Detection of change points in the mean function of spatio–temporal curves

Piotr Kokoszka
Department of Statistics, Colorado State University

Joint work with
Oleksandr Gromenko (Startup, Singapore)
and
Matthew Reimherr (Penn State)
Example of spatial locations

Locations of selected weather stations used in the data example
Example of a functional observation

Grey line represents a transformed raw precipitation record for a single year $n$ and a single location $s$, the black line is the smoothed precipitation $X_n(s; t)$. 
Data model and testing problem

\[ X_n(s_k; t_i) - \text{scalar observations} \]

(possibly transformed and/or smoothed)

\[ 1 \leq n \leq N - \text{year} \]

\[ s_k, 1 \leq k \leq K - \text{location} \quad (s_k \in S) \]

\[ t_i, 1 \leq i \leq 365 - \text{calendar day} \quad (t_i \in T) \]

\[ X_n(s; t) = \mu_n(s; t) + \varepsilon_n(s; t), \quad s \in S, \ t \in T. \]

Each \( \mu_n \) is a function \( S \times T \)

\[ H_0 : \mu_1 = \cdots = \mu_N \quad \text{vs.} \quad H_A : \mu_1 = \cdots = \mu_{n^*} \neq \mu_{n^*+1} = \cdots \mu_N. \]
Examples of mean functions

Red: mean functions before and after the estimated change point.
Grey: curves for two selected years. (one location)
The spatial field showing the $L^2$ distance between the mean log–precipitation before and after 1966. There is an increase in precipitation throughout the year in the area around location 4. Locations 2 and 3 do not show a large change.
Assumptions

Spatial dependence, in \( s \in S \), is estimated. Temporal dependence within each year, in \( t \in T \), is approximated (functional approach).

The errors \( \varepsilon_1, \varepsilon_2 \ldots, \varepsilon_N \) are iid mean zero random fields on \( S \times T \). They have \textit{separable covariances}:

\[
E \left[ \varepsilon_n(s_k; t) \varepsilon_n(s_\ell; t') \right] = C(t, t') \sigma(s_k, s_\ell).
\]

Implication:

\[
X_n(s; t) = \mu_n(s; t) + \sum_{i=1}^{\infty} \xi_{ni}(s) v_i(t),
\]

\[
C(t, t') = \sum_{i=1}^{\infty} \lambda_i v_i(t) v_i(t'),
\]
The algorithm for the estimation of the temporal covariance function \(C(t, t')\) and the spatial covariances \(\sigma(s_k, s_\ell)\) is complex. The procedure must work also if there are (a few) change points. Key elements (require separability):

Use differences \(Z_n(s_k; t) = X_{n+1}(s_k; t) - X_n(s_k; t)\).

Preliminary estimator:

\[
\sigma(s_k, s_\ell) \approx \frac{1}{2(N - 1)} \sum_{n=1}^{N-1} \int Z_n(s_k; t)Z_n(s_\ell, t)dt.
\]

Smooth and make positive–definite (spectral domain, Hall et al. ca. 1990).

Use weights based on the estimated \(\sigma(s_k, s_\ell)\) to estimate the \(C(t, t')\) as a weighted average.
Test statistics

\[ \hat{\Lambda}_1 = \frac{1}{N^2} \sum_{k=1}^{K} \hat{w}(k) \sum_{i=1}^{p} \hat{\lambda}_i^{-1} \sum_{r=1}^{N} \left\langle \sum_{n=1}^{r} X_n(s_k) - \frac{r}{N} \sum_{n=1}^{N} X_n(s_k), \hat{v}_i \right\rangle^2, \]

\[ \hat{\Lambda}_2 = \frac{1}{N^2} \sum_{k=1}^{K} \hat{w}(k) \sum_{i=1}^{p} \sum_{r=1}^{N} \left\langle \sum_{n=1}^{r} X_n(s_k) - \frac{r}{N} \sum_{n=1}^{N} X_n(s_k), \hat{v}_i \right\rangle^2, \]

\[ \hat{\Lambda}_\infty = \frac{1}{N^2} \sum_{k=1}^{K} \hat{w}(k) \sum_{r=1}^{N} \left\| \sum_{n=1}^{r} X_n(s_k) - \frac{r}{N} \sum_{n=1}^{N} X_n(s_k) \right\|^2. \]

The weights \( \hat{w}(k) \) are related to spatial covariances.
Asymptotic null distributions

\[ \hat{\Lambda}_1 \overset{D}{\to} \Lambda_1 = \sum_{k=1}^{K} w(k) \sum_{i=1}^{p} \int_{0}^{1} B_{ik}^2(x) \, dx, \]

\[ \hat{\Lambda}_2 \overset{D}{\to} \Lambda_2 = \sum_{k=1}^{K} w(k) \sum_{i=1}^{p} \lambda_i \int_{0}^{1} B_{ik}^2(x) \, dx, \]

\[ \hat{\Lambda}_2^\infty \overset{D}{\to} \Lambda_2^\infty = \sum_{k=1}^{K} w(k) \sum_{i=1}^{\infty} \lambda_i \int_{0}^{1} B_{ik}^2(x) \, dx. \]

The Brownian bridges \( B_{ik} \) are independent across \( i \) and spatially dependent. Spatial covariances are \( \sigma(s_k, s_\ell) \), and can be estimated.
Estimated empirical size as a function of the captured cumulative variance for the test based on $\Lambda_2$.
We have a data-driven calibration algorithm that chooses optimal CumVar.
Estimation based on $\hat{\Lambda}_2^\infty$. The red vertical line shows the estimated position of the change point, 1966.
Application to Mid-West log-precipitation

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Segment</th>
<th>Decision</th>
<th>$\hat{\Lambda}_2$</th>
<th>P-value</th>
<th>Estimated change-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1941–2000</td>
<td>Reject</td>
<td>0.011719</td>
<td>0.0013</td>
<td>1966</td>
</tr>
<tr>
<td>2</td>
<td>1941–1965</td>
<td>Accept</td>
<td>0.010337</td>
<td>0.0987</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>1966–2000</td>
<td>Accept</td>
<td>0.009997</td>
<td>0.1354</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Segment</th>
<th>Decision</th>
<th>CV,%</th>
<th>$p$</th>
<th>P-value</th>
<th>Estimated change-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1941–2000</td>
<td>Reject</td>
<td>85.01</td>
<td>28</td>
<td>0.0189</td>
<td>1968</td>
</tr>
<tr>
<td>2</td>
<td>1941–1967</td>
<td>Accept</td>
<td>86.00</td>
<td>28</td>
<td>0.8964</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>1968–2000</td>
<td>Accept</td>
<td>85.44</td>
<td>27</td>
<td>0.8350</td>
<td>–</td>
</tr>
</tbody>
</table>
Main paper:
O. Gromenko, P. Kokoszka and M. Reimherr, Detection of change in the spatiotemporal mean function, Journal of the Royal Statistical Society (B), 79, 29-50, 2017

Closely related papers:

References