Guideline on Air Quality Models
March 19-21, 2013
Sheraton Raleigh Hotel, Raleigh, N.C

Pounding Nails with Shoes to Decide Which Shoes to Buy

By
John S. Irwin
Retired NOAA (2004)
www.jsirwin.com

1. Title Slide:

I was invited to "talk about anything I wished to talk about", so those of you who have not left, or do not leave during my talk, unfortunately will have to endure my topic for discussion and strange sense of humor.

For over twenty years I have endured model evaluations that I personally would never present or commit to publication. Today is my day to vent! But besides venting, I hope to provide a solution – or at least a path to enlightenment.

2. Project Prairie Grass:

Conducted in the summer of 1956, this is one of the classic non-buoyant dispersion field experiments. The location is near O'Neil Nebraska which is near the northern border of central Nebraska.

There were 90 receptors on the first three arcs, and 180 receptors on the last (800m) arc. That makes a grand total of 540 receptors, on 5 arcs extending from 50 m to 800 m downwind.

There were 68 experiments with a near-surface release (approximately 0.5m release height for all but the last four (4) experiments), sampled with receptors at 1.5m above the nearly uniform ground cover (mowed wild hay) with a surface roughness of just less than 1cm. The averaging time of the concentration measurements was 10 minutes.

I have shown two cases, one for stable nighttime (left) and one for unstable daytime (right). The red dots are the maximum concentrations seen along each arc (more will be said of these values later).

If you study the two plots, you can see that the right-hand plot (stable nighttime) is only about 20^o wide, while the left-hand plot (daytime unstable) is about 40^o wide. This is

typical for near-surface releases; all factors being equal, the highest concentrations occur during stable nighttime conditions.

Barad, M.L. (Editor) (1958): Project Prairie Grass, A Field Program In Diffusion. Geophysical Research Paper, No. 59, Vol I , Report AFCRC-TR-58-235(I), Air Force Cambridge Research Center, 299 pp.

Barad, M.L. (Editor) (1958): Project Prairie Grass, A Field Program In Diffusion. Geophysical Research Paper, No. 59, Vol I I, Report AFCRC-TR-58-235(II), Air Force Cambridge Research Center, 218 pp.

Haugen, D.A. (Editor) (1959): Project Prairie Grass, A Field Program In Diffusion, Geophysical Research Papers, No. 59, Vol III, AFCRC-TR-58-235(III), Air Force Cambridge Research Center, 686 pp.

3. Electric Power Research Institute (EPRI) Kincaid:

Conducted in the summer of 1980 and spring of 1981, this is an extensive dispersion field experiment involving a buoyant release from a power plant stack. The stack was 187 m in height with a plume rise of about the same magnitude.

The location is 15 miles Southeast of Springfield IL on the southwest extent of the Sangchris Lake State Park. The lake was created in 1950 to provide cooling water for the Kincaid power plant, which was then under construction. The land is fairly flat and homogeneous with a surface roughness of 7cm to 15cm depending on wind direction.

There were concentrations collected on 29 days in 1980 (197 hours) and 21 days in 1981 (175 hours) for a total 2,808 arcs for analysis, extending from 0.5km to 50km downwind, with most of the useable data in the 3km to 20km range.

The experimental procedure was to inject SF6 tracer directly into the stack for a period of 4 to 7 hours, and to activate about 200 samplers in the anticipated downwind direction. The receptors were about 1.5 m above ground and collected 1-hour samples.

I have combined the developmental and evaluation data sets (Hanna and Paine, 1989); sorted the data using objective criteria to 12 arcs, and assigned objective quality codes to the results:
3 = 5 or more nonzero values, with the maximum within the middle 1/3;
2 = 5 or more nonzero receptors, with the maximum outside of the middle 1/3, and
1 = the rest.

In the comparisons to be shown, I have used quality 3 data, of which there are 546 arc-hours of data. I limited my analyses to the 447 arc-hours of data from the 3km – 20km arcs:

3km = 56 hours
5km = 88 hours
7km = 100 hours
10km = 78 hours
15km = 74 hours
20km = 51 hours

We lose 16 arc-hours because AERMET has no hourly surface data for 7 days and Kincaid OBS data are available. Of the remaining 530 arc-hours, ISCST3 computed zero concentrations for 83 arc-hours (plume rise goes above Zi), so I have 447 arc-hours where I have estimates from both ISCST3 and AERMOD.

