Data for testing draft Practice.

On June 14, 1999 an update was made to ACODE.zip. The new version of this zipped file is named acode(4).zip. The old version is named acode(3).zip.

The update consist of updates made to a fortran program named NEWFIT.FOR. See discussion below.

1. CROSSWIND INTEGRATION, GOODNESS OF FIT STATS, OUTPUT DATA
Helge Olesen reported that unusual values were being listed for the crosswind integrated values for Kincaid. In the Feb 23, 1999 version of NEWFIT only one option was provided to compute the crosswind integrated concentration, in which the concentration pattern was 'folded' about the center of mass, and only values >= CMIN were used in the integral. The pattern was 'folded' to provide a more stable integral when concentration patterns become noisy. The use of only nonzero concentrations ignored gaps in the profile. Ignoring gaps and zero values works sometimes, if the gaps are not too large, but when the patterns become very noisy and large gaps appear in the concentration profile then ignoring the gaps inflates the integral, and provides the erroreous values Helge reported.

The problem is not trivial to solve, because when the pattern becomes noisy it is difficult to defend any integral. What I have done in this update is to provide four other results: 2 using a folded pattern and 2 using no folding. In each case, I run one with all values, and then a second using only the values >= CMIN. I also provided listing of all four results, along with ratios of the various pairs.

Then (since the 'folding' was causing more problems then it solved', I used the simple crosswind integration, using all values (even zeros if they are reported). This usually provides the smallest value for the crosswind integral, thus the smallest value for the fitted maximum (Cmax). Using this new version of NEWFIT will alter the values output in STATSUM.DAT for Cmax, Cy, CyU/Q.

To improve the intrepretation, I discarded the attempt to provide a regression analysis, in favor of a simple least-squares fit of the form:

$$\ln(obs) = a \ln(pred) + b.$$

I took the log of the values as the residuals about the fit appear to be somewhat log-normally distributed. I now report the statistics of this fit (slope, intercept, regr2). I also now list to the output file the obs and pred concentration pairs. Using this version of NEWFIT will alter the values output in STATSUM.DAT for IERR, R2.

From a quick review, I surmise the following. When all the ratios of the crosswind integration pairs are within 0.80 to 1.20, and the regression coefficient (regr) is found to be significant at 90% confidence bounds, the concentration profiles look Gaussian, and we can likely trust the

crosswind integrals. If anyone can refine this, have at it, and report you results. The basic question is: how do you compute a crosswind concentration integral with the profile is 'noisy'?

2. ASSIGNMENT OF RECEPTORS TO CERTAIN ARCS

Helge Olesen has noted that for Kincaid, the assignment of some of the receptors to certain arcs (especially true for 3km and 5km arcs) appears to be in error. I did not invent the assignments of receptors to the arcs (I used what was delivered to me). But what Helge reports is correct, it looks like the assignments need to be improved.

3. ARC 13, 'WHAT ARE YOU?'

Please note, that arc '13' is not an arc, for either Kincaid or Indianapolis. It is assigned a distance of '99' in my Kincaid 'input.dat' file, and is assigned a distance of '88' in my Indianapolis 'input.dat' file. Arc '13' are those receptors that no assignment to arcs was made (this is again a convention in the data listing provided to me, and not my invention).

On February 23, 1999 an update was made to two of the archived files (acode.zip and eval.zip). The new versions of these files are now called acode(3).zip and eval(6).zip. The differences are as follows: I deleted an erroneous version of newfit from eval.zip, and I added two data files to acode.zip to provide a complete listing of all modeling results for Kincaid.

Complete data sets for use in testing this practice have not been developed. I have partial data sets (they have the concentration data, but not the meteorology). The meteorology would have to be obtained from data volumes for the various experiments, etc. Such data as I have is in an anonymous ftp directory, that can be accessed via:

http://www.epa.gov/asmdnerl/asmd.html

Go to the left side of the page. Click on ‘Data Access’, follow instructions on this page by clicking on ‘ASMD 'anonymous ftp directory' Go to the 'Irwin' directory. There are three files there that may be of interest

‘Prof(3).zip’ contains a routine I’ve used to compute Monin-Obukhov lengths for the Prairie Grass data. To better understand what is in this zipped file you will want to study the Lotus 123 spreadsheet (pgexp.123) that is contained in eval.zip.

