RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEUHYGIËNE

BILTHOVEN

Confidence Limits for Hazardous Concentrations Based on Logistically Distributed NOEC Toxicity Data

T. Aldenberg, W. Slob

Report nr. 719102002, April 1991

This investigation has been carried out on behalf of and for account of the Directorate General for Environmental Protection within the framework of Project 719102: 'Eco-effecten'.

NATIONAL INSTITUTE OF PUBLIC HEALTH AND ENVIRONMENTAL PROTECTION

BILTHOVEN, THE NETHERLANDS

Verzendlijst

1. Directoraat-Generaal Milieubeheer, Directie Stoffen en Risicobeheersing, t.a.v. drs.C.J.van Kuijen
2. Directeur-Generaal van de Volksgezondheid
3. Directeur-Generaal Milieubeheer
4. Plv. Directeur-Generaal Milieubeheer
5. ir.F. Balk, Bureau BKH
6. ir.A.G. Berends, Duphar
7. mr.A.V. van den Berg, DGM/SR
8. drs.M.M.H.E.van den Berg, Gezondheidsraad
9. dr.C. Bodar, Gezondheidsraad
10. dr.W.A. Bruggeman, RIZA
11. dr.J.de Bruijn, DGM/SR
12. ir.P.J.de Bruijn, DHV Raadgevend Ingenieursbureau
13. dr.S. Broderius, EPA, USA
14. drs.C.A.J. Denneman, DGM
15. drs.G.van Dijk, RIN
16. dr.ir.P.E.T. Douben, RIN
17. dr.W.van Driel, Instituut voor Bodemvruchtbaarheid
18. dr.H.Eijsackers, SPBO
19. dr.R.J. Erickson, Center Lake Superior Environmental Studies
20. dr.J.Evers, DGW
21. ing.R. Faasen, RIZA
22. dr.M.van der Gaag, RIZA
23. dr.ir.H. de Heer, Plantenziektenkundige Dienst
24. dr.G.P. Hekstra, DGM/SR
25. dr.N.van der Hoeven, TNO
26. ir.D.Jonkers, Waterloopkundig Laboratorium
27. drs.F. Klijn, CML
28. ing.L.A. van der Kooij, RIZA
29. prof.dr.S.A.L.M.Kooijman, Vrije Universiteit
30. dr.B.van Leeuwen, NMF-MW
31. dr.C.J.van Leeuwen, DGM/SR
36. dr.P. Leeuwangh, Staring Centrum
37. dr.ir.Th.M. Lexmond, Landbouw Universiteit Wageningen
38. dr.H. Lokke, National Environmental Research Institute, DK
39. prof.dr.J.A.J. Metz, Rijksuniversiteit Leiden
40. dr.G.M. Murphy, OECD
43. ir.N.T.de Oude, Proctor and Gamble
44. mr.F. Pedersen, National Agency of Environmental Protection, DK
45. ir.J.G. Robberse, DGM
46. drs.M.C. van Rossenberg, DGM/DWB
47. drs.M. Ruys, DGW
48. drs.J. Schobben, DGW
49. dr.A.J.M. Schoot Uiterkamp, TNO
50. dr.A.J. Sedee, DGM/SR
51. ir.P. Stortelder, RIZA
52. dr.N. van Straalen, Vrije Universiteit
53. prof.dr.H. Udo de Haes
54. dr.J.J. Vegter, Technische Commissie Bodembescherming
55. ir.W. Visser, RIN
56. ir.M.de Vries, Waterloopkundig Laboratorium
57. dr.W. Welling, Staring Centrum
58. dr.R.J.van Wijk, AKZO
59. drs.J.A.W. de Wit, RIZA
60. ir.P.van der Zandt, DGM/SR
61. Depôt van Nederlandse Publicaties en Nederlandse Bibliografie
62. Directie van het Rijksinstituut voor Volksgezondheid en Milieuhygiëne
63. dr.W. Admiraal
64. dr.ir.C.van den Akker
65. ir.B.A. Bannink
66. dr.P.van Beelen
67. ing.G.P. Beugelink
68. dr.L. Braat
69. drs.J.H. Canton
70. ir.N.D.van Egmond
71. dr.ir.C.A.M.van Gestel
72. drs.A.van der Giessen
73. drs.J.de Greef
74. ir.G.J. Heij
75. drs.S.H. Heisterkamp
76. drs.J.M. Hesse
77. drs.J.A. Hoekstra
78. drs.J.H. Janse
79. drs.J.A. Janus
80. drs.L.H.M. Kohsiek
81. drs.A.G.A.C. Knaap
82. ir.J.M. Knoop
83. drs.P.R.G. Kramer
84. prof.dr.R. Kroes
85. dr.H.A.M.de Kruijf
86. dr.ir.J.W.de Kwaadsteniet
87. hr.R. Lammers
88. ir.F. Langeweg
89. dr.E.van Liere
90. ir.A.M.A. van der Linden
91. ir.J.B.H.J. Linders
92. drs.R. Luttik
93. dr.ir.D.van de Meent
94. dr.G. de Mik
95. drs.A.C.M.de Nijs
96. drs.E.J.van der Plassche
97. drs.P.J.T.M.van Puijenbroek
98. drs.R. Reiling
99. ir.C.J. Roghair
100. drs.C.A.F.M. Romijn
101. drs.A.J. Schouten
102. dr.W. Slooff
103. dr.J. Struijs
104. drs.T.P. Traas
105. drs.G.M.J. Tubbing
106. ir.M.A. Vaal
107. drs.T.G. Vermeire
108. ir.J.F.M. Versteegh
109. drs.F.G. Wortelboer
110. drs.D.de Zwart
111. Auteurs
121. Bibliotheek RIVM
122. Bibliotheek depot ECO/LWD
123. Hoofd Bureau Voorlichting en Public Relations
124. Bureau Projecten- en Rapporten Registratie
125. Reserve exemplaren
135.

Contents

Abstract

This paper deals with the calculation of Hazardous Concentrations of toxic substances from small sets of laboratory toxicity data, e.g. NOECs. A procedure due to Van Straalen and Denneman, as adapted from Kooijman (case n=1), in which one seeks a concentration that protects 95% of the biological species, is modified to account for the uncertainty in the estimates.

New constants are obtained by simulation, that allow the calculation of the one-sided 95% left confidence limit of the Hazardous Concentration, from the mean and standard deviation of a sample of toxicity data. This 95% confidence limit is always lower than the 95% certainty value calculated with the Kooijman(n=1)/ Van Straalen method.

We also derive constants to calculate a one-sided 50% confidence value, that overpredicts as often as it underpredicts. This value may be used as a best guess of the Hazardous Concentration. It will always be higher than the 95% certainty value of the Kooijman(n=1)/ Van Straalen method. However, by using the 50% value, one runs the risk of protecting substantially less than 95% of the biological species.

1. Introduction

This paper falls into the category of estimating safety factors for the extrapolation of laboratory toxicity data to allowable toxic substance concentrations in the field, using statistical methodology. Species differ as to the sensitivity to a toxic substance. The statistical approach focusses on some presupposed distribution of these species sensitivities for a particular substance. In fact, this article will treat some essential modifications to earlier procedures, hence for a motivational introduction we refer to the original articles: Kooijman (1987) and Van Straalen and Denneman (1989).

In Kooijman (1987), a hazardous concentration for sensitive species (*HCS*) is defined, and an algorithm is given for its computation from a sample of LC_{50} values of different test species on the basis of the logistic distribution. Several, more or less independent, components in his theory are: the choice of input data (LC_{50}s), the type of statistical distribution employed (logistic), the definition of hazardous concentration, and the statistical methodology, i.e. algorithm to calculate hazardous concentrations from small samples of toxicity data.

