Statistical Analysis of Climate Changes in East Africa

Matthew Williams and Dong-Yun Kim, Department of Statistics, Virginia Tech

Introduction

The objective of this research is to apply accepted statistical methodology to identify changes in times series climate data. These data have spatial components, so both *when* and *where* the changes occur are of interest. The most common methods for testing change is a simple test of linear trend or non-parametric smoothers. However, these methods do not provide statistical tests or direct estimates of the change point. Instead, methods for detecting changes in mean (Jaruskova, 1998) variance (Chen and Gupta, 1997), or both (Horvath, 1993) can be used in conjunction with standard mapping software to examine precipitation and temperature data for East Africa. There are two datasets discussed here; station data from Tanzania and regional data from the Climatic Research Unit (CRU) dataset. Both contain precipitation and rainfall time series. The discussion begins with a brief overview of the change point methods used, continues with the application to the East Africa climate data, and finishes with some conclusions and suggestions for future work.

Methods

The change point methods used are described in general terms below. The asymptotic and simulation results needed can be found in the corresponding papers. Tests used here are the two-sided versions and the rejection limits are based on Type I error rates of 5%.

Horvath (1993) method:
- -Abrupt change in the mean or variance
- -Independent identically distributed (iid) normal random variables.
- -Likelihood ratio approach

The Chen and Gupta (1997) method:
- -Abrupt changes in the variance only
- -Assumes the data are iid normal.
- -Penalized likelihood ratio - SIC (Schwartz Information Criteria)

The Jaruskova (1998) method:
- -Linear trends in the mean
- -Assumes that the data are iid (not necessarily normal)
- -Likelihood-ratio-like approach.

In order to make inference about the estimated parameters, semi-parametric bootstrapping can be used. The concept is to fit the model and repeatedly re-sample the residuals and fit the model again. The distribution of parameters that results will provide some bounds for the parameters of interest. For the Tanzania data set, we use the basic pivotal method (Carpenter & Bithell, 2000).

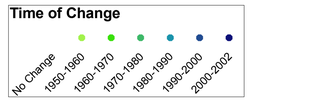
Data

The CRU data (http://www.cru.uea.ac.uk/cru/data/).:
- -"Rectangle" from 26.8ºE to 42.8ºE and 5.8ºN to 12.8ºS
- -Encompasses Kenya, Uganda, Rwanda, Burundi, Tanzania
- -Monthly average precipitation and max and min temperatures

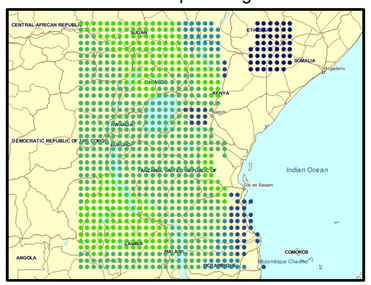
CRU Precipitation data (1950-2002):
- -Changes in mean and the variance are of interest
- -Monthly averages: not iid normal
- -Yearly averages: much closer to iid normal
- -Use Horvath and Chen and Gupta on yearly data
  - --Mean change: Horvath but not Chen and Gupta
  - --Variance change: Horvath and Chen and Gupta

CRU Max and Min temperature (1950-2002)
- -Trends (i.e. global warming) are of interest
- -Monthly averages for daily extreme temperatures
- -Monthly averages: not iid
- -Yearly averages: much closer to iid
- -Use Jaruskova method on yearly data

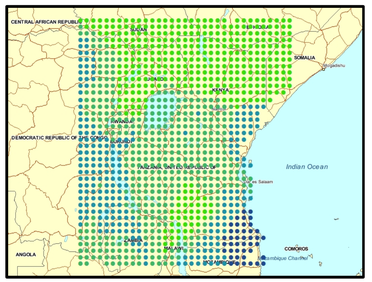
Tanzania station (1961- 2006) data (Tanzania Meteorological Agency):
- -Couple dozen stations
- -Monthly average precipitation and temperature
- -Use yearly data:
  - --Horvath and Chen and Gupta for precipitation
  - --Jaruskova for temperature.



