External Peer Review of the EPA Document
Ten Berge CxT Models, External Draft Version 1.0

Prepared for:
Jeffrey S. Gift, Ph.D.
U.S. Environmental Protection Agency
National Center for Environmental Assessment (B243-01)
Research Triangle Park, North Carolina 27711

Prepared by:
Versar, Inc.
6850 Versar Center
Springfield, VA 22151

Contract No. EP-C-07-025
Task Order 11

Peer Reviewers:
A. John Bailer, Ph.D.
Ralph L. Kodell, Ph.D.
Frederick J. Miller, Ph.D., Fellow ATS

June 30, 2008

Peer Reviewers

A. John Bailer, Ph.D.
Department of Mathematics & Statistics
Miami University
Oxford, Ohio 45056

Ralph L. Kodell, Ph.D.
Department of Biostatistics
University of Arkansas for Medical Sciences
Little Rock, Arkansas 72205

Frederick J. Miller, Ph.D., Fellow ATS
Fred J. Miller & Associates LLC
Cary, NC 27511

Review by
A. John Bailer, Ph.D.

Peer Review Comments on EPA's Draft Document *Ten Berge CxT Models, External Draft Version 1.0*

A. John Bailer
Miami University
June 30, 2008

I. GENERAL IMPRESSIONS

Accuracy of the information presented:

The information presented includes a summary of the $C^N xT$ model that ten Berge originally implemented in Visual Basic and EPA is now porting to C. This recoding would allow EPA to include this model in a future BMDS release. By modeling concentration-duration data (along with other predictor variables), the BMDS software will add important functionality. This is reasonably described and motivated.

Clarity of presentation:

I find that the manual for the use of the ten Berge model is unclear. There is insufficient review of the model being implemented, and notation/terminology is employed that is confusing. The presentation could be improved by a careful statement of the models being fit along with precise definition of the variables that are measured and used to predict response probabilities in this model along with the parameters that are being estimated in these models.

Soundness of conclusions:

The current (C) implementation is compared to the previous (Visual Basic) fit by examining the regression coefficient estimates, variance/covariance estimates, etc. from fitting the two implementations to 3 data sets. The similarity of these estimates suggests that the current implementation is reproducing the output from the previous implementation. Now, similar fits to 3 data sets does not constitute exhaustive testing of the software. I am not an expert at software testing; however, I would suggest that the testing should include cases on the "boundaries" and not simply examples with at least 34 binomial experiments.

II. RESPONSE TO CHARGE QUESTIONS

1. Clarity of Report and Model Output: ***Are documentation and model output associated with the EPA C language ten Berge model clear and consistent with its corresponding Visual Basic version?***

The EPA C language model fits are reported in Appendices B/C/D while the corresponding Visual Basic results are displayed in Figures 1(abcd)/2(abcd)/3(abcd). The output from the two fits is similarly structured. This is not necessarily a compliment. The reader has to conduct a hunt between the two sets of output to compare models – e.g. Appendix B and Figures 1a/b/c/d

for the first example data set. A clearer display of the output with better section headings should be implemented. In addition, I would like to have seen a Table with the two fits explicitly compared. For example, I believe that a table of the form below would be very useful.

Example Data set	# binomial expts	Model fit?	Estimates		
			entry	Visual Basic	C
1	68				
2	34				
3	84	$\text{Logit}(p)=\beta_0+\beta_1\log_e(\text{conc})+\beta_2\log_e(\text{time})+\beta_3\log_e(\text{conc})\times\log_{10}(\text{time})$	b_0 (SE)	-35.01 ()	-35.01()
			b_1 (SE)		
			b_2 (SE)		
			b_3 (SE)		
			$\text{cov}(b_0,b_1)$		
			$\text{cov}(b_0,b_2)$		
			$\text{cov}(b_0,b_3)$		
			…		
			BMD…		

The output from the two implementations could still be included as appendices; however, the comparison should be included in the main body of text.

2. Adequacy of Testing Methods and Results: **The testing process should ensure that the EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version.**

(a) ***Is the record provided in the development and testing reports sufficient to document the testing methods used and results of software testing?***

The document describes the implementation and then compares the fit of the Visual Basic and C implementation to 3 data sets. As I mention both above and in later comments, I believe that more extensive testing should be considered and that better display of the comparison of the results of the tests should be provided.

(b) ***Have appropriate aspects of the EPA C language ten Berge model been tested?***

Not completely. The example cases should include situations where the ten Berge model does not converge or yields "no solution for some data" (see p. 23). Testing on the boundary is an important strategy for assessing code. The 3 examples compared in this report appear to be very "clean" in the sense that there are no missing values, the models converge nicely, etc. It is worth testing with cases that will cause the software to abnormally end execution. Can you cause the ten Berge software to fail? Does it fail in the same way in both Visual Basic and C implementations?

(c) ***Do the test results indicate that EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version?***

It is reasonable to compare a previous implementation with a recoding as is described here. The additional comparison with other implementations of binary response models is critical.

3. Other Issues: ***Are there any aspects of software development and testing, or model documentation, or reporting of model results (output file) that give you special cause for concern? If so, please describe your concerns and recommendations.***

As you can see in the specific comments section below, I believe that the output file could be improved with more detailed descriptions of the material being displayed. The output should include a more explicit description of the model being fit (i.e. an echo of the input arguments and a display of the model being fit). In addition, the hypotheses being tested should be described as part of the output, the section containing parameter estimates should have appropriate section "signposts" and the BMDs estimates (CIs) should have an appropriate section "signpost."

