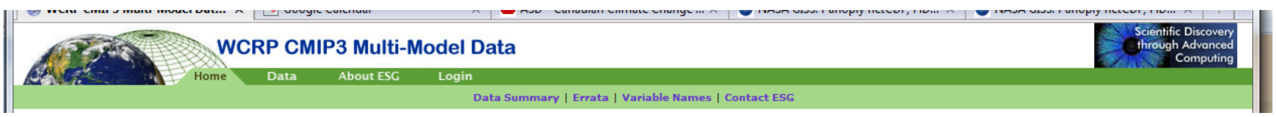
Spreadsheet exercise on uncertainty in climate model downscaling

Ed Maurer
Civil Engineering Department
Santa Clara University
Santa Clara, CA 95050

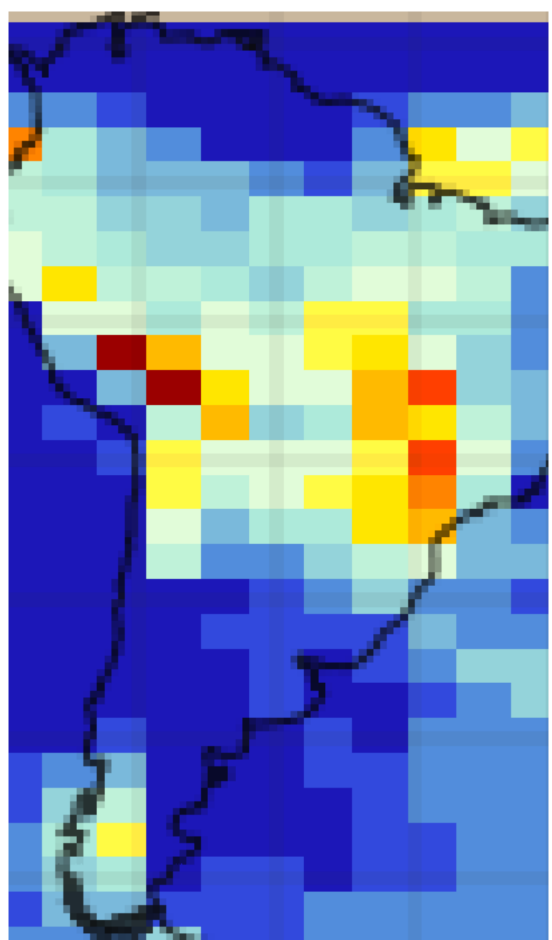
Workshop on Downscaling and Analysis of Uncertainty
June 8-9, 2010
Centro de Cambio Global
Pontificia Universidad Católica de Chile, Santiago, Chile.

Obtaining raw GCM output:

Data are available through the PCMDI CMIP3 data portal (http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php):



These are all in netCDF format. While they can be opened in some form by ArcGIS or using add-ins to Excel, other tools are designed for this, including: scripting environments: R, python/cdat, matlab; command line software such as NCO utilities, grads, ferret, and many others. There are convenient viewers as well, like Panoply, but files cannot be manipulated or exported into other formats with it (yet).



Below is a portion of a plot created with Panoply, showing precipitation for one month simulated by the HadCM3 GCM. It is obvious that the spatial scale is a problem in characterizing local changes.

To explore this raw data, using Panoply is easy. It can be downloaded from http://www.giss.nasa.gov/tools/panoply

The file displayed here, which I clipped down from the global extent to South America, can be downloaded here:

http://www.engr.scu.edu/~emaurer/shared/chile/ukmo_hadcm3_southamerica_pr_A1.nc

As can be seen, at this spatial resolution the Andes are barely resolved, meaning local expressions of climate driven by terrain will not be present in the GCM output.