Hanna, S.A., and R.J. Paine (1989): Hybrid plume dispersion model (HPDM) development and evaluation. J. of Applied Meteorology. (29):206-224.

4. Scatter and Cumulative Frequency Plots of Model Estimates of Maximum Concentration Versus Observed Arc-Maximum Concentration Values:

In the upper scatter plots, the model estimates are along the x-axis and the observations are along the y-axis.

The common assumption in least-square fits for scatter plots is that all of the uncertainty is in the y-axis values. Since model estimates are deterministic (same value is estimated for a given circumstance, and different values are observed for a given circumstance), I have placed the model estimates in all plots on the x-axis.

I could have separated the ISCST3 (Version 02035) and AERMOD (Version 12345) to separate plots, but you can quickly get the impression (which is confirmed in the cumulative frequency plots) that ISCST3 typically has higher maximum concentrations than AERMOD (see shift to the right in blue ISCST3 values in comparison to red AERMOD values).

In the cumulative frequency plots, it is easily seen that ISCST3 (blue) is in closer agreement with the larger observed arc-maxima concentration values, while AERMOD seems to be consistently under-estimating the larger observed arc-maxima concentration values.

Plots and evaluation statistics involving comparisons of observed arc-maxima with model estimates are common in the journal and symposium literature.

Question: Is it valid to compare model estimates with arc-maxima values?

Slide 5. What do air dispersion models use to estimate concentration values? What do dispersion models estimate (or characterize)?

To estimate concentration values, the air dispersion model must estimate the lateral (horizontal) extent of the dispersing plume (top-left figure); the vertical extent of the dispersing plume (bottom-left figure), and the buoyant plume rise (right-side figure).

Lateral and Vertical Dispersion

Shown in left-side figures is a comparison of model estimates (horizontal x-axis) versus observations (vertical y-axis). The model is a version of Roland Draxler's model (1976) that I recommended in Irwin (1983). It performs quite well, with correlation coefficients above 70%, and most of the estimates within a factor of two of the observations.

For those of you not familiar with air dispersion modeling, these comparisons are unbelievably good! Yet there is unresolved scatter, and the scatter in the observations is easily estimated to be about a factor of 2.

Draxler, R. R., 1976: Determination of atmospheric diffusion parameters. Atmos. Environ., (10):99-105.

Irwin JS and Hanna SR. Characterizing uncertainty in plume dispersion models (2004): Proceedings of the 9th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes held in Garmisch-Partenkirchen, June 2004, pages 287-292.

Irwin, J. S., 1983: Estimating plume dispersion-A comparison of several sigma schemes. J. Climate Appl. Meteor., (22):92-114.

Plume Rise

The stringy lines in the plume rise figure are the observations, and the dotted line is the model estimate. The uncertainty in the final plume height is seen to be on the order of the plume rise, which implies a factor of 1.3 uncertainty in plume rise estimates.

Briggs, G.A. (1969): Plume Rise. U.S. Atomic Energy Commission, Office of Information Services. Available as TID-25075 from NTIS, U.S. Department of Commerce, Springfield, VA 22151 82 pages.

Ensemble Averages

So our dispersion models provide very good estimates of the lateral dispersion, vertical dispersion and plume rise – ON AVERAGE. Stated another way, the dispersion model provide an ensemble average of the concentrations to be seen.

An ensemble average is not one 10-minute (or one 1-hour) concentration average; it is determined through the analysis of a collection of 10-minute (or 1-hour) concentration values all have very similar dispersive conditions.

Venkatram, A. (1979): The expected deviation of observed concentrations from predicted ensemble means. Atmos. Environ. (11):1547-1549.

"…we expect the 1-h averaged concentration to deviate from the ensemble mean by more than 100%. …This analysis shows that under unstable conditions, poor comparison of observations with predictions should be expected. …Our discussion brings up the question of model validation. It is clear that the expected deviation can be reduced by averaging several observations under similar conditions. Then, for adequate validation, the predicted concentration should be compared against an average derived from an ensemble of measured concentrations…."

Fox, D.G. (1984): Uncertainty in air quality modeling. Bull. Amer. Meteoro. Soc. (65):27-36.

