Detekce více změn v sezónním chování průtokových řad

HANA HORÁKOVÁ
DANIELA JARUŠKOVÁ

Úvod

- Cílem naší práce je detekce změn v sezónním chování průtokových řad. Předpokládejme, že data jsou denní průměrné průtoky a roční chod je tedy dán vektorem, jehož složky odpovídají denním průměrným průtokům v jednotlivých kalendářních dnech. Změna v sezónním chování odpovídá změně střední hodnoty vektoru o 365 souřadnicích. Detekce takovéto změny je však velmi obtížná, neboť bychom pro její odhalení potřebovali data z období pokrývající mnoho let.
- *Naší snahou bylo sezónní chování charakterizovat vektory s výrazně méně souřadnicemi.*

Úvod

- Navrhli jsme použít tři metody. První metoda spočívá v charakterizaci sezónního chování pomocí měsíčních průměrných průtoků. Druhá metoda je založena na proložení řady lineární kombinací sinů a kosinů s nejmenšími Fourierovými frekvencemi. Sezónní chování je pak charakterizováno Fourierovými koeficienty. Třetí metoda spočívá v nahrazení vektoru o 365 souřadnicích vektorem, jehož složky jsou lineární kombinací složek původních vektorů s největšími rozptyly. Tato metoda je známá pod názvem metoda hlavních komponent.

Úvod

- V druhé a třetí metodě je třeba subjektivně volit délku studovaných vektorů. Pro srovnání s metodou průměrných měsíčních průtoků jsme obvykle volili délku vektoru 12.
- Popsaným způsobem je úloha detekování změny v sezónním chování řady převedena na statistickou úlohu detekce změny střední hodnoty vektoru o *p* souřadnicích.

Test pro detekci jedné změny ve střední hodnotě náhodného vektoru

- Předpokládáme-li, že v řadě došlo maximálně k jedné změně, pak můžeme definovat nulovou hypotézu a odpovídající alternativu *A* následovně:

$$H_0 : \mu_1 = \cdots = \mu_N, i = 1, \ldots N$$
$$A : \exists k^* \in \{1, \ldots, N-1\}\, \textit{takové, že}$$
$$\mu_1 = \cdots = \mu_{k^*},$$
$$\mu_{k^*+1} = \cdots = \mu_N, \; kde\; \mu_{k^*} \neq \mu_{k^*+1}.$$

Testová statistika

$$U(\beta) = \max_{|\beta N| \le k \le N - |\beta N|} \frac{k(N-k)}{N} (\bar{\boldsymbol{X}}_k^{*} - \bar{\boldsymbol{X}}_k^{**})^T \hat{\Sigma}^{-1} (\bar{\boldsymbol{X}}_k^{*} - \bar{\boldsymbol{X}}_k^{**}), \text{ kde}$$

