RESPONSE TO PEER REVIEW COMMENTS ON EPA-CMB8.2 AND ITS DOCUMENTATION

U.S. Environmental Protection Agency
Office of Air Quality Planning & Standards
Air Quality Modeling Group
Research Triangle Park, NC27711

June 2005

Introduction

During October - December 2004 a peer review was conducted of the latest update to the Chemical Mass Balance model (EPA-CMB8.2) and its documentation (Schauer *et al.*, 2005).   In the final report, all three reviewers agreed that EPA-CMB8.2 is a significant improvement over previous versions of the CMB software releases. Reviewers noted that the model effectively takes advantage of Windows systems, enabling users to efficiently create input files, operate the model, and review and analyze model results within the User Interface. Compared to other air quality modeling software, EPA-CMB8.2 is very user friendly and flexible to meet the broad needs of the air quality field. The reviewers noted that the documentation (manuals) provides a solid foundation on the appropriate use of the model, and on the theoretical and statistical basis for the model. The references provided were considered excellent for CMB users who need additional information. Finally, the reviewers thought the CMB input files provided with the model offered an excellent training opportunity for new users, and provide ideal templates for users who are developing new input files.

Critical comments were organized into three broad categories: Critical issues, matters of software inconvenience, and "carry-over" issues and supporting information.

Critical Issues

Ten items (2.1.1 - 2.1.10) were listed as critical. Items 2.1.1, 2.1.2, 2.1.7, 2.1.8, 2.1.9 & 2.1.10 were examined and all determined to be related to the software. Items 2.1.1 & 2.1.2 are actually related, as are items 2.1.7 & 2.1.8. Coding corrections were made to resolve all of these. With respect to items 2.1.7 / 2.1.8, while an arbitrary formulation, the fit measure formula has been modified from:

$$FM = W_1\left(\frac{1}{\chi^2}\right) + W_2 R^2 + W_3\left(\frac{\% \ mass}{100}\right) + W_4 \left(FracEst\right) \qquad \text{\% mass explained} < 100$$

$$FM = W_1\left(\frac{1}{\chi^2}\right) + W_2 R^2 + W_3\left(\frac{100}{\% \ mass}\right) + W_4 \left(FracEst\right) \qquad \text{\% mass explained} > 100$$

to:

$$FM = \frac{W_1\left(\frac{1}{\chi^2}\right) + W_2 R^2 + W_3\left(\frac{\% \ mass}{100}\right) + W_4 \left(FracEst\right)}{W_1 + W_2 + W_3 + W_4} \qquad \text{\% mass explained} \leq 100$$
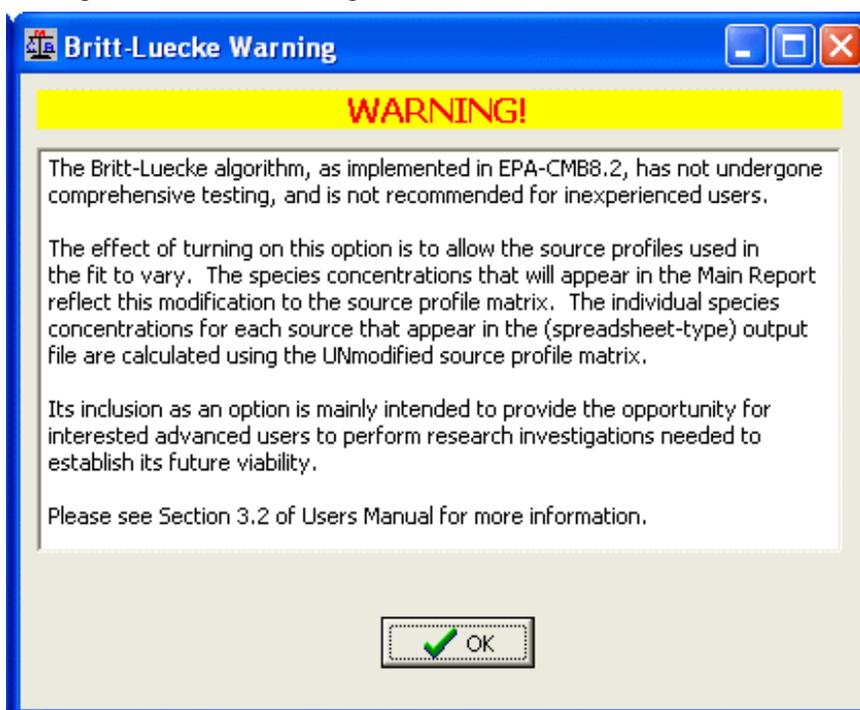
$$FM = \frac{W_1\left(\frac{1}{\chi^2}\right) + W_2 R^2 + W_3\left(\frac{100}{\% \ mass}\right) + W_4 \left(FracEst\right)}{W_1 + W_2 + W_3 + W_4} \qquad \text{\% mass explained} > 100$$

Per peer review recommendations, this formulation is now normalized to the sum of the weights (coefficients).