Our approach essentially follows a modification of Kooijman's theory by Van Straalen and Denneman (1989). Whereas Kooijman considered the probability of harming the most sensitive of a number of species, e.g. 1000, we will follow Van Straalen and Denneman in considering the probability of just harming species. This is the current approach in the Netherlands (Gezondheidsraad, 1988, DGEP, 1988-1989). We will also follow Van Straalen and Denneman in their choice and motivation with regard to the input data: *NOEC* toxicity data, instead of LC_{50} data in Kooijman. Furthermore, we will stick to the choice of the logistic distribution.

However, we will develop an alternative approach to the statistical methodology in calculating the agreed hazardous concentration levels, and this is the main concern of this paper. Hence, our presentation is statistically oriented. Of course, different calculation methodologies lead to different outcomes as regards to what seems a justifiable safety factor, or acceptable concentration, and this is where the environmental implications cannot be easily overestimated. However, these implications are discussed elsewhere.

According to Van Straalen and Denneman (1989), a concentration of a certain compound is considered hazardous for $p\%$ of the species, if the probability of selecting a species with a *NOEC* smaller than this concentration is equal to $p\%$. In other words, above this concentration, called HC_p, $(100\text{-}p)\%$ of the species is relatively safe, while $p\%$ of the species may not function properly or even worse. The general approach is to strive for 95% species protection, i.e. $p = 5$.

Fig.1 shows the logistic probability density function against logarithmic *NOEC* concentration. The logistic distribution is very much like the well-known normal distribution. The logistic has more extended tails, and therefore can be regarded as a more conservative assumption in comparison to the normal distribution. It furthermore has some nice mathematical features, that make certain calculations relatively easy. (We have put most of the technical aspects in an Appendix.) The base of the logarithm by which the raw *NOEC* data are transformed does not matter, as long as the back-transformation of the results to concentrations is done with respect to the same base. Hence, we use the generic term 'log', that may either stand for natural logarithms, or for logs to the base 10, or otherwise.

Also indicated in Fig.1 is '$\log HC_5$', the logarithm of HC_5, below which 5% of the species is in danger (shown shaded). In fact, we are looking for the fifth percentile of the distribution of species *NOEC* toxicity data. The difficulty is how to account for uncertainty in trying to estimate this percentile from a limited data set.

In this paper, we will present improved extrapolation constants that allow straightforward calculation of estimates of HC_5 from mean and standard deviation of a sample of *NOEC* data. The procedure is essentially identical to the one of Van Straalen and Denneman (1989), but we will focus on meeting the required confidence level exactly, in order to protect against overprediction. The previous extrapolation constants are shown to lead to unacceptably high percentages of overprediction of the true HC_5, and therefore do not meet their confidence level. Furthermore, constants are obtained to calculate estimates of HC_5, that can be considered as a best guess, and that overpredict as often as not. As an example, we will recalculate the cadmium data from Van Straalen and Denneman.

2. Estimating Hazardous Concentrations

In order to estimate the agreed hazardous concentration (95% species protection) from a usually small number of toxicity data, we have to develop a statistical procedure to correct

for uncertainty due to small sample size. Hence, we want to quantify the uncertainty of our estimates, and we certainly do not want to overestimate too often. Therefore, a confidence approach seems natural.

Suppose we knew mean, μ, and standard deviation, σ, of the presupposed logistic distribution of log *NOEC* data of test species, as the one depicted in Fig.1. Then the log Hazardous Concentration for 5% of the species is easily calculated as (cf. Appendix):

$$\log HC_5 = \mu - 1.62 \cdot \sigma$$

One can estimate mean and standard deviation from the usual sample mean, $\bar{x}_m$, and sample standard deviation, s_m, of m test species, and estimate the log Hazardous Concentration straightforwardly, i.e. by substituting the sample statistics for the population statistics:

$$Z = \bar{x}_m - k_Z \cdot s_m.$$

With $k_Z = 1.62$, we act as if mean and standard deviation did not come from a sample, but were the true ones, but this would suffer from frequent overprediction. Fig. 2 shows sampling distributions of Z for sample sizes $m = 2, 5, 10$, and 20. These sampling densities are simulated through Monte Carlo sampling (cf. Appendix for details). The respective percentages overprediction are estimated to be 67%, 61%, 57%, and 55%. Note that all of them overestimate by more than 50%. If Z in a particular sample would come out higher than $\log HC_5$, then obviously more than 5% of the species may be affected. In fact, we want a recipe that overestimates $\log HC_5$ in a minority of samples only, so that with large *confidence* we can say that no more than 5% of the species is affected.

2.1 Kooijman / Van Straalen extrapolation constants

The reason for reconsidering this estimation question is that Kooijman (1987) does not intend to construct an estimate with this confidence property --in fact, his equation (16) and subsequent derivations cannot be motivated from a confidence point of view--, while Van Straalen and Denneman (1989) do interpret the results that way.

The final expression (Kooijman 1987, eq.(24), Van Straalen and Denneman 1989, eq.(6)), which we will call K here, looks very similar to Z:

$$K = \bar{x}_m - k_K \cdot s_m$$

only with a different k-value, here called k_K. For an estimate based on a sample, this constant depends on the sample size. The original expression for k_K is given in the references cited and repeated in the Appendix.

Table 1 lists various values of k_K for two certainty levels: 95% and 50%, calculated from Kooijman (1987). The term 'certainty' is ours to distinguish these constants from those to be given in the next paragraph. The reason to consider 50% certainty or confidence will be discussed later. The columns in Table 1 correspond to δ_2=0.05 and 0.5 in Kooijman (1987, Table 1), respectively.

Note that the asymptotic value corresponds to the value 1.62 under complete knowledge about the mean and standard deviation of the distribution at hand. But, if interpreted as constants to calculate the $\log HC_5$ with a certain level of *confidence*, Table 1 is suspect for two reasons. First, for 95% certainty, the constants do not seem to 'blow up' enough for decreasing sample sizes, e.g. m=4, 3, 2. We know this effect for confidence limits of the mean in normal distribution theory: Student's t-values, and expect it to be even worse for confidence limits of a tail value, what $\log HC_5$ essentially is. But, secondly, the 50% column, if interpreted as confidence factors, seems to be on the wrong side of 1.62 anyway. It tells us, that it is better to have 10 *NOEC* values, than 30, which in its turn is better than an infinite number of test data available. This is suspect, because the Z-estimates with k = 1.62 already overpredict for more than 50%, so, how can smaller k-values overpredict less? This does not seem realistic.

In order to test the confidence property of the Kooijman/ Van Straalen extrapolation constants, we simulated the K-95% and K-50% sampling distributions, based on the extrapolation constants in Table 1, in the same way as we simulated the Z densities. Fig.3 displays the K-95% sampling densities for the same set of sample sizes as before: 2, 5, 10, and 20. We observe considerable overprediction of $\log HC_5$. Fig.4 shows the simulated K-50% sampling densities. These indeed seem to overpredict even more than the corresponding Z-densities.

Table 2 summarizes the overprediction percentages for these 4 sample sizes. If K is to be interpreted as a one-sided 95% left confidence limit, the percentage of simulated samples with a K-value *above* $\log HC_5 = -2.94$ should be somewhere in the vicinity of 5%. The percentages estimated (39%, 22%, 20%, and 14%, respectively) seem to be unacceptably high. The same holds for a one-sided 50% confidence value. Overprediction should approximate 50%. These simulated values (83%, 67%, 65%, and 60%) seem to be too high as well.

In the next paragraph, we will calculate extrapolation constants, that lead to estimates of $\log HC_5$ that do have the required confidence interpretation.

2.2 New extrapolation constants on two levels of confidence

In order to construct an expression L that calculates the 95% species protection level with true one-sided 95% and 50% confidence levels, we need not develop an essentially new methodology. In fact, if we stick to the same type of formula:

$$L = \bar{x}_m - k_L \cdot s_m$$

and focus on the new k_L extrapolation constants for different m, it is easy to prove that, for each m, there is just one value of k_L with the required confidence property for any logistic distribution (cf. Appendix). Thus, for each sample size, we determined k_L through Monte Carlo simulation by generating random sample averages and standard deviations for the standard logistic distribution only, and by adjusting k_L in such a way that a pre-specified confidence level was obtained. These are tabulated in Table 3.

