Detekce změny ročního chodu průtokových řad

Ing. Mgr. **Hana Horáková** [1], Prof. RNDr. **Daniela Jarušková**, CSc. [2], Doc. Ing. **Ladislav Satrapa**, CSc. [3]

[1] ČVUT v Praze, Fakulta stavební, Katedra matematiky, Thákurova 7, Praha 6, 166 29, horakovah@mat.fsv.cvut.cz
[2] ČVUT v Praze, Fakulta stavební, Katedra matematiky, Thákurova 7, Praha 6, 166 29, jarus@mat.fsv.cvut.cz
[3] ČVUT v Praze, Fakulta stavební, Katedra hydrotechniky, Thákurova 7, Praha 6, 166 29, satrapa@fsv.cvut.cz

Úvod

Základní data tvoří časová řada, tj. posloupnost vektorů $\boldsymbol{Z_1},\dots,\boldsymbol{Z_{2n}}$, kde $2n$ odpovídá počtu let, pro které máme k dispozici průměrné denní průtoky. Každý vektor $\boldsymbol{Z_i}$ má 365 souřadnic, kde j-tá souřadnice odpovídá průměrnému dennímu průtoku v j-tém kalendářním dni v roce. Základní otázkou je, zda se vektor středních hodnot během pozorování nezměnil, což odpovídá stacionaritě ročního chodu.

Časovou řadu jsme rozdělili na dvě stejně dlouhé části o délce n a použili jsme dvouvýběrový test pro shodnost dvou vektorů středních hodnot. Vzhledem k tomu, že vektory pozorování mají 365 souřadnic, je vhodné nejprve redukovat dimenzi úlohy.

Redukce dimenzionality

Redukce dimenzionality spočívá v nahrazení vektorů $\boldsymbol{Z_1},\dots,\boldsymbol{Z_{2n}}$ vektory $\boldsymbol{V_1},\dots,\boldsymbol{V_{2n}}$, kde $\boldsymbol{V_i}=\left((\boldsymbol{c}(\mathbf{1}))'\boldsymbol{Z_i},(\boldsymbol{c}(\mathbf{2}))'\boldsymbol{Z_i},\dots,(\boldsymbol{c}(\boldsymbol{k}))'\boldsymbol{Z_i}\right)$, $i=1,\dots,2n$, $\boldsymbol{c}(\mathbf{1}),\dots,\boldsymbol{c}(\boldsymbol{k})$ jsou vhodně zvolené vektory konstant. To znamená, že studujeme vektory, jejichž souřadnice jsou určitým způsobem zvolené lineární kombinace složek původních vektorů, přičemž se volí $k<<p$.

Redukce dimenzionality pomocí aproximace denního chodu Fourierovou řadou

Při této redukci volíme $\boldsymbol{c}(\mathbf{1})'=\left(\frac{1}{365},\frac{1}{365},\dots,\frac{1}{365}\right)$, …, $\boldsymbol{c}(\mathbf{2}\boldsymbol{l})'=\frac{2}{365}\left(\cos\frac{2l\pi}{365},\cos\frac{2l\pi\cdot 2}{365},\dots,\cos 2l\pi\right)$, $\boldsymbol{c}(\mathbf{2}\boldsymbol{l}+\mathbf{1})'=\frac{2}{365}\left(\sin\frac{2l\pi}{365},\sin\frac{2l\pi\cdot 2}{365},\dots,\sin 2l\pi\right)$. To znamená, že nahradíme vektor $\boldsymbol{Z_i}=\left(Z_{i1},\dots,Z_{i365}\right)'$ vektorem $\boldsymbol{V_i}=\left(\hat{\mu}_i,\hat{A}_{i1},\hat{B}_{i1},\dots,\hat{A}_{il},\hat{B}_{il}\right)'$, kde

$$\hat{A}_{ij}=\frac{2}{365}\sum_{t=1}^{365}\left(\cos\frac{2j\pi t}{365}\cdot Z_{it}\right) \text{ a } \hat{B}_{ij}=\frac{2}{365}\sum_{t=1}^{365}\left(\sin\frac{2j\pi t}{365}\cdot Z_{it}\right),\ j=1,\dots,l.$$

