Roque Vinicio Cespedes
MPO 524
02/05/2013

homework1

HW 1 assignment: An exploratory multivariate data analysis, oriented toward your term project. Due Tues, Feb. 5, when we will go over them in class.

1. Pick a broad interest, a realm of scientific inquiry, for your term project. It might as well be your research, but it could be something completely different. Your work about it should involve multivariate data (a data set with multiple "variables"). In traditional low-dimensional statistics, this term could include simply space-time data, like a time-longitude section of SST, with the different locations viewed as a "multivariate" set of time series. But that is too little (and we will do that example to death in the homeworks). For project purposes, I'd rather see you explore the relationships among quantities with different units, which forces you to think about normalization choices, and gives statistical results (such as the slopes of fitted lines) evocative physical units. Your exact approach and science questions can be vague at this stage, but this choosing part is necessary to move on to question 2, so I wanted to make clear that it is half the work! Browsing data catalogues (below) is a good way to start: it tickles your neurons and helps you sniff your psyche for curiosities. Just budget an hour or more, spend it imagining a project that will interest you, and jot down a verbal sketch of what that project might look like (for later narrowing and refinement, which Sarah and I can help with).

For the last part of my Master's Thesis, I analyzed how well a subset of the CMIP5 models simulates the Florida historical precipitation, so I am quite familiar with the format of the CMIP5 data. I did not use openDap, because the CMIP5 150-year "historical" global monthly data files can be quite large in file size. Here, I chose to look at the first ensemble (r1i1p1) of the CCSM 4 historical (1850-2005) experiment. Here, I explore the relationship in the CCSM 4 model between the three variables of precipitation, total cloud fraction, and sea-level pressure. For this, I look at the entire global domain of the model, and I start by getting the 1850-2005 temporally-averaged mean of each of the three variables of interest.

2.
a) Choose and load some multivariate data into arrays within Matlab (or your favorite language). Please consider loading data from URLs (remote datasets) using the tools below, rather than downloading files which may get lost making your work less repeatable for you in the future. Extra credit: capture some notes from your explorations of online datasets, including any "neato" items you think are worth a spotlighting in class.

b) Use some of Matlab's (or any other language's) multivariate exploration powers to seek and show structure and relationships: At a minimum, show histograms and joint histograms or scatter plots, perhaps with coloration or symbol size used to denote a third or fourth variable. The plotmatrix command is a quick one line call that does this: it shows all possible scatters and histograms of the columns in a data table (array) down the diagonal of a multi-panel graphic. Perhaps a 3D scatterplot can be telling, just capture some still images (like by screen capture) as

you rotate it this way and that, or post a code with an openDap and 3d scatter calls (just a couple of lines) and we can demo it in class. See the examples from previous years below.

