Preparing data for climate change analysis using MATLAB

Alexandre Martinez (alex.m@uci.edu)
University of California, Irvine.

Introduction

Because of the recent increase in the frequency of extreme weather events widely attributed to climate change, there is a concern in the scientific community about the intensity and duration of droughts. The intensity of drought, also known as drought severity, is defined as the cumulative (or average) deviation of SPI values below a threshold level, while the drought duration is defined as the time period over which drought occurs (J. W. Kwak, 2012). One issue is to have a consistent and long record of data. We are now looking into these issues.

Data and method

We are using CRU (Climate Research Unit) precipitation dataset, which is a monthly gridded (0.5 degree) product. For technical reasons (processing time) we have clipped the data to North America only. We are now exploring the datasets to see if it is suitable for climate change analysis. Climate change refers to long period of time and drought is an extreme event, which means that it doesn't happen every year. Therefore, if we want to study a 100-years return period, we need at least 100 years of data. The CRU datasets starts in 1901 so should be suitable.

Instruction

First, we load the data, and visualize the cumulative precipitation to see how it looks.

```
load([pwd '\dataFolder\prcp.mat']);
load coast
sLat = -30:0.5:60;
sLon = -180:0.5:-60.5;
clab = 'Precipitation [mm/mo]';
colorRange = 'jet';
figure('Name', 'Monthly precipitation');
axesm('MapProjection', 'miller','GLineWidth', 1.0, 'MeridianLabel',
'on','MLabelParallel'...
    , 'south','ParallelLabel', 'on','MLineLocation', 60, 'PLineLocation',
30,...
'MapLonLimit',[-180 -60],'MapLatLimit',[0 59.5],'FFaceColor',[.75 .75 .75]);

pcolorm(sLat, sLon, squeeze(nansum(data,3))');
plotm(lat, long, 'k');
colormap(colorRange);
framem on;gridm on;tightmap;title('Monthly precipitation');
% Colorbar
c = colorbar;
c.Label.String = clab;
```

The map looks like:

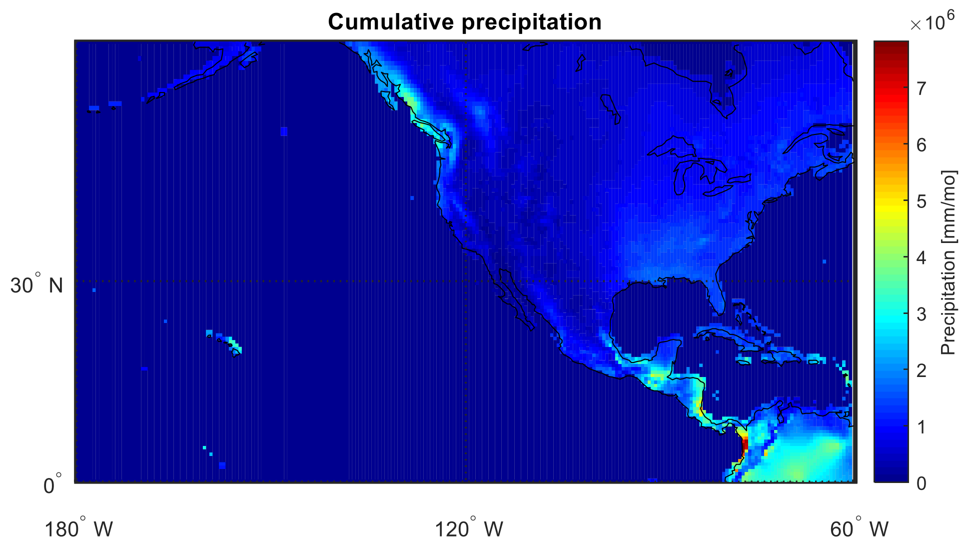

Figure 1 - Global cumulative precipitation

The first thing to notice is that CRU used 0 for ocean values (no data).

