



**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

---

**Fakulta stavební  
Katedra matematiky**

**Stochastické vlastnosti průtokových řad a metody ke zjišťování  
jejich nestacionarity**

**Stochastic properties of discharge series and methods for detecting  
their nonstationarity**

**DISERTAČNÍ PRÁCE**

**Ing. Mgr. Hana Horáková**

Doktorský studijní program: Stavební inženýrství

Studijní obor: Matematika ve stavebním inženýrství

Školitel: prof. RNDr. Daniela Jarušková, CSc.

**Praha, 2017**



## **PODĚKOVÁNÍ**

Děkuji prof. RNDr. Daniele Jaruškové, CSc., za odborné vedení práce, trpělivost a vytváření přátelské atmosféry během našich společných konzultací. Děkuji také kolegům z katedry, zvláště Ing. Stanislavu Olivíkovi, Ph.D., za pomoc s formálními náležitostmi práce. Za podporu děkuji i své rodině a přátelům.



## **PROHLÁŠENÍ**

Ing. Mgr. Hana Horáková

Stochastické vlastnosti průtokových řad a metody ke zjišťování jejich nestacionarity

Prohlašuji, že jsem uvedenou disertační práci vypracovala samostatně pod vedením školitelky prof. RNDr. Daniely Jaruškové, CSc. Použitou literaturu a další materiály uvádím v seznamu použité literatury.

V Praze dne 30. 8. 2017

---

Podpis



## **ABSTRAKT:**

Cílem dizertační práce je navrhnout statistické postupy pro detekování změny střední hodnoty ročního chodu průtokových řad. Pod pojmem "roční chod" se zde rozumí vektor o 365 souřadnicích, jehož komponenty tvoří denní průměrné průtoky. Rozhodnutí, zda můžeme sezónní chování řady považovat za stacionární, je v rámci matematické statistiky nejčastěji založeno na testování hypotéz. Hlavním cílem první části práce je navrhnout statistické testy, které by byly schopny změnu ve střední hodnotě ročního chodu odhalit.

Pokud se detekuje změna v sezónním chování, zajímá nás, k jakému typu změny došlo. To znamená odhadnout střední hodnotu vektoru průměrných denních toků odpovídající jednomu kalendářnímu roku před i po této změně.

Vzhledem k tomu, že jedním z hlavních rysů detekované změny je posun jarní kulminace, je cílem práce také navrhnout statistické metody k odhadu doby, kdy k jarní kulminaci dochází, tj. odhadnout argument maxima vektoru středních hodnot ročního chodu.

Velmi důležité je nejen navrhnout statistické metody, ale také je použít pro zpracování reálných dat. V našem případě jsme zpracovávali údaje z vodoměrných stanic na 18 malých českých tocích, které byly vybrány tak, aby průtoky byly co nejméně ovlivněny lidskou činností. Pro problém detekce více změn jsme navrženou metodu použili na denní průměrné průtoky Dunaje v Bratislavě. Aplikace našich postupů na reálná data pak ukazují, které z navržených metod jsou vhodné a které naopak jsou méně vhodné pro detekci studovaných nestacionarit.

## **KLÍČOVÁ SLOVA:**

statistické modely, časové řady, stacionarita, analýza bodu změny, sezónnost





## **ABSTRACT:**

The aim of the dissertation thesis is to propose statistical methods for analyzing stationarity of discharge series annual cycle. The analyzed data are daily mean values in several years and under the name "annual cycle" we understand a vector of 365 coordinates whose components are daily mean discharge values. The decision on stationarity of seasonal behavior is based on hypotheses testing and the aim of the first part of the thesis is to propose statistical tests suitable for detection of changes in the mean vector of annual cycles. As the time period, when the discharge time series is measured, is relatively short we propose methods for reducing a dimension of the problem. In addition to the two-sample tests we suggest application of change-point methods. If any change in seasonal behavior is detected, we estimate the mean vector of annual cycle before and after the change using non-parametric as well as parametric approach.

Since one of the main features of the observed changes is a shift of the spring time, the aim of the thesis is also to propose statistical methods for estimating the time when the peak culmination occurs, i.e., estimate the argument where the vector of the mean values of the annual cycle attains its maximum.

We do not only propose statistical methods for studying the stationarity of seasonal behavior of time series but we also apply them for a real data analysis. In our case, we analyze the data from 18 water gauge stations of small Czech rivers selected in a way that the discharge is not affected by human activity. As we also suggest a method for detecting more than one change, we apply it to the discharge series of the Danube River in Bratislava.

Application of our methods to the real data sets shows which of the proposed methods are appropriate, or on the other hand, which are less suitable for a detection of the studied nonstationary behavior.