III. SPECIFIC OBSERVATIONS

[Page #, Paragraph #]

[5, 2] Important to emphasize that this add-on to BMDS will only apply to dichotomous responses even though BMDS can apply to other response scales.

[6] The ability to incorporate other variables as well as dose in BMDS is a neat addition.

[7] g() and h() function distinctions are not clear. Is it important to have this distinction?

[7, 3] Since the model is going to apply to only logistic and probit link functions, it would be useful to write out the models that are being fit. In addition, the notation on page 7 looks that is it written to describe arrays vs. standard notation. Compare

$$\text{dose}(1), \text{dose}(2), \ldots, \text{dose}(k)$$

to

$$d_1, d_2, \ldots, d_k$$

[7, 3] confusing terminology notation –
Do you mean to have a distinction between "dose" and "conc" in this discussion? Are these intended to be synonymous here? Is it important to make a distinction here?

Using "n(i)" as the number of responses out of "N(i)" trials is confusing. A better notation would be to let Y(i) = number of responses out of N(i) trials. Even better to explicitly state, $Y_i \sim \text{binomial}(N_i, p_i)$ where $\text{logit}(p_i)=$ or $\text{probit}(p_i)=$

[8, 1] How can a background rate be zero "by design?"

[9, 2] Reference for the "CatReg" method. Is this a method for ordinal responses? Is this fitting some ordinal response model such as proportional odds?

[9, 3] This notation here is confusing. On p. 7, p(dose) = g(dose, α, β, …) [not a great example of clear notation here] but how does this correspond to what is described on p. 9? I believe that h(z) on p. 9 corresponds to p(dose) but you don't link these directly. In your notation, "g" [p. 7]

would correspond to a LINK function in a generalized linear model (GLiM), and "h" would correspond to the INVERSE LINK function [$h = g^{-1}$]. The "z" term on p. 9 should be expanded. This is more commonly known as a linear predictor in a GLiM and is often denoted by η. This needs to be clarified.

Will the BMDS ever support the ten Berge modeling for links other than logit and probit?

[9, footnote 1] Is there any reason other than historic interest to add 5 in the probit link? GLiM links are usually of the form $g(\mu)$ and not $g(\mu + c)$ for some constant c.

[10, 1] z equation: you should explicitly identify C as concentration, T as time, x as other covariates such as mean body weight. In addition, you have used Greek symbols for parameters of the models. The regression coefficients are parameters as well and should be denoted by greek symbols. This equation would be better written as

p = g(η) where $\eta = \beta_0 + \beta_1 f_C(C) + \beta_2 f_T(T) + \beta_3 f_x(x) + \beta_4 r_4(C,T,x) + \beta_5 r_5(C,T,x) + ...$

where:

p = probability of response at concentration C, time T, covariate x.
g(*) = link function (either logit or probit)

$f_i(u)$ = transformation of concentration (i=C), time (i=T) or covariate (i=x). Transformations that have been implemented include:

identity: $f(u) = u$
logarithm: $f(u) = \ln(u)$ [natural logarithm or log-base e]
reciprocal: $f(u) = 1/u$

$r(C, T, x)$ = interactions of the C,T,x variables or functions of them

BMD estimation gets a little trickier with r() defined as products of C,T and/or x. This should be illustrated with a simple example (e.g. a text box?).

[10, 2] "parameter 'x'" – NO! x is not a parameter (in a statistical model sense although possibly from the perspective of a programmer).

[10, 2] Does the number of unique CxT combinations matter is terms of determining the number of parameters in the model?

[10, 5] Now "n" is being used as the power of concentration in a generalization of Haber's Law ($C^N xT$) – although text uses "n" and equation uses "N" and both of these were used for different quantities on p. 7! This notation needs to be cleaned up.

[11, 1] and CIs obtained? Using Fieller's Method? Profiling? Need to comment on this here?

[11, 3] Convergence criteria? Is an absolute deviation of $< 10^{-6}$ sufficiently stringent for convergence?

[11,4] Aren't the tests of coefficients based on estimates divided by standard errors simply Wald tests? What is tested with the X^2 statistic? A goodness-of-fit test? A test of the model; with {C,T,x} predictors versus an intercept-only null model?

[11, 4] User needs to hand calculate the heterogeneity adjusted tests? Surely, this should be implemented as an option for the computer to calculate.

[12, 2] Text describes that Fieller's method is used; however, a Wald interval Y±t*sqrt(Var(Y)) is reported (with a delta-method based Var(Y) presumably). Is Y intended to represent regression coefficients (e.g. b_i) or ratios (e.g. b_1/b_2) or …?

[12, 3] Program should incorporate a look up of the t-critical value vs. requiring user input of the critical value. The argument that follows "use the t critical value because it is bigger than the z and we aren't scaling up variances by a heterogeneity factor" doesn't seem compelling to me.

[12, 4] You say the Wald approximation is less reliable than a profile method and then proceed to use the Wald method?

[13, 1] Use the Wald method since that was in the original ten Berge code. Why not implement better solutions if you believe they exist?