We can also plot the time series of the total precipitation:

```
figure()
plot(nansum(nansum(data,2),3));
title('Total precipitation');
xlabel('Year');
```

And we get the following time series:

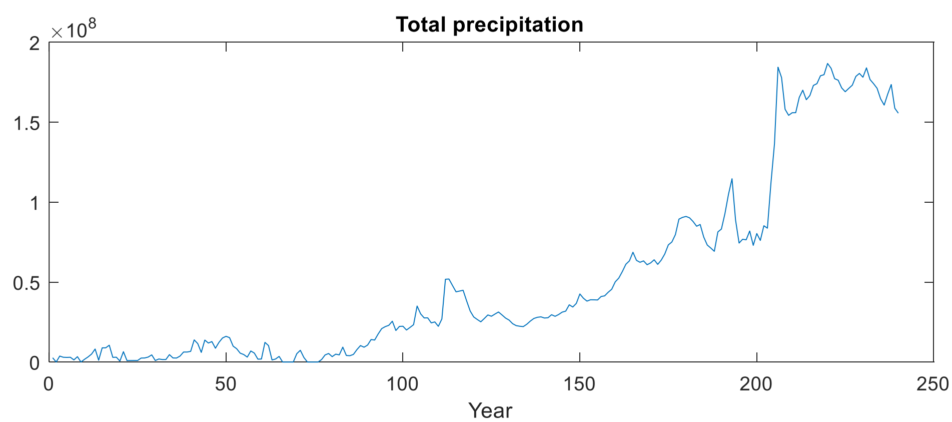

Figure 2 - Time series of the global precipitation

What we immediately notice is that the global precipitation is increasing with time, by a factor 10. Is it reasonable? Do we get more rain than before? Obviously no, so it means that we have a lot of area with missing early values. Looking at CRU datasets, it is ground-based precipitation observation, and not every pixel had a gage in 1901. Therefore, we need to remove those pixels from the later analysis.

Let's look at some random pixels that we can suspect to have missing early data (let say in Grand Canyon).

```
for i=1:20
       figure()
       plot(squeeze(data(240-i,180-i,:)));
       title('Total precipitation');
       xlabel('Year');
end
```

This code is looking around for some time series. Here are three different type of pixels:

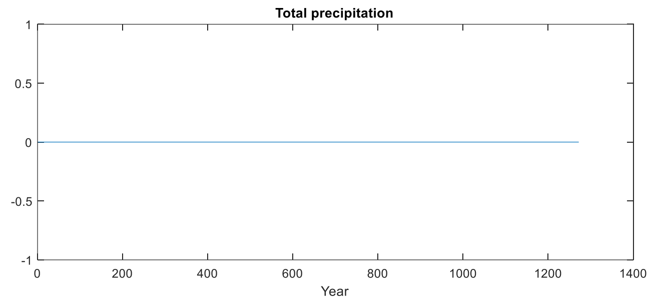

Figure 3 - Example of ocean pixel

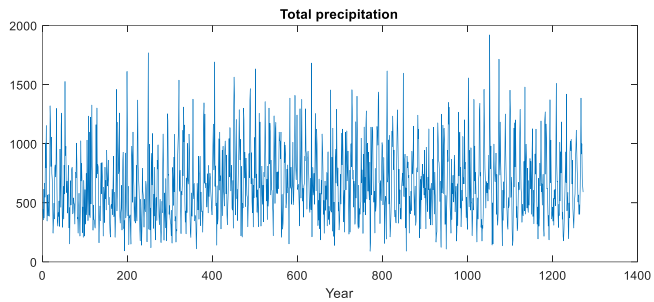

Figure 4 - Example of ideal pixel

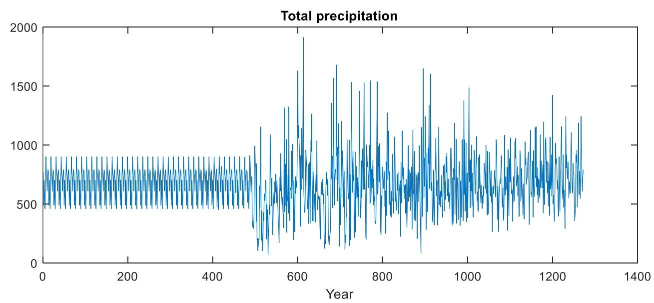

Figure 5 - Example of pixel with missing early data

The last figure shows the problem we are looking for: some pixels have no data, and CRU added climatology data, i.e. long-term average for each month. We need to remove these pixels.

```
%Create a mask of useful pixels (with 100 years or 30 years of records)
[mask, maskHalf]     = f_IO_maskNaN(data,'dataFolder');
```

Activity: **write a function which is taking out the pixel using climatology data.**