## **KEYWORDS:**

statistical models, time series, stationarity, change point analysis, seasonality



# Obsah

<b>1</b>	<b>Úvod</b>	<b>13</b>
<b>2</b>	<b>Popis dat a popis použitých výpočetních prostředků</b>	<b>17</b>
<b>3</b>	<b>Jednoduché testy na zjištění změny v řadě průměrných ročních průtoků</b>	<b>21</b>
3.1	Dvouvýběrové testy pro shodnost parametrů výběrů normálního rozdělení . . .	22
3.2	Existence lineárního trendu ve střední hodnotě časové řady . . . . .	25
<b>4</b>	<b>Detekce změny průměrného ročního chodu - test pro známý bod změny</b>	<b>27</b>
4.1	Test změny ve střední hodnotě - dvouvýběrový problém . . . . .	27
4.2	Použití testu pro shodu středních hodnot k detekci změny ročního chodu . . . .	28
<b>5</b>	<b>Redukce dimenze</b>	<b>31</b>
5.1	Použití dvouvýběrových testů pro průtokové řady . . . . .	32
5.2	Nahrazení denních průměrných průtoků průměrnými měsíčními průtoky . . . .	33
5.3	Redukce dimenze pomocí aproximace Fourierovou řadou . . . . .	34
5.4	Redukce dimenze metodou hlavních komponent . . . . .	37
<b>6</b>	<b>Neparametrický odhad střední hodnoty ročního chodu před a po změně a odhad posunu mezi dobami, kde střední roční chody nabývají svého maxima</b>	<b>39</b>
6.1	Jádrový odhad regresní funkce . . . . .	39
6.2	Odhad střední hodnoty ročního chodu . . . . .	40
6.3	Neparametrický odhad středních hodnot ročního chodu před a po roce 1997 . .	41
6.4	Neparametrický odhad polohy maxima středního ročního chodu . . . . .	42
6.5	Neparametrický odhad posunu mezi dobami, kde dochází k maximu středního ročního chodu . . . . .	47
6.6	Model s odlišným středním chodem pro různé roky . . . . .	49

<b>7</b>	<b>Odhad střední hodnoty ročního chodu před a po změně pomocí periodické funkce a odhad posunu mezi dobami, kde roční chody nabývají svého maxima</b>	<b>53</b>
7.1	Odhad střední hodnoty ročního chodu a odhad argumentu maxima střední hodnoty ročního chodu pomocí periodické funkce . . . . .	53
7.2	Odhad posunu mezi dobami, kde dochází k maximu ročního chodu, v regresním modelu s periodickou regresní funkcí . . . . .	55
<b>8</b>	<b>Testování existence bodu změny použitím metod analýzy bodu změny</b>	<b>59</b>
8.1	Testová statistika maximálního typu . . . . .	59
8.2	Test shody střední hodnoty posloupnosti náhodných vektorů . . . . .	61
<b>9</b>	<b>Detekce více změn</b>	<b>65</b>
9.1	Detekce $d$ bodů změny v případě, kdy je známo, v kterých časových okamžicích by se řada mohla měnit . . . . .	66
9.2	Detekce $d$ bodů změny v případě, kdy není známo, v kterých časových okamžicích by se řada mohla měnit . . . . .	67
9.3	Dvě změny - Dunaj . . . . .	67
<b>10</b>	<b>Odhad společného bodu změny</b>	<b>71</b>
10.1	Metoda pro odhad společného bodu změny . . . . .	71
10.2	Odhad společného bodu změny pro malé české toky . . . . .	72
<b>11</b>	<b>Závěr</b>	<b>75</b>
11.1	Poděkování . . . . .	76
11.2	Reference . . . . .	77

# Kapitola 1

## Úvod

"Globální oteplování" je jev, o jehož existenci pochybuje jen málo klimatologů a hydrologů. V National Geographic z dubna 2017 autoři předkládají jasná fakta. Je prokazatelné, že na světě dochází k oteplování. Vedra v roce 2016 překonala historický rekord v roce 2015, který pro změnu překonal rekord z roku 2014. Průměrné hodnoty teploty na zemském povrchu z roku 2016, které byly sestaveny z měření provedených na tisících meteorologických stanicích, bójích a lodích, byly o 1,69 °F (tj. o 0,94 °C) vyšší než průměrné hodnoty za 20. století. Družice zkoumající atmosféru rovněž zaznamenaly jasný trend oteplování. Navíc El Niño přidává k rekordu z minulého roku dočasné uvolnění tepla z Pacifiku. Žádná přírodní příčina (sluneční cyklus, sopečné erupce apod.) však nevysvětluje trend oteplování, který probíhá minimálně polovinu století. Naproti tomu antropogenní skleníkové plyny, které se hromadí v atmosféře, způsobují ohřívání zemského povrchu. Přestože mezi vědci nalezneme i odpůrce tohoto tvrzení, dá se zjednodušeně říct, že 9 z 10 vědců na základě svých výzkumů existenci globálního oteplování potvrzuje. O skleníkovém efektu víme již od 19. století. V roce 1896 švédský fyzik Svante Arrhenius předpověděl, že oxid uhličitý ze spalovaného uhlí by mohl oteplovat planetu. Jak je tedy tento jev odborníky vysvětlován a co vše způsobuje?