Fig.5 shows the sampling densities of the one-sided 95% left confidence limits (*L*-95%) for m=2, 5, 10, and 20, as determined by the new extrapolation constants. Each one overestimates $\log HC_5$ with 5%, as they should. Fig.6 displays the sampling densities of the one-sided 50% confidence limits (*L*-50%). They overpredict as well as underpredict with 50%.
Clearly, the extrapolation constants of Table 3 would pass the test of Table 2, since they are constructed that way. The percentages overprediction would be 5% and 50%, respectively. Moreover, the new constants do show the expected Student-*t*-like blow-up for small m. Furthermore, contrary to the 50% certainty constants in Table 1, the 50% confidence extrapolation constants for finite samples are higher than the asymptotic value, i.e. 1.62 (k_Z), for 'infinite' samples. This means that a one-sided 50% confidence estimate of $\log HC_5$ must still be lower than the straightforward answer (*Z*), acting as if we knew the logistic parameters.
At the time of publication, we have come into contact with Wagner and Løkke (1990), who derived extrapolation constants for the 95% species protection level, when a *normal* distribution is assumed, by using existing theory that applies to the normal distribution only. The resulting extrapolation constants are very similar to those presented here for the logistic distribution.

3. Example

As an example, we recalculated the HC_5 from 7 NOEC values for toxicity of cadmium to reproductive parameters of various soil animals, corrected for standard soil (Van Straalen and Denneman, 1989, Table 2). The sorted data are: 0.97, 3.33, 3.63, 13.5, 13.8, 18.7, and 154 [µg/g].
After transformation with base 10 logarithms, we have mean $\bar{x}_7 = 0.9712$, and standard deviation $s_7 = 0.7028$, respectively. The Kooijman/ Van Straalen estimate of the HC_5 for 95% certainty is

$$10^{(0.9712 - 2.52 \cdot 0.7028)} = 0.16 \quad [\mu g/g]$$

Note that we directly employed the Kooijman extrapolation constant 2.52 from our Table 1, entry number 7. Secondly, it is easy to show that the base of the logarithm doesn't matter. When using the mean an standard deviation on the basis of natural logarithms, i.e. 2.236 and 1.618, respectively, we arrive at the same result:

$$e^{(2.236 - 2.52 \cdot 1.618)} = 0.16 \quad [\mu g/g].$$

(Van Straalen and Denneman, 1989, Table 3). And this is true in general of course.
By using the new Table 3 extrapolation constants, 3.59 and 1.78, for a sample size of 7, we arrive at the 95% left confidence limit of

$$10^{(0.9712 - 3.59 \cdot 0.7028)} = 0.03 \quad [\mu g/g],$$

while the 50% confidence estimate of HC_5 is

$$10^{(0.9712 - 1.78 \cdot 0.7028)} = 0.53 \quad [\mu g/g].$$

We note that the 95% lower confidence limit (0.03) and the 50% confidence, or 'median', estimate (0.53) embrace the Kooijman/ Van Straalen estimate (0.16). This will *always be the case*, as can easily be seen by comparing the 95% column from Table 1 with the 95% and 50% columns of Table 3. The former k-constant is always between the latter two, for corresponding sample sizes.

It is interesting to observe that if we really want to limit the probability to overestimate HC_5 to only 5%, we have to apply a safety factor

$$T = 10^{(3.59 \cdot 0.7028)} = 333,$$

instead of 59, as estimated by Van Straalen and Denneman (1989, Table 3), for this example. Hence, we may conclude that, if we want to have 95% confidence to not overestimate the 95% species protection level, we have to calculate values that are generally lower than those calculated up to now.

4. Discussion

We have given extrapolation constants on two levels of confidence: 95% and 50%. The larger confidence level of the two can easily be motivated thus: we wish to protect *at least* 95% of the species, hence we want to limit overprediction of the true $\log HC_5$ to 5%. But why calculate a 50% confidence estimate?

First of all, there is a practical reason. We have found it confusing to present one left confidence limit value as *the* single answer to an extrapolation exercise. Users start asking for a confidence interval for it, and forget that it is already a confidence limit. So there is a need for a middle value, that could as easily be too high as it could be too low. Then, in analogy with a classical two-sided confidence interval for the mean of a normal distribution, e.g. a value ± a half-range, we could use the 50% confidence value as the middle value, and the 95% confidence value as a one-sided left confidence limit. Which one of these values to use in a particular situation, with all kinds of practical considerations involved, eventually is a matter of policy or decision making. However, in this decision process, the following, more theoretical, considerations should be taken into account.

The presently followed approach to estimate hazardous concentrations for ecosystems from a small set of single species data illustrates the basic principle of risk analysis in the face of uncertainty. In this situation we have to deal with two levels of risk. The primary risk is what we are interested in and what we want to estimate (or keep low). In the present paper the primary risk is the percentage of species that is actually harmed.

The secondary risk is the risk that our estimate of the primary risk is wrong. In this paper the secondary risk is set by the confidence level. If the results of the analysis are to be used as a basis for action, e.g. to determine a maximum tolerable concentration for ecosystems, the secondary risk should be taken into account. Both Kooijman and Van Straalen felt that the secondary risk should be low (5%). Yet there have been recent discussions on the

necessity of this low value; it has been even suggested to accept a confidence level of 50% as the single answer to work with. However, it does not seem to make much sense to demand a low value for the primary risk, and at the same time allow a high secondary risk.
To illustrate the danger of using a 50% confidence level as a basis for maximum tolerable concentrations for ecosystems, Fig.7 shows the risks of harming larger percentages of species, for several values of *m* (number of species tested). For example, in case of four test organisms, the risk that more than 10% of the species is not protected is almost one third, whereas the risk that even 20% or more of the species is not protected is still 15%.
We therefore suggest to routinely calculate both the 50% and the 95% confidence value. The first value can be regarded as the best estimate of the hazardous concentration, whereas the latter may be taken as the "safe" value (given the assumptions underlying the calculations, of course). Comparison of these values can be used for deciding to examine more species: large differences between both values indicate considerable uncertainty.
The great virtue of regarding the 95% confidence value as the safe value, is that it tends to outweigh ecological and economical interests. If this safe value, as based on the available data, appears to be low enough to have important economical drawbacks, one would not hesitate to investigate more species, since the associated reduction of uncertainty might quite well result in higher values for the safe concentration. On the other hand, using the 50% confidence value as an indication of the safe value results in a strong bias towards economical interests. One would test a minimum number of species, hoping that the coin falls on the right side; if not, one could always extend the number of test organisms afterwards. Obviously, this situation would be quite harmful from an ecological viewpoint.

5. References

DGEP (1988-1989) Premisses for Risk Management. Risk Limits in the context of Environmental Policy. Directorate General for Environmental Protection at theMinistry of Housing, Physical Planning and Environment. Second Chamber of the States General, session 1988-1989, 21137, no. 5.

Gezondheidsraad (1988) Ecotoxicologische risico-evaluatie van stoffen. Advies No. 28, 's-Gravenhage.

Kooijman, S.A.L.M. (1987) A safety factor for LC_{50} values allowing for differences in sensitivity among species. Wat. Res., 21, 269-276.

Van Straalen, N.M., Denneman C.A.J. (1989) Ecotoxicological Evaluation of Soil Quality Criteria. Ecot.& Env.Safety, 18, 241-251.

Wagner, C., Løkke, H. (1990) Estimation of Ecotoxicological Protection Levels from NOEC Toxicity Data. Water Research, subm.

6. Appendix: Compilation of some mathematical aspects

In this Appendix, we will compile some of the more mathematical technicalities.