This function creates a mask of pixel of interests: at least 100 years of data or at least 30 years of data.

```
function [ mask, maskHalf] = f_IO_maskNaN(data,dataFolder)
    t                = clock;
    disp('Task started: Creating a mask of useful pixels');
    [iMax, jMax, tMax]  = size(data);
    %mask = zeros(iMax,jMax);
    pos = 1;
    posHalf = 1;
    for i=1:iMax
        for j=1:jMax
            if (data(i,j,50)==0)
                pos = pos;
                %disp('empty data');
            else
                % else we have some data, is it full dataset or is there
                % any climatologic data at the begining?
                yTemp = squeeze(data(i,j,:));
                for iMo=1:12:(size(yTemp,1)-12)
                    posCli = (iMo+11)/12;
                   yMaxV(posCli)= max(yTemp(iMo:iMo+12));
                end
                iMonth = 1;
                yMax = max(yTemp(1:12));
                while ((yMax-
max(yTemp(iMonth+12:iMonth+23)))/(yMax)<0.05)&&(iMonth<1249)
                    iMonth=iMonth+12;
                end
                iYear = (iMonth-1)/12;
                if iYear <1 %we are good
                    mask(pos,1) = i;    % can go to full dataset
                    mask(pos,2) = j;
                    maskHalf(posHalf,1) = i;    % and to the half dataset
                    maskHalf(posHalf,2) = j;
                    maskHalf(posHalf,3) = iYear;
                    pos=pos+1;
                    posHalf=posHalf+1;
                    %disp('full data');
                elseif iYear < 90
                    maskHalf(posHalf,1) = i;
                    maskHalf(posHalf,2) = j;
                    maskHalf(posHalf,3) = iYear;
                    posHalf=posHalf+1;
                    %disp('half data');
                end
            end
        end
    end
end
```

We can now display some basic statistics to understand this data issue:

```
%Plot some statistics about the missing pixels
figure();histogram(maskHalf(:,3),[0 1:3:60]);
title('Missing year per pixel');xlabel('# of missing year');ylabel('# of
pixel');
maskHalfMap = nan(size(data,1),size(data,2));
for ii=1:size(maskHalf,1)%create a matrix of strting year
   maskHalfMap(maskHalf(ii,1),maskHalf(ii,2)) = maskHalf(ii,3)+1901;
end
plotGlobalMap(maskHalfMap,'startingYear','First year of non-climatology
data');
```

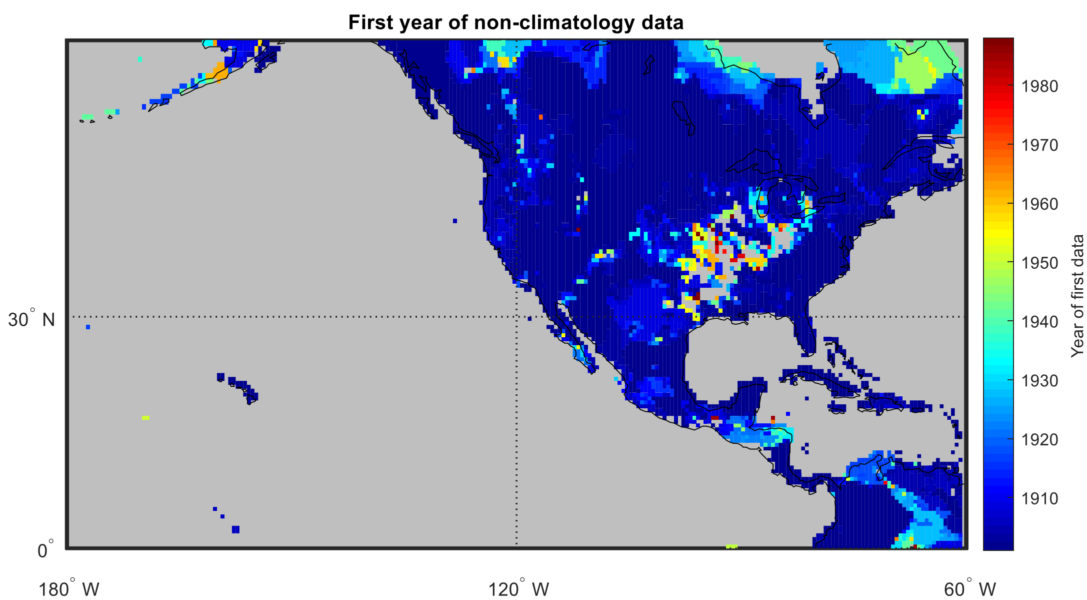

Figure 6 - First year of non-climatology data

What we now see is that a lot of data are missing in higher latitude and Central USA (starting data is about 1940, which leaves only 60 years of data). The main conclusion is that we should exclude these pixels because it doesn't make sense for extreme events to work with a short record of data.