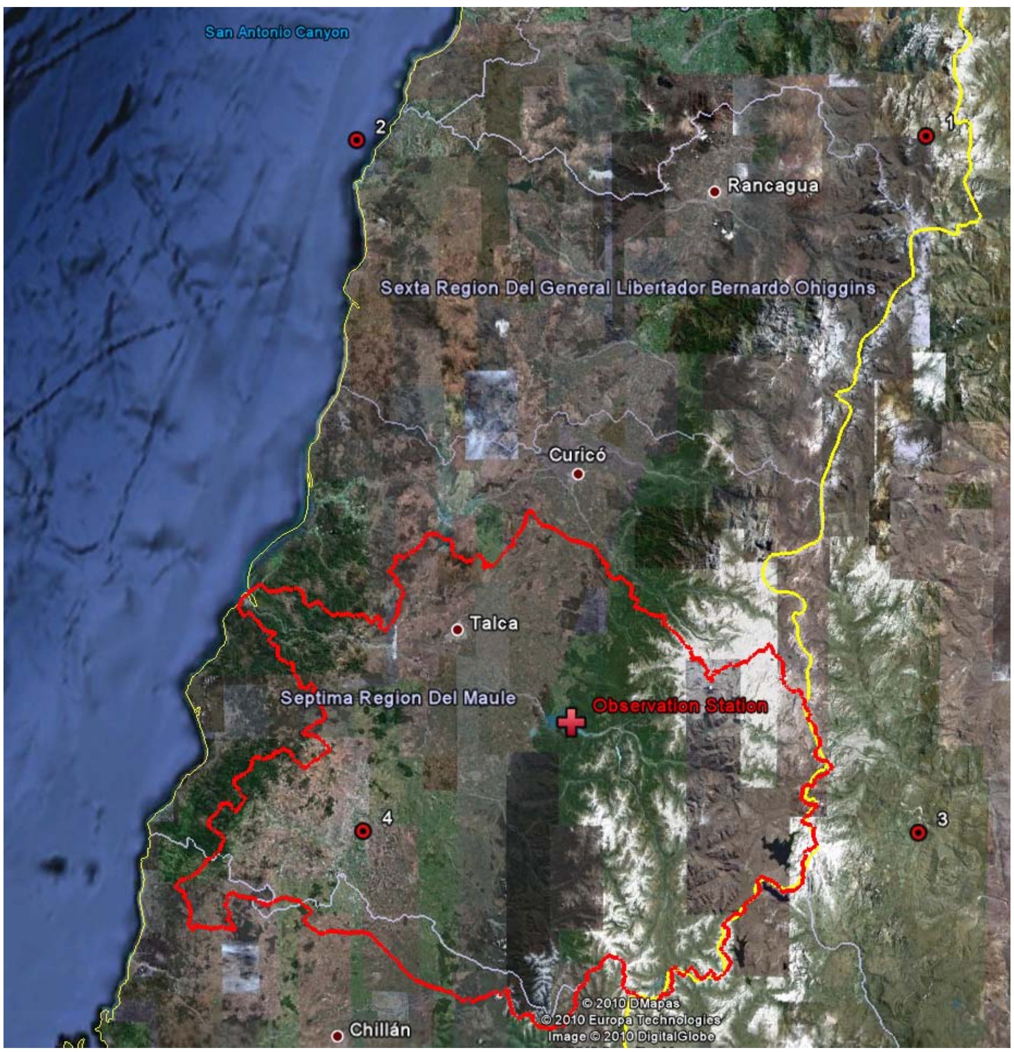
While there are many sophisticated statistical methods for translating these large-scale signals to local changes, we will do one of the simplest methods as an exercise. This will consist of two steps:

1. Interpolating the large-scale signal to a local scale

2. Correcting the bias to recover the historically observed climate patterns.

<u>Step 1: Interpolating the GCM data to a finer grid scale</u>

To illustrate this, I first interpolated the irregular HadCM3 grid to a regular 2-degree grid. Over the Maule basin (outlined in red), the centers of these 2-degree grid points look like this (labeled 1, 2, 3, 4):



If you would like to use these layers in Google Earth, they are available at:

http://www.engr.scu.edu/~emaurer/shared/chile/maule_basin.kml

and

http://www.engr.scu.edu/~emaurer/shared/chile/gcm_2deg_4points.kml

For this exercise, we will use gridded observed data prepared for a 0.5-degree spatial resolution dataset. The data are described in detail in the paper:

> Adam, J. C. and D. P. Lettenmaier, 2003. Adjustment of global gridded precipitation for systematic bias. *J. Geophys Res.* 108:1-14

and is summarized in the online paper:

> Maurer, E. P., Adam, J. C., and Wood, A. W.: Climate model based consensus on the hydrologic impacts of climate change to the Rio Lempa basin of Central America, *Hydrol. Earth Syst. Sci.*, 13, 183-194, doi:10.5194/hess-13-183-2009, 2009, available online at http://www.hydrol-earth-syst-sci.net/13/183/2009/hess-13-183-2009.pdf.