The probability density function (*f*) of species toxicity data is supposed to be logistic:

$$f(x) = \frac{1}{\beta} \cdot \frac{\exp(-(x-\alpha)/\beta)}{(1+\exp(-(x-\alpha)/\beta))^2}.$$

Here, x stands for logarithmic *NOEC* data (the base of the logarithm does not matter); α is the location parameter and β is the scale parameter. The mean (also median), to be called μ, and the standard deviation, to be called σ, can be expressed in α and β:

$$\mu = \alpha,$$

$$\sigma = \beta \cdot \frac{\pi}{\sqrt{3}} = 1.8138 \cdot \beta.$$

So, the standard deviation of the logistic distribution is roughly 2 times as large as the value of β. The standard logistic distribution, used in the simulations, has $\alpha = 0$, and $\beta = 1$, and therefore a standard deviation $\sigma = 1.8138$.

The cumulative distribution (*F*) of species *NOEC* toxicity data describes the probability for those log *NOEC* values to be smaller than x:

$$F(x) = \frac{1}{1+\exp(-(x-\alpha)/\beta)}.$$

One of the advantages of the logistic distribution over the normal distribution is the fact that this distribution can be represented in the explicit form stated. For example, for purposes of simulation, we need to generate many random logistic data. Due to the explicitness of the cumulative distribution, these can be easily generated with:

$$x^{\alpha,\beta} = \alpha - \beta \cdot \ln\left(\frac{1-U}{U}\right).$$

where U is a uniform random number. Note that

$$x^{\alpha,\beta} = \alpha + \beta \cdot x^{0,1}.$$

A second example where the explicitness of the cumulative distribution comes in handy, is the calculation of the log Hazardous Concentration for p% of the species under complete knowledge of the distribution. Then, one can equate $F(x)$ to $p/100$ and solve explicitly for x:

$$x = \log HC_p = \alpha - \beta \cdot C_p^1,$$

where

$$C_p^1 = \ln\left(\frac{100-p}{p}\right).$$

For example, for p=5, that is 95% species protection, we have

$$C_5^1 = -2.9444.$$

But we can also express $\log HC_5$ in μ and σ as follows:

$$\log HC_5 = \alpha - \beta \cdot C_5^1$$

$$= \mu - \sigma \cdot \frac{\sqrt{3}}{\pi} \cdot C_5^1$$

$$= \mu - 1.6234 \cdot \sigma.$$

This expression allows the calculation of the log Hazardous Concentration, if mean and standard deviation of the distribution are known.

In the original approach of Kooijman, the calculations are essentially similar. The probability that the log *NOEC* of the most sensitive of n species, is smaller than x is (Kooijman, 1987):

$$F_n(x) = 1 - (1 - F(x))^n,$$

with $F(x)$ the single species cumulative distribution given before. (Our notation here differs from Kooijman's.) Equating this to $q/100$ (called δ_1 in Kooijman) and solving for x gives the log Hazardous Concentration for Sensitive species:

$$x = \log HCS_q^n = \alpha - \beta \cdot C_q^n,$$

where:

$$C_q^n = \ln\left(\frac{(1 - q/100)^{1/n}}{1 - (1 - q/100)^{1/n}} \right).$$

When we compare C_p^1 with C_q^n, it easily follows that for n=1, $C_p^1 = C_q^1$ if and only if $p = q$. This shows the mathematical relationship between the Van Straalen and Denneman's (1989) hazardous concentration for p% of the species and Kooijman's (1987) hazardous concentration for p% of the most sensitive of 'communities' of *one* species:

$$HC_p = HCS_p^1.$$

In all estimates, the sample mean and sample standard deviation is used to estimate mean and standard deviation of the supposed distribution:

$$\hat{\mu} = \bar{x}_m = \sum_{i=1}^{m} \frac{x_i}{m},$$

$$\hat{\sigma} = s_m = \sqrt{\left(\sum_{i=1}^{m} \frac{(x_i - \bar{x}_m)^2}{m - 1} \right)}.$$

A simple estimate for $\log HC_5$ neglecting uncertainty due to a limited sample size is

$$Z = \bar{x}_m - 1.6234 \cdot s_m.$$

Table 3, the second 50% column, in fact shows that this estimate overpredicts in more than 50% of the cases.

Instead of $k = 1.6234$, other constants may be derived to account for uncertainty. These necessarily depend on m. The extrapolation constant due to Kooijman (1987), as applied by Van Straalen and Denneman (1989) with community size of 1 and 95% species protection is

$$k_K = \frac{3}{\pi^2} \cdot d_m \cdot C_5^1,$$

with d_m as tabulated in Table 1 of Kooijman (1987). These k_K constants are tabulated in Table 1 of this paper for two levels of certainty, that correspond to Kooijman's $\delta_2 = 0.05$, and $\delta_2 = 0.5$. With these constants, the Kooijman algorithm for calculating a left certainty limit (terminology is ours) of $\log HC_5$ becomes:

$$K = \bar{x}_m - k_K \cdot s_m.$$

A new extrapolation constant k_L is tabulated in Table 3 for calculating a one-sided left confidence limit of $\log HC_5$, called L:

$$L = \bar{x}_m - k_L \cdot s_m.$$

L satisfies the required confidence level.

However, the determination of these constants turned out to be a surprisingly hard numerical exercise. Each constant in Table 3 is an average of 20 such simulations with roundabout 250,000 sample points each, e.g. 30,000 samples for $m = 8$ (cf. 500 in Kooijman 1987). That means that each constant is based on roughly 5 million drawings from the standard logistic distribution. We still cannot guarantee every second decimal in k_L, though, but the true confidence level will be closely approximated.

The simulated densities depicted in Figs.2, 3, up to Fig.6 are estimated as follows. We generated 60.000 samples of size $m = 2$ and 5, plus 30.000 of size 10, plus 10.000 of size 20. All data were drawn from the standard logistic distribution. For each sample, the mean $\bar{x}_m$ and standard deviation s_m was calculated, along with Z-, K-, and L-values. These were sorted, and converted to histogram densities with bin width 0.2. The histogram midpoint values were smoothed with three-point running means with weights 1:2:1, and plotted.

Next follows the proof, referred to in the main text, that, if k_L were the proper extrapolation constant for a particular sample size in the case of the *standard* logistic distribution, then $L = \bar{x} - k_L \cdot s$, for that same sample size, would have the correct confidence property for *any* logistic distribution.

Suppose $\bar{x}^{0,1}$ is a standard logistic sample average (sample size m), $s^{0,1}$ is a standard logistic sample standard deviation, and suppose that

$$L^{0,1} = \bar{x}^{0,1} - k_L \cdot s^{0,1}$$

overestimates the true $\log HC_5^{0,1} = -C_5^1$ with known probability. Now, given the sample size, consider the statistic:

$$L^{\alpha,\beta} = \bar{x}^{\alpha,\beta} - k_L \cdot s^{\alpha,\beta},$$

with $\bar{x}^{\alpha,\beta}$ and $s^{\alpha,\beta}$ the sample mean and sample standard deviation respectively for some arbitrary logistic distribution. Then, the probability that it overestimates $\log HC_5$ is

$$\begin{aligned} Pr\{L^{\alpha,\beta} > \log HC_5^{\alpha,\beta}\} &= Pr\{\bar{x}^{\alpha,\beta} - k_L \cdot s^{\alpha,\beta} > \log HC_5^{\alpha,\beta}\} \\ &= Pr\{\alpha + \beta \cdot \bar{x}^{0,1} - \beta \cdot k_L \cdot s^{0,1} > \alpha + \beta \cdot \log HC_5^{0,1}\} \\ &= Pr\{\bar{x}^{0,1} - k_L \cdot s^{0,1} > \log HC_5^{0,1}\} \\ &= Pr\{L^{0,1} > \log HC_5^{0,1}\}, \end{aligned}$$

which was assumed to be known. Hence, L for any arbitrary logistic overestimates the corresponding $\log HC_5$ with that same probability. Therefore, for each m, we have to calculate k_L only once, e.g. for the standard logistic distribution.

