Functional change detection for mapping annual urban dynamics using Landsat Data

Xinyue Chang, Xiongtao Dai, Zhengyuan Zhu

April 10, 2019

Content

Motivation

Urban Dynamics

Testing if there is an urbanization process and detecting when urbanization starts or ends would be helpful to study the urbanizing process, develop urban growth models, and investigate further environmental impacts.

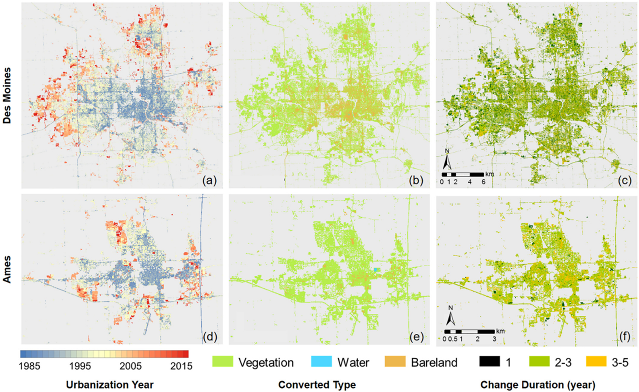
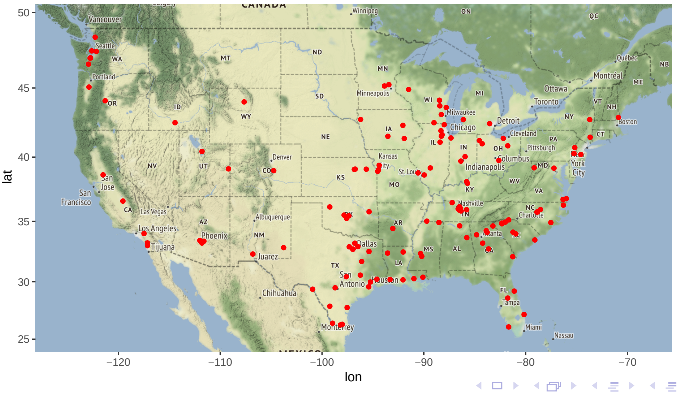


Figure: from Li et al., 2018

Areas and Datasets

- 135 urbanized locations and 124 non-urbanized locations all around United States.

- For each location, time ranges from 1985 to 2015 where at most 74 days and 19 days on average are observed. And B1(Blue), B2(Green), B3(Red), B4(NIR), B5(MIR), B7(SWIR) are included.



The Problem

We want to...

- test if there is urbanization happened for each location.

- estimate the year of urbanizaiton starts and ends for each location.

The change could be vegatation→urban, water→urban or bareland→urban, corresponding to NDVI = (NIR−Red)/(NIR+Red), MNDWI = (Green−MIR)/(Green+MIR), and SWIR(B7).

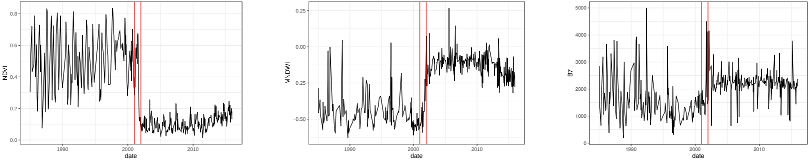


Figure: starting and ending of change in the 74th location

Functional Change Detection

Li et al., RSE (2018)

Regression based temporal segmentation approach:

- a linear regression was fitted to the annual time series, two turning points with maximum deviations are P_1 and P_2.
- dominant conversion type is determined by maximum absolute difference between two turning points.
- change-points are estimated as the turning points of dominant type.