Globální oteplování spočívá ve zvyšování průměrné teploty na zemském povrchu v důsledku zvýšení koncentrací skleníkových plynů v atmosféře. Skleníkové plyny (oxid uhličitý, metan, oxid dusný, freony) propouštějí část infračerveného slunečního záření (10 až 80 %, v závislosti na oblačnosti), které dopadá na zemský povrch a ohřívá ho. V noci naopak brání úniku tohoto tepla zpátky do vesmíru. Zvyšováním koncentrací skleníkových plynů v atmosféře uniká méně tepla do vesmíru a více ho zůstává v atmosféře. V porovnání s rokem 2000 je dnes již téměř všemi vědci tento fenomén uznáván. Objektivním měřením bylo dokázáno, že v atmosféře roste obsah některých skleníkových plynů a průměrná teplota atmosféry stoupá. Tím dochází k tání ledovců, stoupá hladina oceánů, mění se proudění v oceánech, což ovliv-

ňuje celkové klima i počasí, viz Wetherald (2010). Diskuze se nyní vedou především o míře lidského zavinění na tomto stavu a možnostem, jak zamezit dalšímu negativnímu vývoji. Tento jev již zasáhl i politické sféry. Emise antropogenních skleníkových plynů jsou kontrolovány tzv. Kjótským protokolem (podepsán v r. 1997, účinnost od r. 2005) a Rámcovou úmluvou (podepsána 1992, účinnost od r. 1994). Montrealský protokol (podepsaný 1987, účinnost od r. 1989) o látkách poškozujících ozónovou vrstvu kontroluje použití halonů a freonů. Neméně významná je také Pařížská dohoda o snižování emisí oxidu uhličitého z roku 2016.

Podle expertů z MFF UK se globální oteplování nevyhýbá ani České republice. Do roku 2040 má průměrná teplota vzduchu narůst minimálně o 1°C a za pětadesát let má potom stoupnout až o 2,5 stupně. Čechy tak v budoucnosti čekají teplejší zimy, ubude mrazových a ledových dní. Bude také až o dvě třetiny méně sněhu v horských oblastech (Šumava, Krkonoše, Jeseníky atd.). Naopak v létě oteplení přinese zvýšené množství tropických dnů a nocí. V České republice již zaznamenali odborníci některé z klimatických změn. Obvykle bývají na jaře řeky po tání sněhu plné vody, v roce 2014 se průtoky některých řek na jaře pohybovaly na historických minimech (noř. povodí horního Labe, horní Vltavy, Berounky, Ohře a Opavy). Kvůli hrozbě sucha vznikla dokonce i expertní skupina, která s Výzkumným ústavem vodohospodářským (VÚV TGM, v.v.i.) připravuje strategii na udržení vody v krajině.

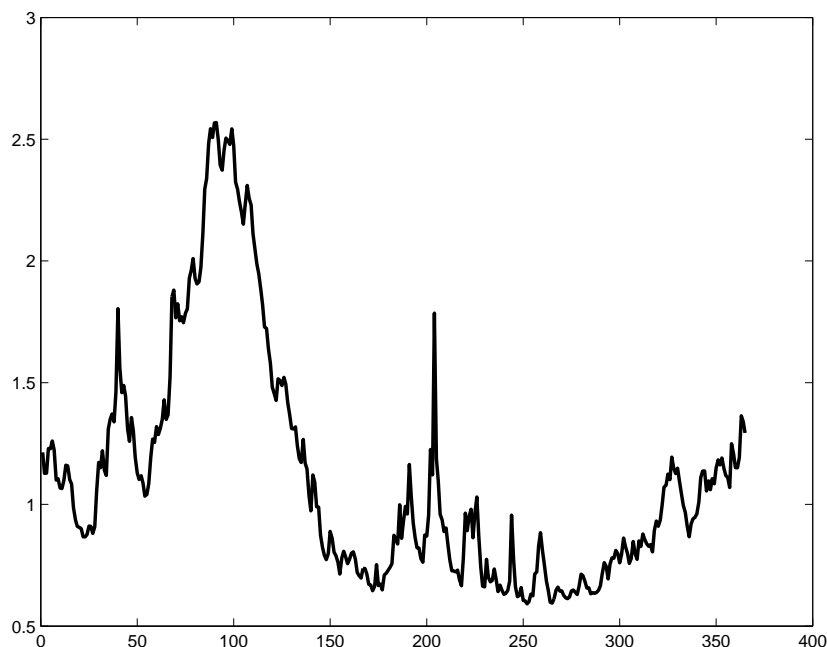
Velké množství vědců ve světě i u nás se zabývá predikcí chování průtokových řad pomocí klimatologických a hydrologických modelů, viz např. Nohara et al. (2006), Hanel et al. (2013), Hanel et al. (2014). Méně z nich se snaží pomocí statistických metod detekovat probíhající změny v již naměřených řadách. Sem patří např. články Milly a Dunne (2005), Dai et al. (2009), Irmak et al. (2012), Pekarova et al. (2003). Nejčastěji se studuje existence trendů buď v ročních nebo měsíčních průměrných průtocích, viz např. Kliment et al. (2011), Ledvinka a Lamačová (2015).