"…There is agreement in the meteorological community that air quality modeling results contain various types of uncertainty (although not state explicitly) such that they represent no more than an estimate within the distribution of possible values. Generally, turbulence must be averaged in time, space, or over a number of realizations of the flow pattern, in order to elicit meaningful information. In doing this, parameters such as the dispersion coefficients are defined by their mean or average values without consideration of the variation around that mean. … The details of atmospheric motion fields are not predictable without uncertainty, nor is the concentration of a pollutant released into any turbulent fluid predictable without uncertainty. In studies of turbulence, it is convenient to introduce the notion of an ensemble, namely a number of repeats of the same 'experiment,' holding external conditions (boundary and initial conditions) fixed…."

Weil, J.C., R.I. Sykes, and A. Venkatram (1992): Evaluating air-quality models: review and outlook. Journal of Applied Meteoro. ((31):1121-1145.

"…Air-quality models predict the mean concentration for a given set of conditions (i.e., an ensemble), whereas observations are individual realizations drawn from the ensemble. The natural variability: is the random concentration fluctuation about the mean and is large (of the order of the mean; section 2). The steering committee considered the natural variability to be very significant in hampering the performance evaluation. …The natural variability, also called the inherent uncertainty (Fox 1984; Venkatram 1982), is caused by PBL turbulence. It arises because the details of the velocity field are not the same in each realization of a turbulent flow…."

Slide 6. A look at some pseudo-ensembles.

The top two figures were formed by combining concentrations from six (6) Project Prairie Grass 10-minute experiments (top-left is near-neutral daytime and top-right is near-neutral nighttime). These results are for concentrations measured along an arc 400-m downwind from the release.

The bottom figures were formed by combining concentrations from six (6) EPRI Kincaid 1-hr experiments (bottom left is near-neutral and bottom-right is verging on stable). These results are for concentrations measured along an arc 5km downwind from the release.

Typical of other results I have seen, the ensembles formed using EPRI Kincaid data exhibit more raggedness than those developed using Project Prairie Grass data. This likely is traceable to the effects of buoyant plume rise and the transport of the dispersing plume by turbulent eddies to the surface.

The solid black lines are Gaussian fits to the combined data, and represents what a very good dispersion model might provide as an ensemble estimate. The little red dots are the arc-maxima from the individual releases.

As indicated in the inset table, arc-maxima are on average 43% larger than the Gaussian fit Cmax to their respective concentration profiles for Project Prairie Grass, and are on average 2 times larger than the Gaussian fit Cmax to their respective concentration profiles for EPRI Kincaid.

So, if arc-maxima differ so greatly from the ensemble maxima, why are we comparing arc-maxima with dispersion model ensemble-average maxima to assess model performance?

Excuse #1: It is an easy-to-compute estimate of the observed ensemble maximum concentration.
Answer: Not likely a good or reasonable estimate seeing the results displayed in this slide.

Excuse #2: EPA uses dispersion models to estimate maximum concentrations, so we are providing air quality managers an assessment of how good the estimates are.
Answer: This tells managers how well the model performs when misused, but says nothing about the true performance of the dispersion model.

MY VENT:
You can use a shoe to pound nails into a board. You can even decide which shoes to buy based on their ability to pound nails into a board. But shoes were never made to pound nails, and their ability to pound nails into a board makes for terrible selection criteria as to which shoes to buy!

Dispersion models were never constructed to estimate short-term maxima. You can misconstrue what dispersion models do and say they estimate individual realization maxima; you can even try to assess dispersion model performance by comparing model estimates with individual realization maxima.

It makes no sense to select shoes on their ability to pound nails, nor does it make sense to assess dispersion model performance through comparisons of modeling results with short-term arc-maxima.

For nearly 20 years, I have observed presentations and read papers whose sole basis for assessing dispersion model performance was comparisons of observed individual maxima (arc-maxima) with model ensemble-average maxima, and I have held my tongue and cringed. No longer will I do so. As of this day, I am speaking out, and hope as a result, to put to an end a practice that has absolutely no science basis.

Dispersion models provide ensemble average estimates of the concentration values; observed concentration averages are individual realizations from some ensembles. Their direct comparison with model estimates has no basis in science or statistics – these comparisons have nothing to do with 'model evaluation.'

Slide 7. What can we do that that makes sense, makes good use of available data, and obeys our understanding of science?

