In the previous section we created a ensemble with multiple meteorological data sets and then we did some statistics on those individual simulations. And also we created an ensemble mean. However there is a more automated way to do at least a statistics component. And we will look at this section, the utility program that is included as part of the graphical user interface to do statistical analysis of ensemble data files, or other data files as well.

In the previous section you would have created five output files that are called hysplit2.001 through . 005, each with different meteorological data. These files should be in your working directory. If not, go back to the previous section, and redo those calculations, or you can run the batch file or script in the previous section to generate these output files. Once you've completed that, we will continue on. We're going to go, first, well the first step we need to make sure is, that the wild card name in the menu for the base name for the concentration in the menu is hysplit2, but I see in this particular case the menu had been reset, so what I would propose is you retrieve the basic CAPTEX CONTROL file and also for the name list. Now when we go to the setup menu we should have a base name here of hysplit2 so that the internal scripts will add the suffix ensemble as needed. Remember we're not running the model here, we're just setting variables within the graphical user interface.

Now to do, or to manipulate let's say, the HYSPLIT concentration output files, there is a utility program that will

work with the binary files, the binary output files from HYSPLIT, and the utility program we want for this is the merge program, want to merge together multiple output files. And we want the input name, in this case to be hysplit2, this will be the wildcard, the base name for the wild card. And let's create filenames. This'll create an INFILE. This is very similar to many of the other menus and you know that worked because it replaced the wild card with INFILE and if I were to look in working directory, we should have now INFILE, which was just created, and we should confirm that it has the files that we are interested in. And in this case it has two additional files that we do not want, and I will delete them, and now we have the correct INFILE for the five simulations that were done in the last step.

The next thing we need to do is give the output file a name, and I'm going to call it hysplit2.006. So we'll make the 6th member of the ensemble or that's what were going to create here. And what we're going to do is, we're going to merge together the five files and create a sixth file, and be merged together, I mean add together, and once we've added them together, what we're going to do is multiply by .2 or divide by five. So if you add five things together and divide by five we get the mean. And that's what we're creating here is the mean or the ensemble mean for those five members. And in this working directory we've now created the sixth file which is the mean for the five others and we could look at the plot at this point if you like. The plot of the mean concentration but I'm not going to do that, I'm just going to quit here.

And then we're going to go to the concentration, display, ensemble, and we're going do statistics now. So once we've created these ensemble files we can do statistics, and hysplit2 is the base name for the wildcard. We need to select the measured data file which will be the entire data set and this is fine for the output file, you can rename it to something else, and we're going to execute, and it's finished processing and what we get here is essentially the same thing that we had in the previous section, except it's all nicely summarized for us into one file, one output file. We've analyzed all the ensemble members and we can see that the highest rank, the overall rank is the number six, the ensemble mean by a quite substantial margin over the others. The next highest was the 9 km WRF. The, and you can see there is quite a difference in bias as well, I mean the ensemble mean has virtually no bias, it has by far the lowest normalized mean square error and highest correlation. This is consistent with what we discussed in the previous section, is that the ensemble mean can at times provide much more robust results. And this is a quick way of doing a statistical analysis of multiple ensemble members.

And this concludes our discussion of ensemble member statistics.