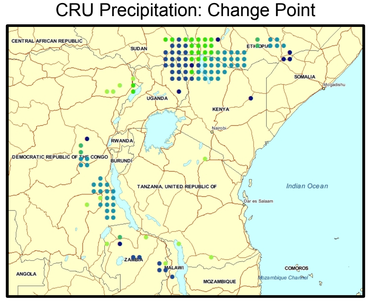
CRU Max Temp: Change Point



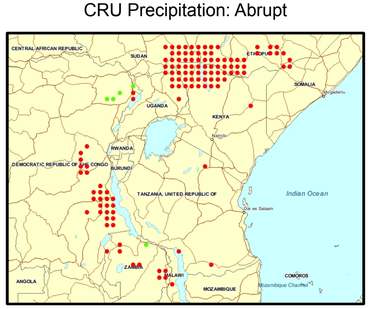
CRU Min Temp: Change Point



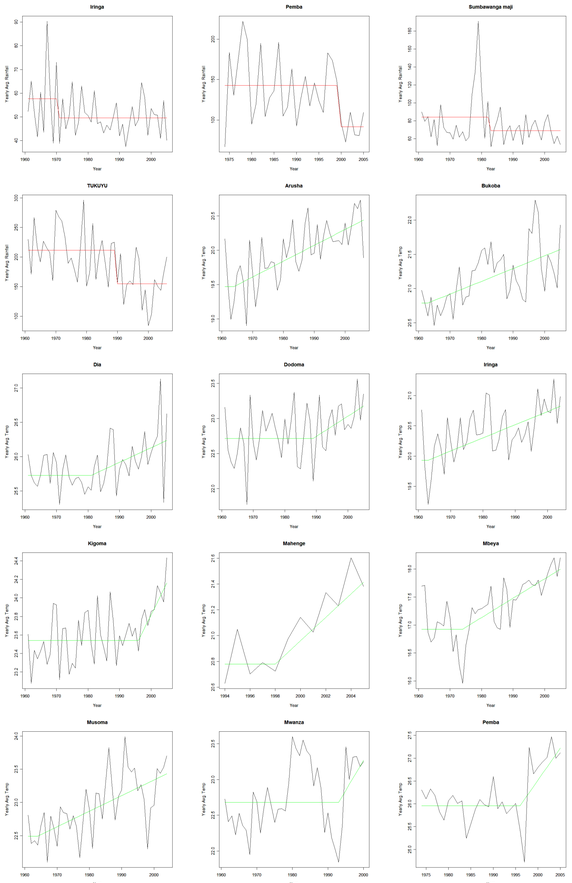
CRU Precipitation: Change Point



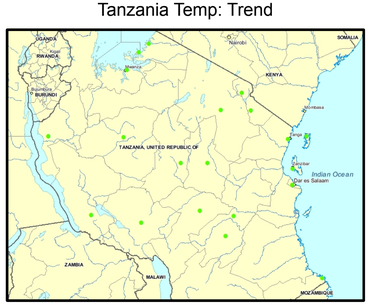
CRU Precipitation: Abrupt



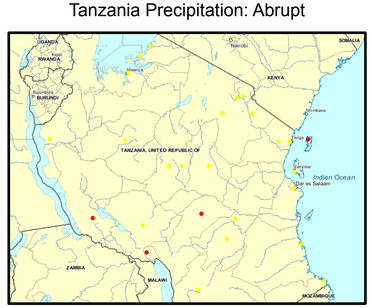
Results



Tanzania Temp: Trend



Tanzania Precipitation: Abrupt



Tanzania: Sites of Significant Change and Bootstrap 95% CI

Precipitation					Changept 95% CI		Mean1 95% CI		Mean2 95% CI	
Site	Lat	Long	Start	# of yrs	L	U	L	U	L	U
Iringa	-7.80	35.72	1961	45	0	18	45.98	65.99	37.11	56.96
Pemba	-5.00	39.75	1974	32	22	32	112.12	164.19	42.55	129.12
Sumbawanga maji	-7.97	31.62	1961	45	1	42	45.92	100.31	26.66	86.67
Tukuyu	-9.25	33.65	1961	46	19	45	194.86	229.18	126.94	179.59