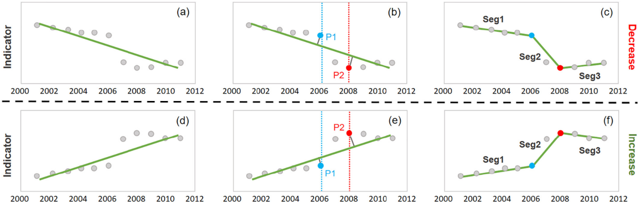


Figure: from Li et al., 2018

Li et al., RSE (2018)

Commets:

- simple and easy to implement
- averaging within each year may result in the loss of functional information
- no testing procedure developed

Berkes et al., JRSSB (2009)

Univariate dense functional observations: $X_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_k \phi_k(t)$, $i = 1, \cdots, N$

$$P_k(t) = \sum_{1 \leq i \leq k} X_i(t) - \frac{k}{N} \sum_{1 \leq i \leq N} X_i(t)$$

Let $x \in [0, 1]$, $\widehat{\xi}_{ik} = \int \{X_i(t) - \bar{X}_N(t)\} \widehat{\phi}_k(t) dt$, $k = 1, 2, \cdots, K$.

$$\int \left\{ \sum_{1 \leq i \leq Nx} X_i(t) - \frac{[Nx]}{N} \sum_{1 \leq i \leq N} X_i(t) \right\} \widehat{\phi}_k(t) dt$$
$$= \sum_{1 \leq i \leq Nx} \widehat{\xi}_{ik} - \frac{[Nx]}{N} \sum_{1 \leq i \leq N} \widehat{\xi}_{ik}$$

Berkes et al., JRSSB (2009)

Let $B_1(\cdot), \cdots, B_K(\cdot)$ denote independent standard Brownian bridges.

$$T_N(x) = \frac{1}{N} \sum_{k=1}^{K} \hat{\lambda}_k^{-1} \left(\sum_{1 \le i \le Nx} \hat{\xi}_{ik} - x \sum_{i=1}^{N} \hat{\xi}_{ik} \right)^2$$

$$S_{NK} = \frac{1}{N} \sum_{i=1}^{N} T_N(\frac{i}{N})$$

Under hypothesis H_0: $\mathrm{E}\{X_1(t)\} = \cdots = \mathrm{E}\{X_N(t)\}$,

$$T_N(x) \xrightarrow{d} \sum_{k=1}^{K} B_k^2(x)$$

$$S_{NK} \xrightarrow{d} \int_0^1 \sum_{k=1}^{K} B_k^2(x) dx \equiv D_K$$

where D_K was derived by Kiefer (1959).

Berkes et al., JRSSB (2009)

Comments:

- works well on detecting changes in the mean of univariate functional data and nice asymptotic properties.
- the statistics is developed under assumptions of dense functional data, asymptotical distribution does not hold for sparse functional data automatically.
- cannot apply to multivariate functional data directly, especially when there are correlations among functions.

Our Method

For each location, model three indicators as multivariate functional data with yearly replications.

$$\mathbf{Y}_i(t) = \mathbf{X}_i(t) + \boldsymbol{\epsilon}_i(t)$$
$$= \boldsymbol{\mu}(t) + \sum_{k=1}^{K} \xi_{ik}\boldsymbol{\phi}_k(t) + \boldsymbol{\epsilon}_i(t)$$

where $\mathbf{X}_i(t) = \{X_{i1}(t), X_{i2}(t), X_{i3}(t)\}$ corresponds to NDVI, MNDWI and B7 indicators, hence $\boldsymbol{\mu}(t) = \{\mu_1(t), \mu_2(t), \mu_3(t)\}$ and $\boldsymbol{\phi}_k(t) = \{\phi_{k1}(t), \phi_{k2}(t), \phi_{k3}(t)\}$ is orthonormal basis in three-dimensional Hilbert space.

$\widehat{\xi}_{ik}$: conditional expectation, Dai et al., Annals (2018)

Testing Statistics

Multivariate Weighted CUSUM (MWC):

$$T_N(x) = \frac{1}{N}\sum_{k=1}^{K}\widehat{\lambda}_k^{-1}\left(\sum_{1\le i\le Nx}\widehat{\xi}_{ik} - x\sum_{i=1}^{N}\widehat{\xi}_{ik}\right)^2$$

$$S_{NK} = \frac{1}{N}\sum_{i=1}^{N} T_N(\frac{i}{N}) \xrightarrow{d} D_K^*, \quad \text{reject if greater than } (1-\alpha)th \text{ percentile}$$

Bootstrap distribution

To simulate distribution D_K^* for $K = 1, 2, \cdots, 10$, we conducted the following procedure.

