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

Cody Carroll, Eric Kalosa-Kenyon, Amy Kim

December 10, 2017

Abstract

1 Principal Component Analysis

We have done the time series analysis for the one location, and we can expand our analysis into global scale. This becomes the analysis of space-time with a large dataset (12818 locations and 1140 time points). In order to extract the underlying trends, we can consider Principal Component Analysis for examining both the spatial and temporal variation here.

Principal component: Temporal pattern (true values x loadings) vs. loadings of each principal components: Spatial Pattern - eigenvectors

Principal Component Analysis

$$\mathbf{Z}(s,t) = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{1}$$

where $\mathbf{U}$ is a $T \times k$ orthogonal matrix with columns $\mathbf{u}_j$, $\mathbf{V}$ is a $S \times k$ orthogonal matrix with columns $\mathbf{v}_j$ and $\mathbf{\Lambda}$ is a $k \times k$ diagonal matrix with diagonal entries λ_j. This $\mathbf{Z}$:

$$\mathbf{Z}(s,t) = \begin{pmatrix} z(s_1,t_1) & z(s_2,t_1) & z(s_3,t_1) & \dots & z(s_{12818},t_1) \\ z(s_1,t_2) & z(s_2,t_2) & z(s_3,t_2) & \dots & z(s_{12818},t_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z(s_1,t_{1140}) & z(s_2,t_{1140}) & z(s_3,t_{1140}) & \dots & z(s_{12818},t_{1140}) \end{pmatrix} \tag{2}$$

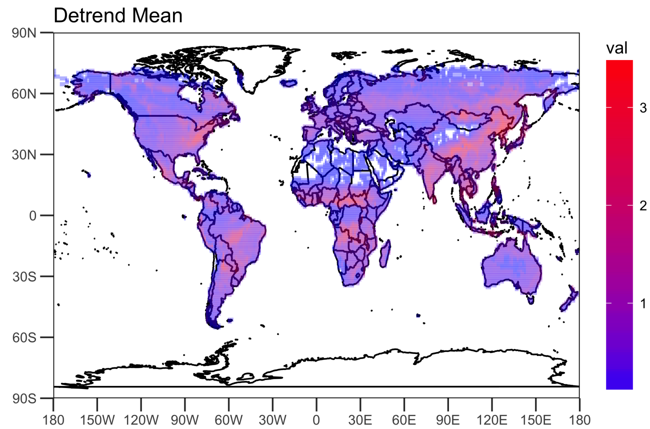


Figure 1: Overall Spatial Mean

We decide to report PCA on a detrend dataset [1] since its first few Principal Components(PCs) have more cumulative explained variances. We mainly examine the first Principal Component(PC) since it explains 43% of variances and others do less then 10% (Table 2) [2].

1.1 Spatial Pattern

Spatial pattern explains how strong the PCs depend on some locations, and it is represented by the loadings of each principal components. The mean spatial structure, Figure 1, indicates locations known forest area higher mean values(red) and desert areas have lower values (blue). We can interpret PC1 implies main variance across the all locations and over the 95 years, Figure **??**. It shows forest areas have negative effects(blue) and infertile lands have positive effects(red) which contracts to mean structure.

1.2 Temporal Pattern

Temporal pattern explains the dominant temporal variation of time series in the all locations, and it is represented by principal components (PCs, a number of time series) of PCA. We can confirm the detrend first three PC are stationary in Figure 3, and the first PC has widest range of oscillation(black) and ranges of oscillation get smaller (Blue is the second PC, and red is the third.)

[1] We used cubic splines for each time series to extract variations

[2] 419 PCs achieve to explain 90% of variations.

