Tree Cover Variability Increases from 2005 to 2100 in Sub-Saharan Africa

Eric Kalosa-Kenyon, Cody Carroll, and Amy Kim
Department of Statistics, University of California, Davis

December 2017

1. CHANGEPOINT ANALYSIS

Instead of dealing with the non-stationarity of our series by differencing and using the ARIMA framework, another approach is to notice that there are specific points in time where the time series changes behavior in one fell swoop. An example of this is around the year 2025, where the series jumps up to a mean level of around 11.75 and stays there. Figure 1 illustrates this "jump" with a red dashed line.

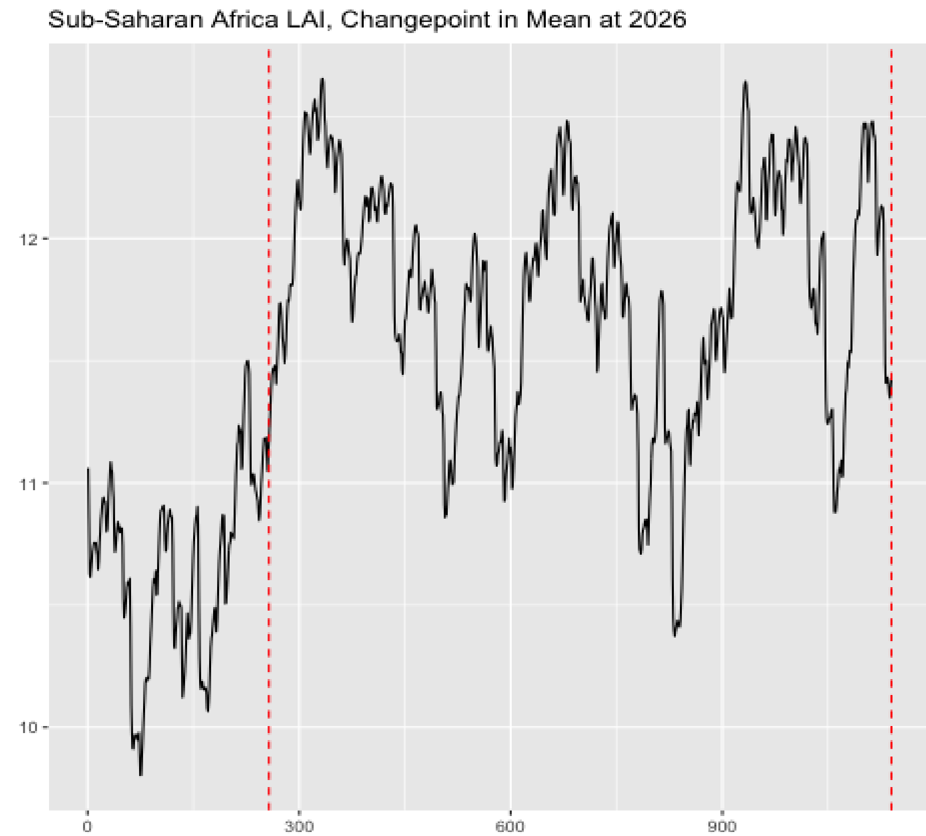

Figure 1: LAI TS

This point in time is called a changepoint. We give a short introduction on changepoint detection methods.

1.1. Changepoint Detection Conceptually, for data $Z_1, ..., Z_n$, a changepoint τ is a point in time, such that $Z_1, ..., Z_\tau$ differ from $Z_\tau + 1, ..., Z_n$. For the sake of illustration, we take a simple example. Assume the parametric form

$$Z_t|\theta_t \sim N(\theta_t, 1). \qquad \text{where } \theta_t = 1 \text{ if } t \leq \tau \text{ and } \theta_t = 0 \text{ if } t > \tau$$

An example of such a process is depicted in Figure 2.