Temperature					Changept 95% CI		Mean 95% CI		Slope 95% CI	
Site	Lat	Long	Start	# of yrs	L	U	L	U	L	U
Arusha	-4.00	36.50	1961	46	0	8	19.21	20.02	0	0.03
Bukoba	-1.83	33.42	1961	45	0	6	20.51	21.33	-0.01	0.02
Dia	-6.83	39.20	1961	45	4	41	25.58	26.08	-0.04	0.03
Dodoma	-6.00	36.00	1961	45	15	45	22.59	23.05	-0.16	0.05
Iringa	-7.80	35.72	1961	44	0	6	19.63	20.66	0	0.03
Kigoma	-5.00	30.00	1961	45	30	45	23.45	23.62	-0.03	0.12
Mahenge	-8.75	36.68	1994	12	2	10	20.6	20.91	0.06	0.13
Mbeya	-8.25	33.50	1961	45	4	27	16.75	17.4	0.01	0.05
Morogoro	-8.00	37.00	1975	31	16	30	24.34	24.65	0.01	0.13
Moshi	-3.37	37.30	1961	45	0	16	22.99	23.85	0	0.03
Mtwara	-10.33	40.33	1961	45	25	45	25.34	25.58	-0.05	0.09
Musoma	-1.50	33.80	1961	44	0	8	22.23	23.13	-0.01	0.03
Mwanza	-2.50	32.97	1961	40	27	40	22.5	23.1	-0.3	0.16
Pemba	-5.00	39.75	1974	32	17	32	25.76	26.17	-0.05	0.23
Same	-4.03	37.63	1961	45	34	45	23.2	23.56	-0.08	0.32
Shinyanga	-6.00	35.00	1987	19	14	26	24.05	24.43	0.15	0.54
Sumbawanga	-7.97	31.62	1989	11	0	6	17.65	18.1	0.03	0.13
Tabora	-5.03	32.83	1960	43	2	22	22.65	23.26	0.02	0.04
Tanga	-5.10	39.03	1961	45	22	45	26.16	26.44	-0.09	0.08
Zanzibar	-6.22	39.22	1961	45	0	12	24.8	26.03	0.02	0.06

Discussion

The research goal was to identify time and location of changes in climate data using statistical change point methods

Horvath/Chen and Gupta and Jaruskova methods:
- -Identify changes for relatively short durations (50 years)
- -Confidence intervals for the change point: wide for short durations

Precipitation data:
- -Little change, but some decreasing
- -Changes scattered across time (with large variance)

Temperature data:
- -Widespread increases
- -Time of change varies (with large variance)
- -Strong evidence increases in region from 1950 – 2000

Overall these methods provide some insight into climatic behavior. Precise inference using bootstrap confidence intervals is dubious: Intervals are wide and often contain illogical values (<0 or > 50). However, these methods are a first step to understanding change points and quantifying their uncertainty.

Conclusions

Statistical change point methods allow for the detection and estimation of abrupt changes or trends and the location in time of the change point. The application of these methods, as demonstrated in this research, can be readily used on climate data. From the data, some local decreases in precipitation and a large scale regional increase in temperature seem to have occurred in the past 50 years. However, more clarity can be obtained with the integration of other geographical information such as topology and land cover. In addition, testing for multiple change points (instead of just one) may add more information. Furthermore, spatial information has been used only for visualization and interpretation, but spatial correlation and its impact on testing and inference has not been explored.

Literature Cited

Carpenter, J. & Bithell, J. (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statist. Med.* 19, 1141-1164.

Chen, J. & Gupta, A. K. (1997) Testing and Locating Variance Changepoints With Application to Stock Prices. *J. Am. Stat. Assoc.* 92, 739-747.

Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.* 21, 671-680.

Jarušková, D. (1998). Testing appearance of linear trend. *J. Stat. Plan. Inf.* 70, 263-276.

Acknowledgments

This research was performed in cooperation with the NSF-CNH project 0709671 entitled "Dynamic Interactions among People, Livestock, and Savanna Ecosystems under Climate Change" (Eaclipse) and is supported by the Institute for Critical Technology and Applied Science (ICTAS) Doctoral Scholar program at Virginia Tech. Special thanks to the Tanzania Meteorological Agency for the station data.