The global data, covering 1950-1999, are freely available at:
http://www.engr.scu.edu/~emaurer/global_data/

The observational station marked on the map is at latitude -35.75, longitude -71.25, and corresponds to the grid cell center on one of the 0.5-degree grid cells. The data for these four surrounding were interpolated to the observational point using an inverse-distance method, where the distances from each of the four stations 1-4 to the observational point was 226, 206, 116, and 73 km, respectively.

While it is not needed for this demonstration, all of the raw data are available at:
http://www.engr.scu.edu/~emaurer/shared/chile/maule_gcm_and_observed_data.zip

It should be noted that in some statistical downscaling methods the GCM data are corrected prior to interpolation, rather than as a first step as we are doing here.

Step 2: Correcting for biases in the GCM output

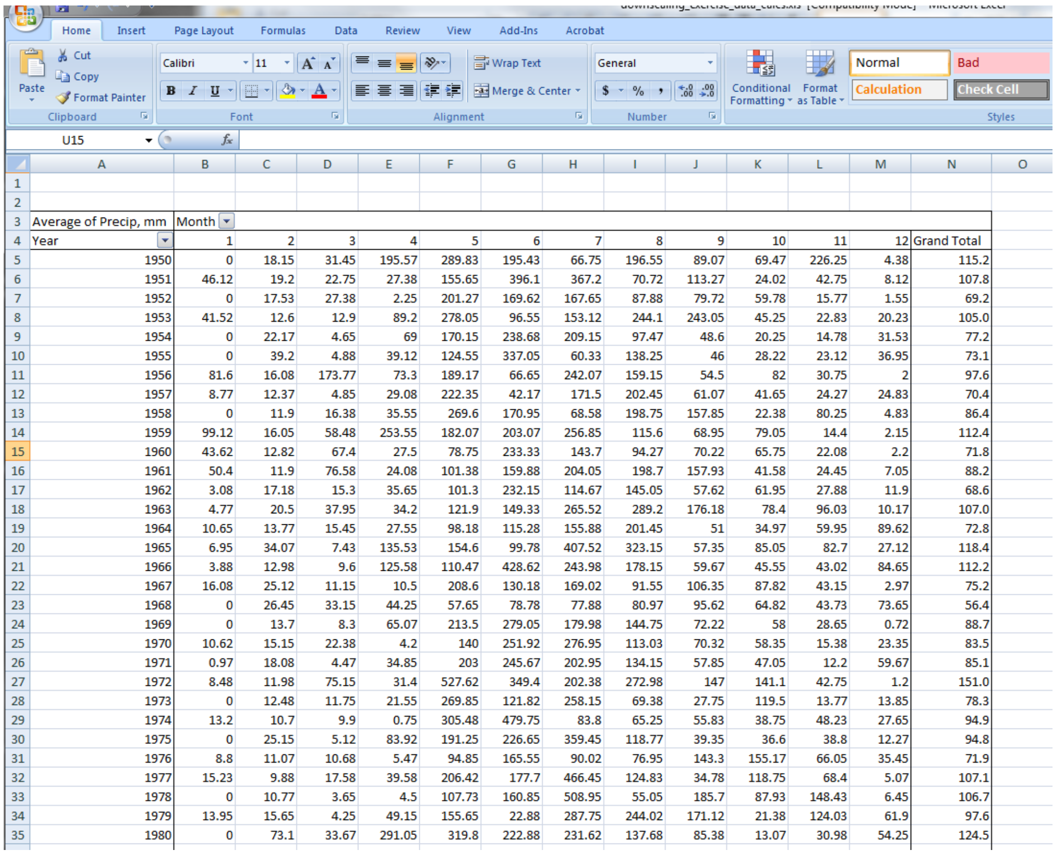
Now we have two sets of monthly data from 1950-2099 for the point at 35.75°S, 71.25°W: one we'll call "observed" and the other an interpolated GCM projection. The two sets of monthly data are in the spreadsheet that can be downloaded from:
http://www.engr.scu.edu/~emaurer/shared/chile/downscaling_exercise_data.xls

A. Simplest delta method

The simplest method of downscaling adjusts a historical, observed record by a "delta," where the delta is derived from the GCM run. In this case we will do this for both temperature and precipitation. This assumes that while the GCM has biases (due to lack of topography and imperfect process descriptions, for example), the sensitivity of the GCM is plausible. In other words, the *changes* are assumed to be simulated reasonably, even though absolute amounts are not.