- estimate change-point $\widehat{P}_1$ and $\widehat{P}_2$ by our proposed ensemble method.
- estimate multivariate mean function before $\widehat{P}_1$ and after $\widehat{P}_2$, substract mean function from standardized functional data: $r_i(t)$,
- resample 100 curves from $r_i(t)$ and do multivariate FPCA, calculate S_{NK} for $K = 1, \cdots, 10$, repeat the procedure 500 times and obtain the bootstrap sample of S_{NK}.

Estimation

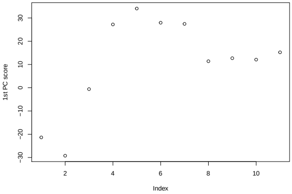
Multivariate Sigmoid CUSUM Ensemble (MSCE): for $k = 1, \cdots, K$,

$$f(x) = \frac{1}{1 + exp(-\beta x)}$$

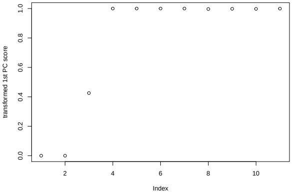
$$T_N^k(x) = \frac{1}{N}\widehat{\lambda}_k^{-1}\left(\sum_{1 \le i \le Nx} f(\widehat{\xi}_{ik}) - x\sum_{i=1}^{N} f(\widehat{\xi}_{ik})\right)^2$$

Sigmoid function

Apply a sigmoid function $1/\{1 + exp(-\beta x)\}$to the estimated $\widehat{\xi}_{ik}$ to reduce the effect of big noise in data.



(a) before sigmoid transformation



(b) after sigmoid transformation

Estimation of starting and ending points

Multivariate Sigmoid CUSUM Ensemble (MSCE):

- for each principle component $k = 1, \cdots, K$, estimate start and end of urbanization

$$\widehat{P}_1^k = \lfloor N \times \min\{x : T_N^k(x) = \max_{0 \le y \le 1} T_N^k(y)\} \rfloor$$

,

$$\widehat{P}_2^k = \lceil N \times \max\{x : T_N^k(x) = \max_{0 \le y \le 1} T_N^k(y)\} \rceil$$

.

- determine the dominant component by $\Delta_k = |\widehat{\xi}_{\widehat{P}_1^k, k} - \widehat{\xi}_{\widehat{P}_2^k, k}|$ Thus,

$$\widehat{P}_1 = \sum_{k=1}^{K} \widehat{P}_1^k I\{\Delta_k = \max_{1 \le k \le 3}\{\Delta_k\}\}, \widehat{P}_2 = \sum_{k=1}^{K} \widehat{P}_2^k I\{\Delta_k = \max_{1 \le k \le 3}\{\Delta_k\}\}.$$

Simulation

Settings

$$
\begin{aligned}
\mathbf{Y}_i(t) &= \mathbf{X}_i(t) + \boldsymbol{\epsilon}_i(t) \\
&= \widehat{\boldsymbol{\mu}}_1(t) + \delta I(i \geq N_c) + \sum_{k=1}^{K} \xi_{ik} \widehat{\boldsymbol{\phi}}_k(t) + \boldsymbol{\epsilon}_i(t) \\
\delta &= c \left\{ \widehat{\boldsymbol{\mu}}_2(t) - \widehat{\boldsymbol{\mu}}_1(t) \right\} \\
\beta(c) &= P(\text{reject } H_0 | H_0 \text{ is false})
\end{aligned}
$$

Applications

Preprocessing Data

- Modified NDVI formula as NDVI $= (-\text{NIR} + \text{Red})/(\text{NIR} + \text{Red})$ such that it increases after change, which is consistent with the other two indicators.
- Savitzky-Golay (S-G) filter was applied to remove abnormal observations in the raw time series.
- Functional normalize three indices: NDVI, MNDWI, SWIR by univariate sparse FPCA: $\{Y_{id}(t) - \widehat{\mu}_d(t)\}/\sqrt{\widehat{G}(t,t) + \widehat{\sigma}^2}$, d denotes indicators.

Cross-validation and choosing K, β

K, the number of principle components, is chosen by variation explained (FVE) in eigen decomposition of covariance function. FVE and β are treated as parameters to fit in training set and selected through a grid by the maximum AUC or accuracy. Then they are used in test set for evaluating performance.
So a repeated k-fold cross-validation was conducted to compare our method with others.