[14, 1] "only run from a command prompt" – can you direct input and output to files?

[14, 3] "exponential.exe" – why is the executable named "exponential?" Should this be "tenberge.exe?"

[14, 4] Extension ".(d)" seems unusual.

[15, 2] An example with 68 different trials is huge. It might be better to have an example with a smaller number of concentration x time combinations. In addition, it is more common studies with a fixed number of concentration-time combinations as opposed to an example when the concentration varied with each trial.

In addition, you are characterized a binomial experiment with a common covariate "x" such as mean weight for a group. Isn't this artificial? If you have individual "x" values/data, then why not model the data as binary, i.e. each individual animal is binomial with N=1 and their own measured "x" value?

[18, 6] Confusing to call "variables" "parameters" especially in light of the statistical use of "parameter."

[19, 12/13/14/15] replace "parameter" with "variable" in each entry.

[21, 37] Any ratio of "β's?"

[22, 3] "color-coded section" was not color coded.

[22, 8] X^2 evaluation of fit? Testing what? Pearson goodness-of-fit test statistic?

[23, 2] Should output a warning message and not just rely on "-1" or "J" or some similar error code.

[23, 4] "response probability" vs. "response rate?"

[23, 5] "variance-covariance correction … nor choice of the deviate …" – how hard would it be to implement this?

[24, 1] Echo the model that is being fit including all of the transformations of the C, T, and x variables.

[24, 4] Here "parameter" is used correctly – regression coefficients of the model (e.g. two-parameter probit model)

[26, 3] There are lots of methods for fitting these models (e.g. SAS PROC LOGISTIC or GENMOD, R/S-Plus "glm", etc.). It is worth comparing the results of your fits/implementation to these alternatives (you do a little of this with R comparisons).

[27, 3] I like seeing the comparison of the tenberge.exe and R code. Would a table of the results of the comparison be worth including?

[28, 1] Display the variance-covariance entries in a matrix vs. a list.

[30] Inconsistent citation formatting. Missing titles on Moerbeek et al. and Nitcheva et al.

[32] Very confusing description – "Probability of correct model" – this should be clarified as part of the output. Could the output be labeled more clearly?

[46] Re-label contents of the output? …Need to have section labels for the output (e.g. Test of model with predictor vs. null model [LR test of full model to null model?] … Parameter Estimates [estimate SE t-stat, P-value] … BMD estimates … etc.

Review by
Ralph L. Kodell, Ph.D.

Peer Review Comments on EPA's Draft Document *Ten Berge CxT Models, External Draft Version 1.0*

Ralph L. Kodell
University of Arkansas for Medical Sciences
June 22, 2008

I. GENERAL IMPRESSIONS

Much of the information is presented clearly and accurately. However, there are some important exceptions. I have discussed the exceptions and some concerns in Section II under Charge Question 3 and in Section III. The intent of EPA was to present the results for the C program exactly as they are presented for the Visual Basic program. I understand that objective, but I think it has led to a carryover of some statistical inaccuracies that need to be addressed before the C program is released for use. I believe the conclusion that the C program's output matches that of the Visual Basic program is basically correct, but I strongly recommend that potential deficiencies in the Visual Basic program's output be examined carefully before EPA is comfortable that the output and interpretations are statistically sound.

II. RESPONSE TO CHARGE QUESTIONS

1. Clarity of Report and Model Output: ***Are documentation and model output associated with the EPA C language ten Berge model clear and consistent with its corresponding Visual Basic version?***

The documentation and model output associated with the EPA C language model are consistent with its corresponding Visual Basic version. However, I do not think the documentation and output are clear because they contain what I believe to be statistical inaccuracies and potentially inappropriate advice for usage. I discuss these concerns under Charge Question 3.

2. Adequacy of Testing Methods and Results: The testing process should ensure that the EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version.

(a) ***Is the record provided in the development and testing reports sufficient to document the testing methods used and results of software testing?***

Yes, the record provided in the development and testing reports is sufficient to document the testing methods used and the results of software testing.

(b) ***Have appropriate aspects of the EPA C language ten Berge model been tested?***

Many appropriate aspects of the EPA C language ten Berge model have been tested. In the first two examples, the model has three explanatory variables and one product term. Two of the three explanatory variables are C and T (log-transformed), whose product also makes up the product

term. The first example uses the probit link function while the second and third examples use the logit link. In all three examples, the log transformation is employed for C and T. In the first example, the log transformation is used for a third variable, body weight, while in the second example, the identity transformation is used for a third variable, sex (coded 0 or 1). The third example has only two explanatory variables but involves very disparate sample sizes, from 2 to 31. In all cases, confidence interval options have been invoked for "dose", response and ratio (ratio of coefficients of lnC and lnT). The input parameters for invoking the graphical functions are illustrated, but that part of the software is still in development, so no results were provided. Although, many appropriate aspects of the EPA C language ten Berge model have been tested, it would have been good to see an example where the reciprocal transformation was used. Also, it would have been good to see an example with just a few treatment groups, say 4 concentrations crossed with 3 times, for a total of 12 groups, to see how well the program implements the probit and logit models in such cases. The chi-squared test will have far fewer degrees of freedom than in the examples provided.

