



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

~~Handwritten signature~~
~~Handwritten signature~~
~~Handwritten signature~~
Bob Cox - last
11/18/91
files

November 7, 1991

MEMORANDUM

SUBJECT: Phelps Dodge--Hidalgo Modeling Protocol
FROM: Dean A. Wilson, Meteorologist
Source Receptor Analysis Branch (MD-14)
TO: James W. Yarbrough, Air Modeling Contact
Region VI (6A-AP)

In response to your request, the Model Clearinghouse has reviewed: 1) your position with respect to the Industrial Source Complex Model Short Term (ISCST) equivalency demonstration, and 2) Section 5.0 of the preliminary draft protocol, dealing with the performance methodology for testing the competing models.

With respect to the modification to ISCST to accommodate the input of hourly emissions, we agree with you that an equivalency demonstration is needed rather than a full performance evaluation, since the code modifications are not intended to produce substantive changes in the model estimates. We have previously discussed the proposed equivalency test scheme with you and have agreed that it is appropriate. Our only additional suggestion is that, while not critical, it might enhance the credibility somewhat if the meteorological data used in the test could come from the same sensor that will be used in the performance evaluation of competing models.

With respect to comments on Section 5.0 of the protocol, Bill Cox has previously reviewed that material. His comments are attached.

If you have any questions, please contact me at 629-5683 or Bill Cox at 629-5563, as appropriate.

Attachment

cc: W. Cox

bcc: Regional Modeling Contact, Regions I-V, VII-X (with copy of incoming memorandum and list of FY-92 Clearinghouse memoranda)

Note To: Dean Wilson
From: Bill Cox
Date: July 11, 1991

I have reviewed the "Phelps Dodge-Hidalgo Modeling Protocol" and have only a few specific comments to offer for their consideration. First, it appears that the protocol as developed adheres pretty closely to our Procedures for Determining the Best Performing Model. Clearly, the authors have been innovative in trying to adapt the procedures to best fit the needs for this source. Second, nothing was said about presentation of the results. I assume that a variety of tables would be used to convey the outcomes along with appropriate graphics for the various data categories and performance components (frequency slopes, RHC, least squares slopes). Also, the protocol should indicate how many bootstrap trials are to be used in determining the standard error for the composite performance measure. In addition, I have a few comments directed at particular pages of the document.

Page 15--Robust Highest Concentrations.

They propose that least squares be used to fit and estimate the Robust Highest Concentration claiming that smelter data behaves in such a way that $\log(1-f)$ plots well vs concentration. If this is the case, then the data is essentially exponential and hence the method described in the "Cox" document should suffice. Also, assuming that the least squares approach is used, are the 25 highest values being used in the fitting process or do all "N" values get used? Following the discussion on page 16 and 17, it appears that the 25 highest are used but this is not clear.

Page 21--Linear least squares correlation of...

A slope of 1 (perfect model) is associated with no (zero) bias while a slope of 0 is assigned a score of 1. If the slope is only slightly negative (e.g. -0.000001) an inconsistency is introduced by scoring the result with 0. Shouldn't the score variable (CFB for slope) be continuous over the range of possible outcomes in such a way that negative values receive a value of (at least) 1? Regarding calculation of the slope, standard least squares is likely to produce very low slopes (0.0-0.1) even though the correlation coefficient is relatively good (e.g. 0.5). A better estimate of the slope might be obtained using the slope of the first principle component applied to the variance/covariance matrix between observed and predicted concentrations. Since the regression slopes of both models will likely be low as a result of this effect, it probably won't have any impact on the overall outcome.

FY-92 MODEL CLEARINGHOUSE MEMORANDA

<u>Date</u>	<u>Region</u>	<u>Subject</u>
10/16/91	IV	Dade County, Florida, Stack Height Increase
11/7/91	VI	Phelps Dodge--Hidalgo Modeling Protocol