'Eval(6).zip' contains tracer data from several experiments. In eval.zip are the input files to a routine called Newfit.for. (See Read2.me which is contained in eval(5).zip).

‘Acode(3).zip’ contains a fortran routine ASTM90.for that implements the current version of the draft ASTM model evaluation procedure. It takes as input a file with the ‘splus.dat’

type information. If you are evaluating 1 to 5 models, then you will also have to provide the modeled centerline concentration (divided by Q) for each arc and experiment. (See Read1.me which is contained in acode(3).zip)

The 'splus.dat' file can be created using Newfit.exe, which is needed for input to Astm.exe (see below). If you are working with other experiments, you can easily create splus.dat, but you will have to create a routine to do this, as the receptor coordinates need to be expressed in terms of the center of mass for the arc, and the concentrations need to be divided by the emission rate.

.....
Copy of draft Practice (astm.pdf)

Astm.pdf is the latest version of Z6849Z (dated April 20, 1998), Standard Practice for Statistical Evaluation of Atmospheric Dispersion Models. It is in adobe-reader format. Adobe-reader is freeware, and is available from any number of locations off the www.

Z6849Z is being provided to you to allow you to participate in the review process. Comments should be sent to me. I can be reached at:

John S. Irwin
U.S. EPA (MD-14)
RTP, NC 27711

irwin.john@epa.gov
tel: 919-541-5682
fax: 919-541-0044

Some History:

Draft 1 Dated April 21, 1997
This version of the draft was sent to ASTM Subcommittee D22.11 for internal review and balloting on May 16, 1997. This first balloting was closed June 16, 1997. Comments received were of an editorial nature, with two negatives. The essence of the negative votes were that the ideas needed to be tested further.

I had a temporary employee for the summer of 1997, who was a graduate student in statistics. Working with him we have discovered that in general the draft practice seems to work, but there are several flaws that need to be addressed.

Flaw one. The practice calls for computing and comparing specific percentiles derived from cumulative frequency distributions. And it calls for bootstrap resampling to be used to allow confidence bounds to be placed on the comparison statistics. In our testing we have found that the histograms generated by the resampling of the 50-th and 90-th percentile values are sometimes uniform, sometimes shewed one way or the other, and never look bell-shaped. We

would prefer to have bell-shaped distributions for these values, because it provides a better basis for subsequent statistical tests.

Idea to correct Flaw one. Use the average. The histograms of the averages computed from the resampling all appear to be bell-shaped.

Flaw two. The practice calls for computing the lateral dispersion for each arc of concentration values, and selecting values deemed to be close to the computed center of mass for the arc, with the selection criteria specified in terms of y/Sy, where y is the distance from the center of mass and Sy is the computed lateral dispersion. This works well for Project Prairie Grass, where the receptors are evenly spaced and close to one another. But it does not work well for Kincaid.

Idea to correct Flaw two. Compute the center of mass for each arc, as before. Then express the receptor coordinates relative to this computed center of mass, as distance from the center of mass or degrees, etc. Now lay each of the arcs on top of one another, placing all the centers of masse on top of one another. Now compute Sy for the entire group of arcs. Note, the arcs used in this manner have been previously sorted into smaller groups (regimes). So the Sy is computed for all the arcs in this subgroup or regime. Now use the 'group' Sy to sort through and select receptors that are 'near' the center of mass. Testing of this on Prairie Grass and Kincaid suggests that this will work. But we have only just started testing these ideas.

Darft 2 Dated April 20, 1998
See paper Irwin and Rosu (1998) proceedings of AWMA/AMS Phoenix, AZ January 1998 pages 6-10.

Was balloted in June 1998. Negative comments received generally called for more testing before accepting. In this regard, the current standard compares the modeled centerline concentration with observations within 0.67Sy of the computed center of mass of the group of experiments collected together. There is a concern that since the observed values are not exactly at the computed center of mass, the modeled values compared with the observations should be simulated at coordinates (relative to the computed center of mass). The current version of ASTM.for does not provide for this to be tested, so if someone wishes to explore this (which I would greatly appreciate), the fortran would have to be altered, etc.

Best Regards,

John S. Irwin