On page 19 under annotation note 11, it is stated that the option to use a background response correction has not been fully tested. I assume this option is something new that was not in the Visual Basic version. Similarly, on page 21 under annotation note 43, it is stated that the graphical functions are in development. I assume this is also something new that was not in the Visual Basic version. If my assumptions are incorrect, then these aspects of the C program have not been tested against their counterparts in the Visual Basic program.

(c) ***Do the test results indicate that EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version?***

With one possible exception, the test results indicate that the EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version. Section 4.2 points out that the results in Appendix B produced by the new C program are to be compared to Figures 1a-1d produced by the Visual Basic program. Similarly, Appendix C outputs are to be compared to Figures 2a-2d, and Appendix D outputs are to be compared to Figures 3a-3d. I have compared those results and have found them to match.

The lone exception occurred in the limited comparison of the C language program to the R language program. This was not a comparison of the C program to the Visual Basic program, but it was included as an independent verification of the C program. The parameter estimates, variance estimates, and test statistics of the C and R programs all match. However, I believe that the Chi-Square value from the C program (114.87 on 64 degrees of freedom) should match the Residual Deviance from the R program (137.65 on 64 degrees of freedom), as I believe these outputs are measuring the same thing. If I'm wrong, and they actually represent different measures of lack of regression fit (failure of the model to sufficiently explain the data) then I don't understand how these two values, Chi-Square=114.87 and Residual Deviance=137.65, have the same degrees of freedom.

3. Other Issues: ***Are there any aspects of software development and testing, or model documentation, or reporting of model results (output file) that give you special cause for concern? If so, please describe your concerns and recommendations.***

The main concern that I have is the definition and use of the "Student T" variate. I do not see how the so-called "Student T" can be an actual T random variable unless the "heterogeneity factor" is used in the variance of the parameter estimate. I don't think it should matter if the chi-squared test of heterogeneity is significant or not. I don't think the "Student T" test can be a true T test unless its denominator includes the square root of a chi-squared random variable divided by its degrees of freedom. I think the discussion that begins with the last paragraph on page 11 and includes most of page 12 needs improvement.

In the second paragraph on page 12, the method of Fieller is cited. I do not think the confidence limit that is defined in that paragraph is the Fieller limit, even if the right variance is used. Finney (1971) discusses the inclusion of an additional function, "g", when defining Fieller limits; this function is not mentioned in the paragraph. I think the proposed confidence limits are valid without the g function, provided the right variance and distribution are used. I just don't think they are Fieller limits. The Wald method is also mentioned. If the Wald method is actually used, then the test statistic will be a Z-statistic (asymptotically), not a T statistic. In this case, the heterogeneity factor seems moot. The Wald test is not a T test, and thus doesn't need a chi-squared divided by its degrees of freedom. However, I do not think a Z-statistic is appropriate.

I believe the stated plan for modifying statements in the output (last paragraph page 23) regarding the application of a "correction factor" when the chi-squared goodness-of-fit test is statistically significant needs to be re-evaluated. I think that Finney himself was unsure about which variate to use and what degrees of freedom to use. His footnote on page 451 in Chapter 17 (Finney, 1971) indicated uncertainty on his part. I think it would be a good idea to run a Monte Carlo simulation study to examine the distribution of the so-called Student T variable and the coverage of the associated confidence limits, with and without the heterogeneity factor. I think it will be an actual Student's T when the heterogeneity factor is applied, irrespective of whether or not the chi-squared test is significant. Regardless of what ten Berge recommends (Visual Basic program), I do not think that recommending the use of percentage points of a standard normal distribution is good advice. To me, the safest approach is to always apply the so-called heterogeneity factor in calculating the variance (significant or nor), and to always use a Student's T variate (with degrees of freedom equal to the total number of independent treatment groups less the number of estimated parameters) for confidence limits. I would not recommend using a Z (normal) statistic because I think this could lead to incorrect results.

III. SPECIFIC OBSERVATIONS

Page 6, paragraph 3: Is the ten Berge model really unlike any other model in BMDS? What about the Multistage Weibull, which has both dose and time as explanatory variables?

Page 8, paragraph 1: I don't understand what is meant by "Whenever the background rate of response is zero (either by design – e.g., using ln(conc) in the h( ) function – or…" Why is the background rate of response zero by design when using ln(conc) in the h() function? The fact that ln(0) is undefined doesn't mean there couldn't be a non-zero background response. And, if a non-zero background response were observed, it might not be justifiable to ignore it. This needs better explanation. Still, for short-term tests with death as the endpoint, the background response

rate might be zero much of the time. In fact, like in the examples tested, I think that most of the time in lethality tests there simply would not be a zero-concentration group in the experiment.

Page 11, paragraph 2: Isn't the optimization routine described the same as Newton-Raphson? If so, I think Newton-Raphson is a more recognizable name.

Page 14, paragraph 1: The second sentence should be moved to the end of the paragraph.

Page 14, paragraphs 3 and 5: What is the exponential.exe file and what is the exponential model? I have not tried to run the program, but I would have trouble understanding these instructions.

Page 16, first sentence: I would say "The example (d) input file…

Page 16: Is Table 3 the input file that is illustrated starting on page 16? Page 27 refers to the first data set shown above in Table 3, but I can't find Table 3.