Předkládaná práce se zabývá metodami pro detekci změn v sezónním chování řad a dále odhady, k jaké změně (změnám) časových řad došlo. Detekce změn sezónního chování byla studována např. v člancích Stewart (2005), Ye et al. (2004), Villarini et al. (2011).

Práce je zaměřena především na studium průtokových řad. Statistické postupy popsané v práci by se samozřejmě daly použít i pro jiné časové řady, nicméně my jsme je aplikovali výhradně na průtokové řady, přičemž se jednalo o časové řady průměrných denních průtoků měřené v určité stanici během několika po sobě jdoucích let.

Vzhledem ke střídání ročních období má vektor průměrných denních průtoků řek v České i Slovenské republice sezónní charakter. Vektor průměrných denních průtoků určuje roční chod. Nejvodnatějším obdobím je jaro, kdy taje sníh. V létě se často vyskytuje období se silnými

srážkami a to také způsobuje poměrně vysoké odtoky. Obrázek 1.1 ilustruje sezónní chování pomocí průběhu průměrů (za celé sledované období) průměrných denních průtoků řeky Bělé. Stejně jako i v dalších obrázcích na ose  $x$  jsou vyneseny kalendářní dny v roce, na ose  $y$  průtoky v  $m^3/s$ .



Obrázek 1.1: Průměrný roční chod řeky Bělé.

Rozhodování, zda studovaná řada nemění v čase svoje stochastické chování, je v rámci matematické statistiky obvykle založeno na testování hypotéz, kde nulová hypotéza tvrdí, že je řada stacionární. Stochastický model stacionárního chování se může volit parametrický nebo neparametrický. My jsme obvykle používali parametrické modely, přičemž jsme předpokládali, že vektory průměrných denních průtoků v jednotlivých letech jsou nezávislé vektory. To je samozřejmě zjednodušující předpoklad, protože denní průtoky na konci kalendářního roku a denní průtoky z počátku následujícího kalendářního roku jsou zcela jistě závislé. Domníváme se ale, že tato závislost není natolik silná, aby ovlivnila výsledky statistické analýzy.

Po stanovení nulové hypotézy je třeba určit alternativní hypotézu. Alternativní hypotéza v detekci bodu změny tvrdí, že stacionarita řady byla porušena určitým daným způsobem. V práci se věnujeme především změně střední hodnoty. Tato změna může být různého typu, a proto existuje celá řada různých testů, přičemž každý může být vhodný pro jiný typ změny.

Odborníci také rozlišují mezi změnou náhlou a změnou postupnou. Při použití modelů s náhlou změnou se předpokládá, že ke změně došlo buď v jednom časovém okamžiku, nebo častěji, v nějakém krátkém (vzhledem k celkové době měření) časovém období. V modelech

pro postupnou změnu se předpokládá, že se začal v určitém okamžiku v časové řadě projevovat nějaký trend. V naší práci se zabýváme především náhlými změnami. Poznamenejme však, že postupy pro detekci náhlé změny často detekují i postupné změny a naopak. Samozřejmě v případě, že typ skutečné změny neodpovídá změně předpokládané v alternativní hypotéze, postup ztrácí svou sílu, tj. je schopen detekovat jen "výrazně větší změny".

Pokud máme alespoň přibližnou představu, v kterém časovém období ke změně mohlo dojít, můžeme pro zamítnutí nulové hypotézy použít testové statistiky známé z dvouvýběrových testů. Je zřejmé, že použitá dvouvýběrová testová statistika zamítá nulovou hypotézu i tehdy, když se skutečný bod změny neshoduje přesně s očekávaným bodem změny. V takovém případě ovšem opět použitý postup ztrácí svou sílu. Jestliže nevíme, kde ke změně došlo, je lepší použít postupy, které byly navrženy v odvětví matematické statistiky, známé pod názvem analýza bodu změny, anglicky "change point analysis". Testové statistiky, které se pro detekci bodu změny používají v rámci analýzy bodu změny, odpovídají maximální hodnotě dvouvýběrových testových statistik. Jiní autoři dávají přednost testovým statistikám, které odpovídají sumám jednotlivých dvouvýběrových testových statistik. V naší práci se zaměřujeme především na testové statistiky maximálního typu. Ze zahraničních prací, které se zabývají analýzou bodu změny, připomeňme alespoň Jandhyala et al. (2012).

V řadě se samozřejmě může vyskytovat i více než jedna změna. Více změn je možno detekovat jen tehdy, pokud je řada dostatečně dlouhá a mezi jednotlivými změnami jsou dostatečně velké časové odstupy. Při detekci více změn se nejčastěji postupuje sekvenčně. To znamená, že odhalí-li se jedna změna, řada se rozdělí na dvě části, tj. před změnou a po ní, a pak se analyzují obě části odděleně. Najde-li se alespoň v jedné z těchto částí změna, řada se opět rozdělí atd. Antoch a Jarušková (2013) navrhli jiný postup, pomocí kterého se detekuje více změn současně.

