A New Strategy for Modeling Time Series Data with Structural Change

Matthew Williams¹, Nicolle Goble², & Dong-Yun Kim¹

¹ Department of Statistics, Virginia Tech
² Department of Statistics, Temple University

4/14/2010
Motivation

- Most climate data are (noisy) time series.
- How can we find/model structural changes in the data?

Figure: Time Series for NDVI and Monthly Temperature for two sites in East Africa
Standard Approach

- Box and Tiao (1975) describe how to incorporate different forms of structural change into ARIMA models.
- Tsay (1988) extends this to variance changes.

Drawbacks

- Methods use diagnostics, not statistical tests to determine a change.
- The diagnostics look like tests, but they treat the location of the change (intervention) as known.
- Diagnostics have no measure of “how big is big enough”. They can only identify “sore thumbs”.
Our Approach

Our Two Stage Approach

- Test for an unknown change (intervention)
  - This is difficult in general, since most tests require independent data
  - Climate data can often be usefully summarized by yearly data (which is much more independent)
  - So we test for change on yearly summary data
- Then fit the intervention model with change and structure estimated from the tests on the yearly data
  - We could use ARIMA or State Space machinery
  - Instead we’ll use non-linear mixed models
Models: Autoregressive Lag 1 with seasonal mean

- Box and Tiao AR(1) model with intervention

\[ Z_t = \frac{1}{(1 - \phi B)} a_t + A + B \sin(2\pi t/w + C) + D_t 1_{\{t \geq \nu\}} \]

- Non-linear mixed model formulation

\begin{align*}
Z_t &= A + B \sin(2\pi t/w + C) + D_t 1_{\{t \geq \nu\}} + \epsilon_t \\
\epsilon_t &= \phi \epsilon_{t-1} + a_t \\
a_t &\sim N(0, \sigma^2) \tag{1}
\end{align*}

where \( \nu \) is time of change, \( A \) is the mean before the change, \( A + D_t \) is the mean after the change. \( B, w, \) and \( C \) give the amplitude, period and phase of the sine function, which describes seasonal effects, and \( \epsilon \) is autoregressive lag 1 (AR(1)) noise.
Testing for Change

There are several tests for change-points in the literature. We will consider two:

- Horváth (1993): abrupt mean and/or variance test (step function).

Both tests assume independent data (Horváth also assumes normality) and yield estimates of the time of change as part of the testing procedure. Our yearly average data should more or less satisfy these conditions.

Based on the results of our tests, we either fit a reduced model \( D_t = 0 \), an abrupt model \( D_t = D \) or a linear trend/hockey stick \( D_t = \delta(t - \nu) \).
Example: Simulated Data with Large Change

We simulate 600 observations of period 24 (so 25 “years”) using model 1:
- $\nu = 300$, $A = 500$, $B = 200$, $w = 24$, $C = 0$, $D_t = 500$, $\sigma = 50$, and $\phi = 0.8$.
- Horvath test statistic is greater than the 5% critical value ($\hat{\nu} = 313$).
- $\hat{A} = 533.6(17.16)$, $\hat{B} = 196.0(10.69)$, $\hat{w} = 24.00(0.028)$, $\hat{C} = 0.100(0.108)$, $\hat{D} = 473.1(24.79)$, $\hat{\sigma} = 95.85$, and $\hat{\phi} = 0.822$.

Figure: Simulated Time Series and Yearly Means
Example: Simulated Data with Small Change

We simulate 600 observations of period 24 (so 25 “years”) using model 1:

- $\nu = 300$, $A = 500$, $B = 200$, $w = 24$, $C = 0$, $D_t = 100$, $\sigma = 50$, and $\phi = 0.8$.
- Horvath test statistic is greater than the 5% critical value ($\hat{\nu} = 265$).
- $\hat{A} = 511.3(15.7)$, $\hat{B} = 208.5(9.68)$, $\hat{w} = 23.98(0.024)$, $\hat{C} = -0.097(.092)$, $\hat{D} = 84.08(20.36)$, $\hat{\sigma} = 86.1$, and $\hat{\phi} = 0.811$.

![Simulated Time Series and Yearly Means](image-url)
Example: Normalized Difference Vegetation Index (NDVI)

We now consider NDVI from a site in East Africa from 1982 - 2006.
- We first use the data up to 2003, then forecast the next three years.
- From our plots, it seems that the NDVI has decreased.

Figure: NDVI Semi-monthly and Yearly Time Series
Example: Normalized Difference Vegetation Index (NDVI)

Using the first 22 years

- Horváth test statistic is greater than the 5% critical value ($\hat{\nu} = 193$).
- $\hat{A} = 468.7(16.6), \hat{B} = 149.4(11.2), \hat{w} = 24.07(0.045), \hat{C} = 3.60(0.14), \hat{D} = -46.76(20.55), \hat{\sigma} = 92.9$, and $\hat{\phi} = 0.73$.

![Figure: NDVI Semi-monthly and Yearly Time Series](image-url)
Forecasting with the intervention \((D_t)\) term improves forecasts by removing bias. However, since the variability of these time series is large, the improvement to short-term forecasts is relatively small.

**Figure:** Forecasts of next three years vs observed values. Intervention model (left), no-intervention model (right)
Example: Temperature

We now consider monthly temperatures from Tanga, Tanzania from 1961-2005.

- We analyze the data up to 2002 and then forecast the next 3 years.
- From plots of the data, it seems that the Temperature has increased.

Figure: Monthly Average and Yearly Average Temperatures for Tanga, Tanzania
Example: Temperature

Using the first 41 years

- Jarušková test statistic is larger than the 5% cut-off ($\nu = 97$)
- $\hat{A} = 26.13(0.079)$, $\hat{B} = 2.05(0.048)$, $\hat{w} = 11.998(0.0037)$, $\hat{C} = 0.622(0.047)$, $\hat{D} = 0.482(0.146)$, $\hat{\sigma} = 0.535$, and $\hat{\phi} = 0.578$.

Figure: Monthly Average and Yearly Average Temperatures for Tanga, Tanzania
Forecasting with the intervention ($D_t$) term improves forecasts by removing bias.

**Figure:** Forecasts of next three years vs observed values. Intervention model (left), no-intervention model (right)
What we've done (extendible to ARMA models)
- Screening process of statistical tests on reasonable units of agglomeration (yearly averages).
- This process seems to detect small changes in the mean (good sensitivity)
- The models fitted with interventions (after testing) have improved prediction.

Why not test or estimate the unknown change-point in the non-linear mixed model framework?
- Asymptotics are complicated (Not simple Normal or Chi-square)
- Asymptotic distribution of the change is not Normal either

What about variance changes?
- Incorporating variance change into mixed models is possible.
- However testing for such a change is not as straight-forward.
Acknowledgements

- The first author is supported as a doctoral scholar through the Institute for Critical Technology and Applied Science (ICTAS) at Virginia Tech.
- This work has been sponsored in part by the National Science Foundation Biocomplexity of Coupled Human and Natural Systems Program, Award No. BCS-0709671.