Page 18, lines 15, 17, 19, annotated numbers 26, 33 and 39: The value 1.998 is said to be the value of Student's T with 63 degrees of freedom associated with 95% confidence (page 23). Presumably, this "deviate" should be used in the calculation of confidence limits. However, the output file in Appendix B corresponding to this input file has 1.96 instead of 1.998 in all three places. Is the output file from an initial run, in which it is determined whether or not the heterogeneity chi-squared test is significant? Are the values of 1.998 presumably for a follow-up run to get the right confidence limits? This needs to be clarified. See also page 23, paragraph 3 : : "Dose" estimation. Here it states that the values 1865 and 3029 can be taken as a 95% confidence interval, which is incorrect because they are based on 1.96 instead of 1.998 (Figure 1b, page 32 and Appendix B, page 46). As I said above, I think it is safest and statistically correct to always include the "heterogeneity factor" in the variance of the test statistic, and to use a Student's T distribution.

Page 18, annotation note 1: I would phrase it "Number of possible input parameters (possible explanatory variables)." In other words I would repeat the word "possible" to emphasize that not all input variables in the data file are necessarily used in the model. This was not clear to me at first.

Page 18, annotation note 2: Insert "to" at the end of the first line.

Page 19, annotation note 15: I think it should read "Number of input parameters (transformed or untransformed explanatory variables) to include in the model." Otherwise, it implies that this note refers only to explanatory variables that have been transformed.

Page 22, paragraph 4: Finney (1977) is cited. However, elsewhere in the document, Finney (1971) is cited. Is there a 1977 version of Finney's book? The annotated output on page 44 also has Finney (1977) under Dose-Response Analysis. Whether there is a 1977 edition or not, the citations need to be consistent.

Page 24, second paragraph: In lines 10-12, it seems to be implied that having a negative lower bound for the ratio B_1/B_2 is problematic. Is there a reason to always expect the power of C to be larger than the power of T? I don't see why. Note that the point estimate in the second example is negative, which seems valid and is not questioned. I think more explanation is needed. Also, the word "case" needs to be inserted at the end of the paragraph.

Page 24, last paragraph: This is a better discussion of background response than on page 8. However, it's not clear why the user couldn't choose a background correction when using the identity transformation, which is operational now. Nevertheless, given that use of the log transformation seems more likely than the identity, it would be good to have the background-adjustment implemented as quickly as possible.

Page 46, "Student t" values: To what are these values to be compared for statistical significance? Even if the output of the Visual Basic program was deficient in loosely defining these values and in not giving sufficient information for determining their significance, EPA's C program should provide the necessary information. The document mentions that the statement "Probability of correct model…" isn't accurate and that it will be changed once the C program is sufficiently tested. There are other aspects of the output that also need to be changed or enhanced to aid in the interpretation of results.

Review by
Frederick J. Miller, Ph.D., Fellow ATS

Peer Review Comments on
EPA's Draft Document
Ten Berge CxT Models, External Draft Version 1.0

Fred J. Miller, Ph.D., Fellow ATS
Fred J. Miller and Associates LLC
June 11, 2008

I. GENERAL IMPRESSIONS

For the most part, the authors have provided a well organized and clear presentation of the methodology for the modeling, the nature of the models available, and the structure of the output file generated by the software. The major weakness in the current document concerns the description of the structure and sequence of records that comprise the data input file. Given the record lines are significantly separated from the annotation comments, it will be difficult for the reader to understand easily exactly what is required where in the input file. Currently, the reader would need to spend a lot of time scrolling back and forth between records and annotation comments or else print this section so that the hard copy can be rearranged. This reviewer found it easier to develop a new table that has annotation comments directly associated with a given record (see Attachment 1). The authors have explained concepts clearly and frequently note where future versions can contain improvements. However, as they correctly comment, the exercise for this document was to provide a means to compare the Visual Basic program and output originally developed by ten Berge with the translation of the program to C executed using Windows XP. The conclusions reached by the authors throughout the document are defensible.

II. RESPONSE TO CHARGE QUESTIONS

1. Clarity of Report and Model Output: ***Are documentation and model output associated with the EPA C language ten Berge model clear and consistent with its corresponding Visual Basic version?***

While there are some differences in the structure of the output, the C version output is preferable to the Visual Basic version. However, the two are sufficiently similar that it was easy for this reviewer to determine that both output files give the same values for parameter estimates, confidence limits, variances, covariances, etc.

2. Adequacy of Testing Methods and Results: The testing process should ensure that the EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version.

(a) ***Is the record provided in the development and testing reports sufficient to document the testing methods used and results of software testing?***

Basically yes.

(b) ***Have appropriate aspects of the EPA C language ten Berge model been tested?***

The only thing that was not made clear was whether or not all possible modeling scenarios had been tested and that this document only contained representative cases. While it may not be practical to test all potential modeling scenarios, the authors should have commented on the degree to which they "stressed" the C version by examining various models. That being said, the examples that were presented are representative of the way the software will most frequently probably be used.

(c) ***Do the test results indicate that EPA C language ten Berge model is at least as reliable, accurate and clear as the original Visual Basic version?***

The C language ten Berge model is absolutely at least as reliable and accurate as the original code.

