Odhad změny ročních maxim průtokových řad

Hana Horáková, Daniela Jarušková

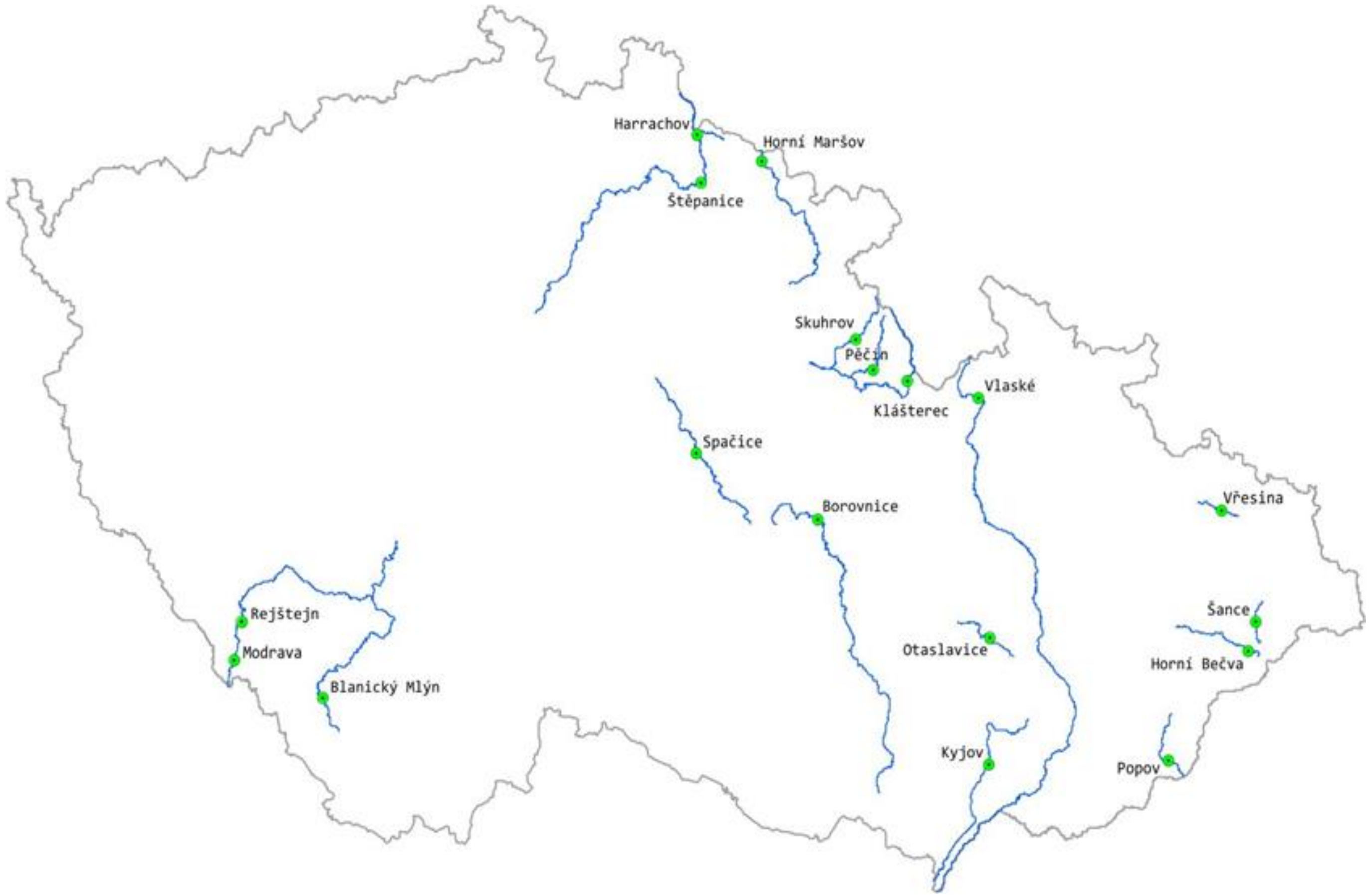
Úvod

- Vstupní data: MATICE, v řádcích hodnoty, které odpovídají jednotlivým kalendářním dnům, sloupce - jednotlivé roky
- $m \times 365$, m… počet let
- Naše data – průměrné denní průtoky

Statistický problém

- Otázka, jestli průměrná hodnota vektoru o 365 souřadnicích zůstává stejná během celé doby pozorování.
- Zkoumali jsme stacionaritu ročního chodu.