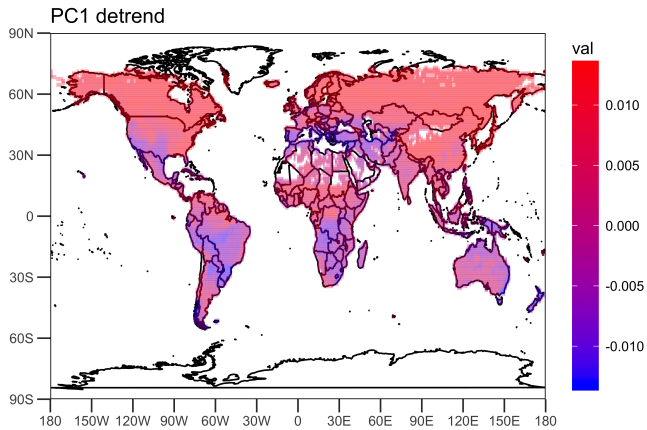
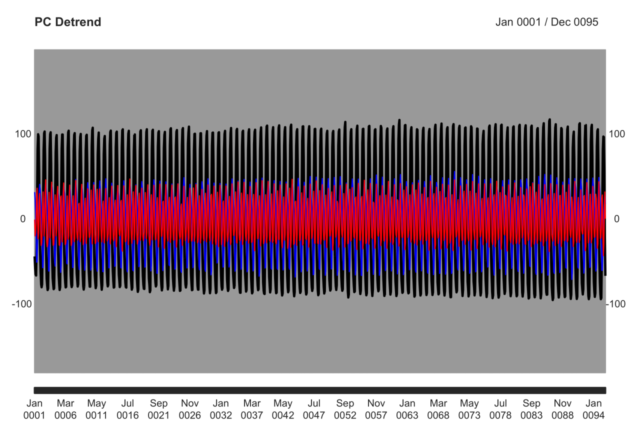


Figure 2: Spatial Pattern of PC 1



(a) Detrened Data

Figure 3: PC temporal patterns

We have found the PCs have seasonality through ACF in Figure 7. PC1 and PC2 have annual seasonality, PC3 and PC4 have semi-annual, and PC5 and PC6 have quarterly seasonality, which get supported by periodograms, Figure 10. It has positive and negative sides, which makes sense because the dataset has covered both north and south hemispheres. Additionally, PC1 and PC2 seems to have similar structures behind as well as PC3 and PC4 and PC5 and PC6 via the Cross-Covariance Functions in Figure 9. Since those have the obvious seasonality, we conduct spectral analysis on PC 1.

Spectral Analysis We model the first principal component (PC1) which has annual seasonality as:

$$\begin{aligned} X_t &= A\cos(2\pi\frac{1}{12}t) + B\sin(2\pi\frac{1}{12}t) && (3)\\ &= R\sin(2\pi\frac{1}{12} + \varphi) && (4)\\ \gamma(h) &= \sigma^2\cos(2\pi\frac{1}{12}h) && (5) \end{aligned}$$

where $R^2 = A^2 + B^2, \varphi = \arctan(\frac{A}{B})$.

Here is our fitted model:

$$\begin{aligned} \hat{X}_t &= -32.652\cos(2\pi\frac{1}{12}t) - 97.1938\sin(2\pi\frac{1}{12}t) && (6)\\ &= 102.5319\sin(2\pi\frac{1}{12}t + \frac{\pi}{10}) && (7)\\ \hat{\gamma}(h) &= 13.93^2\cos(2\pi\frac{1}{12}h) && (8) \end{aligned}$$

This model can explains 96% variances (Adjusted R^2 is 0.9644). The PC 2 can be fitted by

$$\hat{X}_t = 48.96085\sin(2\pi\frac{1}{12} - \frac{2\pi}{5}),\ \hat{\sigma}^2 = 8.17 \qquad (9)$$

which is shifted and smaller oscillations with respect to PC1 model.

Interestingly, we find another cycles from the residuals of the PC1 model. If we allow to add more cyclic variables, we could end up:

$$\begin{aligned} X_t = {} & c_1\cos(2\pi\frac{1}{12}t) + c_2\sin(2\pi\frac{1}{12}t) + c_3\cos(2\pi\frac{1}{6}t) + c_4\sin(2\pi\frac{1}{6}t)\\ & + c_5\cos(2\pi\frac{1}{4}t) + c_6\cos(2\pi\frac{1}{3}t) + c_7\sin(2\pi\frac{1}{3}t) + c_8\sin(2\pi t) \qquad (10) \end{aligned}$$

this model can explain 99.62% variances.