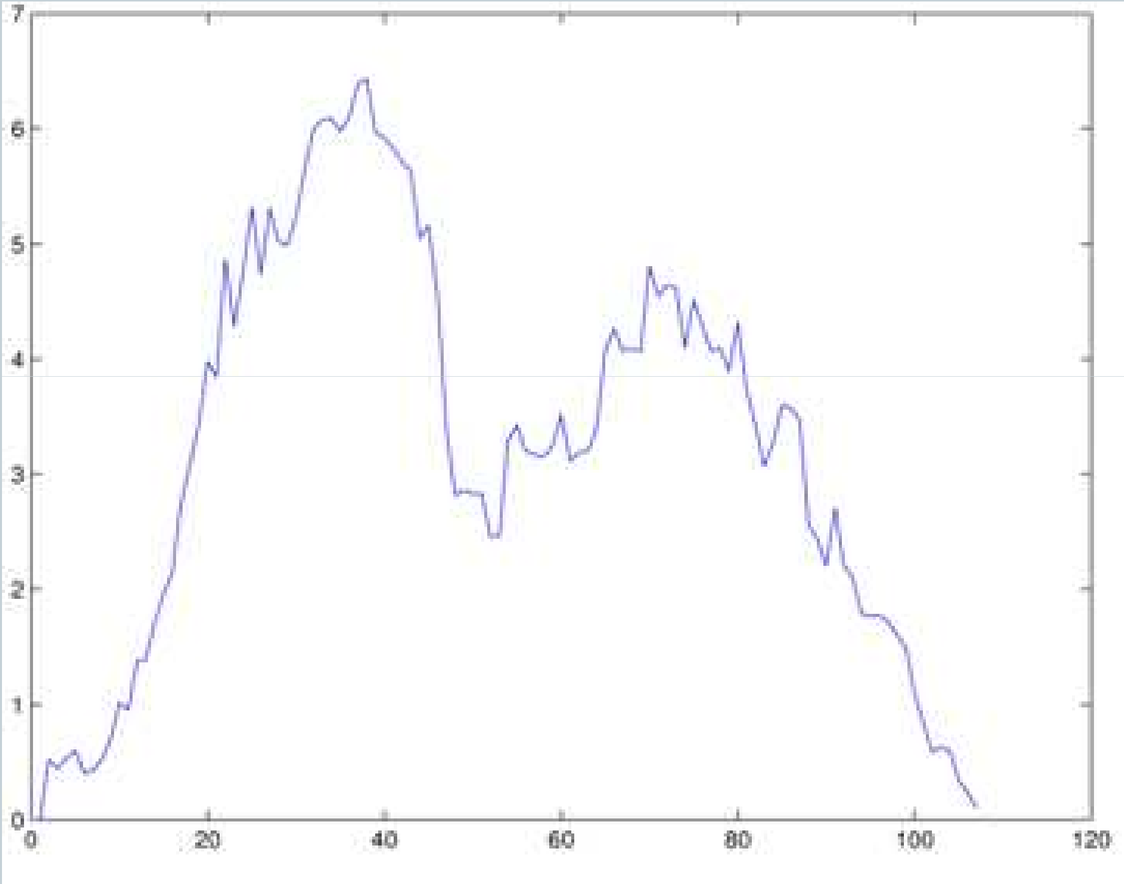
$$\bar{\boldsymbol{X}}_k^{*} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{X}_i, \qquad \bar{\boldsymbol{X}}_k^{**} = \frac{1}{(N-k)} \sum_{i=k+1}^{N} \boldsymbol{X}_i,$$

$$\hat{\Sigma} = \frac{1}{N} \left\{ \sum_{i=1}^{N} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \right\}.$$

