Comments on: Extensions of some classical methods in change point analysis

Piotr Kokoszka

Published online: 9 May 2014

© Sociedad de Estadística e Investigación Operativa 2014

The authors must be congratulated on an extensive and at the same time detailed and very informative survey of the most recent developments in change-point analysis. In my comments, I will focus on an extension of the results discussed in their Section 7: detection of a change point in functional observations. This contribution is based on the work presented in [Gromenko et al. \(2014\)](#).

Climatic data, like temperature and rainfall, are collected at a number of locations, so they can be viewed as sequences of annual curves with an additional index referring to the spatial location of the measurement station; denote a measurement made in year n , at spatial location $\mathbf{s} \in \mathcal{S}$, and time (of year) $t \in \mathcal{T}$, as $X_n(\mathbf{s}, t)$. The interest of climatologists is in trends and changes in climate over a region, the United States is typically divided into 5–10 climatic regions. [Gromenko et al. \(2014\)](#) work with precipitation data at $K = 59$ locations in 12 states in the midwestern US, over the course of $N = 60$ years, there are 365 measurements per year at each location. These values are typical of such data sets: depending on the part of the world, records may reach 50–150 years back and there are several dozen stations in a climatic region.

The statistical model for such data is $X_n(\mathbf{s}; t) = \mu_n(\mathbf{s}; t) + \varepsilon_n(\mathbf{s}; t)$, where $\mu_n \in L^2(\mathcal{S} \times \mathcal{T})$ is a deterministic mean function and $\varepsilon_n \in L^2(\mathcal{S} \times \mathcal{T})$ is a zero mean random error term. The goal is to test the null hypothesis that the mean functions μ_n are the same across n . If this hypothesis is true, the long-term annual pattern over a region does not change. The testing problem can thus be stated as $H_0 : \mu_1 = \dots = \mu_N$ vs. $H_A : \mu_1 = \dots = \mu_{n^*} \neq \mu_{n^*+1} = \dots \mu_N$, for some $1 < n^* < N$. In this spatio-temporal setting, it is necessary to specify the spatial covariances as well. It is

This comment refers to the invited paper available at doi:[10.1007/s11749-014-0368-4](https://doi.org/10.1007/s11749-014-0368-4).

P. Kokoszka (✉)

Department of Statistics, Colorado State University,
Statistics Building, Fort Collins, CO 80522, USA
e-mail: Piotr.Kokoszka@colostate.edu

convenient to postulate the covariance factor as follows:

$$\text{Cov}[X_n(\mathbf{s}_k; t), X_m(\mathbf{s}_\ell; t')] = E[\varepsilon_n(\mathbf{s}_k; t) \varepsilon_m(\mathbf{s}_\ell; t')] = \delta_{nm} C(t, t') \sigma(\mathbf{s}_k, \mathbf{s}_\ell),$$

where $C(t, t')$ and $\sigma(\mathbf{s}_k, \mathbf{s}_\ell)$ are, respectively, purely temporal and spatial covariances. Under this assumption, each function $X_n(\mathbf{s})$ admits the Karhunen–Loève expansion

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

in which the annual functions at each location have the same temporal functional principal components v_i and the spatial dependence is contained in the scores $\xi_{ni}(\mathbf{s})$. One can define several CUSUM-type test statistics, an extension of the statistic $\mathcal{H}_N$ in Section 7 of Horváth and Rice (2014) is

$$\hat{\Lambda}_1 = \frac{1}{N^2} \sum_{k=1}^K w(k) \sum_{i=1}^p \hat{\lambda}_i^{-1} \sum_{r=1}^N \left\langle \sum_{n=1}^r X_n(\mathbf{s}_k) - \frac{r}{N} \sum_{n=1}^N X_n(\mathbf{s}_k), \hat{v}_i \right\rangle^2. \quad (2)$$

In case of a single location ($K = 1, w(1) = 1$), statistic (2) reduces to the test statistic of Berkes et al. (2009). The weights $w(k)$ reflect the intuition that spatially close records contribute similar information, and so should be given smaller weights, while

Fig. 1 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, decrease in the first half of the year in the area around location 1. Locations close to 2 and 3 do not show a large change nor a consistent pattern

records at isolated locations contribute more information and should be given larger weights. These weights are computed from the data using an optimality criterion. The null limit of $\hat{\Lambda}_1$ is

$$\Lambda_1 = \sum_{k=1}^K w(k) \sum_{i=1}^p \int_0^1 B_{ik}^2(x) dx,$$

where B_{ik} are Brownian bridges, independent across i . For each i , the vector of Brownian bridges $\mathbf{B}_i = [B_{i1}, \dots, B_{iK}]^T$ has the covariances $\sigma(\mathbf{s}_k, \mathbf{s}_\ell)$, $1 \leq k, \ell \leq K$.

The asymptotic test applied to the mid-west precipitation data detects a change point around 1966. Figure 1 shows a heat map of the magnitude of the change. Such maps are often used in climate research to show a change over a region. The change-point test we discussed allows us to conclude that the pattern of change seen in the map is statistically significant.

References

- Berkes I, Gabrys R, Horváth L, Kokoszka P (2009) Detecting changes in the mean of functional observations. *J R Stat Soc (B)* 71:927–946
- Gromenko O, Kokoszka P, Reimherr M (2014) Detection of change in the mean function over a spatial region. Colorado State University, Technical Report
- Horváth, L., Rice, G.: Extensions of some classical methods in change point analysis. Test