Turner (1970) provides the definition of the cross-wind concentration (Cy), and it is seen to be essentially a function of the vertical dispersion and plume rise (UCy/Q). If we substitute Cy into our equations for dispersion, we can derive the last equations (in the red box) for the centerline concentration (Cmax) being equal to Cy divided by the lateral dispersion (Sy) and divided by $\sqrt{2\pi}$.

To save breath, I will not continuously state the obvious that Cmax, Cy and Sy are all ensemble averages.

This expression for Cmax shows that in order to obtain the correct value for Cmax, we must get the correct value for Cy and Sy. Anyone disagree?

I disagree! Because, you can overestimate Cy and overestimate Sy and yet get the correct value for Cmax. Anyone disagree?

I disagree! Because, you can underestimate Cy and underestimate Sy and yet get the correct value for Cmax.

So why do we spend so much effort assessing the ability of models to estimate centerline concentration values? There are three ways to get good agreement, and only one is correct. Worse yet, you have no idea if you get good agreement, whether it is for the correct estimation jointly of Cy and Sy.

Seems to me that we should ignore Cmax, and focus on comparing observed and estimated group averages of Cy and Sy.

Slide 8. Is this difficult to do?

Shown in these three figures are the crosswind integrated concentration values (Cy) observed and estimated by ISCST3 and AERMOD at 100m downwind from the release.

Shown are all 68 experiments. The three shown in red occurred during such low-wind and stable conditions, that the SO2 gas behaved as a dense gas, and pooled on the ground following the gentle drainage channels in the nearly flat terrain.

You will see people who have ‘selected’ 43 of the Project Prairie Grass experiments for use in a model evaluation exercise. I typed in all of the concentration values into a file for use; double-checked the values; plotted the results in various ways, I know these data.. I am the one who provided people with my files.

There are three (3) experiments to omit from use (experiments 4, 13 and 14), all the rest are valid and should be used. I know that some do not fit a preconceived notion of a lateral Gaussian cross-section. So! Who says they have to ‘look’ a certain way before you can use them in a model evaluation exercise? If you only challenge a model to do well on the ‘nice and pretty’ cases, how do you think it will perform on the more typical nasty cases?

I did not make a direct comparison of the Cy values, because we need to create group averages (pseudo-ensembles). I took a direct and simple route, I averaged the 10 most unstable cases, then the next 10, and so forth.

Problem: When I plugged in my average Cy and average Sy into our formula for Cmax, the result did not agree with the 10-value average I computed for Cmax. Then I remembered all the papers I have written, describing the distribution of raw centerline concentrations as having what looks like a log-normal distribution, with a geometric standard deviation of order 1.5 to 2.0.

So I recomputed my group averages as geometric averages. When I plugged these 10-value geometric averaged Cy and Sy values into the formula for Cmax, the result was nearly identical to what I computed directly for the geometric mean of the 10 Cmax values.

This is a handy result, because it says our group geometric means of Cy, Sy and Cmax all obey our relationship for how Cmax is related to Cy and Sy (in a Gaussian dispersion model). It also means that if you are provided with the observed Cy, Sy and Cmax values, you can vary the grouping criteria, and you do not have to recompute the observed Cy, Sy and Cmax values for the individual experiments. They can be computed once, and used in various ways.

We now have a methodology for performing an assessment of dispersion model performance that makes sense, makes good use of available data, and obeys our understanding of science.

1. We focus on comparisons of group geometric mean values of observed and modeled Cy and Sy. The emphasis is on Cy and Sy, not Cmax.

2. Once we understand whatever biases the models have for Cy and Sy, we can look at the Cmax results, and likely will understand why the results are as they are.

Slide 9. Project Prairie Grass – ISCST3.

1. I am using log scales on both the x- and y-axis so we can see what is happening with the small values, as well as the large values.

2. The blue line is an ordinary least-square fit forcing the intercept to be zero. This provides a quick overall average assessment of bias, but as you will see, it is not always useful.

3. As I mentioned before, modeled values are on the x-axis since in ordinary least-squares fits, all the uncertainty is assumed to be in the y-axis values, which in our case are the observations, which we know should have larger variances than the deterministic model values due to unresolved stochastic processes.