Je všeobecně známo, že $\{\hat{A}_{ij}\}$, resp. $\{\hat{B}_{ij}\}$, jsou odhady metodou nejmenších čtverců v modelu

$$Z_{it}=\mu_i+\sum_{j=1}^{l}\left(A_{ij}\cos\frac{2\pi jt}{365}+B_{ij}\sin\frac{2\pi jt}{365}\right)+e_{it},\ t=1,\dots,365,\ i=1,\dots,2n.$$

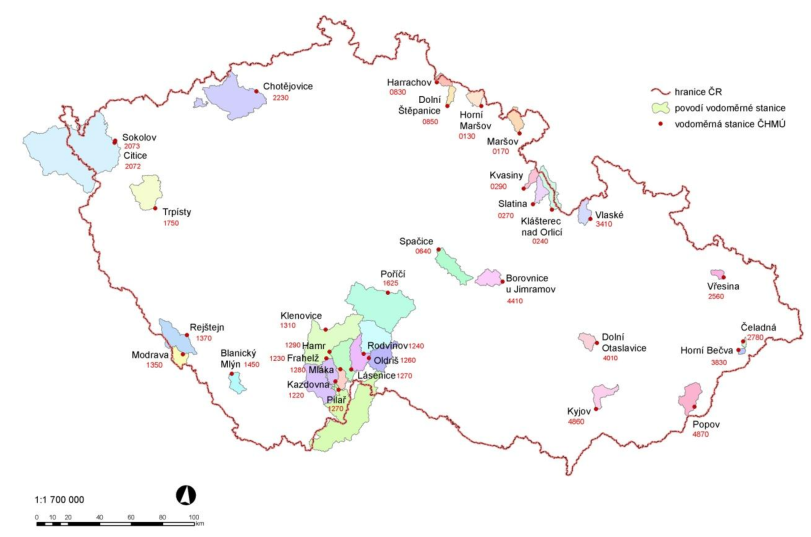
Metodu můžeme tedy interpretovat tak, že nahradíme řadu denních hodnot v jednom roce její aproximací pomocí konečné Fourierovy řady a testujeme, zda se shodují střední hodnoty příslušných koeficientů, tj. $\mu_1=\dots=\mu_{2n}$, $A_{11}=\dots=A_{2n,1}$, $B_{11}=\dots=B_{2n,1}$, $A_{1l}=\dots=A_{2n,l}$, $B_{1l}=\dots=B_{2n,l}$. Jestliže nás nezajímá změna v ročním průměru, můžeme pak testovat pouze zda $A_{11}=\dots=A_{2n,1}, B_{1l}=\dots=B_{2n,l}$.

Můžeme se ptát, proč tato metoda může odhalit změnu ve střední hodnotě. Je zřejmé, že metoda bude dobře fungovat, pokud střední hodnota $E\boldsymbol{Z_i}$, $i=1,\dots,2n$ bude mít komponenty $EZ_{it}=\mu_i+\sum_{j=1}^{l}\left(A_{ij}\cos\frac{2\pi jt}{365}+B_{ij}\sin\frac{2\pi jt}{365}\right)$, $t=1,\dots,365$, nebo alespoň se budou výrazu na pravé straně velmi blížit. Jinými slovy můžeme očekávat, že metoda bude dobře fungovat tehdy, když pro libovolné $i=1,\dots,2n$ bude střední hodnota periodická hladká funkce. Záleží potom na daném problému, jak volit počet uvažovaných největších Fourierových frekvencí. V našem případě se ukázalo jako nejvhodnější volit 3 frekvence.

