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

AUG 17 1993

MEMORANDUM

SUBJECT: Draft Protocol for the UAM-V--A Correction

FROM: Joseph A. Tikvart, Chief
Source Receptor Analysis Branch, TSD (MD-14)

TO: Brenda Johnson, Meteorologist
Air Programs Branch, Region IV

In a July 22 model clearinghouse communication, we inadvertantly attached comments to you on the subject protocol which reflected some initial thoughts on the protocol and your initial draft suggested response. As a result of our subsequent discussions with you, we revised our initial comments somewhat, and should have sent the attached set of comments instead. Please disregard the comments which were attached to our July 22 clearinghouse response, and replace them with the attachment to this memo. I regret any confusion this error may have caused.

Attachment

cc: Regional Modeling Contact, Regions I-X
Ozone Modeling Contact, Regions I-X

I. Comments on the Draft Protocol

General. Preparation of a protocol for a head to head comparison between UAMIV and UAMV breaks new ground. In general, we feel that a number of the concepts presented in this document are good ones. We also believe inclusion of tests which allow for some relaxation of spatial and temporal pairing (pp. 22-23) is a good idea. However, we believe the protocol needs to address or elaborate upon several additional issues.

In order for UAMV to be accepted for application in Atlanta, in place of UAMIV, the **Guideline on Air Quality Models (Revised)** requires that it be subjected to a statistical performance evaluation, and the results must show that it performs better than UAMIV. The **Guideline** indicates that the **Interim Procedures for Evaluating Air Quality Models** should be used, as appropriate, in designing the protocol for such an evaluation. However, we recognize that the **Interim Procedures** were not designed specifically for use with episodic models such as the UAM. Thus, not all of the individual stipulations in the **Interim Procedures** document necessarily apply even though the principles contained in that document should be followed.

One apparent conflict between the protocol and existing guidance is the protocol's provision that if performance of the two models is comparable, UAMV should be the model of choice for use in the State implementation plan (SIP). As previously noted, the normal procedure for determining the most appropriate model is that if the proposed model does not perform clearly better than the reference model, then the reference model (in this case, UAMIV) should be used. However, the **Interim Procedures** do allow for the use of other technical criteria to make a decision in the case of comparable performance. A legitimate criterion would be scientific merit of the two approaches. To be consistent with the guidance, the protocol should make a strong case that the UAMV is scientifically superior to the UAMIV. We believe that this point needs to be addressed more specifically in the draft protocol. For example, the first paragraph on p.2 contains a list of new features within the UAMV. However, most of these are unaccompanied by explanations as to why the UAMV treatment is superior.

The EPA also has guidance which applies specifically to the use of the Urban Airshed Model to demonstrate attainment of the ozone standard in an ozone SIP. This guidance, embodied in **Guideline for Regulatory Application of the Urban Airshed Model**, requires explanations/justification for deviations. Procedures which have merit for comparing UAMIV vs. UAMV may not necessarily be consistent with those recommended for use in SIP applications and vice versa. For example, it may make sense to use the SAIMM meteorological model for comparing the two models, but this method of generating meteorological input is not the recommended procedure

in the Guideline for Regulatory Application of the UAM. Conversely, use of technical and management committees to reach consensus is the procedure we recommend for SIP modeling, but the same committees may or may not be the most appropriate means for reaching a conclusion regarding use of a non-guideline model. For the reasons outlined above, we believe it would be more appropriate to have separate protocols for the UAMIV/UAMV comparison and for the application of the chosen model in the Atlanta SIP.

Finally, once a model is chosen, attention should be paid to a potential problem posed if the model predictions are biased low. As a working group at the May 12-14 Atlanta UAM workshop concluded, this problem should be addressed on a case by case basis. Our concern over this issue becomes even greater if UAMV were to be the chosen model, despite predicting a lower episodic peak concentration than UAMIV for one or more episodes (see pp.41-42 of the Interim Procedures).

Specific Comments

1. p.4--**Schedule.** For devising the schedule, we would like to give you our latest estimates regarding availability of base case supporting ROM data. The July 7-8, 1988 episode should be available by the end of July. The July 29-August 1, 1987 episode should be available by the end of September. You should be advised however, that there are major unresolved contract uncertainties regarding ROM support and that these estimates are subject to change.

2. p.5--Are the technical working groups identified in the SIP demonstration protocol to be used in assessing/approving results and procedures in the UAMIV/UAMV comparisons? Their role is unclear. In any event, other participants should recognize Region IV representatives as the EPA spokespersons for decisions having to do with the model evaluation protocol.