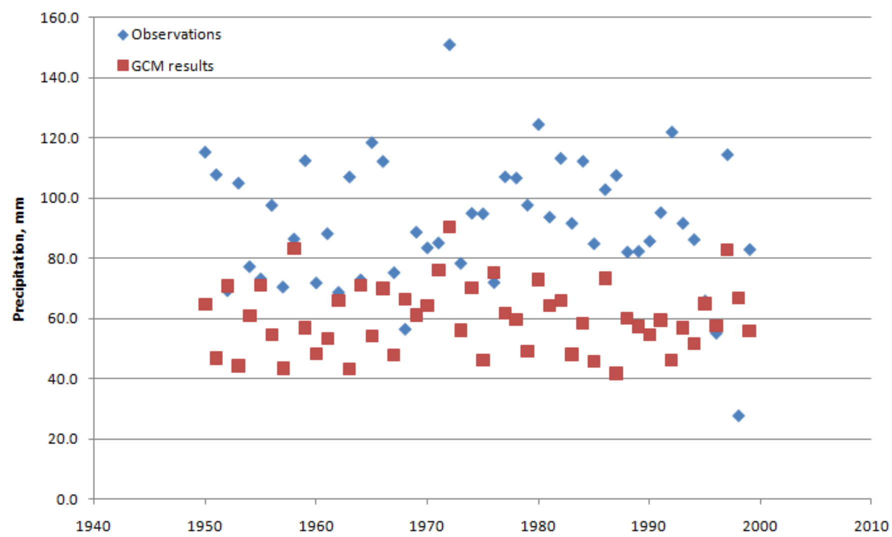
As a first step, create a pivot table of the observed precipitation data for 1950-1999:

This will produce output that looks something like this:

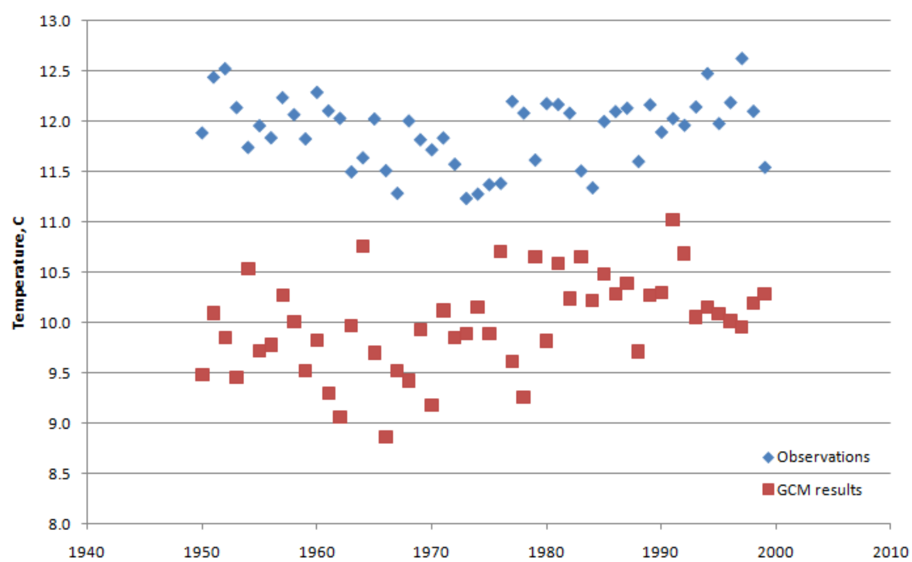


Average of Precip, mm	Month												
Year	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
1950	0	18.15	31.45	195.57	289.83	195.43	66.75	196.55	89.07	69.47	226.25	4.38	115.2
1951	46.12	19.2	22.75	27.38	155.65	396.1	367.2	70.72	113.27	24.02	42.75	8.12	107.8
1952	0	17.53	27.38	2.25	201.27	169.62	167.65	87.88	79.72	59.78	15.77	1.55	69.2
1953	41.52	12.6	12.9	89.2	278.05	96.55	153.12	244.1	243.05	45.25	22.83	20.23	105.0
1954	0	22.17	4.65	69	170.15	238.68	209.15	97.47	48.6	20.25	14.78	31.53	77.2
1955	0	39.2	4.88	39.12	124.55	337.05	60.33	138.25	46	28.22	23.12	36.95	73.1
1956	81.6	16.08	173.77	73.3	189.17	66.65	242.07	159.15	54.5	82	30.75	2	97.6
1957	8.77	12.37	4.85	29.08	222.35	42.17	171.5	202.45	61.07	41.65	24.27	24.83	70.4
1958	0	11.9	16.38	35.55	269.6	170.95	68.58	198.75	157.85	22.38	80.25	4.83	86.4
1959	99.12	16.05	58.48	253.55	182.07	203.07	256.85	115.6	68.95	79.05	14.4	2.15	112.4
1960	43.62	12.82	67.4	27.5	78.75	233.33	143.7	94.27	70.22	65.75	22.08	2.2	71.8
1961	50.4	11.9	76.58	24.08	101.38	159.88	204.05	198.7	157.93	41.58	24.45	7.05	88.2
1962	3.08	17.18	15.3	35.65	101.3	232.15	114.67	145.05	57.62	61.95	27.88	11.9	68.6
1963	4.77	20.5	37.95	34.2	121.9	149.33	265.52	289.2	176.18	78.4	96.03	10.17	107.0
1964	10.65	13.77	15.45	27.55	98.18	115.28	155.88	201.45	51	34.97	59.95	89.62	72.8
1965	6.95	34.07	7.43	135.53	154.6	99.78	407.52	323.15	57.35	85.05	82.7	27.12	118.4
1966	3.88	12.98	9.6	125.58	110.47	428.62	243.98	178.15	59.67	45.55	43.02	84.65	112.2
1967	16.08	25.12	11.15	10.5	208.6	130.18	169.02	91.55	106.35	87.82	43.15	2.97	75.2
1968	0	26.45	33.15	44.25	57.65	78.78	77.88	80.97	95.62	64.82	43.73	73.65	56.4
1969	0	13.7	8.3	65.07	213.5	279.05	179.98	144.75	72.22	58	28.65	0.72	88.7
1970	10.62	15.15	22.38	4.2	140	251.92	276.95	113.03	70.32	58.35	15.38	23.35	83.5
1971	0.97	18.08	4.47	34.85	203	245.67	202.95	134.15	57.85	47.05	12.2	59.67	85.1
1972	8.48	11.98	75.15	31.4	527.62	349.4	202.38	272.98	147	141.1	42.75	1.2	151.0
1973	0	12.48	11.75	21.55	269.85	121.82	258.15	69.38	27.75	119.5	13.77	13.85	78.3
1974	13.2	10.7	9.9	0.75	305.48	479.75	83.8	65.25	55.83	38.75	48.23	27.65	94.9
1975	0	25.15	5.12	83.92	191.25	226.65	359.45	118.77	39.35	36.6	38.8	12.27	94.8
1976	8.8	11.07	10.68	5.47	94.85	165.55	90.02	76.95	143.3	155.17	66.05	35.45	71.9
1977	15.23	9.88	17.58	39.58	206.42	177.7	466.45	124.83	34.78	118.75	68.4	5.07	107.1
1978	0	10.77	3.65	4.5	107.73	160.85	508.95	55.05	185.7	87.93	148.43	6.45	106.7
1979	13.95	15.65	4.25	49.15	155.65	22.88	287.75	244.02	171.12	21.38	124.03	61.9	97.6
1980	0	73.1	33.67	291.05	319.8	222.88	231.62	137.68	85.38	13.07	30.98	54.25	124.5