Redukce dimenzionality pomocí metody hlavních komponent

Při této redukci nahradíme nejprve data $\boldsymbol{Z_1}=\left(Z_{1,1},\dots,Z_{1,365}\right)'$,…, $\boldsymbol{Z_{2n}}=\left(Z_{2n,1},\dots,Z_{2n,365}\right)'$ standardizovanými hodnotami $\boldsymbol{Z_1^S}=\left(Z_{11}-\bar{Z}_{\cdot 1},\dots,Z_{1,365}-\bar{Z}_{\cdot 365}\right)'$,…, $\boldsymbol{Z_{2n}^S}=\left(Z_{2n,1}-\bar{Z}_{\cdot 1},\ \dots,Z_{2n,365}-\bar{Z}_{\cdot 365}\right)'$ kde $\bar{Z}_{\cdot l}=\frac{1}{2n}\sum_{i=1}^{2n}Z_{il}$. Vektory $\boldsymbol{c}(1),\dots,\boldsymbol{c}(\mathrm{k})$ volíme následovně: $\boldsymbol{c}(1)=\hat{\boldsymbol{u}}_1$, …, $\boldsymbol{c}(\mathrm{k})=\hat{\boldsymbol{u}}_{\mathbf{k}}$, kde $\hat{\boldsymbol{u}}_1,\dots,\hat{\boldsymbol{u}}_{\mathbf{k}}$ jsou vlastní vektory matice $\hat{\Sigma}_{\boldsymbol{A}}$ odpovídající k největším vlastním číslům $\hat{\lambda}_1\ \geq\dots\geq\dots\geq\ \hat{\lambda}_k$. To znamená, že nahradíme vektor $\boldsymbol{Z_i^S}=\left(Z_{i1}^S,\dots,Z_{i365}^S\right)'$ vektorem $\boldsymbol{V}_i=\left(\hat{\boldsymbol{u}}_1'\boldsymbol{Z}_i^S,\dots,\hat{\boldsymbol{u}}_k'\boldsymbol{Z}_i^S\right)'$. Počet hlavních komponent se volí subjektivně, v našem případě jsme volili 7-12 vlastních čísel.

Můžeme se ptát, proč metoda hlavních komponent pro detekci změny docela dobře funguje. Jestliže existuje relativně velká změna v průměru některých lineárních kombinací denních měření, pak její výběrový rozptyl je velký. Proto má tato změna velkou šanci být zahrnuta do souboru k lineárních kombinací s největším rozptylem.