In item 2.1.3 the reviewers were concerned that no Control File was provided with the NARSTO-NE (should be NFRAQS) & COAST data sets. We never had any Control Files for these data sets to provide, they are not necessary to run EPA-CMB8.2, and - through coding revisions - we resolved all known operational issues with using the NFRAQS data set. Finally, these data sets were provided to the Peer Review panel for exploratory purposes only; no warrant was made as to the validity or completeness of the data.

In item 2.1.4 the reviewers suggested more explanation of the testing status of Britt-Luecke algorithm in EPA-CMB8.2. The reviewers thought it was unclear if the insufficient testing is for the EPA-CMB8.2 software code or the application to the CMB calculations.

The warning screen has been embellished to make clear that the Britte-Luecke algorithm has not undergone extensive testing within the EPA-CMB8.2 environment:

Britt-Luecke Warning

WARNING!

The Britt-Luecke algorithm, as implemented in EPA-CMB8.2, has not undergone comprehensive testing, and is not recommended for inexperienced users.

The effect of turning on this option is to allow the source profiles used in the fit to vary. The species concentrations that will appear in the Main Report reflect this modification to the source profile matrix. The individual species concentrations for each source that appear in the (spreadsheet-type) output file are calculated using the UNmodified source profile matrix.

Its inclusion as an option is mainly intended to provide the opportunity for interested advanced users to perform research investigations needed to establish its future viability.

Please see Section 3.2 of Users Manual for more information.

OK

The information provided in the Users Manual reflects our current understanding of this algorithm.

In item 2.1.5 (mislabeled as 2.1.4) the reviewers suggested more discussion on the appropriate settings for *maximum source uncertainty* and *minimum source projection*. Specific guidance on adjusting these settings for the VOC applications was not available to include in the Users Manual; as such becomes available it may be added to a future revision of the CMB Protocol (which may be a better place for it).

In item 2.1.6 the reviewers thought the discussion of the EST parameter in Appendix G should be moved up to the main body of the manual. This information, while perfectly valid, is anecdotal and probably best left to an appendix of notes. In fact, we frankly state that "We currently don't know how those settings translate into useful information ..." The purpose of this appendix was to provide diagnostic notes, including observations from limited sensitivity testing, for users to peruse. These notes should be used to supplement user judgment.

Software Inconveniences

Seven items (2.2.1 - 2.2.7) were listed as software inconveniences. In item 2.2.1 the reviewers complained about EPA-CMB8.2's forgivingness for input data formatting issues. As the reviewers stated, many formatting problems with the input files are not detected by EPA-CMB8.2, and it's impractical for EPA to revise EPA-CMB8.2 to catch all of these errors. Several "guard rails" have been installed in the User Interface to prevent the uninitiated or unwary user from introducing a bad input condition for the model. We have also embellished Section 4 of the Users Manual to make more explicit formatting instructions (e.g., for certain header formats) that are critical for a successful CMB session. At present, input formats are not as flexible as many would like, but we feel that the Users Manual provides adequate guidance on constructing these files. Users are also urged to use existing input files as templates.

In item 2.2.2 the reviewers questioned whether the 5 data sets provided for their review were peer reviewed. Firstly, with the exception of the San Joaquin Valley Fine data set that served as the developmental data base for EPA-CMB8.2, the data sets were provided for exploratory purposes only. Again, no warrant was made as to the validity or completeness of the other data. There is also debate as to whether the data sets in fact haven't been peer reviewed, and this status is beyond the scope of this project to update the model.

In item 2.2.3 the reviewers objected to cryptic messages that appear during certain fitting conditions. We agree that these messages, which originate in the Fortran/C++ DLL, are rather cryptic. However, it's also not clear whether the alternative messages suggested by the reviewers are entirely appropriate either. For now, these two messages remain unaltered.