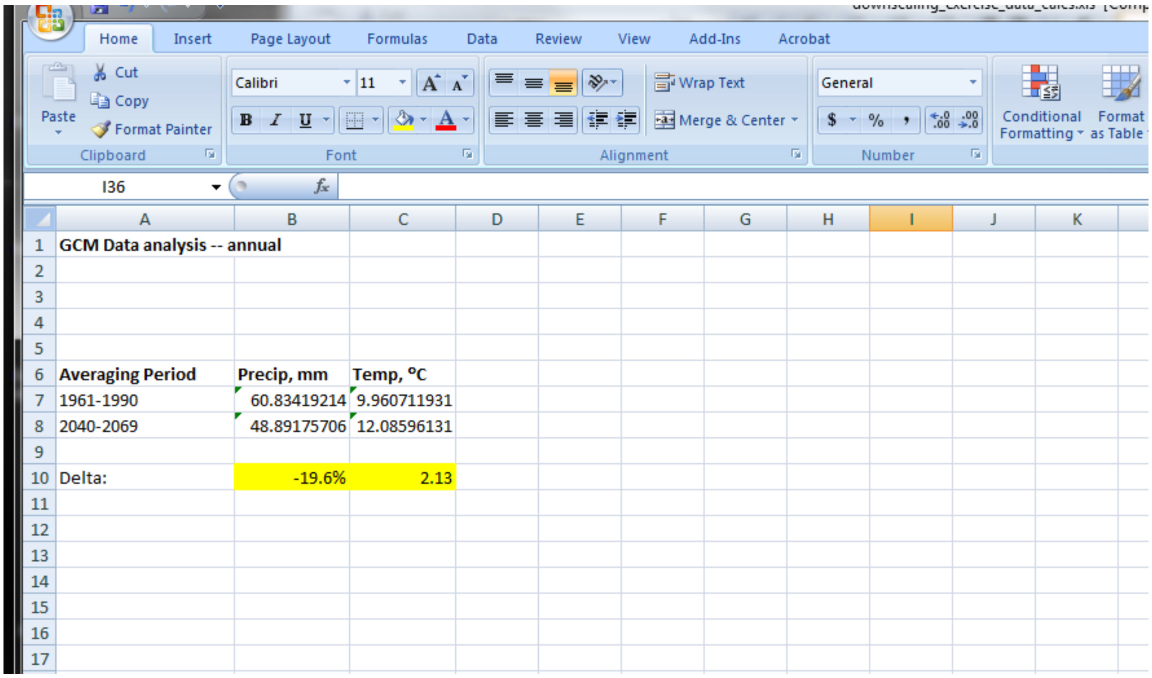
Do the same for the GCM for the 1950-1999 period. This will allow you to do a quick assessment of precipitation bias:



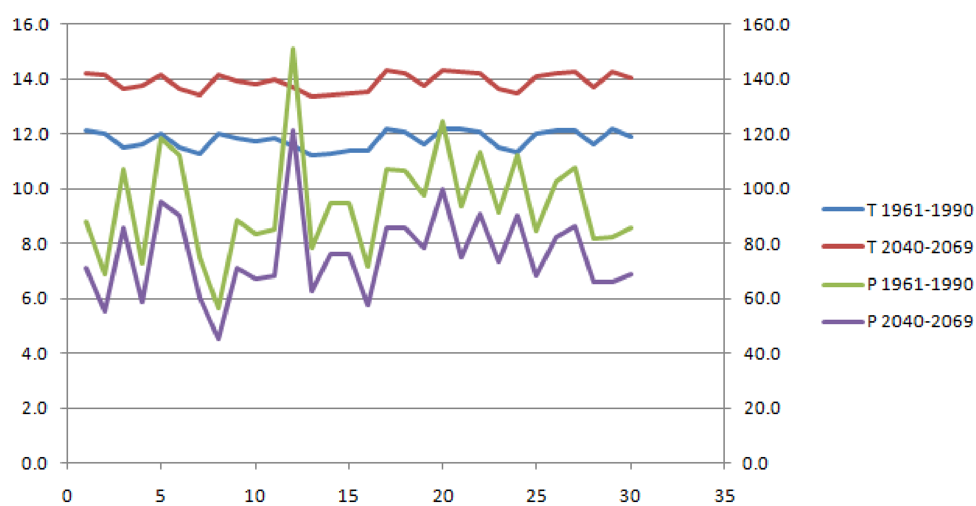
It is obvious that there is a bias, both in mean and inter-annual variability in the GCM data. The same can be done for temperature to look at annual average temperature bias.



To implement the delta method, open the downscaling_exercise_data.xls spreadsheet. Since to define a climatic state 30-year periods are typically used, we'll take the average 1961-1990 for the GCM, and compare it to the average for a future time slice, say 2040-2069, for the GCM.



