaster. pp. 137-151.

Halfon, E. 1985a. Is there a best model structure? III Testing the goodness of fit. *Ecological Modelling*, 27, 15-23.

Halfon, E. 1985b. Regression method in ecotoxicology. A better formulation using the geometric mean functional regression. *Environmental Science and Technology*, 19: 747-749.

Halfon, E. and Reggiani, M.G. 1978. Adequacy of ecosystem models. *Ecolical Modelling*, 4: 41-50.

Mankin, J.B., O'Neill, R.V., Shugart, H.H. and Rust, B.W. 1977. The importance of validation in ecosystem analysis. In G.S. Innis (Ed.) New Directions in the Analysis of Ecological Systems Part 1. Simulation Counc. Proc. Ser. 5(1). Society for Computer Simulation. La Jolla, CA pp. 63-71.

Neely, B.C. 1979. Estimating rate constants for the uptake and clearanche of chemicals by fish. Environ. Sci. Technol., 13: 1506-1510.

Neely, B.C. 1984. An analysis of aquatic toxicity data: water solubility and acute LC50 data. Chemosphere, 13: 813-819.

O'Neill, R.V. Gardner,R.H. and Mankin, J.B. 1981. Analysis of parameter error in a nonlinear model. Ecological Modelling, 8, 297-311.

Quenouille, M. 1949. Approximate tests of correlation in time series. J. Royal Statist. Soc. Ser. B, 11: 18-84

Teisser, G. 1948. La relation d'allometrie: sa significance statistique et biologique. Biometrics, 4: 14-48.