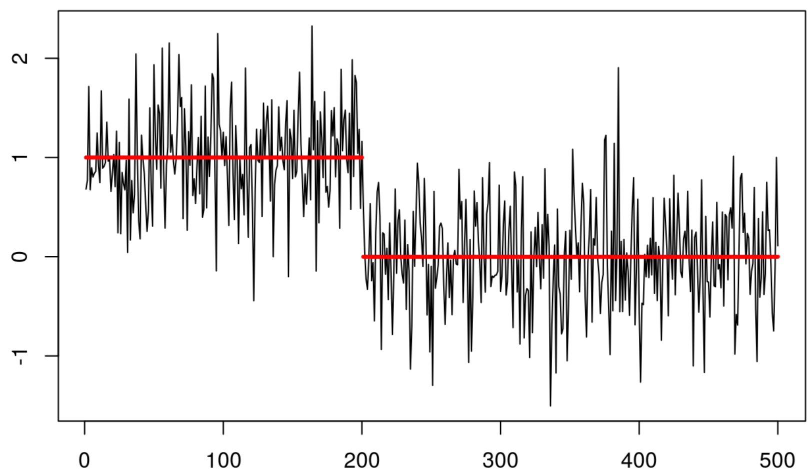

Figure 2: Example of a Changepoint

It's intuitively clear from the picture that the change occurs at $\tau = 200$, but how do we detect where a changepoint occurs, mathematically? One approach is to use a Likelihood Ratio Test. We state the hypotheses as:

$$H_0 : \theta_t = \theta \; \forall \; t \quad \text{vs.} \quad H_1 : \theta_t = \theta_1, \; t \leq \tau; \; \theta_t = \theta_2, \; t > \tau$$

To construct the likelihood ratio statistic, we need to maximize the likelihood under the null and alternative. We let $p(\cdot)$ be the probability density function associated with the distribution of the data and $\hat{\theta}$ is the maximum likelihood estimate of the parameters. Under H_0, the maximum log-likelihood is $\log p(y_{1:n}|\hat{\theta})$. Under H_1 with a changepoint τ, the maximum log-likelihood is $ML(\tau) = \log p(y_{1:\tau}|\hat{\theta}_1) + \log p(y_{\tau+1:n}|\hat{\theta}_2)$. Therefore the likelihood ratio statistic is:

$$\Lambda(\tau) = \frac{\max_\tau ML(\tau)}{\log p(y_{1:n}|\hat{\theta})}$$

A more commonly used version of this statistic is the log-LR statistic:

$$2 \log \Lambda(\tau) = 2 \left[\max_\tau ML(\tau) - \log p(y_{1:n}|\hat{\theta}) \right]$$

We compare this statistic with a cut-off value, λ. If $LR > \lambda$, then we reject H_0 and estimate the changepoint as

$$\hat{\tau} = \arg \max_\tau LR(\tau)$$

There are many different ways to select λ. Selecting an optimal value of λ remains an open research topic in changepoint analysis. We use the package 'changepoint' to investigate if our LAI time series has a changepoint, using a specific λ value that takes into consideration the inherent correlation between observations of a time series. We find an estimated changepoint of $\hat{\tau} = 237$ months after Jan. 2005, i.e. 2026 is when LAI experiences a changepoint in mean. Looking at the series after 2026 though, we suspect there may be more time points. So we segment the series and iterate the changepoint estimation. We need to consider multiple testing since we are doing several tests now..

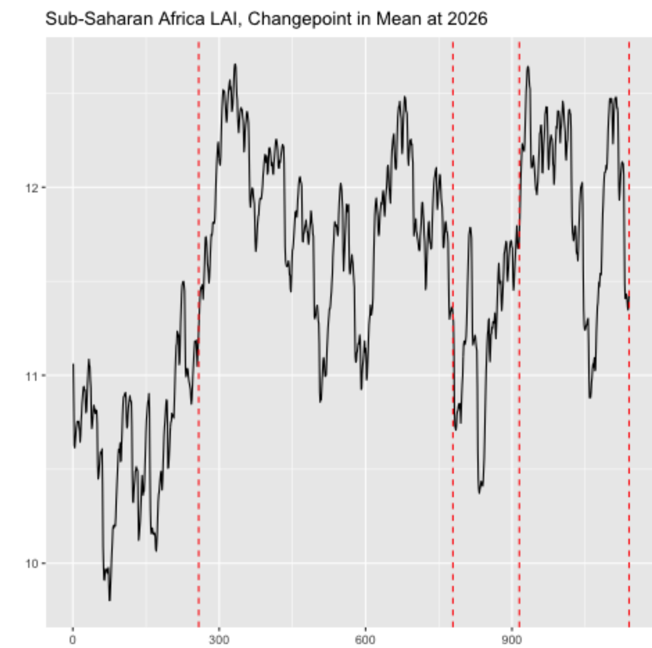

Figure 3: LAI TS