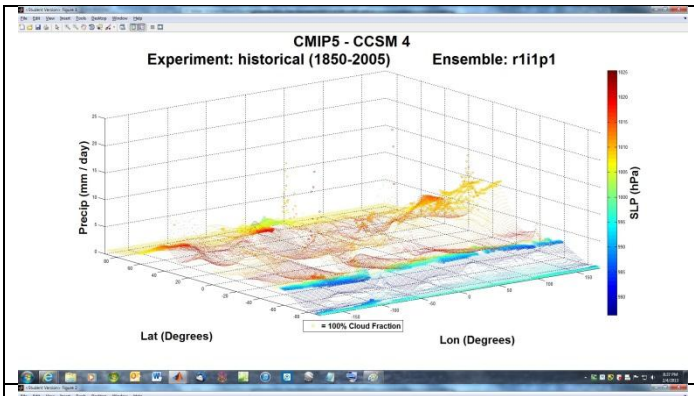

FIG. 1

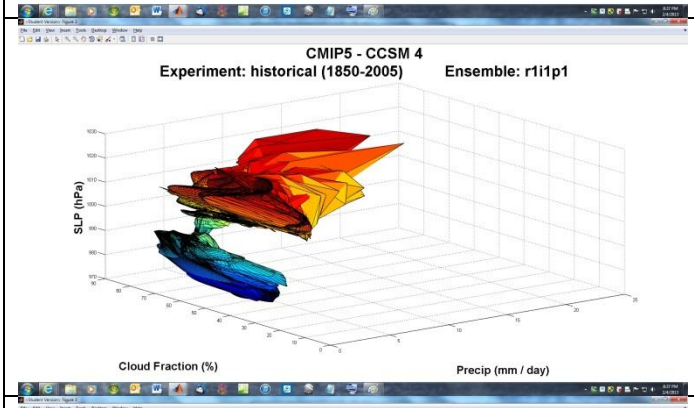

FIG. 2

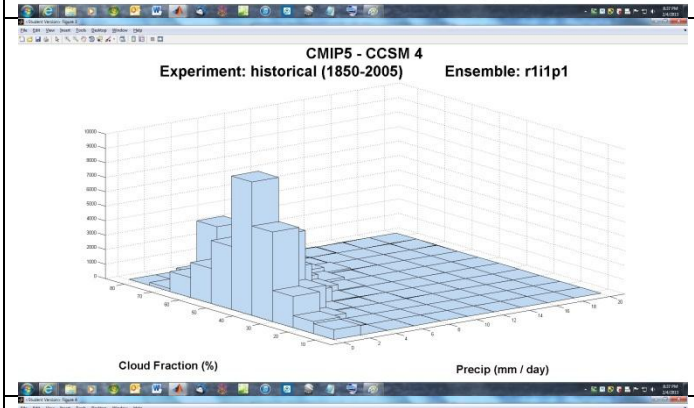

FIG. 3

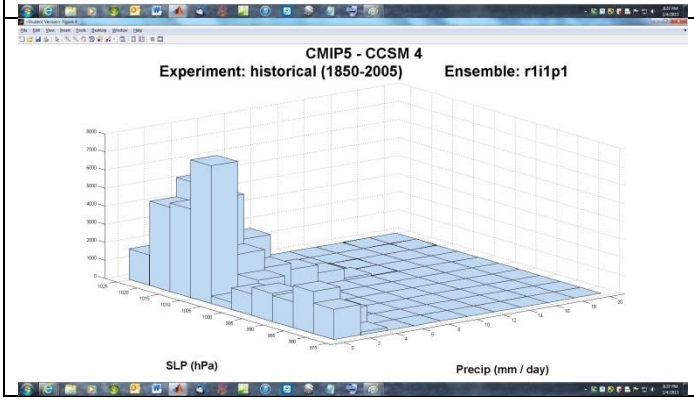

FIG. 4

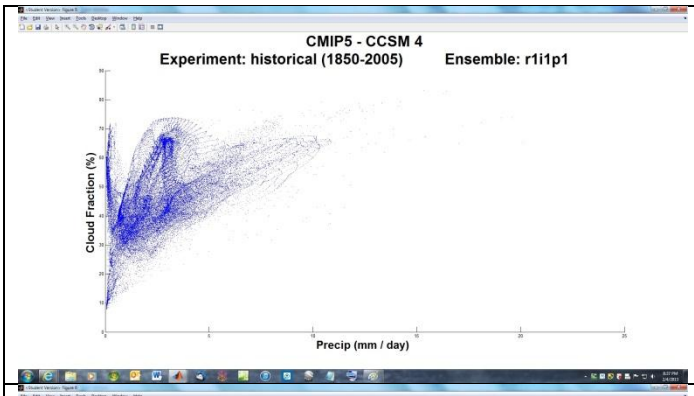

FIG. 5

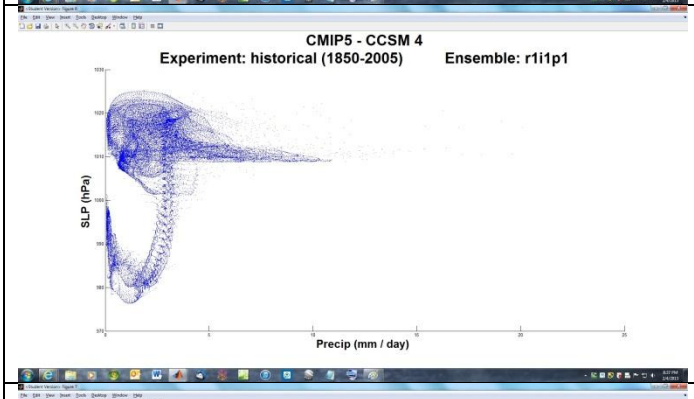

FIG. 6

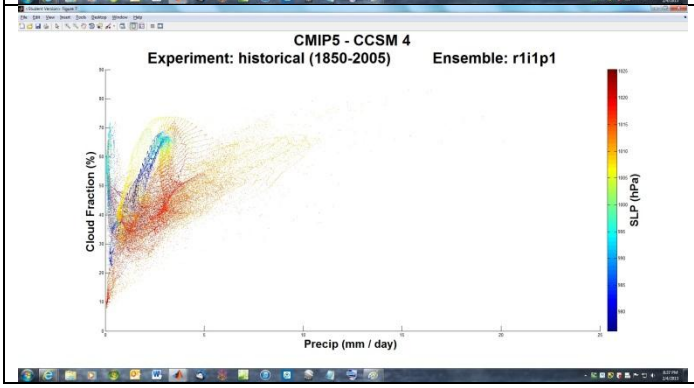

FIG. 7

3. Write a coherent paragraph about the scientific meaning of your activity in part 2. What relationships and non-relationships do you see? How does that fit with expectations? Does anything surprise you? Or simply intrigue you? Does the exercise raise questions (hypotheses) that can be used to guide your further work toward a more meaningful set of conclusions in the term project?

Future work: do the same relationships among the three variables hold true in other climate models and/or in observational data sets???

```
clear;

lat=nc_varget('pr_Amon_CCSM4_historical_r1i1p1_185001-200512.nc','lat');
lon=nc_varget('pr_Amon_CCSM4_historical_r1i1p1_185001-200512.nc','lon');

lon(0.5*size(lon,1)+1:size(lon,1),1)=lon(0.5*size(lon,1)+1:size(lon,1),1)-360;

[lon,lat]=meshgrid(lon,lat);

precip=nc_varget('pr_Amon_CCSM4_historical_r1i1p1_185001-200512.nc','pr');
precip=(24*60*60)*squeeze(nanmean(precip,1));

slp=nc_varget('psl_Amon_CCSM4_historical_r1i1p1_185001-200512.nc','psl');
slp=(1/100)*squeeze(nanmean(slp,1));

cc=nc_varget('clt_Amon_CCSM4_historical_r1i1p1_185001-200512.nc','clt');
cc=squeeze(nanmean(cc,1));

%-------

fig1=figure('Color',[1 1 1]);

scatter3(lon(:),lat(:),precip(:),0.031*cc(:),slp(:));

xlim([-180 180]);
ylim([-90 90]);

xlabel('Lon (Degrees) ','FontSize',24,'FontWeight','bold');
ylabel('Lat (Degrees) ','FontSize',24,'FontWeight','bold');
zlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');

cb1=colorbar('Location','EastOutside');
ylabel(cb1,'SLP (hPa)','FontSize',24,'FontWeight','bold');

legend1=legend('= 100% Cloud Fraction');