3. Other Issues: ***Are there any aspects of software development and testing, or model documentation, or reporting of model results (output file) that give you special cause for concern? If so, please describe your concerns and recommendations.***

Relative to page 22 of the document, the fact that the output file would be overwritten if the input file is used more than once is understandable but also troublesome. A prompt comment should be added at the beginning of the output file instructing the user that before using the input file in other modeling scenarios, the current output file must be renamed to avoid over-writing the current output file.

On page 11, the authors point out that iterative changes $< 10^{-6}$ reflect convergence, but they provide no indication that this was arrived at by appropriate numerical convergence testing. The tolerance level specified (i.e., $< 10^{-6}$) is 2 orders of magnitude less stringent than the convergence that was originally required in the Particulate Matter (PM) Generalized Additive Models solved using Splus software, where $< 10^{-8}$ was shown not to be strict enough under some conditions and that led to significant complications for the Agency in their last review of the NAAQS for PM.

Also, relative to the text on page 11, most users will not understand the need for a correction to the value of the t statistic and to the degrees of freedom when the test for heterogeneity of variance is significant. This reviewer strongly urges that in this instance (i.e., heterogeneity of variance test is significant) the program be modified to present both the corrected and uncorrected results.

III. SPECIFIC OBSERVATIONS

<table>
<tr><th></th><th>Comment</th></tr>
<tr><td>Title</td><td>The document title should have “ten Berge” and not “Ten Berge”</td></tr>
<tr><td>p. 7, l. 1</td><td>Interpretation of p(concentration, time, X) would be clearer if it was written either as p(concentration, time, X1,X2,…) or by using vector notation for X, particularly since the software provides for the potential for a vector of explanatory variables being able to be used. The authors get around to X as representing a vector of variables on p. 10, so why not use vector notation to begin with?</td></tr>
<tr><td>p. 7, l. 22</td><td>It might be clearer to the reader if first the product of the likelihood function for the data was given and then the log likelihood expression presented with an explanation of why the fitting is done with the log likelihood of the data rather than the likelihood of the data.</td></tr>
<tr><td>p. 9</td><td>Section 2.1 Theoretical Development This material does not really provide a good motivation for the theory. It should be pointed out that Haber’s Law is merely a special case of the general power law family of curves relating concentration and duration of exposure to a fixed level of response for a given endpoint (see Miller et al. Toxicol. 149:21-34, 2000).</td></tr>
<tr><td>p. 9</td><td>Footnote Recommend rewording the last sentence of the footnote so that it reads “There is no toxicological or mathematical reason for probits to be positive in value”.</td></tr>
<tr><td>p. 10, l. 6</td><td>The repeated use of $f_i(u)$ is confusing and is not needed. Just list the 3 possible forms of u and the transformation they represent. For example, a simple table like
<table>
<tr><th>Choice of f_c, f_t, or f_x</th><th>Transformation</th></tr>
<tr><td>u</td><td>Identity</td></tr>
<tr><td>ln(u)</td><td>Logarithmic</td></tr>
<tr><td>1/u</td><td>Reciprocal</td></tr>
</table>
</td></tr>
<tr><td>p. 10, l. 22</td><td>What is meant by floor (3/2)=1? Most users will not be familiar with “floor” used in this context. Why not just say something like, “So for this example, the maximum number of product terms is one”?</td></tr>
<tr><td>p. 10</td><td>Note paragraph This is good but the discussion in the paragraph should be expanded to note the difference between using terms to obtain a mathematical curve as opposed to terms that represent a biological process as well as the desire to have the least complicated model that provides a good fit to the data. Are various goodness of fit tests going to be made available in future versions?</td></tr>
<tr><td>p. 10, l. 31</td><td>The formulation of $C^N xT$ is inconsistent because both small and capital n is used.</td></tr>
<tr><td>p. 11, l. 20</td><td>Most users will not know or care about what a “Hessian matrix” is. The authors can make the definition clear by rewording the sentence to read “The expected Fisher information matrix is defined as the negative of the matrix of 2^{nd} derivatives (i.e., the Hessian matrix) of the log likelihood with respect to the model parameters. ”</td></tr>
<tr><td>p. 12</td><td>For the 3 indented sentences, remove the comma after response in the 2^{nd} sentence.</td></tr>
<tr><td>p. 12, l. 20</td><td>The authors only point out Student T deviates being recommended when there are significant Chi-squares. Student T deviates are needed also if the population variance is unknown, which is really the situation in most real world cases. A</td></tr>
</table>

	valuable addition to future versions would be to eliminate reliance on the standard normal distribution.
p. 14, l. 3	Delete “in’ after “tested”.
p. 22, l. 13	Delete “merely”.
p. 23	2nd Notes paragraph This reviewer agrees completely with the future intended path described by the authors in this paragraph.
p. 24, l. 22	Insert “case” after “in that”.