Mapa stanic


Harrachov
Horní Maršov
Štěpanice
Skuhrov
Pěčín
Klášterec
Vlaské
Spačice
Borovnice
Vřesina
Rejštejn
Modrava
Blanický Mlýn
Otaslavice
Šance
Horní Bečva
Kyjov
Popov

18 stanic – malé vodní toky

- Malé vodní toky, jejichž měřící stanice se nacházejí nad přehradami, proto u nich můžeme vyloučit vliv lidské činnosti
- Pokud bychom zjistili nestacionaritu, je možné spojovat ji s klimatickými změnami

První krok – dvouvýběrová analýza

- Data jsme rozdělili na 2 části (subjektivní volba – dělení v roce 1997).
- H_0: $\mu_1=\mu_2$
- A: $\mu_1\neq\mu_2$
- vektor pozorování je rozměrný (365 souřadnic), silná korelovanost mezi sousedními hodnotami

Snížení dimenze úlohy

- Zjednodušení modelu - Fourierovy koeficienty pro 3 frekvence (6 koef.)
- Metoda hlavních komponent
- Měsíční průměry

Change point

- Od dvouvýběrového problémy přecházíme k change point problému

- Zobecnění na situaci, kdy nerozdělujeme na dvě fixní části, ale bod změny neznáme. Vezmeme maximum testové statistiky přes všechna možná dělení.

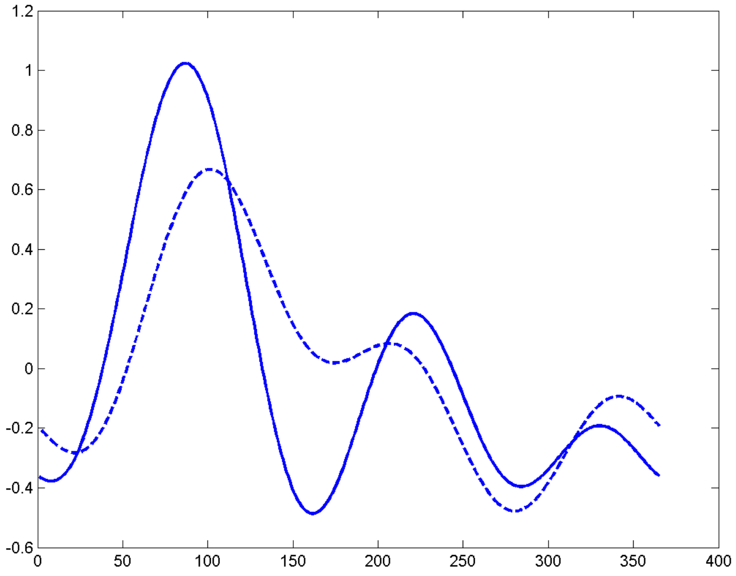
Stacionarita vs. nestacionarita?

- V 6 případech našich řad jsme zamítli nulovou hypotézu o stacionaritě ročního chodu.
- Co je důvodem zamítnutí?