3. p.7--**Episode Selection.** Are there not 5 rather than 3 primary episode days? Why aren't July 30, 31 and August 1, 1987 all considered to be primary days?

4. p.10--**Modeling Domain Specification.** The protocol mentions that it may be appropriate to use a fine mesh (2 km x 2 km) nested horizontal grid. Since this is a new feature of UAMV, it would seem appropriate to use the feature if the data base warrants. Where would this finer grid be located? The protocol is completely silent about the vertical resolution to be assumed in UAMV. Is it, like UAMIV, to be 5 cells, or will the resolution be finer to more closely reflect resolution available in the wind model? If fine horizontal and vertical resolution are used in UAMV, some concern arises over costs. It would not be appropriate to use the fine

resolution with UAMV to improve its performance over that of UAMIV, unless it were practical to take advantage of this capability in performing the SIP analysis.

5. **p.15--Input Procedures.** We agree that use of the mesoscale meteorological model with both UAMIV and UAMV is the best way to compare the models while at the same time taking advantage of new features offered by UAMV. However, the performance of prognostic models in photochemical modeling studies thus far has been less than overwhelming. What happens if both UAMIV and UAMV perform poorly as a result? Do the model comparisons with each other mean anything in such a case? The protocol should include a contingency plan to run UAMIV (and, if feasible, UAMV) with the Diagnostic Wind Model if performance of both models is poor.

6. **p.15--Input Procedures.** It is unclear whether four dimensional data assimilation or some other approach to use observed data to nudge SAIMM predictions is to be used. If so, would this be done consistently for UAMIV and UAMV? To what extent, if any, would diagnostic UAM analyses be used to revise the wind models? Remember our UAM applications guidance requires some physical justification for adjusting wind fields or other inputs, not just improved model performance.

7. **p.17--PTSOURCE.** The protocol needs to be more explicit about which sources it will treat as "major point sources" for plume-in-grid (PiG) treatment. Will PiG treatment apply to VOC sources as well as NOx? Are the cutoffs the same? Is there an upper limit to the number of sources which can practically be treated with the PiG algorithm in the Atlanta application?

8. **p.17--HEIGHT.** see comment #4 on vertical resolution.

9. **p.22--Statistical Measures of Performance.** The difference between observations and predictions should be computed by subtracting predictions from observations (Oi - Si), rather than as shown. This will provide signed numbers which are consistent with EPA Guidance and performance evaluations for other demonstrations.

10. **pp.22-23--Statistical Measures.** Greater effort needs to be made to be very precise about the definitions of these measures. This can be done by greater use of equations/subscripts in illustrating what is meant by tests 4-6. In addition, use of very simple examples might be made to illustrate the calculations for each of the tests (particularly (4) - (10)) so that the reader would have a clearer understanding of what these are. The following are examples of the sorts of ambiguities arising from the current descriptions:

in tests (1)-(3), the protocol needs to explain more clearly what "N" is.

test (4): are we talking about 1 "S" and "O" per primary day per monitor, or 1 value of each for an entire episode?

test (5): is the bilinear interpolation used, or is the prediction used the one from the 9 cells which agrees most closely with the observation?

tests (7) - (10): pictures and illustrations would be immensely helpful.

11. pp.23-24--While we concur with the basic approach for selecting a model, we feel that there is merit in comparing the performance of models based on the use of the fractional bais. Since we are not specifically asking that you incorporate measures based on fractional bias, we would like to obtain a copy of the final data set so that we may independently evaluate other comparative techniques based on the fractional bias. In particular, we would like to obtain (electronic), all of the houlrly observed and predicted concentrations for each model being evaluated for each day used in the evaluation.

12. pp.23-24--Determination of Acceptable Performance. We recommend that the two models be scored by combining (i.e., summing) the scores from each of the 5(?) primary days, so that there is one test score for each model. We also recommend dropping the notion that if the difference in scores is less than "1.5", it is too close to call. We feel that a better interpretation is, "if the score for UAMV is the same as or better than that for UAMIV, UAMV is the model of choice". (This assumes that the protocol contains a convincing argument that UAMV is scientifically superior). Our underlying rationale for these changes is that they enable one to take account of and weight an episode where one model performs very much better than the other. Further, we feel that the scores on many of the individual measures are likely to be zero. Thus, any non-zero difference in the test scores reflects one or more decisive differences in the performance of the two models.

II. Reaction to B. Johnson's Comments

A number of the comments reflect concerns over distinguishing between procedures which are appropriate for the model comparisons but not for the SIP application protocol, unless further justification is provided. We generally agree with these comments.