Pokud v řadě byla odhalena změna v jejím sezónním chování, je důležité charakterizovat, v čem tato změna spočívá. Z tohoto důvodu používáme neparametrické i parametrické metody pro odhad průměrného ročního chodu před a po změně. Vzhledem k tomu, že častým rysem nestacionárního chování řad je posun jarních kulminačních průtoků směrem k počátku kalendářního roku, zabýváme se v naší práci i postupy, jak bodově a intervalově odhadnout, kde střední roční chod nabývá svého maxima a také posun poloh maxim pro dva časové úseky.



## Kapitola 2

# Popis dat a popis použitých výpočetních prostředků

V naší práci jsme analyzovali průměrné denní průtoky 18 malých českých řek. Profily, v kterých bylo měření prováděno, byly vybrány tak, aby průtoky nebyly příliš ovlivněny lidskou činností. To znamená, že se nad měřícími stanicemi už nenacházely žádné přehrady či nádrže. V takovém případě mohou být změny zjištěné v jejich stochastickém chování nejspíš připsány právě klimatickým změnám. Před tím, než jsme řady začali studovat my, byla data analyzována z hydrologického hlediska v rámci grantu KLISPO podporovaného Národní agenturou pro výzkum v zemědělství Ministerstva zemědělství ČR (2007-2011), viz Satrapa a Zukal (2012).

Tabulka 2.1 udává názvy toků i stanic a také časové období, pro které jsme data analyzovali. V prvním sloupci je uveden název řeky spolu s názvem stanice. Druhý a třetí sloupec ukazují počáteční a konečné roky pozorování, čtvrtý sloupec udává roky, kdy údaje chyběly. Pátý sloupec udává počet let, po který byly údaje analyzovány. Délka studovaných řad se pohybuje mezi 50 až 90 lety. Obrázek 2.1 ukazuje polohu všech 18 stanic.

Základními zpracovávanými údaji byly denní průměrné průtoky měřené v  $m^3/s$ . V naší práci jsme pracovali s kalendářními roky, to znamená, že každý rok začal 1. ledna a skončil 31. prosince. Vynechali jsme data odpovídající 29. únorům, abychom měli 365 hodnot pro každý rok. Zpracovávaná data tak měla formu matice, s 365 sloupci a s počtem řádků, který odpovídal počtu let, v kterých bylo měření prováděno.

TOK - STANICE	poč. rok	posl. rok	chybějící roky	počet roků
<i>BĚLÁ - KVASINY</i>	1941	2007		67
<i>BLANICE - BL. MLÝN</i>	1953	2008		56
<i>BRODEČKA - OTASLAVICE</i>	1941	2007		67
<i>ČELADENKA - ČELADNÁ</i>	1953	2007		55
<i>DOUBRAVA - SPAČICE</i>	1952	2006		55
<i>JIZERKA - DOLNÍ ŠTĚPANICE</i>	1923	2008		86
<i>KYJOVKA - KYJOV</i>	1951	2008		58
<i>MORAVA - VLASKÉ</i>	1950	2008		59
<i>MUMLAVA - JANOV-HARRACHOV</i>	1941	2008		68
<i>DIVOKÁ ORLICE - KLÁŠTEREC N. O.</i>	1938	2006		69
<i>OTAVA - REJŠTEJN</i>	1911	2007	1920-1930; 1937-1947	75
<i>PORUBKA - VŘESINA</i>	1953	2008		56
<i>ROŽNOV. BEČVA - HORNÍ BEČVA</i>	1955	2008		54
<i>SVRATKA - BOROVNICE</i>	1925	2007		67
<i>ÚPA - HORNÍ MARŠOV</i>	1931	2006	1940-1948	67
<i>VLÁRA - POPOV</i>	1956	2008		53
<i>VYDRA - MODRAVA</i>	1931	2007	1940-1948	68
<i>ZDOBNICE - SLATINA N. Z.</i>	1945	2007		63

Tabulka 2.1: Přehled stanic

V oddílu 9.3, který je věnován vícenásobným změnám, jsme pak zpracovávali denní průměrné průtoky Dunaje v Bratislavě z let 1901 - 2009. Tato data byla poskytnuta Slovenským hydrometeorologickým ústavem.

Abychom byli schopni popsané metody k analýze dat použít, museli jsme připravit řadu výpočetních programů. Programy byly napsány jako makra v prostředí Matlab. Zpracování pak probíhalo na osobních počítačích.