4. You will see little labels on the plot symbols. 1, 2, 3, …5 to denote which arc: 1 = 50m and 5 = 800m. The decimals denote stability, so 1.1 is the 50m arc and the most unstable group, and 4.6 is the 800m arc and the most stable group. To further help, I have used solid black symbols for the most stable two groups on each arc.

The use of geometric means insures a science basis in the relationship between Cy, Sy and Cmax. You should not deviate on this. The use of log scales on x- and y-axis; least-square fits forcing the intercept to be zero, and labeling the symbols are what I call ‘good-practices’ developed from experience; trial and error. They are worth using, unless you think you can do better.

ISCST3 is doing very well both in estimating Cy and Sy. The slight underestimation of Cy is offset by the slight over-estimate of Sy, such that Cmax is well characterized.

Slide 10. Project Prairie Grass – AERMOD

Here we see that AERMOD is consistently under-estimating Cy and over-estimating Sy. This obviously leads to under-estimating Cmax.

If you study the plots some, you will see that Cy is under-estimated for the most stable cases, and Sy is over-estimated for stable cases, which means the cases for a surface release when we should have our maximum Cmax concentrations is consistently under-estimated.

Ths bias is so obvious, it likely is easily corrected.

Slide 11. EPRI Kincaid – ISCST3

Here we see what I had hoped to show.

Note, the thin black line in the Sy plot is the 1:1 line.

1. ISCST3 grossly under-estimates Cy for near-stable conditions, but these do not result in high surface concentrations. ISCST3 significantly under-estimates Sy for the near-stable cases.

2. ISCST3 has a tendency to over-estimate only a few of the Cy values for arcs 1-3 (3-, 5- and 7-km) where the highest observed concentrations occur. ISCST3 significantly under-estimates Sy for the unstable cases.

3. The biases in Cy and Sy compensate for one another, such that ISCST3 overestimates Cmax during unstable conditions (when the highest observed concentration occur).

A clear explanation of why ISCST3 correlates so well with arc-max concentrations!

Slide 12. EPRI Kincaid – AERMOD

1. Study of the plots shows that AERMOD is over-estimating Cy for near-stable conditions, and under-estimating Sy for these near-stable cases. This leads to over-estimates of Cmax for near-stable cases.

2. AERMOD over- and under-estimates Cy for the unstable conditions. The same happens for Sy with over- and under-estimates of Sy.

3. If you flip back and forth, comparing ISCST3 and AERMOD on the EPRI Kincaid data, you may see what I see, there seems to be some structure in the model's behavior.

a. This is what I call 'good behavior,' as it suggests that a minor adjustment in the algorithms for the dispersion parameters and/or plume rise, might well resolve the biases seen.

b. However, I would make adjustments following a logical sequence:

(1) We have intensive field experiments where the releases are non-buoyant: Project Prairie Grass, Round Hill, Copenhagen and Cabauw. The first two involve 10-minute average concentrations. Copenhagen has two consecutive 15-minute averages, and Cabauw has two consecutive 30-minute averages. Using these data, we can attempt to see what needs to be improved in the characterizations of the lateral and vertical dispersion, without plume rise complicating the picture.

(2) Round Hill has a unique set of 10 experiments having joint measurements of 0.5-min, 3-min and 10-min concentrations along arcs. Combined with the data just mentioned, we have an interesting set of data to teach AERMOD how to adjust the dispersion as a function of averaging time.

(3) EPRI Kincaid and EPRI Bull Run are similar in that they involve tall stacks with buoyant plume rise and 1-hr SF6 sampling. Once you have adjusted the vertical and lateral dispersion in AERMOD using (1) and (2) above, you can then look to see if the plume rise needs to be adjusted, and/or the interaction of the buoyant plume with the convective mixed layer height.

Slide 13. Summary of lessons learned (I hope):

1. All air-quality models (ISCST3, AERMOD, ADMS, CMAQ, CAMx, etc.) provide estimates of the ensemble average concentration. Direct comparisons of short-term (1-hour or less) observations with modeling results is highly questionable. There are large stochastic fluctuations affecting short-term concentration values, which prohibit meaningful comparison between modeled ensemble averages with short-term concentration values.

