# HYSPLIT Air Concentration Uncertainty Analysis

15 July 2008

## 1. Introduction

A typical uncertainty analysis consists of four different procedures to investigate possible variability in model concentration predictions: 1) an ensemble to evaluate gradients in the meteorological data not properly captured by the discrete data grid points, 2) using different vertical velocity assumptions to evaluate the possible under-sampling in space and time of the vertical motion field, 3) testing different physical representations of the dispersion process that determine how much of the mixing is sub-grid scale versus how much is explicitly defined by the meteorological data, and 4) the effect of turbulence in evaluating one realization from a process that has been defined by its ensemble (turbulence computed from mean fields).

Each of these analyses configured through a script which follows a pre-defined series of steps. First, HYSPLIT is run for each model variation, either automatically (in the ensemble and variance modes only) by creating a concentration output file with a suffix ending in .001, .002, etc., or manually (physics or vertical motion options) where the base file name is renamed with the proper suffix in the script. Second, the probability analysis program is called which reads all files with three digit numeric (.000) variations from the input base file name. All concentration data files are loaded into memory and sorted in ascending order of concentration at each grid point. Based upon the sorted values, output grids are created for concentrations at various probability levels or probabilities at various concentration levels. These grids can then be displayed individually using *concplot* or through the *boxplots* program at a specified location.

## 2. Meteorological Data Ensemble

In the ensemble form of the model, the meteorological grid is offset in either X, Y, and Z for each member of the ensemble. The model automatically starts each member on its own processor (*hymodele*) in a multi-processor environment or cycles through the simulations on one processor (*hymodels*) in the single-processor environment. The calculation offset for each member of the ensemble is determined by the grid factor as defined in the *SETUP.CFG* file. The default offset is one meteorological grid point in the horizontal and 0.01 sigma units in the vertical. The result is 27 ensemble members for all offsets. Because the ensemble calculation offsets the starting point both directions in the vertical, the starting point height should be set at least 0.01 sigma (about 250 m) above ground for all height variations to be valid. The model simulation will result in 27 concentration output files named according the file name setting in the control file "{base_name}.{001 to 027}" with a suffix equivalent to the ensemble member number. On a single processor system, the calculation may take some time to cycle through all the members. Only one offset (+ or -) in either X, Y, or Z is applied per member. Starting and ending member numbers can be selected where the range 1-9 applies for no vertical shift, range 10-18 applies for a positive shift, and range 19-27 applies to a negative shift.

## 3. Turbulence Ensemble

Another ensemble variation is the variance option, which also creates 27 ensemble members, but the differences are due to variations in turbulence rather than variations due to gradients in the gridded meteorological input data. The variance ensemble should only be run in the 3D particle mode and with fewer particles, in proportion to the number of ensemble members to the number of particles required for a single simulation. For instance, if 27,000 particles are required to obtain a plume representation that is not sensitive to particle number, then each member should be run for 1000 particles. Normally the same random number seed is used when computing the turbulent component of the particle motion. However, in the variance ensemble, the seed is different for each member, resulting in each member representing one realization of the ensemble. The rationale for this approach is that the relationship between meteorological turbulence and plume dispersion has empirical constants that have been derived from a large number of experiments. The results from any one experiment must be within the bounds of the base case.

## 4. Model Physics

The basic dispersion computational method uses either the 3D particle approach or the puff assumption. In the particle approach, the turbulent path of each particle is explicitly defined while in the puff approach rate of growth of the particle distribution is modeled. In terms of the representing the underlying physical processes, we really don't know the true path of each particle, only that of the ensemble of all particles which in theory should be similar (but not identical) to the puff simulation which uses the velocity vector only at the puff center position. The growth rate of the puff or the separation rate of the particles from the mean is a function of the size of the puff in relation to the turbulent spectra. If all turbulent scales are larger than the puff, then the puff is just moved about by its entirety by the winds. If the turbulent scales are smaller than the puff, the particles are just randomly moved about within the puff. This is expressed initially as linear growth with time progressing to square-root growth (random walk) as the puff grows in size beyond all the turbulence scales. In the 3D particle model the turbulence is defined through the auto-correlation of the turbulence between time steps for a single particle as the model knows nothing about all the other particles. In the puff mode, the original HYSPLIT assumption was that because puffs split when they reach the size of the meteorological grid, all sub-grid growth is assumed to be linear and the square-root growth is implicitly modeled through the splitting process which is controlled by the actual meteorological data. In case this assumption is not valid, the puff growth computation can also be done using the square root assumption. These options exist because in the current HYSPLIT configuration, the turbulent length scales are fixed, at 3600 s and 100 s for the horizontal and vertical directions, respectively. Although there are approaches that permit the computation of the turbulent length scales from mean fields, that would add another level of variability into the calculation and which suggests a knowledge of the turbulence spectra that is not justified by the nature of the data (meteorological model output fields) driving HYSPLIT.