2 modely k modelování chodu (Blanice)

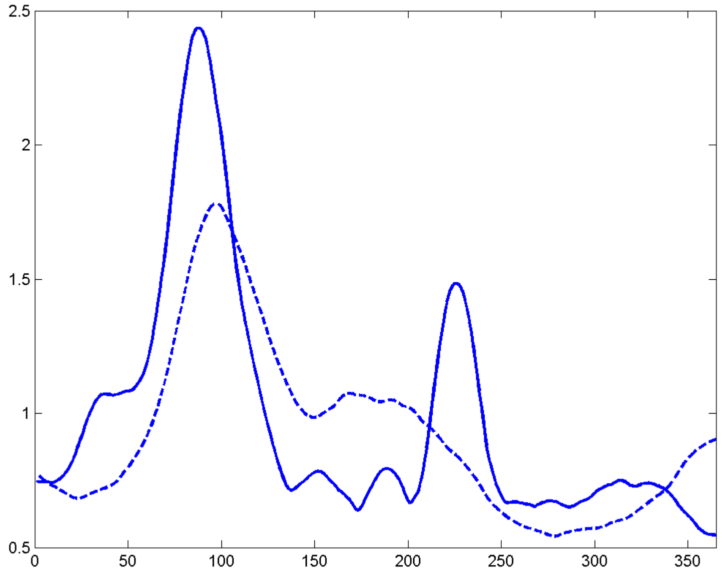
Aproximace Fourierovou řadou

(pro 3 frekvence)

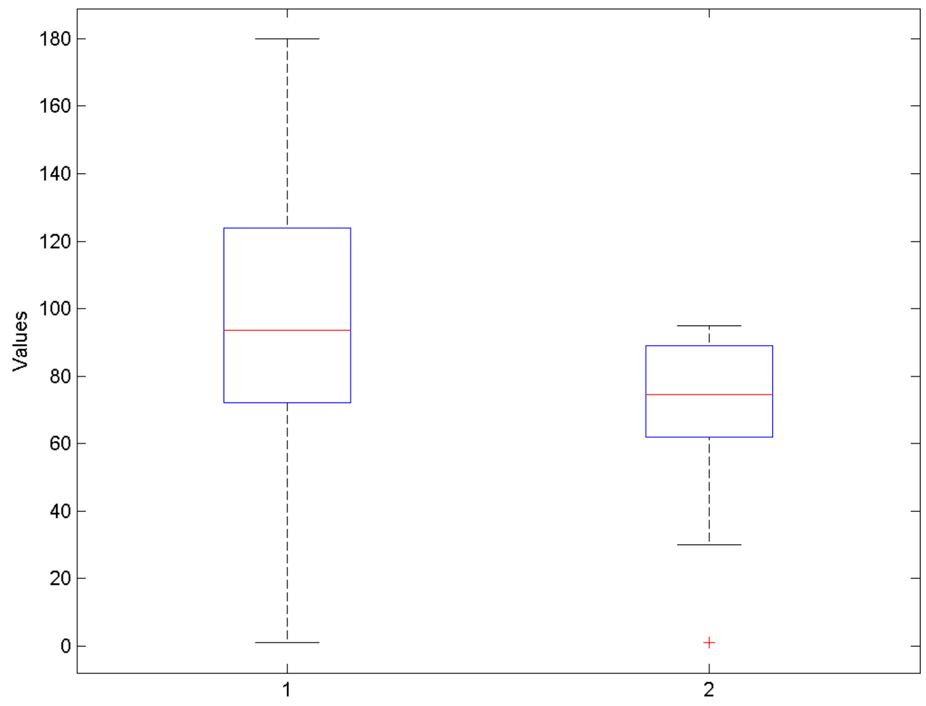


Neparametrický odhad

(vážené průměry s Epanechnikovým jádrem)



Blanice - rozdělení lokace maxim průtoků



Posun jarní kulminace

- Studovali jsme sezónní chování 18 průtokových řad českých řek
- Řady jsme rozdělili na 2 části a porovnávali „roční chod“ před r.1997 (čárkovaná) a po r. 1997 (plná)
- K popisu používali 2 modely

- Posun jarní kulminace - odhad bodový a intervalový

Model I

- Uvažujeme funkci
-
- $f(t) = \mu + A_1 \cos(2\pi t) + B_1 \sin(2\pi t) + A_2 \cos(4\pi t) + B_2 \sin(4\pi t) + A_3 \cos(6\pi t) + B_3 \sin(6\pi t),\ t \in \langle 0,1 \rangle$
- Předpokládáme, že data $\{(Y_{j1}, \dots, Y_{j365}), j = 1, \dots, m\}$ odpovídají modelu
-
- $Y_{j,i} = f\left(\frac{i}{365}\right) + e_{ji},$
- $i = 1, \dots, 365; j = 1, \dots, m\ (počet\ let),$