Attachment 1. Suggested Table for Linking Input Records with Annotation Comments

Record Value	Ann. #	Description
3	1	Number of input parameters (possible explanatory variables). This number will be equal to the number of fields in an input record minus 2 (the last 2 being for the sample size and number responding).
68	2	Number of observations/data records. This number tells the program how many lines of data read (see the lines below labeled "etc."
Conc mg/m3	3	Name of the first input parameter. This name will be used to identify the variable considered to be input parameter 1.
Minutes	4	Name of the second input parameter. This name will be used to identify the variable considered to be input parameter 2.
BW grams	5	Name of the third input parameter. This name will be used to identify the variable considered to be input parameter 3. *NOTE: there will be as many of these "name" lines as the number on line 1 ("Number of input parameters")*
Exposed	6	Name of the parameter corresponding to the sample size. This name should indicate the total number of individuals examined for any given data record.
Dead	7	Name of the parameter corresponding to the number of individuals having the response of interest. This name may be used to indicate what that response is (e.g., "Dead" in this example).
952 15 200 10 1		
1278 15 200 10 4		
1403 15 200 10 6		
1631 15 200 10 7		etc. The data record lines. There should be exactly the same number of lines as the number on the second line of the input file ("Number of observations/data records"). The order of the fields must be the same on each line and must be exactly in the order given in the "Name" lines immediately preceding. And the last two fields on each line must correspond to the total number of individuals examined and the number responding, in that order.
1767 15 200 10 9		
2028 15 200 10 6		
2349 15 200 10 9		
653 30 200 10 0		
886 30 200 10 0		
1006 30 200 10 3		
1033 30 200 10 6		
1267 30 200 10 9		
1359 30 200 10 10		
435 60 200 10 0		
544 60 200 10 1		
653 60 200 10 4		
740 60 200 10 8		
544 15 23 10 2		
707 15 23 10 4		
903 15 23 10 7		
946 15 23 10 7		
1060 15 23 10 6		

1153 15 23 10 9 1256 15 23 10 8 1658 15 23 10 9 1958 15 23 15 15 381 30 23 10 2 489 30 23 10 3 636 30 23 10 6 653 30 23 10 5 761 30 23 10 8 788 30 23 10 8 903 30 23 10 9 952 30 23 10 10 190 60 23 10 1 256 60 23 10 2 337 60 23 10 5 408 60 23 10 9 914 15 10000 4 0 1098 15 10000 4 1 1631 15 10000 4 2 1958 15 10000 4 2 2409 15 10000 4 4 555 30 10000 4 1 816 30 10000 4 2 1033 30 10000 4 2 1213 30 10000 4 3 1370 30 10000 4 3 1490 30 10000 4 4 343 60 10000 4 0 598 60 10000 4 1 696 60 10000 4 2 778 60 10000 4 4 924 60 10000 4 4 897 15 3700 4 0 1049 15 3700 4 1 1223 15 3700 4 3 1822 15 3700 4 3 2148 15 3700 4 3 1077 30 3700 4 0 1185 30 3700 4 2 1283 30 3700 4 4 631 60 3700 4 0		

663 60 3700 4 1 761 60 3700 4 1 1028 60 3700 4 2 1169 60 3700 4 2 1213 60 3700 4 4		
	8	Include a blank line, with no spaces, tabs, or any other delimiter following the lines of data. This separates the data section from the user-specified modeling control sections.
modeling	9	Just the one string "modeling" should be on this line to indicate the section that defines the parameters and link functions that are to be included in the model.
1 1 1 1 1	10-14	10. Type of link function (see Section 2.2 above): =1: Probit link function = 2: Logit link function 11. Background response correction: =1: No background response correction =2: Background response correction *Currently, it is recommended that the user set this control parameter to 1 (no background response correction), as the other option has not been fully tested.* 12. Transformation for input parameter 1. Even if the model subsequently defined (see items 15-21) does not include the first input parameter, a transformation for the parameter must be chosen. The identifiers for the 3 transformations included in this software are: =1: logarithmic =2: reciprocal =3: identity (none) 13. Transformation for input parameter 2. Even if the model subsequently defined (see items 15-21) does not include the second input parameter, a transformation for the parameter must be chosen. The identifiers for the transformations are the same as for the first input variable (see item 12 above). 14. Transformation for input parameter 3. Even if the model subsequently defined (see items 15-21) does not include the third input parameter, a transformation for the parameter must be chosen. The identifiers for the transformations are the same as for the first input variable (see item 12 above). *NOTE: It is possible to have more transformation identifiers on this line. There should be one for each of the input parameters and so the number of them should be equal to the number on the first line of the input file ("Number of input parameters").*
3	15	Number of transformed input parameters to include in the model. This number must be less than or equal to the "Number of input parameters" field (item #1). It indicates the number of single (not product) terms to be included in the model. This value can be zero, but in that case the next line should be skipped altogether
1 2 3	16-18	The numbers corresponding to the input variables to be included in the model. These numbers correspond to the order in which those parameters were entered on the data record lines and to the order of the names given by items #2, #3, #4 (and possibly more) specified above. Remember, these input parameters will be entered into the model transformed however that parameter was specified to be transformed (see items #12-14). The total number of entries on this line will equal the number on the immediately preceding line (unless that number was zero in which