In item 2.2.4 the reviewers complained that the data box which accompanies graphs of sources or samples won't display all species if you have a large number of species. Unfortunately, a scroll feature to accommodate more species is not easily feasible (and should be explored in a future upgrade). We have added a warning to Section 3 of the Users Manual that some distortion may occur if there are > ~25 species.

In item 2.2.5 the reviewers complained that missing data (-99) in ambient sample or source profiles is displayed numerically as -99 in View Graph. We confirmed this behavior and regret the effect it has on the display. Ideally, the species having missing values should be removed from this display, and should be explored in a future upgrade.

In item 2.2.6 the reviewers were disappointed that CMB8.0's capability for displaying time-series charts and source contribution pie charts was removed. They also noted that the *Present Computed Averages* was removed. Since the late 1990s, spread sheet applications such as Microsoft's Excel® have come to the fore and are widely available for making first-rate graphics

displays never possible in CMB8.0. It was thus a conscious decision to remove the internal graphics capability in the User Interface. During the development of EPA-CMB8.2 we did not consider the *Present Computed Averages* to be critical and it was removed for reasons of functionality.

In item 2.2.7 the reviewers were disappointed that *similarity clusters* were removed. The uncertainty/similarity clusters were replaced with a singular value decomposition eligible space treatment (Henry, 1992) that allows the user to define an acceptable error and an acceptable collinearity among weighted source profiles. We believe this feature improves the collinearity diagnostics in EPA-CMB8.2. In the Eligible Space Collinearity Display section of EPA-CMB8.2's Main Report, values for source projections are shown that, when compared to the minimum source projection value (set in Options), confirm whether a source is ESTimable or not. This is a quantitative display.

Carry-over Issues

Four items (2.3.1 - 2.3.4) were listed as carry-over issues. In item 2.3.1 the reviewers suggested more discussion of the appropriate use of Source Elimination in the Users Manual. We agree with some of the embellishments suggested by the reviewers, though we did not have definitive guidance to add for this option. We believe these points are best made in the CMB Protocol and such enhancements were beyond the scope of the upgrade effort in this project.

In item 2.3.2 the reviewers suggested more clarification and technical basis for acceptable criteria to be use for R-square and Chi-square. Criteria for these are taken from the CMB Protocol and have been long-standing practice. Calculated/measured ratios should be understood by experienced users to be performance criteria. The target of 100% $\pm$ 20% is also taken from the CMB Protocol and have been long-standing practice. To change this target is beyond the scope of this upgrade project and we don't currently have definitive information with which to revise this target.

In item 2.3.3 the reviewers were critical of guidance in Section 4.2.3 of the Users Manual (now Section 4.2.2) for naming variables in the ambient data input file. We have embellished the manual to add more clarification on this nomenclature.

In item 2.3.4 the reviewers complained that the column headings in the (spreadsheet-type) output file are not immediately obvious. We agree with this observation and regret that this derives from the internal coding of the model algorithms. Revision of the code to address this cosmetic problem was beyond the scope of this upgrade project and should be deferred to a future update.

Conclusion

EPA is grateful for the panel's diligent work. In its review, the peer review panel provided critical comments that were both rigorous and insightful. Several unknown software bugs were identified, as were errors in the documentation. Part of the Peer Review Final Report included individual comments from the panel members, most of which weren't captured in the summary to which this response is mainly directed. Considerable attention was paid to those individual comments and the vast majority of those comments and suggestions were implemented. Those comments/suggestions included both revisions of the model code and the documentation (Users Manual and CMB Protocol). Regrettably, some of the suggestions were beyond the scope of this update, and should be deferred to the next major update cycle for the CMB series. Comments on the CMB Protocol were interesting and well taken. For some of them, immediate revisions were made of the document. For others - ones that required the intimate knowledge of its prime author[1], efforts were made to solicit resolution. At the time of this writing, all available information was used to resolve issues with the CMB Protocol.

We believe that the model EPA-CMB8.2 is stable and well-behaved and will serve CMB users well for years to come. We also believe that the documentation is significantly enhanced and well suited for users.

References

Henry, R.C., 1992. Dealing with near collinearity in chemical mass balance receptor models. *Atmos. Environ.* **26A:** 933-8.

Schauer, J.J., D.M. Kenski and R.D. Willis, 2005. Peer Review of the Chemical Mass Balance Model (EPA-CMB8.2) and documentation. EPA Contract No. 4D-6097-NTSX.

---

[1]Dr. John G. Watson, Desert Research Institute; Reno, NV.