A PCA

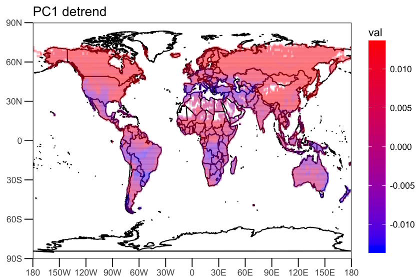
A.1 Tables

	PC1	PC2	PC3	PC4	PC5	PC6
StDev	66.650	28.003	24.356	20.348	17.832	12.301
Prop. of Var.	0.347	0.061	0.046	0.032	0.025	0.012
Cum. Prop.	0.347	0.408	0.454	0.486	0.511	0.523

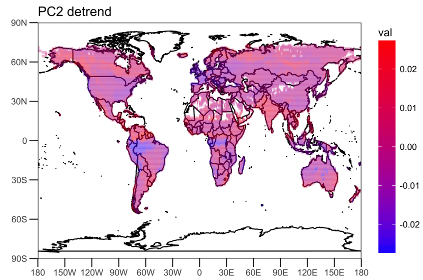
Table 1: Explained Variations of PC: Raw Dataset

	PC1	PC2	PC3	PC4	PC5	PC6
StDev	73.857	35.586	23.852	14.581	11.859	10.669
Prop. of Var.	0.426	0.099	0.044	0.017	0.011	0.009
Cum. Prop.	0.426	0.524	0.569	0.585	0.596	0.605

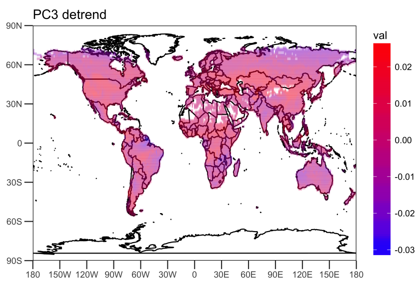
Table 2: Explained Variations of PC: Detrendset