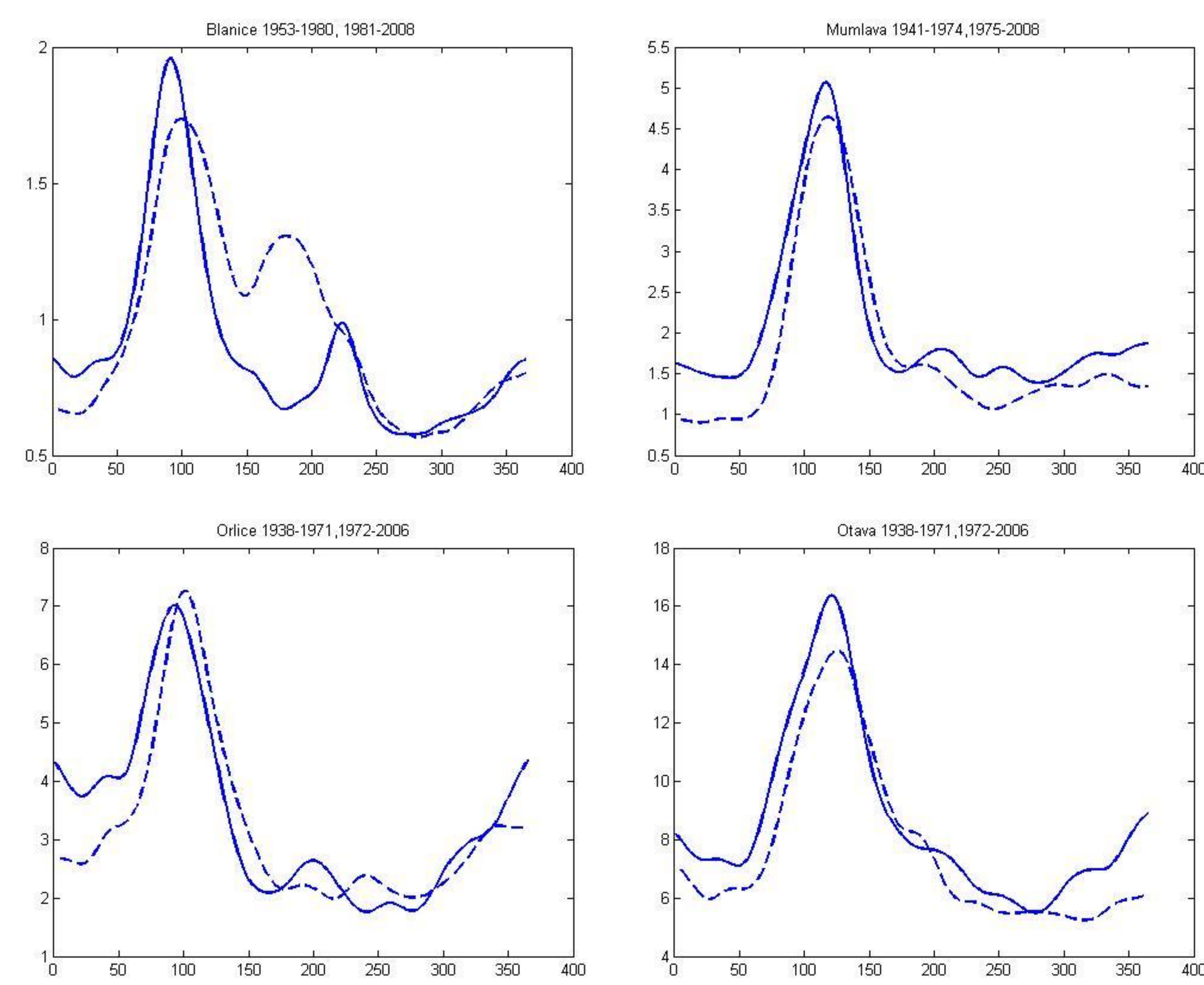
Mapa stanic

	χ2 - stat. (p-hodnota)
BELA	5,9242 (0.4317)
BLANICE	19,6011 (0.0033)
BRODECKA	2,4446 (0.8746)
CELADENKA	2,9026 (0.8210)
DOUBRAVA	4,5397 (0.6040)
JIZERA	6,7926 (0.3405)
KYJOVKA	3,2428 (0.3405)
MORAVA	6,0380 (0.4189)
MUMLAVA	10,7054 (0.0979)
ORLICE	11,9143 (0.0639)
OTAVA	11,1898 (0.0827)
PORUBKA	3,0853 (0.7981)
ROZNOV.BECVA	3,6236 (0.7275)
SVRATKA	4,4339 (0.6182)
UPA	5,4549 (0.4869)
VLARA	6,5175 (0.3678)
VYDRA	13,4804 (0.0360)
ZDOBNICE	5.1780 (0.5212)