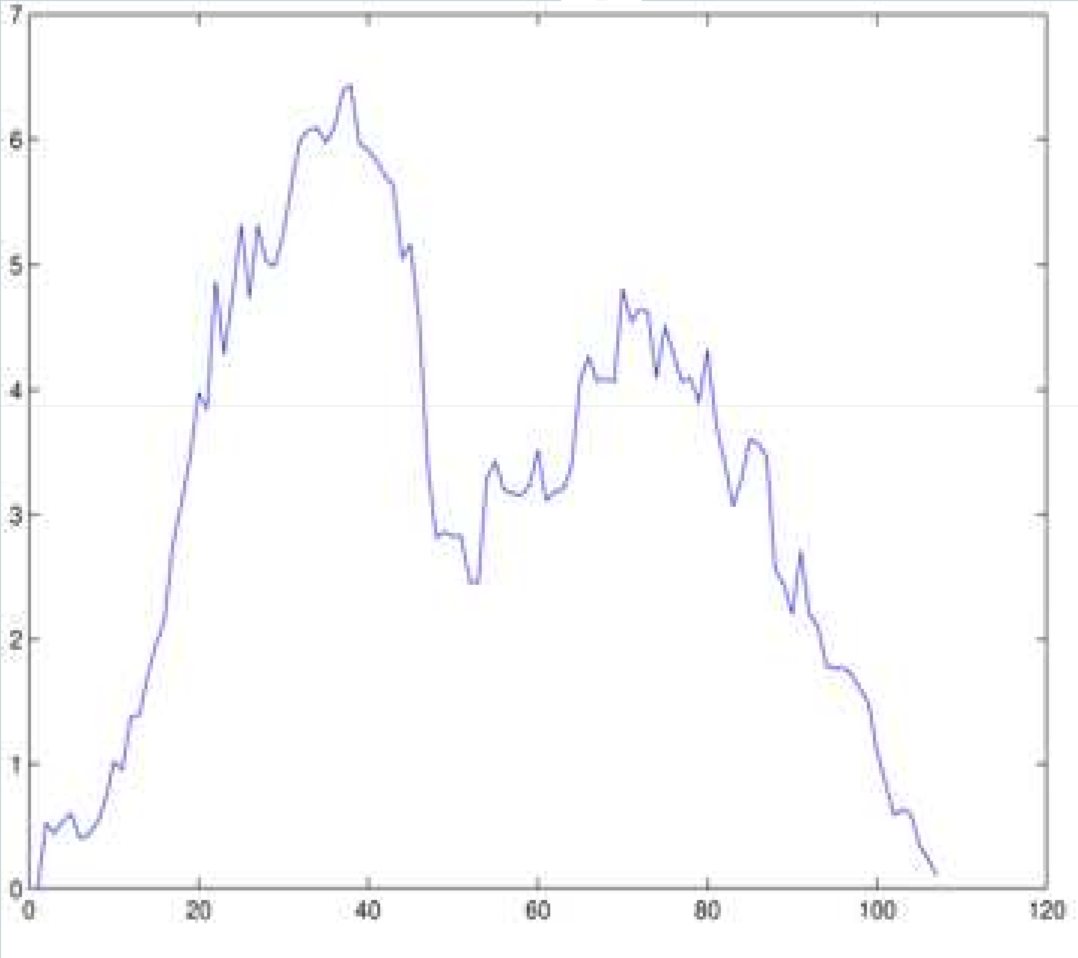
$U(\beta)$

- Pokud došlo ke změně ve střední hodnotě náhodného vektoru, pak statistika *U(β)* nabývá velkých hodnot. Najít přesnou *α.100%* kritickou hodnotu je však velmi obtížné. K nalezení přibližné kritické hodnoty se používá:
- *Bonferroniho nerovnost,*
- *asymptotické rozdělení,*
- *permutační princip*.

Obr. 1: Průběh statistik pro změny Fourierových koeficientů (Dunaj)



Obr. 2: Průběh statistik pro změny koeficientů v metodě hlavních komponent (Dunaj)



TEST PRO DETEKCI DVOU ZMĚN VE STŘEDNÍ HODNOTĚ NÁHODNÉHO VEKTORU

- Pokud máme představu, ke kolika změnám v posloupnosti došlo, změníme testovací problém následovně

$$H_0: \mu_1 = \cdots = \mu_N\,,\ i = 1, \ldots N$$

$$A: \exists k_1^* < k_2^* < \cdots < k_d^*\ takové, že$$

$$\mu_1 = \cdots = \mu_{k_1^*},$$

$$\mu_{k_1^*+1} = \cdots = \mu_{k_2^*},$$

$$\cdots$$

$$\mu_{k_d^*+1} = \cdots = \mu_N,$$

$$kde\ \mu_{k_1^*} \neq \mu_{k_1^*+1}, \mu_{k_2^*} \neq \mu_{k_2^*+1}, \ldots, \mu_{k_d^*} \neq \mu_{k_d^*+1}$$

Testová statistika

Zavedeme-li částečné součty souřadnic jednotlivých vektorů

$$S_{ij} = \sum_{m=1}^{i} X_{mj}, \quad i = 1, \ldots, N; j = 1, \ldots, p,$$

$$U_j(k_1, k_2) = \frac{1}{\sqrt{k_1 k_2 (k_2 - k_1)}} \left(k_1 S_{k_2 j} - k_2 S_{k_1 j}\right),$$