2. I am recommending comparisons of group geometric mean values of observed and estimated Cy and Sy values. Once you understand what is happening with Cy and Sy, you can look at Cmax. We place too much emphasis on the importance of Cmax in our model evaluations; Cmax is dependent upon Cy and Sy (not the other way around).

a. I suggest using log scales on both the x- and y-axis so we can see what is happening with the small values, as well as the large values.

b. I suggest using an ordinary least-square fit forcing the intercept to be zero, to provide a quick overall average assessment of bias, but as you will see, it is not always useful.

c. I suggest placing the modeled values on the x-axis since ordinary least-square fits assume the uncertainty to be in the y-axis values, which in our case are the observations, which we know should have larger variances than the deterministic model values due to unresolved stochastic processes.

c. I suggest you use labels on the plot symbols. 1, 2, 3, …5 to denote which arc: 1 = 50m and 5 = 800m. The decimals denote stability, so 1.1 is the 50m arc and the most unstable group, and 4.6 is the 800m arc and the most stable group. To further help, I have used solid black symbols for the most stable two groups on each arc.

The use of geometric means insures a science basis in the relationship between Cy, Sy and Cmax. You should not deviate on this. The use of log scales on x- and y-axis; least-square fits forcing the intercept to be zero, and labeling the symbols are what I call ‘good-practices’ developed from experience; trial and error. They are worth using initially until we devise something better.

3. It is my contention that there is no science or statistical basis for assessing model performance using observed short-term arc-maxima values. You can pound nails with shoes to decide which shoes to buy and you can attempt to assess dispersion model performance through a comparison of model estimates with observed arc-maxima values.

Neither of these activities makes good sense!

Selecting only the short-term arc-maxima from intensive field data sets for evaluation of dispersion model performance makes very poor use of available data and does a poor job of revealing the underlying model biases.

Slide 14. AERMOD Biases

Using ratios of Cy and Sy, we can look a bit further into where we might look in AERMOD to improve its performance. It is interesting to see that the bias in Cy varies mostly as a function of downwind distance, and the bias in Sy varies mostly as a function of stability. I will leave further detective work to those familiar with AERMOD's and AERMET's features and estimation procedures. Of course, we can not rule out that I may have not properly run AERMET or AERMOD, so I have tried to explain in the Appendices how I created my estimates.

Slide 15. Research ideas

1. Averaging Time. Since we have near-surface and elevated intensive field experiments with concentrations taken with various averaging times, we should be able to augment AERMET/AERMOD to provide ensemble average concentration values for 1-hr or less.

2. Short-Term Concentration Fluctuations: Concentration values with averaging times of 1-hr of less are heavily affected by stochastic effects. It is yet a question in my mind of whether we can characterize short-term concentration fluctuations through direct adjustments in AERMET/AERMOD or through post-processing 1-hr AERMOD estimates.
a) Any short-term ambient air standard would require development of a capability to estimate the distribution of fluctuations about the ensemble-average concentration value. This also suggests to me that short-term ambient air quality standards would need to be probabilistic (e.g., probability of exceeding some maximum acceptable concentration).
b) Given the capability to estimate the frequency distribution of concentration fluctuations for averaging times of 1-hr or less, opens up the possibility of estimating in a probabilistic sense potential for odor and flammability.

Appendix A: Analysis Steps

1. Compute Cy, Sy and Cmax from observations
a. While listing out the computed Cy, Sy and Cmax values, I converted the Project Prairie Grass concentrations from milligrams/m^3 to micrograms/m^3.
b While listing out the computed Cy, Sy and Cmax values, I converted the Kincaid concentrations from PPB to micrograms/m^3, and I listed to a separate file the Kincaid stack parameters in AERMOD's format of hourly emission rates. Note, ISCST3 and AERMOD share the same hourly emission input file format.
c. I wrote a program to read my master file of quality codes and from this I determined that there were six arcs where I likely had sufficient data to form and use group averages sorted by stability, namely: 3, 5, 7, 10, 15, and 20km.

2. Define Dates –
a. Since I am using all 68 experiments initially, I just listed to a side file the 68 dates and times.
b. For EPRI Kinciad, I only need ISCST3 and AERMOD modeling results for those hours when I have valid observed values for Cy, Sy and Cmax. I created a program to read the Kincaid file of Cy, Sy and Cmax results computed in Step 1, which then listed out the hours for which I need modeling results.

3. AERMET meteorology –Note I have fixed the wind direction to be coming from the west, as this simplifies Steps 5 and 6. More discussion in Appendix B.