***Tab.1** Výsledné hodnoty při použití metody Fourierových frekvencí*

			χ2 - statistika (p-hodnota)			
počet vlastních čísel	**7**	**8**	**9**	**10**	**11**	**12**
BELA	4.1153 (0.7664)	4.2803 (0.8310)	5.9319 (0.7467)	8.1585 (0.6134)	14.5432 (0.2044)	14.5762 (0.2654)
BLANICE	14.1824 (0.0480)	17.4900 (0.0254)	17.9354 (0.0359)	18.2342 (0.0511)	18.7648 (0.0654)	20.4121 (0.0597)
BRODECKA	4.2745 (0.7477)	4.8690 (0.7715)	6.0900 (0.7309)	6.0971 (0.8070)	6.1486 (0.8633)	6.5323 (0.8869)
CELADENKA	6.0421 (0.5348)	7.1553 (0.5200)	7.9099 (0.5433)	8.1119 (0.6179)	8.9659 (0.6250)	14.4888 (0.2706)
DOUBRAVA	5.6228 (0.5844)	7.0848 (0.5275)	8.2124 (0.5129)	8.2159 (0.6078)	8.3768 (0.6792)	10.1440 (0.6033)
JIZERA	7.6297 (0.3664)	10.4386 (0.2356)	10.4501 (0.3153)	12.1304 (0.2764)	12.3221 (0.3399)	12.3369 (0.4190)
KYJOVKA	5.0108 (0.6586)	5.3286 (0.7219)	6.8798 (0.6496)	8.0830 (0.6207)	8.3868 (0.6783)	8.6413 (0.7332)
MORAVA	10.0564 (0.1854)	11.7999 (0.1604)	11.8782 (0.2203)	16.0745 (0.0975)	19.7325 (0.0491)	20.3545 (0.0607)
MUMLAVA	16.6833 (0.0196)	17.1027 (0.0291)	17.4546 (0.0421)	17.9703 (0.0555)	18.0041 (0.0815)	18.8567 (0.0920)
ORLICE	7.7327 (0.3567)	7.7371 (0.4596)	7.7464 (0.5599)	8.4871 (0.5814)	16.2054 (0.1337)	16.7005 (0.1612)
OTAVA	10.7944 (0.1478)	11.0788 (0.1973)	11.0893 (0.2696)	11.1440 (0.3464)	11.7296 (0.3843)	11.7335 (0.4673)
PORUBKA	6.7359 (0.4569)	10.6075 (0.2249)	10.7152 (0.2957)	11.3431 (0.3314)	11.6329 (0.3919)	12.0769 (0.4395)
ROZNOV.BECVA	4.9917 (0.6610)	6.5629 (0.5844)	10.9195 (0.2813)	11.7738 (0.3005)	12.2054 (0.3484)	13.8705 (0.3091)
SVRATKA	3.0177 (0.8834)	3.2377 (0.9186)	3.6206 (0.9346)	4.4708 (0.9236)	4.4956 (0.9531)	5.9885 (0.9167)
UPA	6.9100 (0.4383)	7.0184 (0.5347)	7.0276 (0.6342)	9.4851 (0.4868)	9.9573 (0.5342)	10.7439 (0.5510)
VLARA	9.1310 (0.2434)	10.3181 (0.2434)	10.5452 (0.3082)	11.4597 (0.3228)	11.6880 (0.3875)	12.4812 (0.4078)
VYDRA	8.3236 (0.3049)	11.2454 (0.1882)	13.6099 (0.1369)	13.8708 (0.1790)	14.7466 (0.1944)	16.0820 (0.1875)
ZDOBNICE	5.1371 (0.6432)	6.8762 (0.5500)	7.0328 (0.6337)	10.9401 (0.3622)	13.1479 (0.2838)	17.4549 (0.1333)

***Tab. 2** Výsledné hodnoty při použití metody hlavních komponent*



Vyhlazené roční chody u vybraných stanic
(první období čárkovaná čára, druhé období plná čára)

Vyhodnocení

Uvedené metody jsme aplikovali na 18 stanic (viz mapa), v nichž byly měřeny denní průtoky po dobu 50-90 let. Jedná se o následující stanice: Běla - Skuhrov (1940-2007), Blanice – Blanický Mlýn (1952-2008), Brodečka - Otaslavice (1940-2007), Čeladenka – Šance (1940-2007), Doubrava – Spačice (1952-2007), Jizera – Štěpanice (1922-2008), Kyjovka – Kyjov (1950-2008), Morava – Vlaské (1949-2008), Mumlava – Harrachov (1950-2008), Orlice – Klášterec (1938-2006), Otava – Rejštejn (1910-1920, 1930-1937, 1947-2007), Porubka – Vřesina (1952-2008), Rožnovská Bečva – Horní Bečva (1954-2008), Svratka – Borovnice (1954-2008), Úpa – Horní Maršov (1954-2008), Vlára – Popov (1955-2008), Vydra – Morava (1930-1940, 1948-2007), Zdobnice – Pěčín (1944-2007).

Na hladině významnosti α=0,05 vyšly při použití metody Fourierových frekvencí (pro 3 frekvence) jako významné stanice Blanice a Vydra, při volbě α=0,10 dále stanice Mumlava, Orlice a Otava. Při použití metody hlavních komponent jsme volili počet vlastních čísel 7-12. Při volbě hladiny významnosti α=0,05 lze za statisticky významné považovat hodnoty stanic Blanice (pro 7-9 vlastních čísel), Mumlava (pro 7-9 vlastních čísel) a Morava (pro 11 vlastních čísel), na hladině významnosti α=0,10 se výsledky významných hodnot rozšíří ještě o stanici Blanice (pro 10-12 vlastních čísel), pro Mumlavu (pro 10-12 vlastních čísel) a pro stanici Morava (pro 10 a 12 vlastních čísel).

Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT v Praze č. **SGS12/004/OHK1/1T/11** „Stochastické vlastnosti denních průtokových dat“.
Data pocházejí z projektu NAZV QH71201 „Spolehlivost a bezpečnost vodohospodářských děl v měnících se klimatických podmínkách“.