7. Tables

Table 1. Extrapolation constants $k_K = \frac{3}{\pi^2} \cdot d_m \cdot C_5^1$ for 95% species protection (community size: n = 1), calculated from Kooijman (1987) for various values of m, the number of test species for which log(*NOEC*)s are available. The resulting log Hazardous Concentration is $K = \bar{x}_m - k_K \cdot s_m$, where $\bar{x}_m$ and s_m are mean and standard deviation respectively for a sample of size m. The two columns refer to 95% and 50% certainty, respectively.

m	95%	50%
2	3.33	1.00
3	3.04	1.26
4	2.88	1.40
5	2.74	1.48
6	2.62	1.50
7	2.52	1.50
8	2.43	1.51
9	2.37	1.51
10	2.32	1.52
11	2.29	1.52
12	2.26	1.53
13	2.25	1.53
14	2.24	1.54
15	2.23	1.54
20	2.18	1.58
30	2.06	1.58
∞	1.62	1.62

Table 2. Monte Carlo simulation of the one-sided 95% and 50% (left) confidence limit property of K (see text), i.e. ln HCS in Kooijman (1987). Extrapolation constants are from Table 1. The percentages overprediction should approximate 5% for 95% certainty and 50% for 50% certainty. These correspond to areas below the curves in Figs.3 and 4, to the *right* of $\log HC_5$.

sample size m	extrapol. constant (95%)	percentage overprediction (95%)	extrapol. constant (50%)	percentage overprediction (50%)
2	3.33	39%	1.00	83%
5	2.74	22%	1.48	67%
10	2.32	20%	1.52	65%
20	2.18	14%	1.58	60%

Table 3. Extrapolation constants for the calculation of one-sided left confidence limits for the logarithmic Hazardous Concentration for 5% of the species on the basis of the logistic distribution. Tabulated values are k_L such that a one-sided left confidence limit L for $\log HC_5$ is given by $L = \bar{x}_m - k_L \cdot s_m$. Here $\bar{x}_m$ and s_m are mean and standard deviation respectively of a sample of $\log(NOEC)$ test data of size m. Constants are tabulated for two levels of confidence: 95% and 50%.

m	95%	50%
2	27.70	2.49
3	8.14	2.05
4	5.49	1.92
5	4.47	1.85
6	3.93	1.81
7	3.59	1.78
8	3.37	1.76
9	3.19	1.75
10	3.06	1.73
11	2.96	1.72
12	2.87	1.72
13	2.80	1.71
14	2.74	1.70
15	2.68	1.70
20	2.49	1.68
30	2.28	1.66
50	2.10	1.65
100	1.95	1.64
200	1.85	1.63
500	1.76	1.63
∞	1.62	1.62

8. Figures

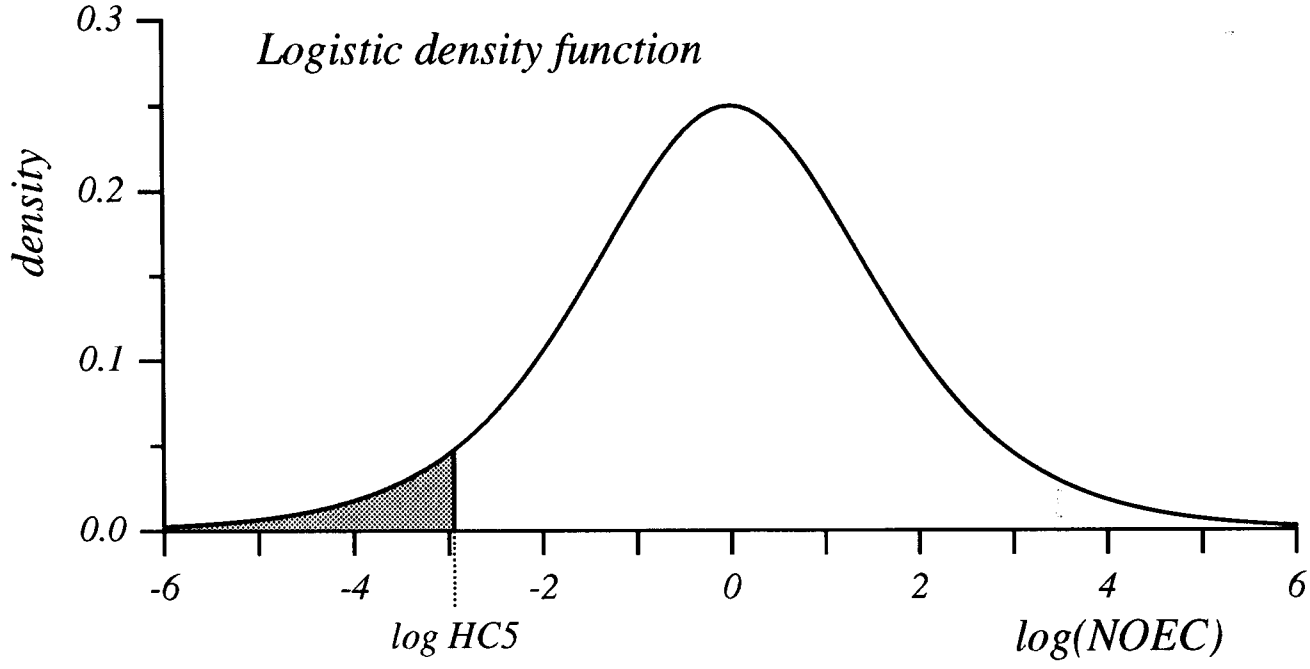


Fig.1. The standard logistic distribution of log(NOEC) values. $\log HC_5 = -2.94$ is the log Hazardous Concentration to be estimated. The fraction of the species harmed is shown shaded.