Obrázek 2.1: Mapa stanic

toky	průměr	medián	minimum	maximum	směrodatná odchylka
<i>BĚLÁ</i>	1.1162	0.7040	0.0460	51.9000	1.3554
<i>BLANICE</i>	0.9331	0.5680	0.0200	64.7000	1.2469
<i>BRODEČKA</i>	0.3323	0.1420	0.0000	21.8000	0.6864
<i>ČELADNÁ</i>	0.8241	0.4500	0.0200	46.5000	1.4090
<i>DOUBRAVA</i>	1.6949	1.0100	0.0430	56.9000	2.3331
<i>JIZERKA</i>	1.3348	0.8000	0.1020	39.5000	1.6844
<i>KYJOVKA</i>	0.2711	0.1750	0.0200	9.5000	0.3799
<i>MORAVA</i>	1.9617	1.4000	0.1900	92.0000	1.9045
<i>MUMLAVA</i>	1.9251	1.1200	0.1000	63.8000	2.3468
<i>ORLICE</i>	3.2949	2.0700	0.1700	108.0000	4.1106
<i>OTAVA</i>	8.1370	5.8500	0.4400	164.0000	7.6899
<i>PORUBKA</i>	0.2522	0.1100	0.0010	26.0000	0.6322
<i>ROŽNOV.BEČVA</i>	0.3093	0.1460	0.0008	16.0000	0.5498
<i>SVRATKA</i>	1.4812	0.9000	0.0290	55.4000	1.9326
<i>ÚPA</i>	2.6053	1.7300	0.2200	75.3000	2.6950
<i>VLÁRA</i>	1.4893	0.6800	0.0060	106.0000	2.8713
<i>VYDRA</i>	3.3674	2.2500	0.2000	82.0000	3.5874
<i>ZDOBNICE</i>	1.9258	1.1900	0.0760	55.8000	2.3955

Tabulka 2.2: Popisná statistika 18 českých řek.

tok	průměr	medián	minimum	maximum	směrodatná odchylka
<i>DUNAJ</i>	2052.6	1848	580	10285	972.52

Tabulka 2.3: Popisná statistika pro Dunaj.

## Kapitola 3

# Jednoduché testy na zjištění změny v řadě průměrných ročních průtoků

Jak jsme uvedli v úvodu, je cílem naší práce navrhnout statistické postupy pro detekci změn v sezónním chování řad. Myslíme si však, že dříve než se začne analyzovat stacionarita sezónního chování řad, je nutné se zamyslet nad tím, zda se nemění řada průměrných ročních průtoků. Z tohoto důvodu zde uvedeme tři jednoduché testy, které bývají součástí základního kurzu matematické statistiky a které je možno použít pro detekci případných nestacionarit řad průměrných ročních průtoků. První dva testy jsou dvouvýběrové a třetí je test na existenci lineárního trendu ve střední hodnotě. Ve všech třech testech se předpokládá, že studované veličiny, zde průměrné roční průtoky, tvoří posloupnost nezávislých náhodných veličin.

Pro detekci nestacionárního chování řady průměrných ročních průtoků by samozřejmě bylo lepší použít "sofistikované" testy vytvořené pro detekci bodu změny za předpokladu, že polohu bodu změny neznáme. Tyto testy jsou však mnohem komplikovanější. Vzhledem k tomu, že námětem předkládané práce není detekce nestacionarit v řadě průměrných ročních průtoků, tyto testy zde neuvádíme, a ani jsme je pro naše data nepoužili.

Jak už bylo řečeno, může nastat situace, kdy se celá řada  $X_1, \dots, X_N$  rozdělí na dvě části:  $X_1, \dots, X_n$  a  $X_{n+1}, \dots, X_{n+m}$ , ( $N = n + m$ ), přičemž první část je stacionární (její členy se řídí rozdělením s distribuční funkcí  $F_1(x)$ ) a rovněž její druhá část je stacionární (její členy se řídí rozdělením s distribuční funkcí  $F_2(x)$ ), přičemž však rozdělení první a druhé části se liší, tj.  $F_1(x) \neq F_2(x)$  pro nějaké  $x \in R^1$ . Tomuto typu změny se říká náhlá změna.

Pokud známe bod změny  $n$ , můžeme použít nějaký dvouvýběrový test. Dvouvýběrový test je schopen často detekovat případnou změnu, i když k ní dochází v jiném čase než v čase  $n$ . V takovém případě má ale menší sílu. Pokud čas, v kterém může dojít ke změně, neznáme, je nejlépe volit  $n = N/2$ . Tak jsme postupovali v případě průměrných ročních průtoků i my.

Pokud máme nějakou představu, že by ke změně mohlo dojít v nějakém jiném čase  $n_0$ , pak samozřejmě rozdělíme řadu v tomto bodě.

Nejčastějším typem změny je změna střední hodnoty, případně změna rozptylu. Odpovídající problémy testování hypotéz pro případ, že průměrné roční průtoky považujeme za výběry z normálního rozdělení, uvádíme níže. Pro velké hodnoty  $n$  a  $m$  se dají testy použít (díky centrální limitní větě) i pro výběry z jiného než normálního rozdělení. Připomeňme, že při aplikaci dvouvýběrových testů postupujeme tak, že první výběr tvoří řada  $X_1, \dots, X_n$  a druhý řada  $X_{n+1}, \dots, X_{n+m}$ .