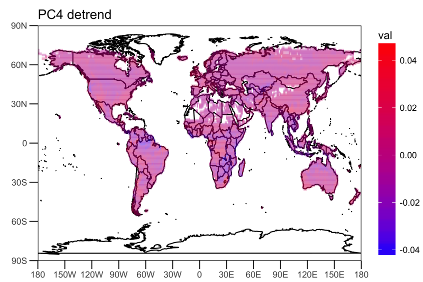
(a) PC 1



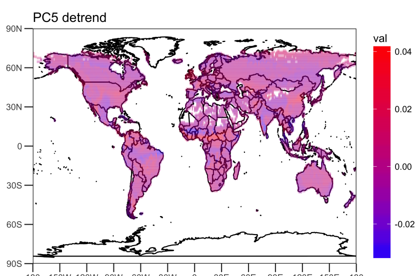
(b) PC 2



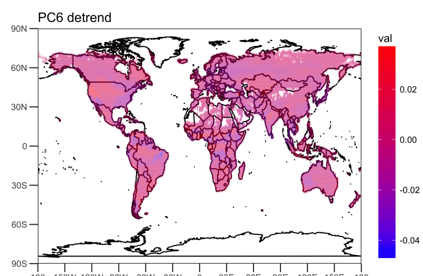
(c) PC 3



(d) PC 4



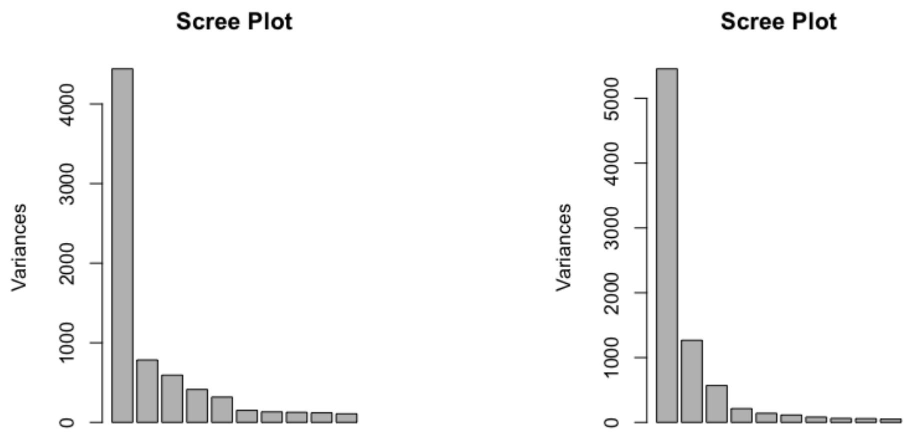
(e) PC 5



(f) PC 6

Figure 4: Spatial Patterns

A.2 Plots



(a) Raw Data (b) Detrended Data

Figure 5: Scree Plots

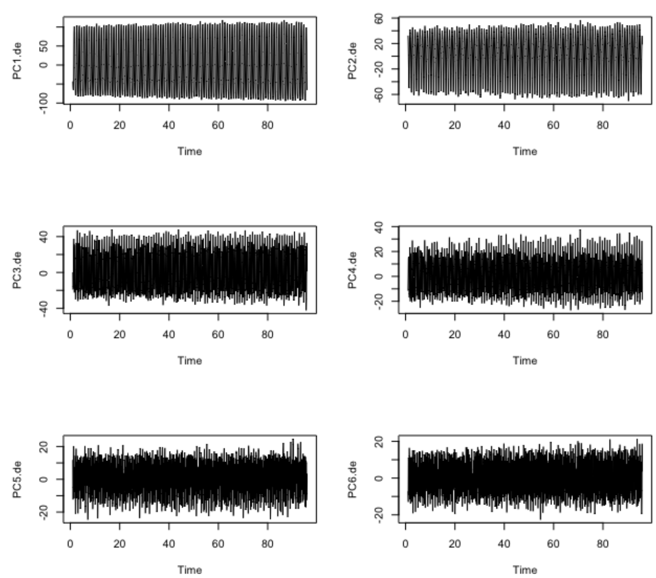


Figure 6: Time Series Plots