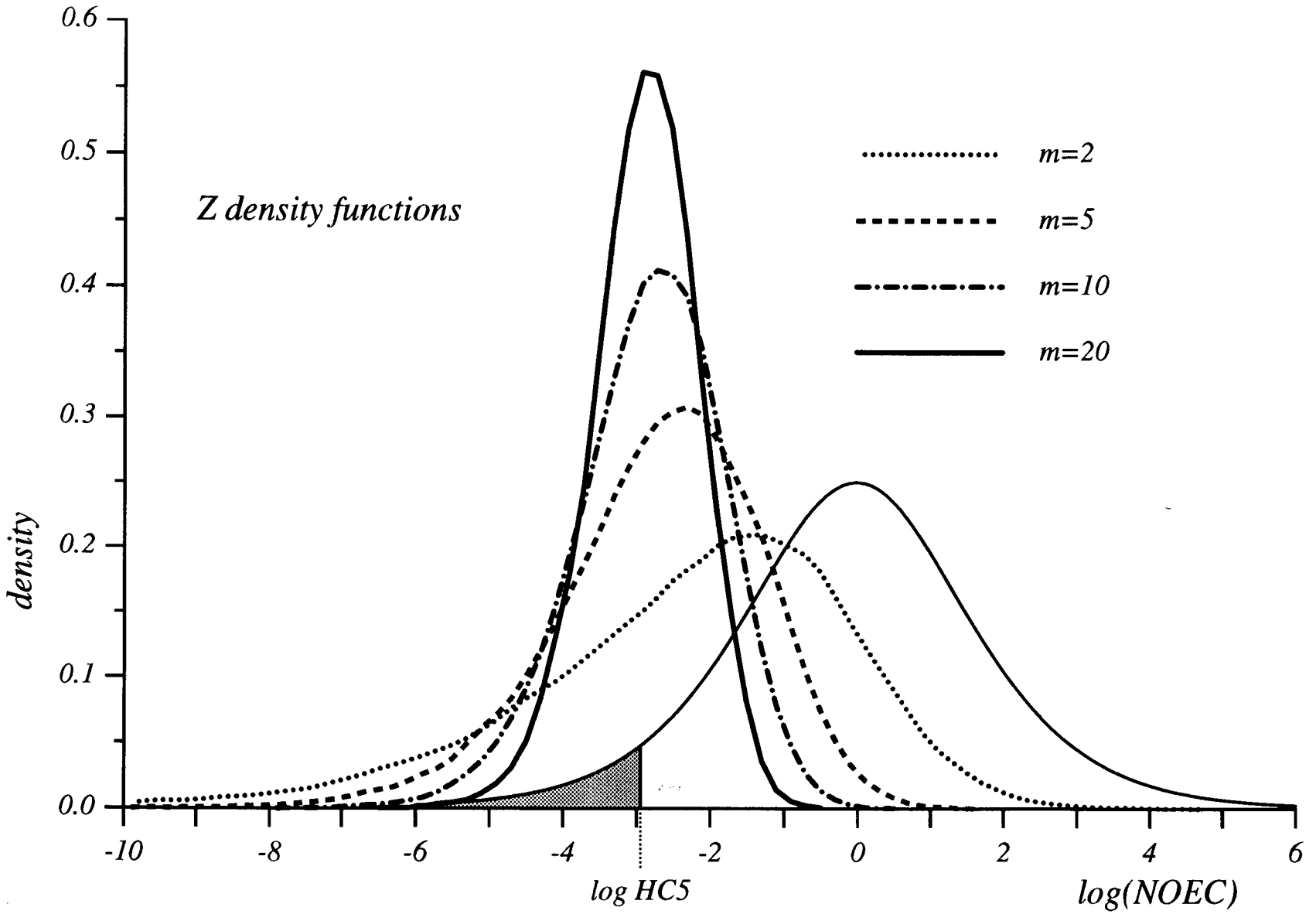


Fig.2. Simulated Z sampling density functions to estimate $\log HC_5$ for 4 different sample sizes, in relation to the standard logistic density function. $Z = \bar{x}_m - 1.62 \cdot s_m$.

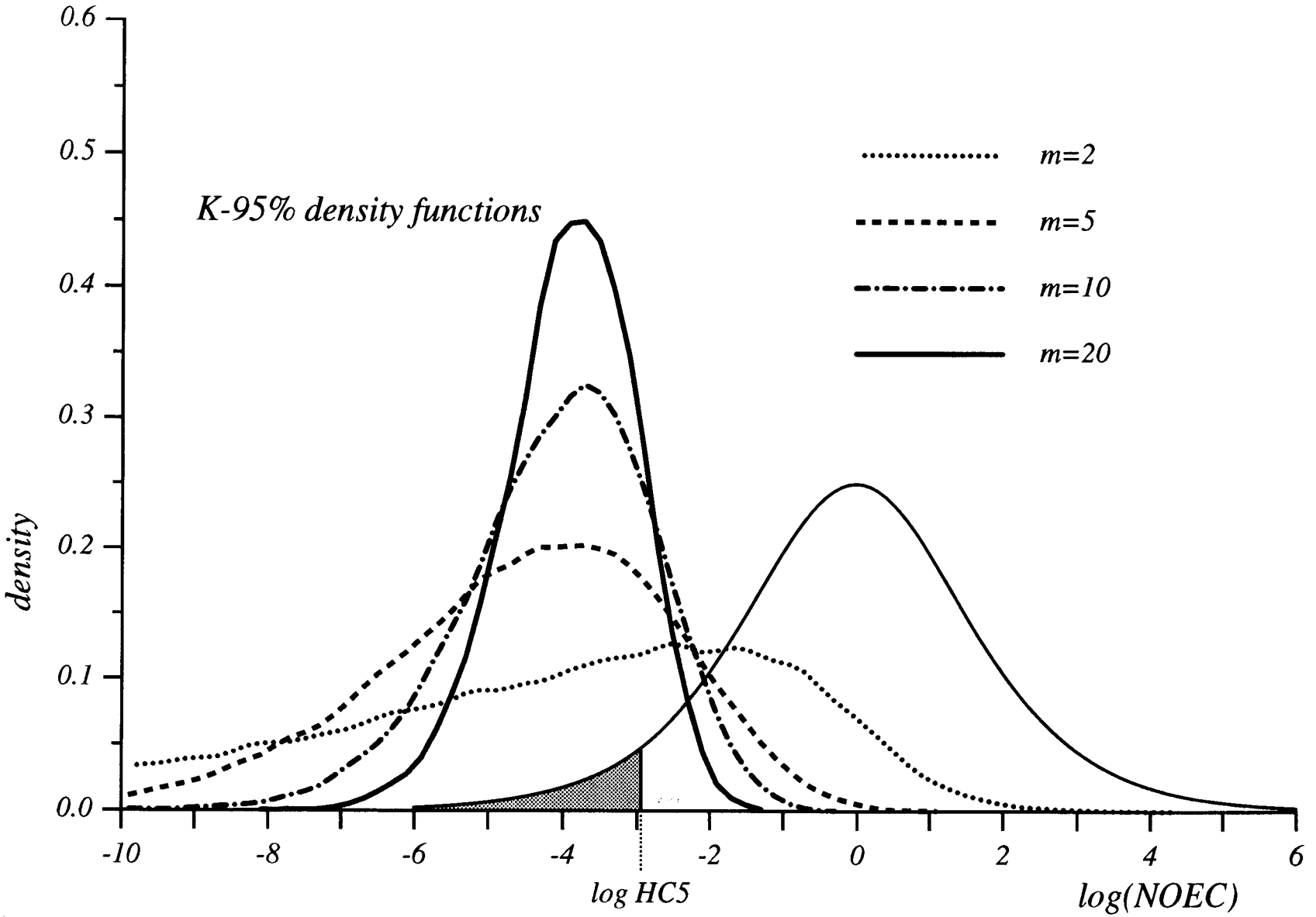


Fig.3. Simulated *K-95%* sampling density functions to estimate $\log HC_5$ for 4 different sample sizes, in relation to the standard logistic density function. $K = \overline{x}_m - k_K \cdot s_m$, with k_K from Table 1 (95%).

