Changepoint Detection in Climate Series Paper Replication Documentation

Kristen L. Gore
Columbia University Department of Statistics

March 8, 2011

This document contains details on how to *replicate my* replication of "Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model" by Robert Lund and Jaxk Reeves.

1 Downloading the Data

There were 3 primary datasets used in this study–1) Chula Vista temperature data 2) Reference temperature data for Chula Vista (created by averaging temperature data from 5 nearby sites) and 3) Mauna Loa CO_2 data. Below is the breakdown of the data files needed to replicate this paper.

- Chula Vista:
  - Chula Vista – chulavista.csv
  - Avalon (ref) – avalon.txt
  - Cuyamaca (ref) – cuyamaca.txt
  - Indio (ref) – indio.txt
  - Redlands (ref) – redlands.txt
  - Irvine (ref) – DOES NOT EXIST YET* (i.e. I can't find it)
- Mauna Loa:
  - Mauna Loa – ml-5.csv

* Because I haven't been able to find the Irvine, CA data, the Chula Vista reference dataset is currently the average of 4 sites instead of 5. I'm still working on finding this last dataset. It only seemed to affect the results in one figure, though.

2 Software and Packages

I used R. You don't need anything beyond the basic stats packages that R loads automatically.

3 General Structure of the Code

The overall flow of the code is as follows:

1. General functions (used for both Chula Vista and Mauna Loa analyses)
2. Chula Vista-specific functions + Chula Vista Analysis
3. Mauna Loa-specific functions + Chula Vista Analysis

Description of each figure:

Chula Vista

- Figure 1: Temperature vs Year
- Figure 2: Fc vs Year
- Figure 3: Residuals of Piecewise Regression Model with c=1975
- Figure 4: Residual Fc vs Year
- Figure 5: Chula Minus Reference Fc vs Year

Mauna Loa:

- Figure 6: Fc vs Year
- Figure 7: Residuals of Piecewise Regression with c=1989
- Figure 8: Residual Fc vs Year

4 Other Comments

I have a lot of comments in the code itself, so hopefully everything is understandable. Make sure you change your working directory from mine to yours. (This line is found at the top of the code.) The figures are identified in the code. If you want to run code for just one of the primary sites, just run the "General Functions" section and the code for that site. You don't have to run the code for the other site. Ok I think that's it. Have fun replicating my replication! Feel free to let me know if you have any question/issues. My email address is klgore@stat.columbia.edu. –KLG