Model I

- Naším cílem je odhadnout a, tj. maximum funkce.
- Za použití **Δ-metody** $\sqrt{m}\,(\hat{a} - a)$ má asymptoticky normální rozdělení s asymptotickým rozptylem $\sigma_{\hat{a}}^2$, který může být spočítán za pomoci variančně-kovarianční matice $\sqrt{m}\left(\widehat{A}_1 - A_1, \dots, \widehat{B}_3 - B_3\right)$ a může být odhanut nahrazením $\widehat{A}_1$ za A_1,..., $\widehat{B}_3$ za B_3. Odhad $\sigma_{\hat{a}}^2$ jsme odznačili $\widehat{\sigma_{\hat{a}}^2}$.
- Asymptotický $(1-\alpha)100\%$ interval spolehlivosti pro $a_1 - a_2$ má tvar
-
- $\widehat{a_1} - \widehat{a_2} \mp u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_{a_1}^2}{m_1} + \frac{\sigma_{a_2}^2}{m_2}}$

Polohy maxim (ve dnech) s intervaly spolehlivosti

River	max1	max2	difference	±2*standard deviation of the difference
Běla	95,4527	85,5618	9,8909	19,7818
Blanice	101,3115	86,7186	14,5929	29,1858
Brodečka	87,4994	84,2744	3,225	6,45
Čeladenka	105,0468	90,5304	14,5164	29,0328
Doubrava	83,5152	73,4117	10,1035	20,207
Jizera	111,7988	92,2169	19,5819	39,1638
Kyjovka	90,171	85,7563	4,4147	8,8294
Morava	108,2486	93,5227	14,7259	29,4518
Mumlava	117,7065	102,774	14,9325	29,865
Orlice	100,2189	82,8374	17,3815	34,763
Otava	123,0629	106,0233	17,0396	34,0792
Porubka	84,0055	79,931	4,0744	8,1488
Bečva	94,4049	80,6521	13,7529	27,5058
Svratka	89,4916	75,9302	13,5614	27,1228
Úpa	121,0079	109,071	11,9369	23,8738
Vlára	73,2751	70,219	3,0561	6,1122
Vydra	125,1269	114,217	10,9099	21,8198
Zdobnice	97,8212	85,8595	11,9616	23,9232

Model II

- V jednovýběrovém problému předpokládáme, že průměrný roční chod může být pro jednotlivé roky odlišný. Uvažujeme soubor neznámých funkcí $\{r_j(t), -1/2 \leq t \leq 1\}$ tak, že $\max_{0 \leq t \leq \frac{1}{2}} r_j(t) = r_j(a)$, tj. všechny funkce dosahují svého maxima na intervalu $[0,1/2]$ ve stejném časovém bodě a.

- Předpokládáme, že naše data $\{Y_{ij}, i = 1, \dots, 365, j = 1, \dots, m\}$, kde m označuje počet roků, splňují
- $Y_{ij} = r_j\left(\frac{i}{365} + \theta_j\right) + e_{ij}$

Model II

- Postupovali jsme tak, že pro každé j jsme neparametricky odhadli maximum funkce $r_j(t+\ \theta_j)$, které nastává v časovém bodě $\{a+\theta_j\}$. Odhad $\widehat{a+\theta_j}$ je bodem, v němž je odhadovaná derivace $\widehat{r'_j}(t+\ \theta_j)$ rovna nule. Odhad funkce r'_jje získán jádrovým vyhlazením:
- $\widehat{r'_j}(x) = \frac{1}{b_n^2}\sum_{i=1}^{n}\int_{\frac{i-0,5}{n}}^{\frac{i+0,5}{n}} K'_0\left(\frac{t-x}{b_n}\right)dx\, Y_{ij},$
- b_n=0,05.