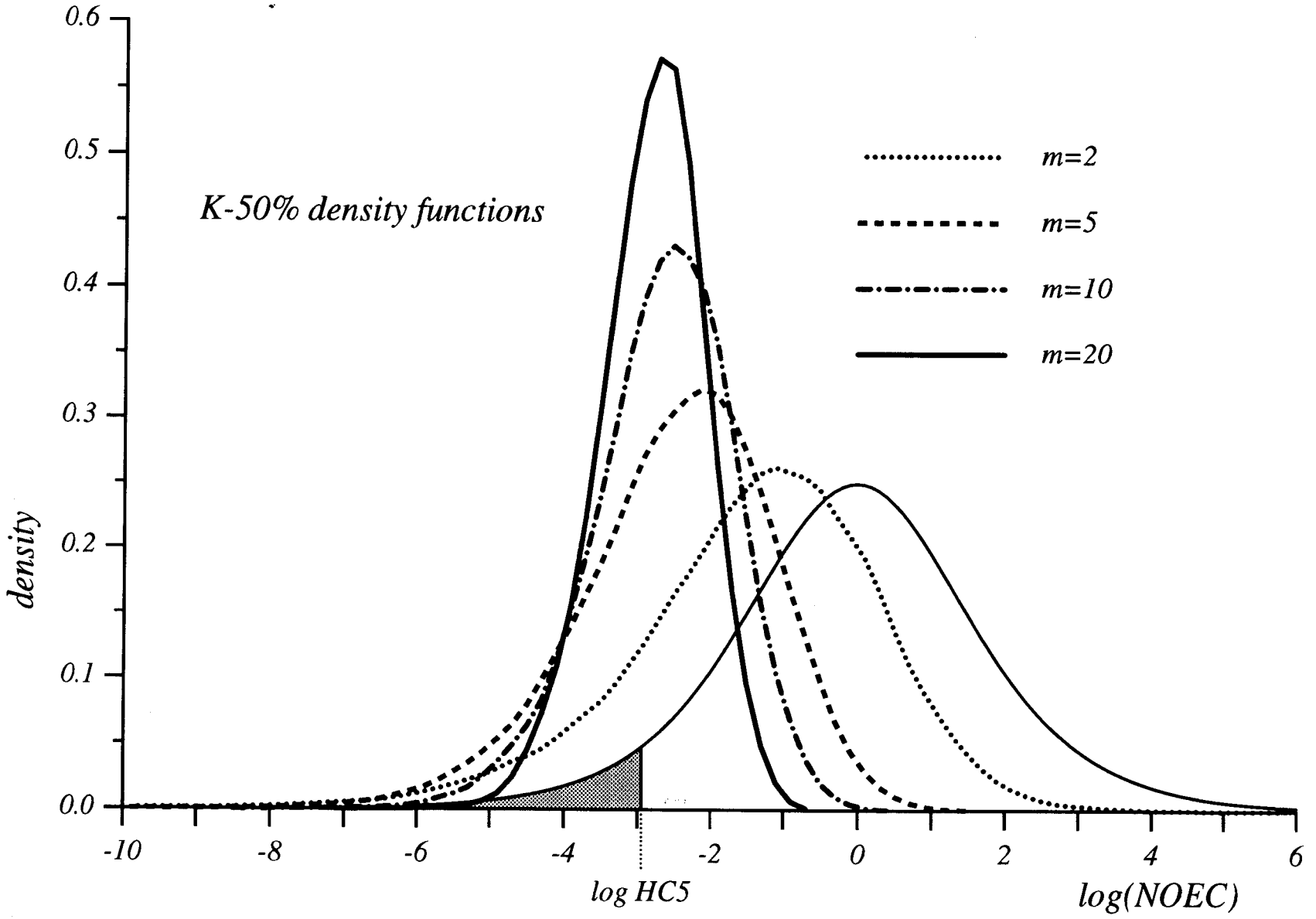


Fig.4. Simulated *K-50%* sampling density functions to estimate $\log HC_5$ for 4 different sample sizes, in relation to the standard logistic density function. $K = \overline{x}_m - k_K \cdot s_m$, with k_K from Table 1 (50%).

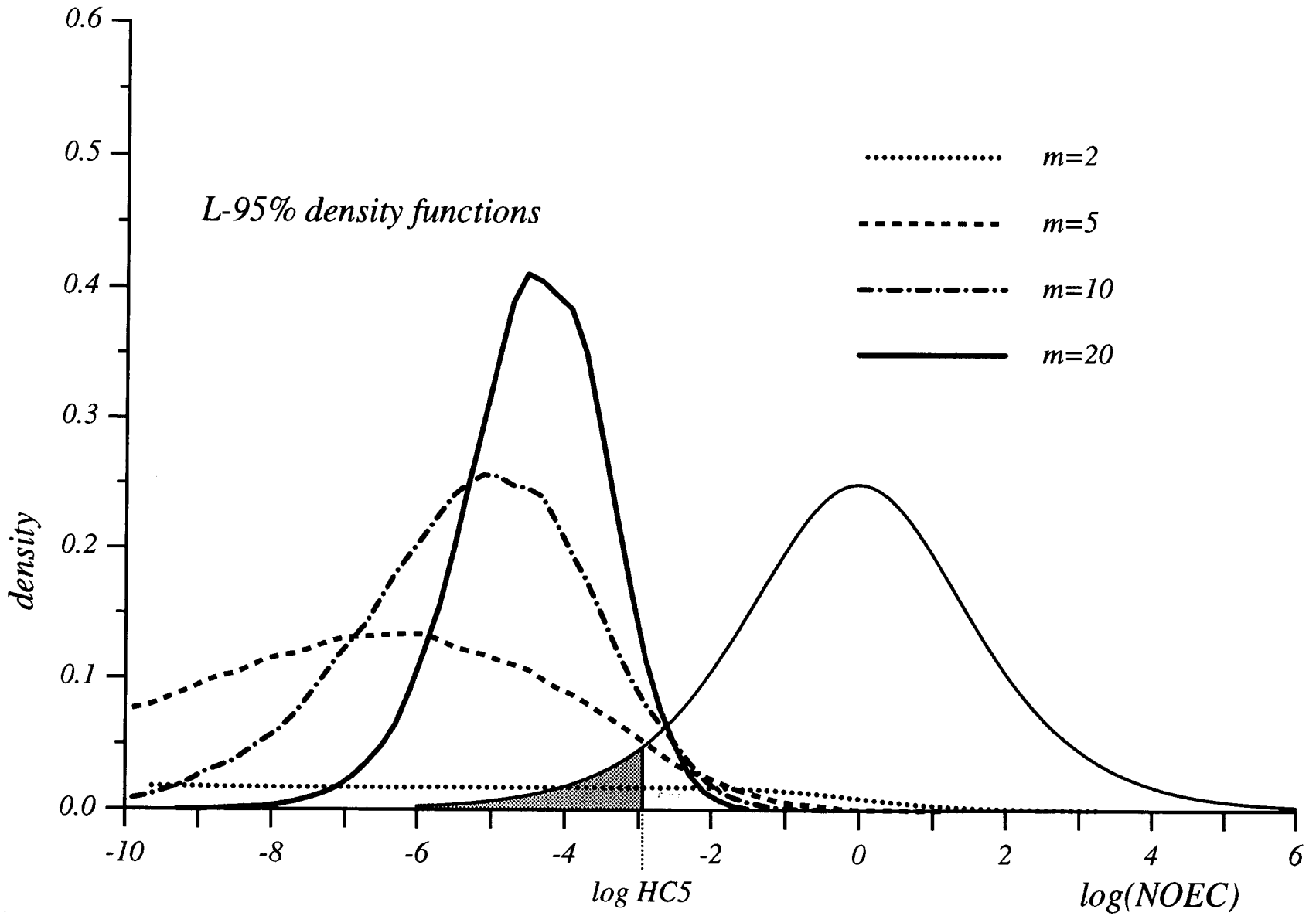


Fig.5. Simulated *L-95%* sampling density functions to estimate $\log HC_5$ for 4 different sample sizes, in relation to the standard logistic density function. $L = \bar{x}_m - k_L \cdot s_m$, with k_L from Table 3 (95%).