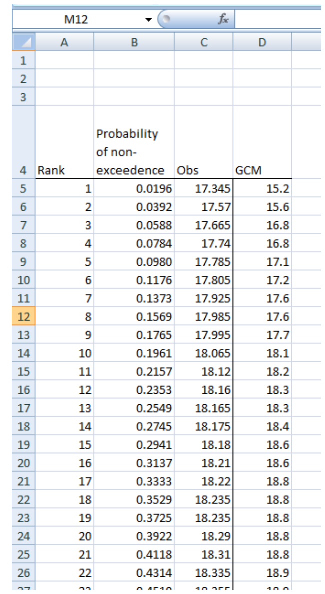
Typically, precipitation delta is expressed as a percent, and temperature as degrees. These are then applied to the 1961-1990 series of observations to obtain a shifted 30-year period that represents climate for 2040-2069. So, since the average precipitation will decrease by 19.6% and the temperature will increase on average of 2.3°C between 1961-1990 and 2040-2069, we can simply adjust the time series to obtain the plausible sequence of actual precipitation and temperature vales at this site for 2040-2069:



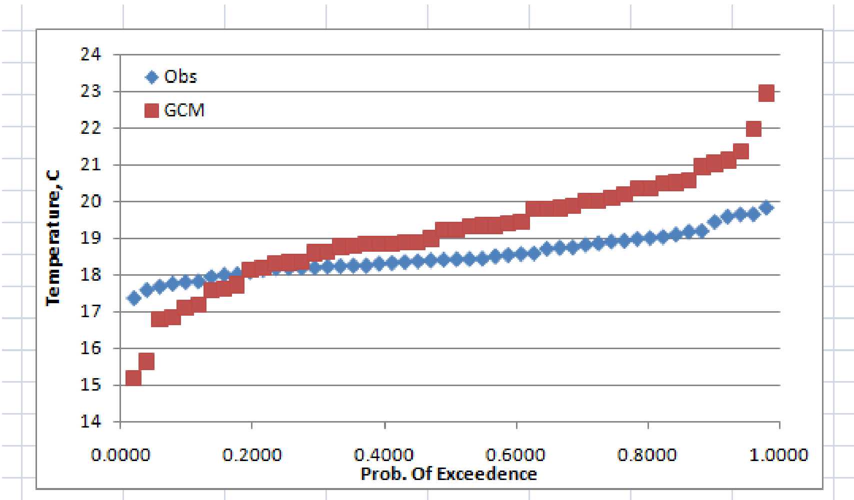
The major shortcoming in this method is evident: variability does not change. Also, implicitly each month changes identically, which is not realistic. Adapting this method to work with 12 deltas each for precipitation and temperature, one for each month, is very straightforward, and we will not do that exercise now.

Finally, we will look at the whole probability distribution of GCM projections. For this example we will use January temperatures. Place all January temperatures for 1950-1999, for both observations and GCM simulations into a new worksheet. These have already been separated out by the pivot table.

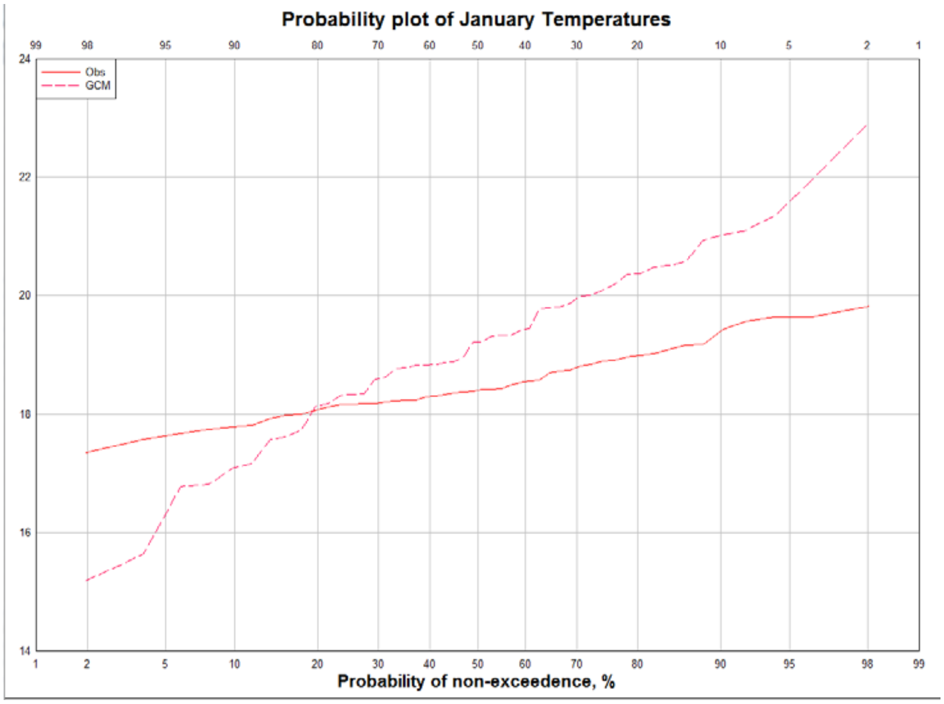
Notice that sorting is done separately for each column, so there is no correspondence of years to the data. Now calculate a probability of non-exceedence for each value, as the rank/(1+N) where N=50.