4. CutMet - Using the listing of dates defined in Step 2, I extracted from the AERMET surf and prof data files, those hours for which I needed modeling results. As I did this, I created the meteorology input file for ISCST3, which required me to convert the Monin-Obukhov length to a Pasquill stability category. More discussion in Appendix B.

5. Define Receptors – I created a program where you tell it the distances downwind in meters; the starting azimuth, the spacing between receptors in degrees, and the number of receptors on each arc. The program creates the receptors, one arc at a time, and lists the coordinates in a format that can be used directly in ISCST3 and AERMOD. For Project Prairie Grass, I created 120 receptors with 1-degree separation on 5 arcs (50, 100, 200, 400, and 800m). For Kincaid, I created 300 receptors with 0.5-degree separation on 6 arcs (3, 5, 7, 10, 15, 20-km). I did not use actual ground elevations for the receptors, i.e. I assumed a flat world.

6. Model runs – Since I am dividing all concentration by the emission rate, I set the emission rate to be 1000 g/s, to insure that I would have nonzero concentrations even out to 50km. The concentration listed in the plot files of AERMOD and ISCST3 are in micrograms/m^3. More discussion in Appendix C.

7. Compute Cy, Sy and Cmax from modeling results – The format of the plot files created by ISCST3 and AERMOD are slightly different, but the first line of their respective files provides the model name and version number, so I can use one program to process the

modeling results from both models. I know from Step 5 how many receptors NUM are along each arc, so I read NUM concentrations in: compute Cy, Sy and Cmax, list out the results, and go to the next NUM of receptors to process. More discussion in Appendix D.

8. Combine Results –
a. For Project Prairie Grass, I computed observed values of Cy, Sy and Cmax for all 5 arcs and all 68 experiments, and I computed Cy, Sy and Cmax for all 5 arcs and all 68 experiments. I combined the observation results and modeling results in an Excel spreadsheet. I pulled the AERMET surf file into an Excel spreadsheet, and extracted AERMET's Monin-Obukhov length, which I then copied and pasted into my file of combined results for Project Prairie Grass.
b. For Kincaid, I computed Cy, Sy and Cmax from the observations results for all arcs having a Quality Code of 3, regardless of how far downwind the arc was. Since I have modeling results for 6 arcs, but may or may not have results for any or only some of the arcs, I wrote a program to combine the observed, ISCST3 and AERMOD Cy, Sy and Cmax values into one file. While doing this, I listed out AERMET's Monin-Obukhov length for each combined listing. Once I had the combined listing, I pulled the results into an Excel spreadsheet.

9. Form Groups – I use Golden Software Grapher, which allows me to create all my plots using data stored in Excel spreadsheets.
a. I started with Project Prairie Grass, doing a combined sort first on distance and then on 1/OBK. I then inserted columns where needed and created 10-value averages of observed and modeled Cy, Sy and Cmax. I had 65 values with which to work, but only used the first 60, which left out the 5 most stable values.
b. I then sorted the Kincaid combined data, and computed 10-value averages, leaving out a few of the most stable values, as in processing the Project Prairie Grass data into group averages.
c. While looking at the results, I decided to add the least square fits forced to have a zero intercept. I could not see what was happening for the low values, so I changed all the plots from linear to log axes. Then I added the labels to the symbols, to see whether I could see any trends. It was after all of the above, that I saw that the arithmetic 10-value averages were not obeying the relationship between Cmax, Cy and Sy. I tried geometric 10-value averages, and I found that the relationship between Cmax, Cy and Sy was preserved. This is a brief summary of the redoes I went through to come to a final set of plots and analyses.

Appendix B: Meteorology Processing

a. Source for data - Integrated Surface Database (ISD), Hourly, Global
http://www.ncdc.noaa.gov/most-popular-data#dsi-3505

b. Reformatting of data – AERMET can not read the ISD data as delivered, so I wrote a program that fakes it into the CD144 format, which AERMET can read. I set the wind direction to 270 degrees, to simplify defining the receptor rings for use in AERMOD and ISCST3. (**CDO2CD144.exe**)

c. I noticed that I had hours with zero wind speeds here and there in the data that I had created in CD144 format. My purpose is to demonstrate a methodology for evaluation of plume dispersion models, not make the definitive assessment. Hence, I use linear interpolation and filled in calm wind hours. (**FillCD144.exe**)

d. I have to run consecutive hours through AERMET to generate the input meteorology for AERMOD, because AERMET using 3-hr averaging to define the stable mixing height. However once the meteorology is developed, I can omit all hours except those for which I need AERMOD results. I created a program that reads a user define file of dates (and hours) for which I want AERMOD results, and the program: 1) list to a file the needed surface and profile data for each needed hour, and
2) creates the ISCST3 input meteorology data. AERMET had all the needed meteorology except for a Pasquill stability category, which I could estimate using the AERMET Obukhov length and the site roughness length. (**MyCutAERISC.exe**)