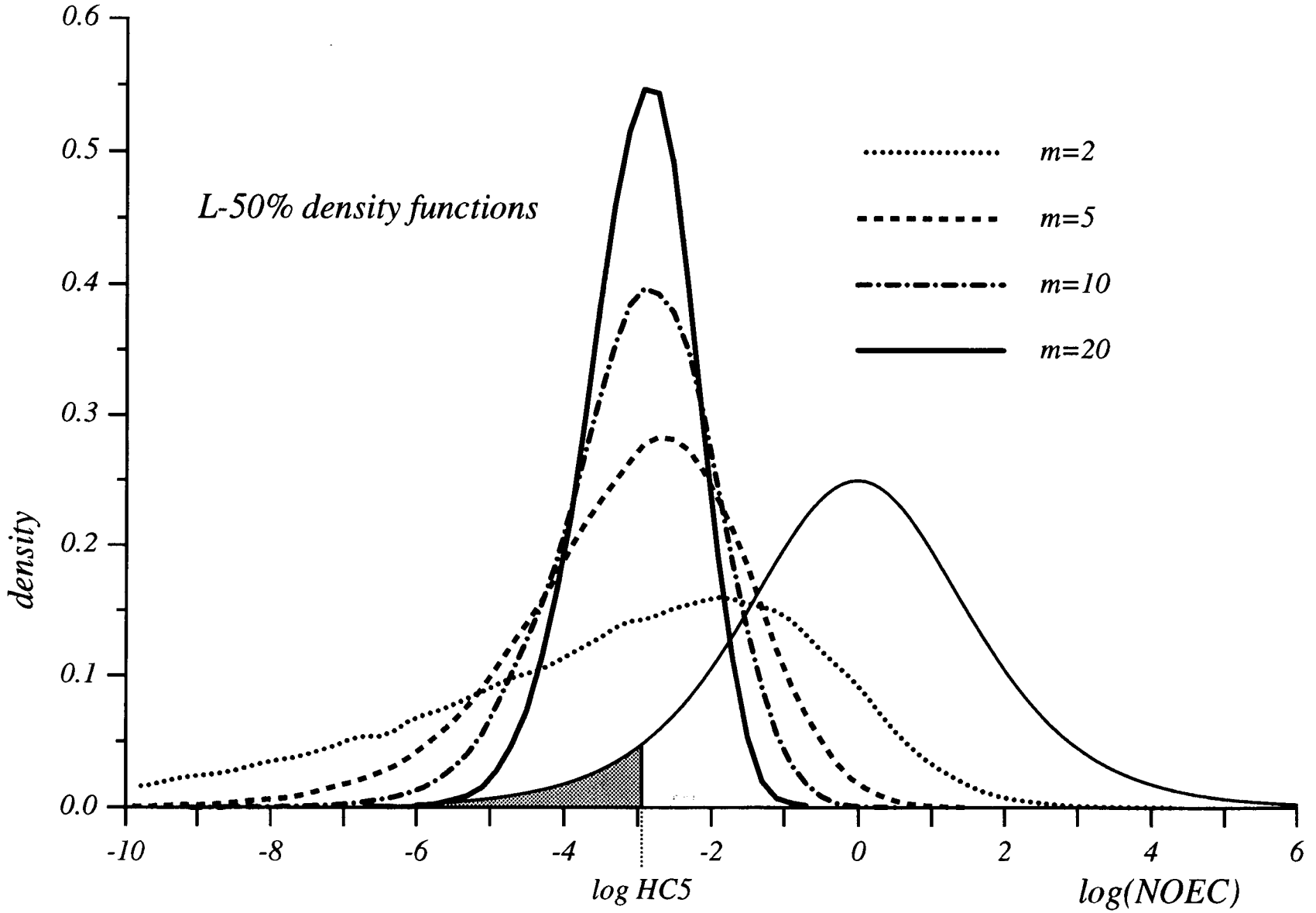
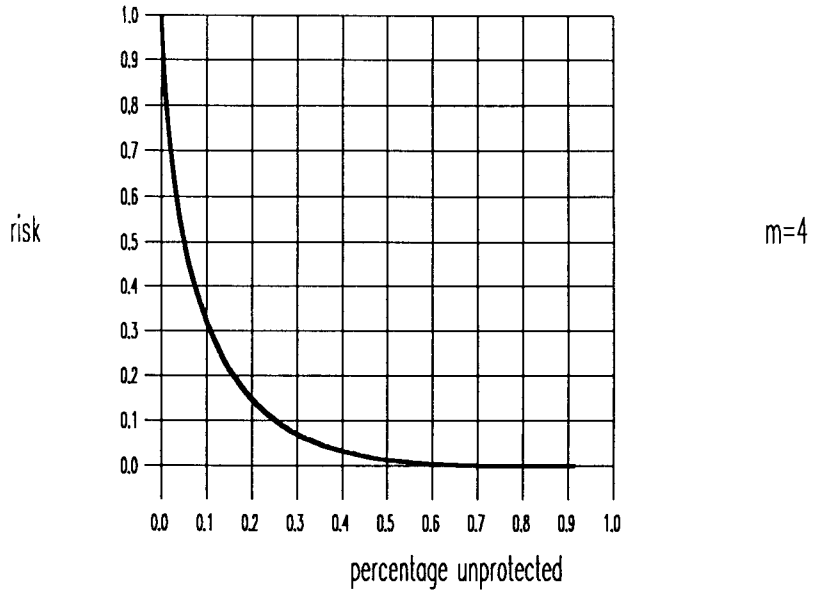
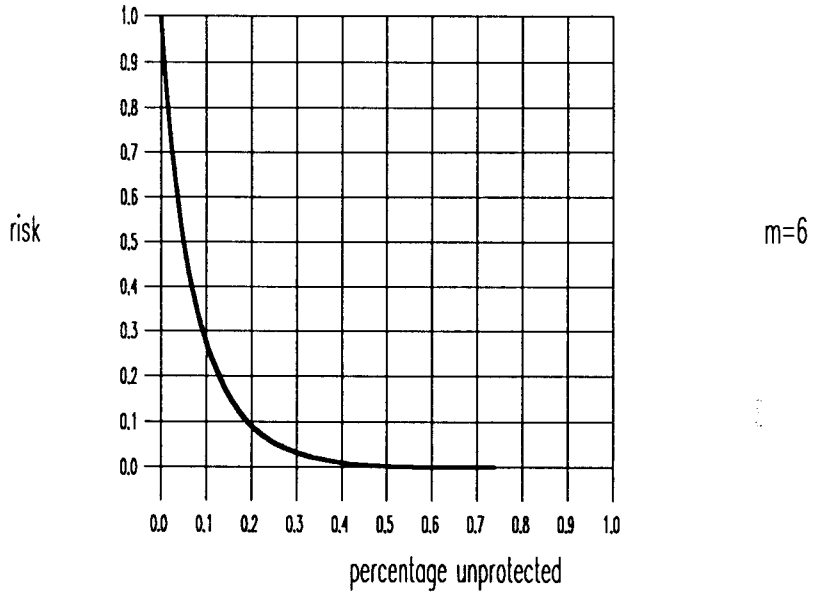


Fig.6. Simulated *L-50%* sampling density functions to estimate $\log HC_5$ for 4 different sample sizes, in relation to the standard logistic density function. $L = \overline{x}_m - k_L \cdot s_m$, with k_L from Table 3 (50%).





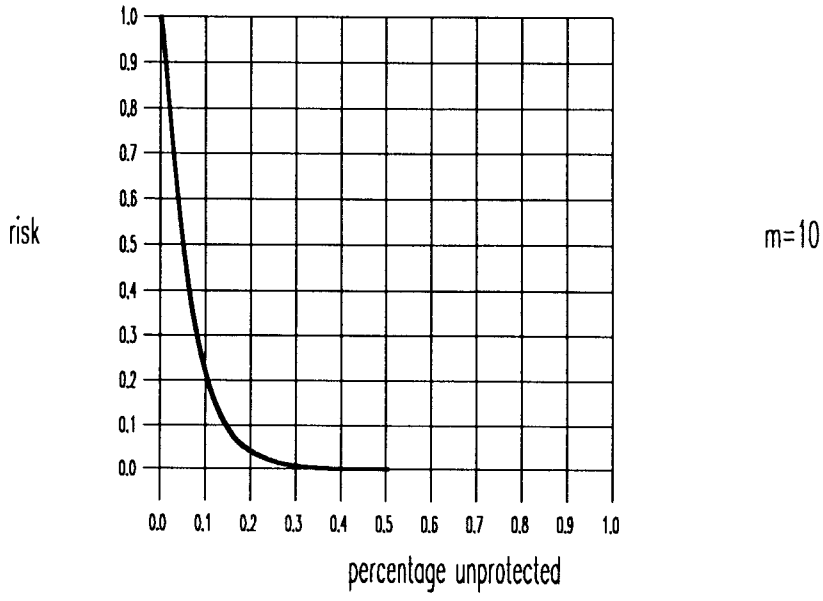


Fig.7. Risk (ordinate) that the percentage of unprotected species is exceeded (abscissa) in case of the 50% confidence estimate of the Hazardous Concentration.