From the last three sections on air concentration uncertainty, we found that the meteorological grid ensemble provided some of the largest uncertainty values. Therefore, we will take a closer look at using multiple meteorological data to generate an ensemble. We will do this by rerunning the simulation for CAPTEX, but using five different meteorological data sets. Unlike the other ensembles this processes is not automated within the graphical user interface and the meteorological input data need to be changed manually. So before we get started, regardless of where you are coming from, you should do a reset on the graphical user interface, and then start by loading or retrieving the previously saved CONTROL file and name list files for the basic CAPTEX simulation.

Now before we close the name list menu, let’s make these run a little faster and reduce the particle number from 50000 to let’s say 10000, and that would be menu number 4. And the rest is relatively straightforward; we will go to the concentration menu, and let's start with the default simulation for NARR. The only thing we should do is with each new simulation, we should change the output name. And because we know that the ensemble programs use the three digit suffix, we'll need to supply this manually, so the first run, for the NARR, will be number one. And we will go through this in sequence, run one to run five, using the ERA40, the Global Reanalysis, and the 27km WRF, and the 9km WRF.

So the first run, and as you can see with 10,000 particles these do not take very long. We should after each run also compute the statistics. If you remember those steps, it is simply the utilities, Convert to DATEM , and we will use the measured data, the full measured data set, because we are running 68 hours, and we will use the convert to picograms. And we should also rename the statistical output to make sure it's identified because we're going to do several of these, so let's call it NARR.

So first step, create a DATEM file, compute statistics, and you can see we get .59 and a rank of 2.8. And let's just look at the scatter plot, because you have to compute statistics, exit, scatter plot. Oh right, because it's not called dataA.txt. So I guess it's an error in the GUI. If we don't rename the output, compute statistics, and then to scatter plot, that works because it's looking, looking to plot the data, dataA.txt file. But we had renamed that with the NARR output and that is flaw the graphical user interface. But for now we don't really care about the plots, so I'm going to redo this, exit and then quit. And what we should've done, and this is the file that we tried to plot, but the graphical user interface did not do that, but this is the output file that we did want, the statistics.

Okay so the next step is to actually go through and do this for the remaining four data sets. The ERA40 is now number two. Convert to DATEM to get the statistics. Go on to the global, which is number three. Convert to DATEM. Number three which is GBLR, exit. Fourth data set, which will be the 27 km WRF and rename. And then convert to DATEM. And lastly the 9 km WRF, which would be number five.

Now clearly, the easier way to do this is to run the script or the batch file, which automatically would go through a loop here, for the five simulations, and depending upon which loop element, it would set the meteorological data file variable accordingly. So you could easily set this up to do other types of, or include additional meteorological data or other variations, perhaps with different meteorological data. As you, it then does, it runs HYSPLIT, it converts to the data format, does the statistics, renames the statistical output file, and then we will discuss the rest of this in a few moments.

Now that the last simulation is completed, we can actually go into the probability display programs, like we did with the other ensembles. Now before I do that, if you remember when we created the other ensembles, the graphical user interface script automatically appended the three digit ensemble member identifier to the end of the output file name. But we had to do this manually. So the ensemble scripts within the graphical user interface, does not know that we did this manually, so what we need to do is go back into the setup menu, and remove the three digit appendix from the output file name. So now these programs will look for the three digit appendix after the base name.

So if you recall, we need to now go into concentration, display, ensemble, and the first step will always be view map, and we could look at the 50th percentile, for instance, but we're just go click on this. And remember there're only five members, so we're looking here at the median concentrations. We're not so much interested in looking at this map, but we want to go on now and look at the boxplots for Little Valley New York, so we can compare that with the previous plots that were done. So display, ensemble, boxplot, and if you recall the location of Little Valley New York was 42.2 and 70, 78.8, was that correct? I'm thinking that might that be correct. Yeah, it is correct. Actually 42.25 and 78.80, and we will execute this. And you can see here, this plot looks a little different because we did the full 68 hour simulation and we have output actually from the beginning until the end. But you can see that the variability is quite large, except for one 3 hour period, which is interesting, and this actually contains an additional time which is plotted on the next frame. But these variabilities are much larger than the variabilities that we have seen in the earlier sections, which is consistent with our findings, and you can of course see the member numbers, which shows member number two having some of the highest concentrations. Member number two, if you recall was the ERA40.

So the last step in this, since we've generated the statistics, you can actually go and look at individual runs, and then lay them out and see which is best. And I can open up the working directory and sort by name here, and you can see the statistical results for the NARR. I'm just taking a look at the correlation which is 0.59, the ERA40 0.79, the global 0.62, the WRF27 0.66, and I think we forgot to do the 9 km WRF. So we Convert to DATEM and that was 0.67. Okay. Let’s put this back to where it was.

And the last step is we've created five of these ensemble members. Now these results are slightly different than our table results because this is with the full 10,000 particles. So you see there's a difference between running 50000 versus 1000. These were with the 50,000 particle simulations.

But the next step and the last step here is that with these five members we've created the ensemble statistics and we can now plot that ensemble. We plotted the maps and we looked at the boxplot. But one of the outputs is the the mean concentration. So when you do the probabilities you also generate the mean concentration. So we could for instance, do statistics on the mean concentration, that is the ensemble mean. So what we would do is change this to cmean, for instance. Or we could do it in the utilities, Convert to DATEM menu, and force it to be cmean right here in this field. And we can create a, well let's not rename this yet. We can create the DATEM file and I'm going to compute statistics, and the correlation is now .81, and if I do a scatter plot, you'll see this is really quite decent and this is probably one of the the best and you can confirm that by looking at the statistics we just generated, the mean statistics. Let's go to stat, if I look at the clean, I get a correlation of .81, a rank of 3.3, that's really pretty decent, compared to let's say with the 27 km WRF simulation, which we know was also very good. And again the ensemble mean gave us much better results in terms of correlation, in terms of calculated ratio of, calculated to mean, one versus .7. Fractional bias, hardly any, and the final overall ranking was also much higher. So the end result here is that the ensemble mean of these five models gave us much better results than the performance of any of the individual simulations using just one meteorological data set and that is the main point of this section, is that an ensemble mean can be a much more robust simulation results for many different applications.

And this concludes the meteorological ensemble discussion.