Preliminary Comments on the REA from Dr. James Boylan

Ambient Air Concentrations (Chapter 3)

4. Does the Panel find the description of the three study areas and their key aspects (section 3.3) to be clear and technically appropriate?

The REA Planning document identified nine candidate study areas that meet the air quality, design value, and population criteria. However, the draft REA added a couple of new criteria for selecting individual study areas. None of the locations selected in the draft REA were mentioned in the REA Planning document. The REA should document the other areas that were considered and the reasons they were not selected. For example, Savannah, GA seems to meet all the criteria listed in the draft REA.

5. Does the Panel find the description of the air quality modeling done to estimate the spatial variation in 1-hour concentrations (section 3.2) to be technically sound and clearly communicated?

The air quality modeling used in the draft REA appears to follows standard modeling procedures to estimate 1-hour concentrations.

6. To simulate air quality just meeting the current standard, we have adjusted model predicted 1-hour SO2 concentrations using a proportional approach focusing on the primary emissions source in each area to reduce the modeled concentrations at the highest air quality receptor to meet the current standard (section 3.4). Considering the goal of the analyses it to provide a characterization of air quality conditions that just meet the current standard and considering the associated uncertainties, what are the Panel's views on this approach?

This approach seems reasonable. Figure 3-3 seems to be missing two columns of receptors to the left and right of the fine grid.

7. A few approaches were used to extend the existing ambient air monitoring data to reflect temporal patterns in the study area (section 3.5). Does the Panel find the approaches used below to be technically sound and clearly communicated?

a. Data substitution approach for missing 1-hour, 5-minute maximum, or 5-minute continuous ambient air monitor concentrations (section 3.5.1).

Seems reasonable.

b. Estimating pattern of within-hour 5-minute continuous concentrations where 1-hour average and 5-minute maximum are known (section 3.5.2).

Seems reasonable.

c. Combining pattern of continuous 5-minute concentrations within each hour from monitors in or near the study area with the modeled 1-hour concentrations (section 3.5.3).

Seems reasonable.

Characterization of Uncertainty and Representation of Variability (Chapter 6)

b. To what extent have sources of uncertainty been identified and their implications for the risk characterization been assessed (section 6.2)?

Table 6-3 discusses multiple sources of uncertainty. The first category is "AERMOD Inputs and Algorithms". However, AERMOD Model Outputs" should be added to the first category or added as a new category. Specifically, the spatial and temporal uncertainty associated with the modeled 1-hour SO_2 concentrations should be discussed. See detailed discussion below related to "Modeled Air Quality Evaluation" in Appendix D.

<u>APPENDIX D</u>

The first paragraph of Appendix D states:

"AERMOD output for the three study areas was evaluated using three methods. First, comparison of the 99th percentile of daily 1-hour maximum concentrations for each and subsequent 3-year design values were compared at each monitor. Second, simple QQ-plots were generated to provide a quick visual performance of the model for 1-hour, 3-hour, and 24-hour averages. The QQ-plots are comparisons of the observed and modeled concentrations, unpaired in time and space, consistent with regulatory evaluations of AERMOD (U.S. EPA, 2003; Venkatram et al., 2001). Third, for a more rigorous comparison, the EPA Protocol for determining best performing model, or sometimes called the Cox-Tikvart method (U.S. EPA, 1992; Cox and Tikvart, 1990) was used. Normally, this protocol is used to determine which model or model scenarios among a suite of models or scenario is the better performer for regulatory application and focuses on the higher concentrations in the concentration distribution as these are the concentrations of interest in most regulatory applications (State Implementation Plans and Prevention of Significant Deterioration)."

The ISA states, "For models intended for application to compliance assessments (e.g., related to the 1-h daily max SO_2 standard), the model's ability to capture the high end of the concentration distribution is important. Measures such as robust highest concentration (RHC) (Cox and

Tikvart, 1990), and exploratory examinations of quantile-quantile plots (Chambers et al., 1983) are useful. The RHC represents a smoothed estimate of the top values in the distribution of hourly concentrations**. In contrast, for dispersion modeling in support of health studies where the model must capture concentrations at specified locations and time periods, additional measures of bias and scatter are important.**"

All three of the model evaluation methods used in Appendix D are associated with using the model for regulatory compliance assessments. For example, the model's ability to capture the high end of the concentration distribution is evaluated with Q-Q plots where the highest data point from the model is compared to the highest data point from the observations even if they occur at different locations, time of day, and season of the year. In the REA, the model is being used to support health studies where spatial and temporal accuracy is much more important compared with compliance assessments. Since the APEX model uses the model results paired in time and space, the model results need to be evaluated against observations paired in time <u>and</u> space. Appendix D does include absolute fractional bias (AFB) paired in space and presents Q-Q plots paired in space. However, there is no detailed discussion on the model performance paired in space <u>and</u> time. The last sentence in appendix D states "Given the lack of temporal variability of source emissions in the model and the fact that a monitor does pick up temporal variability of emissions not seen by the model, the performance of AERMOD is acceptable for the purposes of this exposure assessment." It is not clear how the conclusion that "AERMOD is acceptable for the purposes of this exposure assessment" was determined.

To examine model performance paired in time and space, EPA should generate scatter plots and hourly time series plots for each year at each monitor. In addition, seasonal average diurnal time series plots should be developed. Finally, model statistics should be calculated for each monitor across all hours (8760) as well as broken down by hour-of-day (24) and season-of-the-year (4). These plots can be used to evaluate temporal biases in the SO_2 concentrations generated by the air quality model.

The REA needs to discuss "acceptable" model performance criteria for the statistics presented in the tables. For example, what are acceptable (or typical) values for composite performance metrics (CPM), absolute fractional biases (AFB), and percent difference between observed and modeled 99th percentile daily 1-hour maximum concentrations and 3-year design values? Add references to support. What if the model does not meet acceptable model performance criteria?

Finally, options for correcting the model results if there are significant biases in the modeling results should be discussed. EPA should consider adjusting the REA model results up/down to match the SO_2 observations. This approach would keep the relative spatial distributions identified by the model in place, but would adjust the concentration levels to match the observations. This would minimize the impacts from poor model performance on the ambient SO_2 concentrations used in the exposure estimates. For example, high SO_2 concentrations at Fall River are significantly under-predicted in 2013 leading to modeled exposures in APEX that are

biased low. Alternately, a series of sensitivity runs should be performed with the adjusted REA model results to see if the model under- and over-predictions significantly impact the exposure results and conclusions in the Planning Assessment.