Another physics option is to change the way the turbulence is computed from the mean fields. The default option is to use the fluxes of heat and momentum to compute a boundary layer stability parameter and in conjunction with the turbulent length scales, a turbulent velocity is computed. However, not all meteorological models produce the same flux fields, some may be

instantaneous, some may be averaged, and some may be integrated over the duration of the meteorological model simulation. For this reason it is also possible to specify for HYSPLIT to use the low-level profile of winds and temperatures to compute the stability. This option works best when 10 m winds and 2 m temperatures are also available in the input data file. Another turbulence option would be to use the turbulent kinetic energy field if available. The model will automatically partition the turbulence between the horizontal and vertical components, or the partition can be forced. Note another complication is that NCEP model TKE fields are only for the vertical component.

The last option, not fully implemented prior to HYSPLIT 4.9 is to select the method to compute the boundary layer depth. The default is to use the value of the field provided by the meteorological model if available, otherwise the depth is computed to be the height at which the potential temperature first becomes 2 degrees larger than the minimum value (usually near the ground). Although this approach appears overly simplistic and perhaps inferior to more sophisticated Richardson number methods, the mixed layer depth is not just the capping layer of low-mixing but it should also represent the top of the layer from the ground up where vigorous mixing is occurring. Ri and other local gradient approaches are very sensitive and may lead to large mixing depth variations between time steps. In V4.9, it is also possible to set a fixed height.

## 5. Vertical Motion

When vertical motions are controlled by synoptic scale processes and are on the order of 1 cm/s (36 m/h) the under-sampling the vertical motion field (output every 3-6 hrs) should not be important. However as spatial resolution increases, the vertical motion field no longer provides the hydrostatic compensation for the synoptic scale but would also include the effects of smaller scale processes related to convection, terrain, and other land-use variations. Resulting vertical motions may be much larger (10 cm/s or more) and could approach boundary layer depths easily within one or more hours. This suggests the need for more frequent meteorological data output intervals or a different approach to computing a particle's vertical motion. Therefore there are several different options available within HYSPLIT to compute vertical motion from the meteorological data fields. These options include maintaining a particle on constant surfaces of pressure, density, sigma, and potential temperature. Another option is to compute the vertical motion by integrating the horizontal velocity divergence.

## 6. Scripting Options

Four example scripts are provided as a starting point to easily configure various uncertainty simulations. In some cases, the master *CONTROL* and *SETUP.CFG* files are provided and in other cases either the *CONTROL* or *SETUP.CFG* files are created within the script.

| | |
|---|---|
| Meteorological Ensemble | - *run_metd.scr* |
| Turbulence Ensemble | - *run_turb.scr* |
| Model Physics Ensemble | - *run_phys.scr* |
| Vertical Motion | - *run_vert.scr* |

In addition to the HYSPLIT executable and the concentration plotting program, the scripts call two additional programs, one to create the probability files, and the other to create a box plots from those files.

*conprob* – From the concentration file base name, the program opens and reads all files with the three digit numeric suffix (.000).

Usage: conprob [-options]

| | |
|---|---|
| -b[base] input file name | without the suffix |
| -c[concentration] value set | maximum value for cmax00 |
| d[dignostics = true] | |
| -p[pollutant] index number | for files with multiple pollutant |
| -t[temporal] aggregation period | treats time variation as an ensemble |
| -v[value] below which equals zero | |
| -z[level] index number | for multi level input files |

Output files include the following: prob05, *prob10, prob25, prob50, prob75, prob90, prob95* for concentrations at various probability levels (5% to 95%), and *cmax01, cmax10, and cmax00* for probabilities at various concentration levels (1%, 10%, and 100% of the maximum), and finally the special files *cmean, cnumb, cvarn*, representing the concentration mean, number count, and variance. For example, in a plot created from the *prob90* file the concentration contours would represent the region in which only 10% of the ensemble members have air concentrations greater than the contour value.

*boxplots* - Once the probability files have been created, the concentration distribution at a single location can also be viewed by entering a latitude-longitude location. The nearest grid point will be selected and the concentration values are retrieved from the *prob*{xx} files, which represent the various points on the box plot. The limits of the box represent the quartiles (25th and 75th percentiles), the whisker extensions show the 10th and 90th percentiles, and the circles show the 5th and 95th percentiles. The median value is shown by the line through the box, while the mean is shown by the plus symbol. Up to 12 time periods can be shown in one plot.

USAGE: boxplots [–options]
-x[longitude]
-y[latitude]