APPENDIX R:

RECOMMENDATIONS FOR PHASE II

APPENDIX R: RECOMMENDATIONS FOR PHASE II

The goal for Phase II of the project is to locate the sources identified in Phase I of the project. At best, the locations can be identified as general regions of the country. There are several methods that might be suited to using the source strength output from Phase I together with meteorological data and meteorological modeling to address this problem. Discussed below are the strengths and weaknesses of the ATAD (Atmospheric Transport and Diffusion) model developed by NOAA, the HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) model developed by NOAA, the CAPITA Monte Carlo model developed by Bret Schichtel of Colorado State University, Cooperative Institute for Research in the Atmosphere, and a hierarchical modeling approach being developed at this time by Battelle. Each of these methodologies is being used in on-going research (see Kenski, 2001). HYSPLIT, the Monte Carlo model, and the hierarchical modeling are being updated and developed further. Moreover, there is research under way in comparing the models. The results of these activities may make the recommendations in this report somewhat obsolete before the end of the year. Also discussed below are some recommendations on the procedures that might be used with the recommended model.

The current status of the hierarchical modeling methods makes this the least desirable of the methods for this project. It is a very powerful set of techniques, but they are not as well suited for this particular problem *at this time*. Successful use of these techniques would simply require much more development work than the others. In spirit, it would be similar to the Monte Carlo modeling, but with a different statistical basis. In its favor, the output can be a probability field rather than a series of trajectories. Also in its favor is that the development would likely be done in SAS or S-Plus and, hence, make the data management issues easier to work around. However, since the development has not been completed, there is no basis for comparison to the other models.

The remaining three models produce similar output and are reasonably well known. Any of the three could be used for ensemble back trajectory analysis. To make a recommendation, one should consider:

- Accuracy,
- Consistency with methodologies already used, and consistency with methodologies that may be used in the near future,
- Cost, and
- Availability.

Accuracy is difficult to assess. There have been some cross comparisons done between the models and they are continuing to be studied. Schichtel and Wishinski (1996) have compared HYSPLIT and the Monte Carlo model and found that when trajectories are treated in

ensemble they produce very similar results, at least when taken in aggregate. EPA and others have also compared HYSPLIT with observable data and found good agreement (see Draxler, 1996). Hence, there is some evidence, beyond the theoretical development of the models, for trusting in the accuracy of the output. Moreover, since the use in this case will be with aggregate sets of trajectories, only accuracy on average is needed. However, if there is doubt as to which is the correct model to use, then multiple versions should be compared.

Since this study is likely to be, at least in part, a model for similar analyses, consistency with previous studies and likely future studies should be a consideration. HYSPLIT is the clear winner in this category, since it is probably the most used and most scrutinized of the models. Moreover, EPA is considering developing a database of back trajectories with HYSPLIT for the year 2000. Consistency with this database should be considered as an important factor, so that they can be combined for future use.

The cost of the technique includes the cost in time to obtain and preprocess the input, the amount of time needed to use the model, the time needed to quality assure the input, and the cost of teaching future users how to use the input and/or output. In discussions with Bret Schichtel, the time needed for quality assurance and preprocessing of the input data for the Monte Carlo simulation for 1995 through 1999 is a concern. The time needed for quality assurance of the input data for ATAD should be similar. The input data for HYSPLIT are already being preprocessed and quality assured in the Reanalysis Project of the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR).

The availability of the models is probably not an issue for MARAMA unless multiple models are desired. In this case, the use of at least two models should be stated as a requirement.

Overall, the single model that seems to be best suited for the project is HYSPLIT, mostly because good input data are readily available and should continue to be available. If a pair of models is used, then HYSPLIT should be one of them to make the results more comparable with other studies.

Given that HYSPLIT will be included as a model, there is still some concern for operational parameters. Multiple starting heights spread between 50 m and 500 m should be used both to assess the consistency and as a way to smooth the resulting probability field. It may be desirable to use lower starting heights for sites that seem to have more local sources, and higher ones, say at least 150 m, for the rural sites that do not appear to have local sources.

There is also flexibility in how the trajectories and source apportionment output are combined and used. Bootstrapping methods and various metrics have been and continue to be explored. (See Hopke, Cizek, Li, and Landsberger, 1995 and Kenski, 2001.) There are a variety of other statistical techniques that would also apply for associating probability fields with the possible grid of locations. However, Rich Poirot's[1] success with using these types of data to

[1] Richard Poirot, Vermont Department of Environmental Conservation.

locate source regions is a very compelling model to follow. (See Poirot, Wishinski, Hopke, and Polissar, 2000.) Any work done should build on his methodologies. For example, smoothing or modeling the probability field may prove useful in dealing with the edge effects. It may also be useful to weight trajectories with local meteorological data. For example, one could down-weight the trajectories when it rained.

The main departure from what Rich Poirot has done should be the use of all or multiple site solutions simultaneously. This assumes that common sources can be identified between the source apportionment solutions. By having multiple receptor sites, the problems associated with the fact that all the relevant trajectories converge to the receptor site are reduced. Much better triangulating should be possible.

The overall recommendation is to use HYSPLIT with newly released data, to use Rich Poirot's methods for analyzing the trajectories together with the source apportionment output, and to use multiple sites simultaneously in addition to site-specific analyses. Priority for funding should go toward a simultaneous, multiple site analysis.

(This page intentionally left blank)