### 3.1 Dvouvýběrové testy pro shodnost parametrů výběrů normálního rozdělení

Uvažujme náhodný výběr  $(X_1, \dots, X_n)$  o rozsahu  $n$  z rozdělení  $N(\mu_1, \sigma_1^2)$  a náhodný výběr  $(Y_1, \dots, Y_m)$  o rozsahu  $m$  z rozdělení  $N(\mu_2, \sigma_2^2)$ . Necht' náhodné výběry  $(X_1, \dots, X_n)$  a  $(Y_1, \dots, Y_m)$  jsou nezávislé.

a) Test shodnosti parametru středních hodnot dvou nezávislých výběrů

Nulová hypotéza:  $H_0 : \mu_1 = \mu_2$

Alternativa:  $A : \mu_1 \neq \mu_2$

Nulovou hypotézu  $H_0$  zamítáme (za předpokladu  $\sigma_1^2 = \sigma_2^2$ ), jestliže:

$$T = \frac{|\bar{X} - \bar{Y}|}{\sqrt{(n-1)s^2(x) + (m-1)s^2(y)}} \sqrt{n+m-2} \sqrt{\frac{nm}{n+m}} > t_{\frac{\alpha}{2}}[n+m-2], \quad (3.1)$$

$t_{\frac{\alpha}{2}}[n+m-2]$  je 100  $\frac{\alpha}{2}$ % horní kvantil  $t$  rozdělení o  $n+m-2$  stupních volnosti, kde  $\alpha$  označuje hladinu významnosti testu. Ve shora uvedených vzorcích jsme značili  $\bar{X} = \sum_{i=1}^n X_i(n)/n$ ,  $\bar{Y} = \sum_{i=1}^m Y_i(m)/m$ ,  $s^2(x) = \sum_{i=1}^n (X_i - \bar{X})/(n-1)$ ,  $s^2(y) = \sum_{i=1}^m (Y_i - \bar{Y})/(m-1)$ . Pokud je  $n$  a  $m$  velké, můžeme použít asymptotické zamítací pravidlo, kde nahradíme kvantil  $t$ -rozdělení odpovídajícím kvantilem normálního rozdělení. Jak bylo řečené, pokud případný bod změny neznáme, je nejlepší při testování volit  $n = m$  a dělit časovou řadu na dvě stejně dlouhé části. Tak jsme postupovali v našem případě. Pokud  $n + m$  bylo liché, volili jsme první část o 1 rok delší než část druhou. Výsledky pro jednotlivé toky jsou uvedeny v Tabulce 3.1. V této tabulce  $t$  označuje hodnotu testové statistiky  $T$ ,  $p$  označuje  $p$ -hodnotu příslušného testu, tj.  $p = P(T > t)$  a  $t_{kv}$  označuje příslušný 2.5% horní kvantil Studentova rozdělení o  $n + m - 2$  stupních volnosti.

toky	$t$	$t_{kv}$	$p$	$\bar{X}$	$\bar{Y}$	počet dat
<i>BĚLÁ</i>	-0.8437	1.9966	0.4019	1.0917	1.1580	67
<i>BLANICE</i>	1.0554	2.0040	0.2959	0.9710	0.8940	56
<i>BRODEČKA</i>	1.0579	1.9966	0.2940	0.3539	0.3128	67
<i>ČELADENKA</i>	0.0087	2.0049	0.9931	0.8394	0.8389	55
<i>DOUBRAVA</i>	-0.7583	2.0057	0.4516	1.6404	1.7513	55
<i>JIZERKA</i>	1.3463	1.9883	0.1818	1.4002	1.2858	86
<i>KYJOVKA</i>	2.0764	2.0025	0.0266	0.3040	0.2425	58
<i>MORAVA</i>	0.8507	2.0017	0.3984	2.0015	1.9014	59
<i>MUMLAVA</i>	-3.0516	1.9960	0.0033	1.7717	2.0738	68
<i>ORLICE</i>	-1.2980	1.9960	0.1987	3.1572	3.4366	69
<i>OTAVA</i>	-0.9811	1.9908	0.3296	8.0809	8.5419	75
<i>PORUBKA</i>	2.1189	2.0040	0.0386	0.2875	0.2197	56
<i>ROŽNOV.BEČVA</i>	-0.5669	2.0057	0.5732	0.3047	0.3159	54
<i>SVRATKA</i>	0.2213	1.9893	0.8254	1.4882	1.4684	83
<i>ÚPA</i>	0.8085	1.9960	0.4217	2.6690	2.5147	67
<i>VLÁRA</i>	2.2634	2.0066	0.0278	1.6320	1.3246	53
<i>VYDRA</i>	-0.3048	1.9949	0.7614	3.3568	3.4293	68
<i>ZDOBNICE</i>	0.4314	1.9990	0.6677	2.0640	1.9924	63

Tabulka 3.1: Výsledky testování - test shodnosti středních hodnot (pro roční průměrné průtoky).