Available Methods for Comparison

Univariate Weighted CUSUM Ensemble (UWCE): For each of three indices (NDVI, MNDWI, B7), calculate univariate principle component scores and weighted CUSUM statistics: $T_{Nd}(x)$ and S_{NKd} where d denotes indicators. Testing and estimations ensemble over three indices. Bootstrap distribution of S_{NKd} can be obtained similarly for each indicator.

Regression Ensemble (RE): fit a linear regression to the annual time series for each indicator, t-test for three slopes and calculate two turning points with maximum deviations. Testing and estimations ensemble over three indices.

Other Possible Methods for Comparison

Multivariate CUSUM Ensemble (MCE): same $T_N^k(x)$ with MSCE but applying no sigmoid functions, $S_N^k = \dfrac{1}{N} \sum_{i=1}^N T_N^k(\dfrac{i}{N}) \xrightarrow{d} D_1^*$ is used for testing: reject if any k rejects, estimation ensembles over principle components in the same way as MSCE.

Multivariate Sigmoid Weighted CUSUM (MSWC): same $T_N(x)$ with MWC but applying sigmoid functions, testing procedure is the same as MWC, change-points are estimated as $\widehat{P}_1 = \lfloor N \times \min\{x : T_N(x) = \max_{0 \le y \le 1} T_N(y)\} \rfloor$, $\widehat{P}_2 = \lceil N \times \max\{x : T_N(x) = \max_{0 \le y \le 1} T_N(y)\} \rceil$

Testing Results

Repeated 5-fold cross validation, AUC based on tuning α.

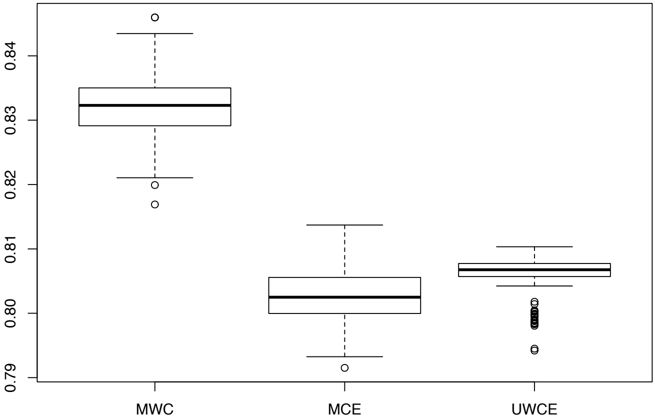
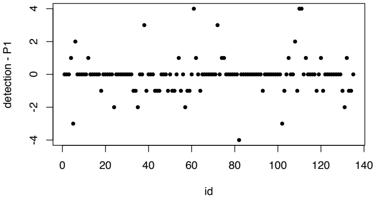
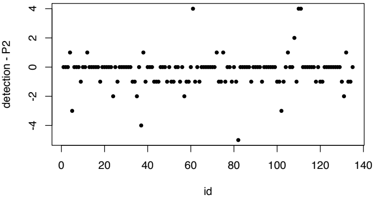


Figure: MWC: 0.8322, MCE: 0.8026, UWCE: 0.8061, RE: 0.7493.

Estimation Results

For a fair comparison with the regression method, we also selected principle components with i between 2001 and 2011 for calculating statistics.
$\#$ of correct detection for P_1: $121/135 = 0.8963$, and $\#$ of correct detection for P_2: $123/135 = 0.9111$.





Regression estimation $\widehat{P}_2$: $114/135 = 0.8444$

Comparison

repeated 4-fold cross-validation: (135 cases in total)

Mean	MSCE	MCE	MSWC	UWCE
# of $\lvert P_1 - \widehat{P}_1 \rvert \leq 1$	120.735	114.710	88.785	99.455
# of $\lvert P_2 - \widehat{P}_2 \rvert \leq 1$	122.735	116.710	92.800	104.410

MSCE: multivariate sigmoid cusum ensemble, MCE: multivariate cusum ensemble, MSWC: multivariate sigmoid weighted cusum, UWCE: univariate weighted cusum ensemble.

Q & A

Thank you!