Rank	Probability of non-exceedence	Obs	GCM
1	0.0196	17.345	15.2
2	0.0392	17.57	15.6
3	0.0588	17.665	16.8
4	0.0784	17.74	16.8
5	0.0980	17.785	17.1
6	0.1176	17.805	17.2
7	0.1373	17.925	17.6
8	0.1569	17.985	17.6
9	0.1765	17.995	17.7
10	0.1961	18.065	18.1
11	0.2157	18.12	18.2
12	0.2353	18.16	18.3
13	0.2549	18.165	18.3
14	0.2745	18.175	18.4
15	0.2941	18.18	18.6
16	0.3137	18.21	18.6
17	0.3333	18.22	18.8
18	0.3529	18.235	18.8
19	0.3725	18.235	18.8
20	0.3922	18.29	18.8
21	0.4118	18.31	18.8
22	0.4314	18.335	18.9



This illustrates inter-annual variability biases in January temperatures between the GCM data and the observations. By putting these on a probability plot (created with the free excel add-in Dplot), we see that they are close to straight lines, meaning the temperatures in both the GCM and observations are normally distributed. In this case, we will use this to simplify the exercise.



The summary statistics for the GCM and observed data for 1950-1999 are:

	Mean	Standard Deviation
GCM	19.14	1.51
Obs	18.50	0.56

Now, for each January temperature in the GCM simulation, calculate its standard normal deviate, z:

$z = \frac{x - \bar{x}}{s_x}$ with $\bar{x}$ being the mean and s_x the standard deviation of the raw GCM January T data for 1950-1999,.

Its quantile, F(z), can be found using the excel function NORMSDIST(z). F(z) is the probability of non-exceedence for that value.

The bias-corrected T value for this month is calculated using:

$x_{bc} = \bar{x}_{obs} + z(s_{x-obs})$ where the "obs" subscripts indicate that the mean and standard deviation are for the January temperature observations for 1950-1999. In this way, a new sequence of bias-corrected January temperatures are obtained. These will have the exact same mean, interannual variability, skew and every other statistical property as the observations for 1950-1999. They will evolve into the future in a way that assumes the biases seen during 1950-1999 will remain the same into the future.

	Year	GCM January T	z	F(z)	Bias-corrected T
4	1950	20.94	1.19	0.88	19.17
5	1951	18.17	-0.64	0.26	18.14
6	1952	18.34	-0.54	0.30	18.20
7	1953	18.32	-0.55	0.29	18.19
8	1954	20.37	0.81	0.79	18.96
9	1955	17.72	-0.94	0.17	17.97
10	1956	16.82	-1.54	0.06	17.63
11	1957	18.84	-0.20	0.42	18.39
12	1958	20.35	0.80	0.79	18.96
13	1959	19.31	0.11	0.54	18.57
14	1960	18.77	-0.25	0.40	18.36
15	1961	17.61	-1.02	0.15	17.93
16	1962	17.09	-1.37	0.09	17.73
17	1963	19.22	0.05	0.52	18.53
18	1964	22.94	2.52	0.99	19.92
19	1965	20.57	0.95	0.83	19.04
20	1966	16.77	-1.57	0.06	17.62
21	1967	18.86	-0.19	0.43	18.40
22	1968	18.13	-0.68	0.25	18.12
23	1969	18.63	-0.34	0.37	18.31
24	1970	17.17	-1.31	0.09	17.76