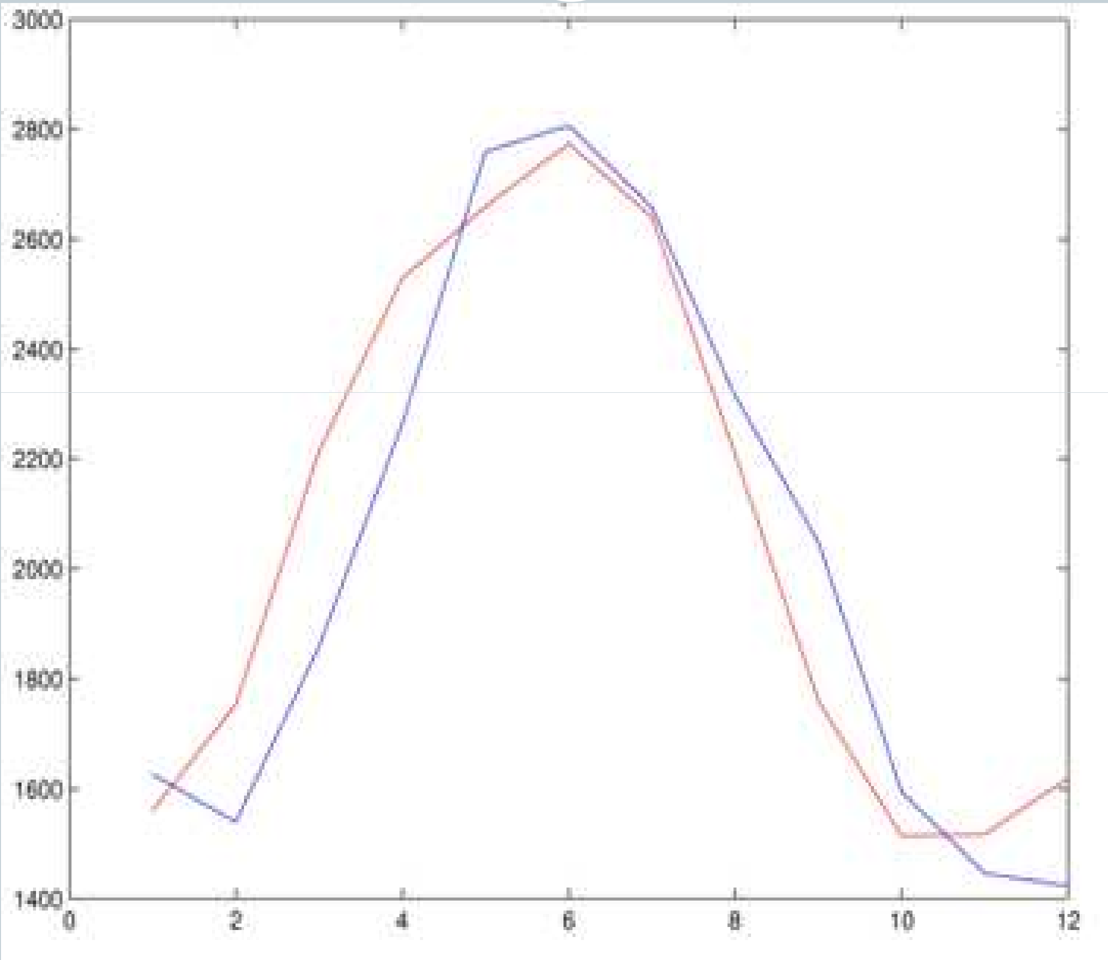
$$U_j(k_2 N) = \frac{1}{\sqrt{k_2\, N (N - k_2)}} \left(k_2 S_{Nj} - N S_{k_2 j}\right),$$

pak testová statistika bude mít tvar maxima následující sumy, kde je maximum počítáno pro $k_1 \geq [\beta N], k_2 - k_1 \geq [\beta N], N - k_2 \geq [\beta N]$:

