Preliminary Comments on the ISA from Dr. Steven Hanna

Comments on Chapter 2

Note that my expertise is primarily in atmospheric transport and dispersion modeling and analysis of observed concentrations, and my comments focus on those areas. I was asked to comment on Chapter 2 of the 2nd draft ISA on the areas under "Ambient Air Concentrations".

General comment 1 – I see that two sections of interest to me have been enhanced in response to our comments a year ago. These are section 2.5.4 "Relationships between hourly mean and peak concentration" and section 2.6.1 "Dispersion modeling. These revisions lead to much better justifications for subsequent analysis. However I have some specific addition suggestions as described below.

General Comment 2 – One of our earlier comments was that it is not consistent with the basic theory to compare the peak 5 min concentration C with the hourly averaged concentration in the same hour where that 5 min C occurred. If you look at the basic theory of peak to mean, it first defines a longer sampling time period over which the mean is calculated. For example, a day could be used as the sampling time (longer time period) for calculating the mean. Then the peak 5 minute C and the peak 1 hr C (or the peak 8 hr C or whatever) can be calculated for that one day period. Thus the basic theory refers to the ratio (peak x min C)/(peak y min C) over the entire sampling period and this is proportional to $(y/x)^p$. Turner suggested $p = 0.2$ and that has been widely used in applications and works fairly well. I suggest that the above definition be used in the EPA analysis.

General Comment 3 – Although the EPA made the suggested revisions to some sections, they did not put the revised concepts into practice in other major sections describing the general processes of plume dispersion, and I see that the revised methodology is not included in the ISA (separately reviewed).

General Comment 4 – I realize that this is an internal EPA document, but it is still a scientific study and the concentration analyses and dispersion modeling sections should review the literature outside of the EPA. The document is being made available for "public" review. There is much work that has gone on in other U.S. agencies and across the globe on topics such as statistical analysis of meteorological and air quality data and development and application of dispersion models for all averaging times. I would like to see a comprehensive review done and then pros and cons listed for alternative methods and rationale for choosing a specific method.

Specific Comments:

p 2-17 ff – Section 2.3 on atmospheric chemistry and fate – I see that there are many equations and references in this section, plus detailed discussions and justifications. This format should be used for all sections.

p 2-29, Fig 2-10. It has been stated that there are 438 sampling sites but this figure contains far fewer than 438 dots. This is probably because there are several areas in the US where sampling sites are concentrated, such as around major industrial areas and in metropolitan areas, and so the points fall on top of each other. As suggested earlier, it would be helpful to include zoomed in figures to show some of these concentrated areas.

Section 2.5.2 -Spatial Variability (with many tables and figures) – As commented a year ago, this analysis of spatial variability appears to have been done without reviewing the U.S. research in other agencies and the international literature on this topic. It is known that atmospheric variables are influenced by the full spectrum of motions, which are characterized by random turbulence variations and effective space scales. For example, SCIPUFF parameterizes this space scale of turbulence in its formulations. The EPA should also look at its own literature such as the 1970s-1980s RAMS study of SO_2 in the St Louis metropolitan area, where there were many samplers operating and several comprehensive analysis reports published.

Pasquill's book "Atmospheric Diffusion" provides the mathematical basis for the space and time variations of meteorological variables and pollutants.

There have been several attempts to fit basic statistical distribution functions to sets of observations such as these (e.g., the data in Table 2-6 on p 2-35). "Long-tail" distributions are usually found, and can be fit by, for example, and exponential distribution that accounts for intermittency. The Robust High Concentration Method used by OAQPS is based on a specific distribution. As stated in later report sections where the text responds to my earlier comments, the variations are also functions of meteorological parameters (e.g., stability, wind speed) plus nearness to large point sources and the stack height. Those concepts should be tested with these data.

p 2-50 (Fig 2-19 for 1 hr avg) – Here is an example of where the "distance scale" could be estimated, although it would be better to have ln C on the vertical axis. My eyeball estimates of the distance scale are about 50 km for all sites except NY, where 150 km might be better. The scientific reason for the larger scale at NY is that NY is the largest metropolitan area. Plus NY is in the middle of the northeast US SO_2 polluted area and has input coming from upwind power plants and other urban area.

p 2-52 top par – These conclusions are truisms that would be expected after a review of the literature on the topic.