Note: I created a program to list the needed Dates and Times for EPRI Kincaid (see Appendix D b iv).

Appendix C: Dispersion Modeling Processing

a. AERMOD Plume Meander – One of the 'tricks' I am using to perform my analyses is to compute the modeled crosswind integrated concentration, Cy, and the lateral dispersion, Sy, using concentrations along receptor rings. In default mode, AERMOD's plume meander will compute nonzero concentrations essentially in all directions, which does not realistically replicate what we see during the intensive field experiments. So I used the non-default model option LOWWIND1 to turn off AERMOD's plume meander. All my AERMOD input files contained the following statement:

MODELOPT BETA CONC NOCHKD FLAT LOWWIND1

b. Receptor Field – I did not consider the ground elevation to be of any significance for the two intensive field experiments that I used in my analyses. However, I needed an easy way to generate receptors along rings at specified distances and separation. AERMOD and ISCST3 use the same flagpole receptor format, so I create a program to list out the needed receptors in the proper format, which I then cut and pasted into the input files for AERMOD and ISCST3. (**MakeReceptors.exe**)

c. Both AERMOD and ISCST3 will list out the concentrations generated at each file in a 'plot' format. All my input files contained a statement similar to the following:

POSTFILE 1 ALL PLOT KINSF6-1981.PST

The only thing that changed was the filename. If you have 120 receptors, then for each hour, the ISCST3 and AERMOD list out 120 concentrations. I created a program that reads the plot files, and computes the Cy, Sy and Cmax values for each arc from the modeling results. (**AZRingAnalysis.exe**)

Appendix D: Observations Processing

Observations – The format of the Project Prairie Grass and EPRI Kincaid observations are slightly different, with different data included in the headers for each arc. I used the same subroutines to compute Cy, Sy and Cmax, but created separate programs for each field data. The input files have the actual emission rates, so when listing out Cy and Cmax, I divided these values by the emission rates.

a. Project Prairie Grass. The Project Prairie Grass results are recorded in milligrams per meter cubed, which I converted to micrograms per meter cubed (for comparison to the modeling results). (**PGRAnalysis.exe**)

b. EPRI Kincaid.

(i) The EPRI Kincaid results are recorded in PPB, which I converted to micrograms per meter cubed (for comparison to the modeling results). The conversion is in the header record for each hour and arc.

(ii). There is plume rise for the EPRI Kincaid experiments, so while listing out the observed values of Cy, Sy and Cmax, I created an hourly emissions file in AERMOD's format (which is the same for ISCST3).

(iii) I used only those arcs that had a quality code of 3 (which means the observed maximum was in the middle third of the observed concentrations along the arc, the observed maximum looked to be in accord with concentrations seen along arcs in front and behind, and there were at least 5 nonzero concentrations along the arc). (**KINAnalysis.exe**)

(iv) Kincaid Modeling Dates. Once I computed the observed values for Cy, Sy and Cmax, I created a program that used the file created KINAnalysis to create a listing of the dates and hours for which I need modeling results for the Kincaid experiments. This listing of dates and times was used by MyCutAERISC (see Appendix B d 2). (**KINDefineDates.exe**)

Appendix E: Final Processing

a. Project Prairie Grass. Once the AERMOD and ISCST3 modeling results were processed by AZRingAnalysis, I combined these results with those generated by PGRAnalysis in Excel, where I formed 10-group averages for plotting and analysis.

b. EPRI Kincaid.

(i) Once the AERMOD and ISCST3 modeling results were processed by AZRingAnalysis, I combined these results with those generated by KINAnalysis using a program I created. (**KINCombineAll.exe**)

(ii) I converted the combined file into an Excel file, where I formed 10-group averages for plotting and analysis.