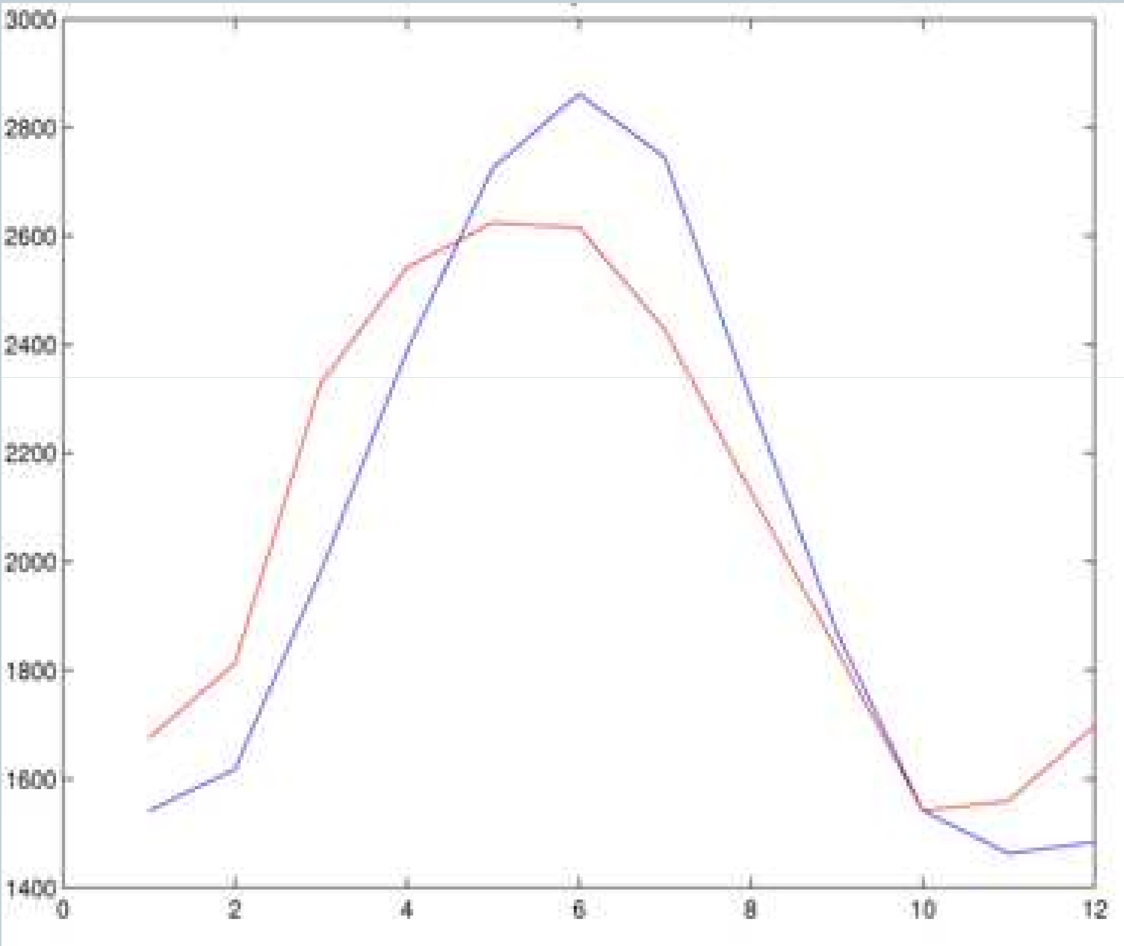
$$U_1^2(k_1, k_2) + U_2^2(k_2, N) + \cdots + U_p^2(k_1, k_2) + U_p^2(k_2, N).$$

- Přibližné kritické hodnoty můžeme opět nalézt pomocí asymptotického rozdělení. Přibližné asymptotické kritické hodnoty lze odvodit pomocí postupu navrženého v článku (Antoch, Jarušková, 2013).
- Postup lze zobecnit i pro *d* změn.

Obr. 3: Průměrné měsíční průtoky Dunaje – 1901-1938(modrá čára),1939-2009 (červená čára).



Obr. 4:Průměrné měsíční průtoky Dunaje – 1901-1979(modrá čára),1980-2009 (červená čára).



Závěr

- Velmi často se stává, že se ve studované řadě vyskytuje více než jedna změna. V takovém případě se obvykle postupuje tak, že se řada rozštěpí v místě detekované změny a další změny se hledají v první a druhé části řady odděleně. Nicméně je zřejmé, že metody, které byly navrženy pro případ nejvýše jedné změny, ztrácejí sílu, a tedy schopnost detekce v případě, že je v řadě více změn. Z tohoto důvodu byly také navrženy metody pro detekci vícenásobných změn, které lze použít i při detekci změn sezónního chování řady. Samozřejmě s rostoucím počtem očekávaných změn roste i počet odhadovaných parametrů, což opět snižuje sílu testu. Zdá se tedy, že tyto metody je vhodné aplikovat pouze v případě velmi dlouhých řad.

- DĚKUJI VÁM ZA POZORNOST