Model II

- Jádrová funkce byla zvolena:
-
- $K_0(x) = \frac{105}{96}\left(1 - x^2\right)^3,\ x \in \langle -1,1 \rangle$
- = 0, $x \notin \langle -1,1 \rangle$
-
- $K'_0(x) = \frac{105}{96} 3\left(1 - x^2\right)^2(-2x),\ x \in \langle -1,1 \rangle$
- = 0, $x \notin \langle -1,1 \rangle$

Model II

- $\widehat{a+\theta_j}$ přibližně normální rozdělení se střední hodnotou $a+\theta_j$ a přibližný rozptyl může být odhadnut takto:
-
- $\widehat{\sigma}_j^2 = \dfrac{\widehat{\sigma_e^2}\int_0^1 \left[K_0'(t)\right]^2 dt}{\left[r_j''(\widehat{a+\theta_j})\right]^2 n b_n^3}$
- Druhá derivace může být odhadnuta následovně
-
- $\widehat{r_j''}(x) = \frac{1}{b_n^3}\sum_{i=1}^{n}\int_{\frac{i-0,5}{n}}^{\frac{i+0,5}{n}} K_0''\left(\frac{t-x}{b_n}\right) dx\, Y_{ij}$

Model II

- Ve dvouvýběrovém problému předpokládáme, že budeme porovnávat dva výběry $\{Y_{ji}\}, \{Z_{ji}\}$ o rozsahu m_1 a m_2.Předpokládáme, že oba výběry splňují model, ale s odlišnými hodnotami a_1 a a_2. Cílem je odhadnout rozdíl $a_1 - a_2$. Předpokládáme, že může být odhadnut pomocí
-
- $\widehat{a_1} - \widehat{a_2} = \frac{1}{m_1} \sum_{j=1}^{m_1} \widehat{a_1 + \theta_j} - \frac{1}{m_2} \sum_{j'=1}^{m_2} \widehat{a_2 + \theta_{j'}}$
- nebo
- $\widehat{a_1} - \widehat{a_2} = \frac{\sum_{j=1}^{m_1} w_j \widehat{a_1 + \theta_j}}{\sum_{j=1}^{m_1} w_j} - \frac{\sum_{j'=1}^{m_2} w_{j'} \widehat{a_2 + \theta_{j'}}}{\sum_{j'=1}^{m_2} w_{j'}}$

Model II

- 90% interval spolehlivosti pro $a_1 - a_2$ můžeme získat metodou bootstrapu [4]. Připravili jsme 1000 bootstrapových výběrů ze $\{\widehat{a_1 + \theta_j}\}$ a stejný počet z $\{\widehat{a_2 + \theta_j}\}$ a spočítali některé charakteristiky (průměry, směrodatné odchylky, 5% horní a 5% dolní empirický kvantily). Tabulka 3 ukazuje výsledky pro některé stanice.

Průměry, směrodatné odchylky a směrodatné odchylky rozdílů spočítané metodou bootstrap.

-1997	mean	standard deviation	corrected sd	
Běla	90,1	4	4,03	
Brodečka	86,7	4,6	5,45	
Blanice	92,1	3,54	3,68	
Čeladenka	100,24	3,57	3,7	
1998-	mean	standard deviation	corrected sd	
Běla	73,2	8,92	9,02	
Brodečka	70,25	7,97	8,16	
Blanice	75	3,54	3,73	
Čeladenka	89,5	6,37	7,02	
		standard deviation of the difference	lower limit	upper limit
Běla		16,9	-36,65	2,9
Brodečka		16,42	-36,04	3,21
Blanice		17,14	-30,9	-3,29
Čeladenka		10,74	-26,6	5,12

Shrnutí

- 1. model: Δ-metoda (parametrický)
- 2. model: bootstrap (neparametrický)

Závěr

- U obou modelů se ukázalo, že intervaly jsou hodně velké díky velké variabilitě a tomu, že máme málo dat. Na druhé straně bodový odhad vyšel téměř vždy kladný.

Děkuji Vám za pozornost.