set(legend1,'Location','SouthOutside','FontSize',16,'FontWeight','bold');
```

```
title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------

fig2=figure('Color',[1 1 1]);

surf(precip,cc,slp);

xlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');
ylabel('Cloud Fraction (%) ','FontSize',24,'FontWeight','bold');
zlabel('SLP (hPa) ','FontSize',24,'FontWeight','bold');

title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------

fig3=figure('Color',[1 1 1]);

hist3(horzcat(precip(:),cc(:)));

xlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');
ylabel('Cloud Fraction (%) ','FontSize',24,'FontWeight','bold');

title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------

fig4=figure('Color',[1 1 1]);

hist3(horzcat(precip(:),slp(:)));

xlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');
ylabel('SLP (hPa) ','FontSize',24,'FontWeight','bold');
```

```
title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------

fig5=figure('Color',[1 1 1]);

scatter(precip(:),cc(:),0.01);

xlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');
ylabel('Cloud Fraction (%) ','FontSize',24,'FontWeight','bold');

title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------

fig6=figure('Color',[1 1 1]);

scatter(precip(:),slp(:),0.01);

xlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');
ylabel('SLP (hPa) ','FontSize',24,'FontWeight','bold');

title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------

fig7=figure('Color',[1 1 1]);

scatter(precip(:),cc(:),0.01,slp(:));

xlabel('Precip (mm / day) ','FontSize',24,'FontWeight','bold');
ylabel('Cloud Fraction (%) ','FontSize',24,'FontWeight','bold');

cb7=colorbar('Location','EastOutside');
ylabel(cb7,'SLP (hPa)','FontSize',24,'FontWeight','bold');
```

```
title({'CMIP5 - CCSM 4','Experiment: historical (1850-2005)          Ensemble:
r1i1p1'},'FontWeight','bold','FontSize',32);

%-------
```