		case this line will be skipped entirely).
1	19	Number of product terms to include in the model. This integer must be less than or equal to the number of input parameters (item #1) divided by 2 (rounded down). It indicates the number of product terms to be included in the model. This number can equal zero, in which case the next line is skipped entirely.
1 2	20, 21	The numbers corresponding to the input parameters included as product terms in the model. The numbers must be entered in pairs, the number of such pairs equaling the integer on the previous line (item #19) (unless the preceding number was zero in which case this line is skipped entirely). These numbers correspond to the order in which those parameters were entered on the data record lines and to the order of the names given by items #2, #3, #4 (and possibly more) specified above. Remember, these input parameters will be entered into the model transformed however that parameter was specified to be transformed (see items #12-14), even for the product terms. The input parameter identifier numbers used in the product terms need *not* be restricted to those used in the single-parameter terms of the model.
1 68	22, 23	22. The first record number to be included in the analysis. This number must be greater than or equal to 1 and less than or equal to item #2 (number of observations in the data set). Together with the next item (#23), this allows the user to restrict attention to (model only) observations in a certain range of the entire data set. 23. The last record number to be included in the analysis. This number must be greater than or equal to the number given by item #22, and less than or equal to item #2 (number of observations in the data set). Together with the previous item (#22), this allows the user to restrict attention to (model only) observations in a certain range of the entire data set.
dose	24	Only the string "dose" should appear on the next line to indicate that the user wishes to calculate the value of one input parameter that, for specified values of the other input parameters gives a user-specified response value. If the user does not desire to do such a calculation, this line and the following line should be skipped entirely.
1 1.998 95 1 20 200	25-30	25. Confidence intervals calculated? = 0 for no (even if this value =0, the remaining items on this line, items #26-30, should still be entered in the appropriate order) = 1 for yes 26. Deviate corresponding to the confidence level of choice. See Section 2.2 for a more detailed description of how to select this deviate. Enter this value even if item #25 equals zero. 27. The response of interest. This is given as a percent and so can have values between 0 and 100. X percent response corresponds to a probability of response of X/100. 28. The number identifier corresponding to the input parameter the user wishes to estimate. That number identifier must be one of those included in the list given by items #16-18. 29-30. The values for the other input parameters. These are the values assumed to be known (fixed). For the response of interest (item #27), and for the fixed values, one can determine what value of the input parameter given by the identifier in item #28 gives that response. The values entered here must be in the same order as given in items #16-18, except that the value for the parameter corresponding to the identifier given in item #28 will not be given (because that is the value to be estimated). For example, if parameters 1, 2, and 3 are entered in the model, and we wish to estimate the value of parameter 1 (Conc mg/m3 in our example data set) that give a response of 95% when Minutes is equal to 20 and BW grams is equal to 200, then the line in question will appear as shown above, with the value for Minutes (20) preceding the value for BW grams (200) because Minutes is the second input

		parameter and BW grams is the third input parameter.
`response`	31	Only the string "response" should appear on this line to indicate that the user wishes to calculate the response percentage for user-specified values of all the selected input parameters. If the user does not desire to do such a calculation, this line and the following line should be skipped entirely.
`1 1.998 500 20 200`	32-36	32. Confidence intervals calculated? = 0 for no (even if this value =0, the remaining items on this line, items #33-36, should still be entered in the appropriate order) = 1 for yes 33. Deviate corresponding to the confidence level of choice. See Section 2.2 for a more detailed description of how to select this deviate. Enter this value even if item #32 equals zero. 34-36. Values for the all the input parameters included in the model. These values should be entered in the order corresponding to their number identifier (e.g., in this example, the value for Conc mg/m3 first, followed by a value for Minutes, followed by a value for BW grams).
`ratio`	37	Only the string "ratio" should appear on this line to indicate that the user wishes to calculate the ratio of specified B_i coefficients estimated in the model. If the user does not desire to do such a calculation, this line and the following line should be skipped entirely.
`1 1.998 2 1 2`	38-42	38. Confidence intervals calculated? = 0 for no (even if this value =0, the remaining items on this line, items #39-42, should still be entered in the appropriate order) = 1 for yes 39. Deviate corresponding to the confidence level of choice. See Section 2.2 for a more detailed description of how to select this deviate. Enter this value even if item #38 equals zero. 40. Number of variables selected. This value should always be set equal to 2, as the ratio will be of the coefficients for two input parameters. 41-42. The number identifiers of the pair of coefficients for which the ratio is desired. As above, these number identifiers correspond to the order that the input parameters are entered in a data record. The identifiers must be from among the set of identifiers entered in items #16-18 (and any additional identifiers on that line) as input parameters used as single terms in the model.
`graph/response`	43	Only the string "graph/response" should appear on this line to indicate that the user wishes to produce plots of the model results. *Currently, the graphical functions of this software are in development, so this line and the next merely indicate right now how such plotting will be done.* If the user does not desire to do such plotting, this line and the following line should be skipped entirely.
`1 0.95 1 0 1000 2 20 3 200`	44-52	44. Confidence intervals calculated? = 0 for no (even if this value =0, the remaining items on this line, items #45-52, should still be entered in the appropriate order) = 1 for yes 45. Deviate corresponding to the confidence level of choice. See Section 2.2 for a more detailed description of how to select this deviate. Enter this value even if item #44 equals zero. 46. X-axis variable. This is identified by identifier number as in the previous references to parameters. 47-48. The range for the x-axis. The response will be calculated and plotted for each value of the x-axis variable

		between item #47 and item #48. 49-52. In pairs, the identifier number and value to assume for the remaining input parameters included in the model. Each pair will have the identifier number for the parameter and the value to assign to that parameter when the plots are created. There will be as many pairs as there are input parameters included in the model.
end	53	Only the string “end” should appear on this line to indicate that the program has reached the end of the input file. This must be the last line of each input file.