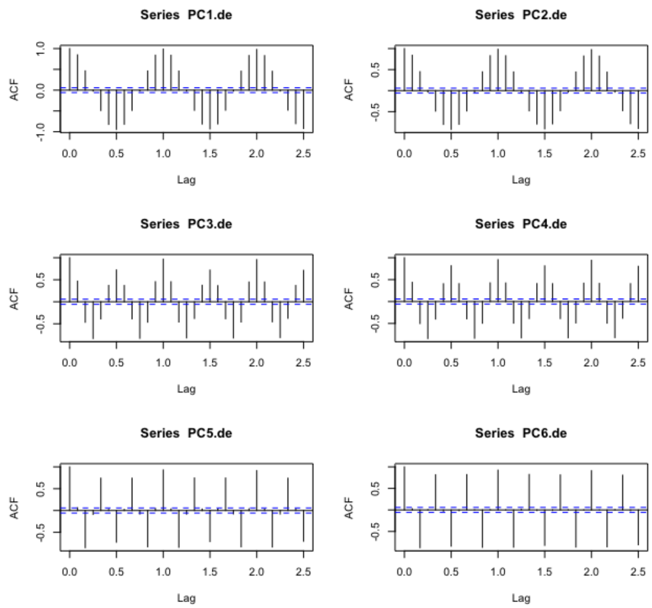


Figure 7: Auto-Covariance Function

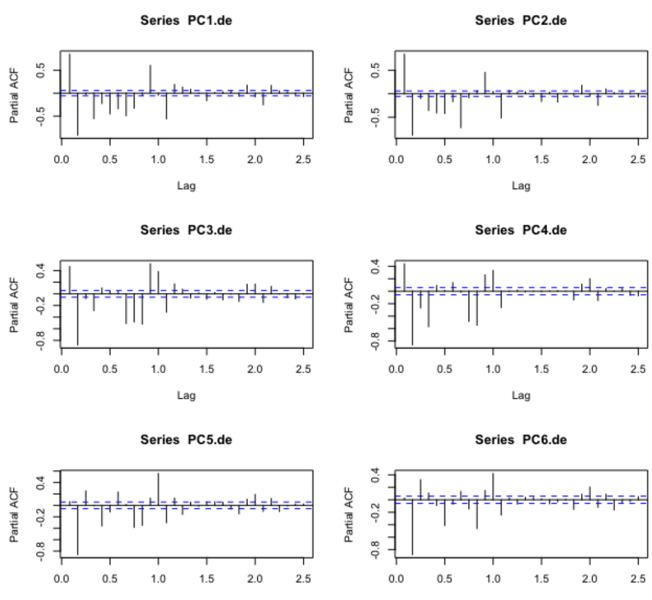


Figure 8: Partial ACF

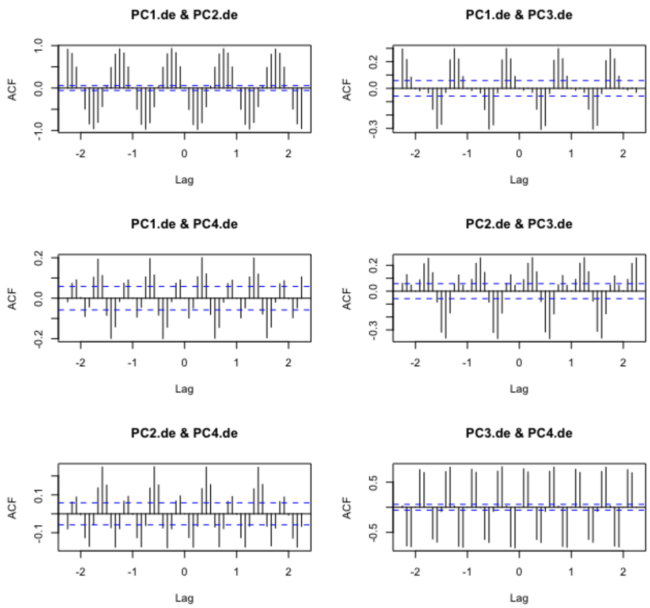


Figure 9: Cross-Covariance Function

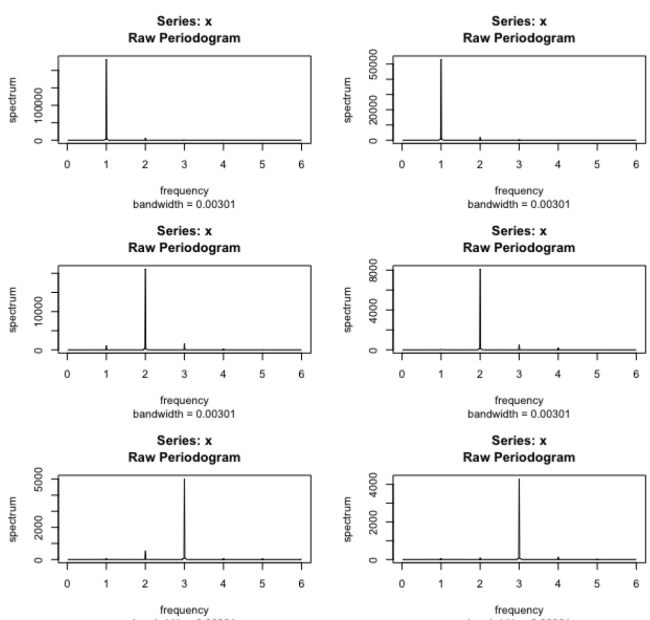


Figure 10: PCA Periodogram

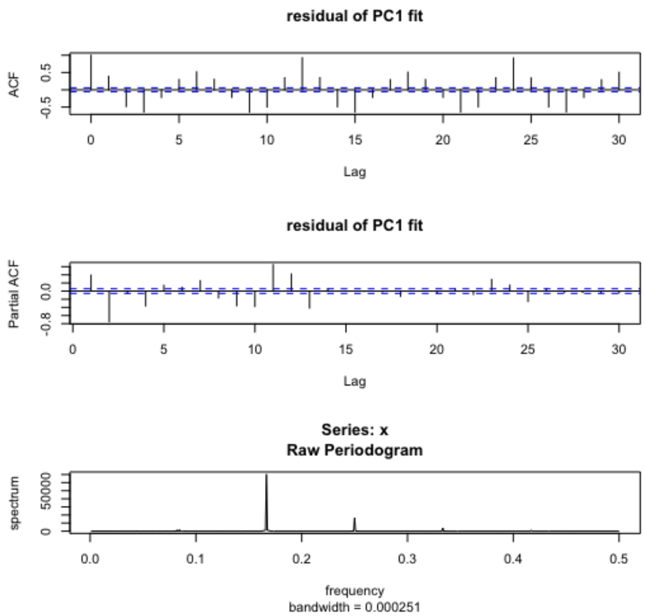


Figure 11: Analysis on Residuals PC1 Fit