Section 2.5.3 – Temporal variability – Again, there are no non-EPA references. The topic of time variability would have a much broader literature that the topic of spatial variability since it is easier for researchers to obtain the data and analyze them from one sampler. Ralph Larsen of the EPA spent much time during his career on this topic and his work should be summarized. His suggested distribution functions should be compared with these new results.

The conclusions about seasonal variability etc, should be compared with those of the many similar analyses.

Section 2.5.3.3 - Define "diel" in the section caption. Figs 2-23 and 2-24 – I would help my comparison of sites if the same vertical scale were used.

Section 2.5.4 – Relationships between hourly mean and peak concentrations – See my general comment 2 above. The max hourly average <u>in the 24 hr day</u> should be used in the ratio. The bottom par on p 2-62 and the scatterplots are erroneously comparing the peak 5 min C in a certain hour to the hourly C <u>for that hour</u>.

Thank you for adding the expanded references and discussions to this subsection.

p 2-62, line 8 – Clarify what exactly is "longer". I think that you are referring to the overall sampling time for the data. The current report is using the 5 min averaged peak C in a 24 hr period and the 1 hr averaged peak C in a 24 hr period. If you expanded the sampling time to one week, the 5 min peak during a week and the 1 hr peak would both likely be larger (they would never be smaller). Also, I would discard the possibility that the 5 min peak equals the 1 hr peak, since that would never happen with good data.

In the scatterplots, there are so many points that they overlap each other and it is not possible where the "best fit line" would pass. Can a best fit line be shown? Mention that, the way you define it (for a given hour) the max possible value of the ratio is 12, Please include the well-known Turner Workbook 0.2 power law prediction, which would be a line of $12^{0.2} = 1.64$ (however this would be using the peak 1 hr avg C in the 24 period).

p 2-65 top and Table 2-9 – What variables are used in calculating the correlation coefficient?

Section 2.5.5 – Background concentrations – Don't we know what SO_2 concentrations are occurring at west coast stations on the coast? Don't SO_2 sources in Canada and Mexico cause transport to the US? How exactly is "background" defined? For example, earlier in this report, you defined six regions in the US with dimensions of about 50 to 100 km. Don't each of these regions have a "background" SO_2 concentration on their upwind boundary?

p 2.6.1 Dispersion modeling – This section is much improved (see general comment 1). However, it would help readers' understanding to point out that the basic dispersion modeling

theory is not locked into any averaging time. AERMOD uses 1 hr because that was mandated by the Clean Air Act. The basic Taylor, Batchelor, and Pasquill references often include averaging time as a variable and can show how concentrations vary with averaging time. European air quality regulatory models such as ADMS allow any averaging time. SCIPUFF (the most widely used dispersion model in the U.S. outside of the EPA) allows any averaging time. AERMOD's basic boundary layer turbulence parameterizations would also allow smaller averaging times than 1 hr. All that is needed is some modifications to the program statements (I confirmed this two days ago with Drs. Jeff Weil and Akula Venkatram, the two main developers of AERMOD).

p 2-70, first line of last par – Insert "by EPA" after "model". Use this qualifier elsewhere in the section, too (such as lines 3, 4 14, and 29 of p 2-71).

p 2-72 line 36 – Did Weil call these "errors"? Usually natural variability is a large contributor, too.

p 2-73, lines 13-17 – Mention the other agencies who are now using WRF for inputs to dispersion models (NRC, DOE, DOD). This is a worldwide trend as computer power increases. There are found to be a few problems in applications – such as a tendency of WRF (or any NWP model) to have wind direction errors of 30 degrees or more, and a failure of the WRF grid (4 km or 12 km) to account for subgrid terrain effects.

p 2-73 lines 33-34 – The RHC is based on a distribution assumption. Why can't this approach be used in the current report?

p 2-74 and 2-75 – As stated above, AERMOD could be modified to include any averaging time (as done in most other dispersion models).