Porovnáme-li  $p$ -hodnoty s hodnotou 0.05 nebo hodnotou testové statistiky  $t$  s příslušným kvantilem  $t_{kv}$ , vidíme, že při testování shodnosti středních hodnot byla pouze pro 4 stanice (Kyjovka, Porubka, Vlára, Mumlava) nulová hypotéza zamítnuta na hladině významnosti  $\alpha = 0.05$ .

b) Test shodnosti rozptylů dvou nezávislých výběrů:

Nulová hypotéza:  $H_0 : \sigma_1^2 = \sigma_2^2$

Alternativa:  $A : \sigma_1^2 \neq \sigma_2^2$

Nulovou hypotézu  $H_0$  zamítáme, jestliže:

$$f = \frac{\max(s^2(x), s^2(y))}{\min(s^2(x), s^2(y))} > F_{\frac{\alpha}{2}}[\nu_1, \nu_2], \quad (3.2)$$

kde  $\nu_1 = n - 1$  a  $\nu_2 = m - 1$ , jestliže  $s^2(x) > s^2(y)$ ,  $\nu_1 = m - 1$  a  $\nu_2 = n - 1$ , jestliže  $s^2(x) < s^2(y)$ . Zde  $F_{\frac{\alpha}{2}}[\nu_1, \nu_2]$  je 100  $\frac{\alpha}{2}$ % horní kvantil  $F$  rozdělení o  $\nu_1$  a  $\nu_2$  stupních volnosti,

a  $\alpha$  označuje hladinu významnosti testu. Výsledky testu shodnosti dvou rozptylů pro dvě části časové řady, kde jsme opět provedli stejné dělení jako při testování shody středních hodnot, jsou shrnuty v Tabulce 3.2. V Tabulce 3.2 označuje  $f$  hodnotu testové statistiky  $F$ , dále  $p$  označuje  $p$ -hodnotu, tj.  $p = P(F > f)$ ,  $F_{kv}$  označuje 2.5% horní kvantil  $F$  rozdělení o  $n - 1$  a  $m - 1$  stupních volnosti.

toky	$f$	$F_{kv}$	$p$	$s^2(x)$	$s^2(y)$
<i>BĚLÁ</i>	2.2375	2.0023	0.0234	0.1451	0.0649
<i>BLANICE</i>	1.4212	2.1397	0.3607	0.0629	0.0894
<i>BRODEČKA</i>	1.9239	2.0023	0.0644	0.0338	0.0176
<i>ČELADENKA</i>	1.9961	2.1609	0.0780	0.0785	0.0393
<i>DOUBRAVA</i>	1.0636	2.1839	0.8771	0.3030	0.2848
<i>JIZERKA</i>	1.0902	1.8415	0.7805	0.1637	1.1502
<i>KYJOVKA</i>	1.0834	2.2821	0.8453	0.0100	0.0108
<i>MORAVA</i>	1.0778	2.1010	0.8415	0.1998	0.2153
<i>MUMLAVA</i>	1.0922	1.9953	0.8015	0.1763	0.1614
<i>ORLICE</i>	1.1996	1.9953	0.6028	0.8710	0.7261
<i>OTAVA</i>	1.1193	1.8907	0.7267	5.2904	4.7267
<i>PORUBKA</i>	2.2624	2.1512	0.0370	0.0201	0.0089
<i>ROŽNOV.BEČVA</i>	1.3693	2.1839	0.4258	0.0062	0.0045
<i>SVRATKA</i>	1.2539	1.8604	0.4720	0.1881	0.1500
<i>ÚPA</i>	1.7102	1.9953	0.1264	0.7895	0.4617
<i>VLÁRA</i>	1.8432	2.1943	0.1254	0.3227	0.1751
<i>VYDRA</i>	2.6152	1.9744	0.0061	1.4453	0.5527
<i>ZDOBNICE</i>	2.3320	2.0486	0.0212	0.6181	0.2650

Tabulka 3.2: Testování hypotéz - test rovnosti rozptylů (pro roční průměrné průtoky).

Porovnáním  $p$ -hodnoty s hodnotou 0.05 nebo porovnáním hodnot  $f$  s příslušným kvantilem  $F_{kv}$  vidíme, že nulová hypotéza o shodnosti rozptylů byla zamítnuta pro 4 stanice (Bělá, Porubka, Vydra, Zdobnice).



## 3.2 Existence lineárního trendu ve střední hodnotě časové řady

Další úvaha nás vede k tomu, že se můžeme ptát, zda střední hodnota naměřené řady nevykazuje trend. Nejjednodušším typem trendu je lineární trend. Opět je zřejmé, že test na existenci lineárního trendu ve střední hodnotě bude často zamítat hypotézu o stacionaritě i tehdy, když trend nebude lineární, ale bude například rostoucí nebo klesající. Test, který uvádíme, je nejjednodušším testem existence trendu v jednoduché lineární regresi, kde předpokládáme normální rozdělení uvažovaných veličin, přičemž jejich rozptyl se s časem nemění. Opět připomeňme, že pokud je počet dat velký, můžeme použít test i pro data, která nejsou výběrem z normálního rozdělení.

Předpokládejme, že jsme naměřili v ekvidistantních časech  $t_1, \dots, t_N$  hodnoty  $X(t_1), \dots, X(t_N)$ . V našem případě  $t_1, \dots, t_N$  odpovídají jednotlivým rokům a  $X(t_1), \dots, X(t_N)$  ročním průměrným průtokům. Předpokládejme, že data odpovídají modelu lineární regrese

$$X(t_i) = b_0 + b_1 t_i + e_i, \quad i = 1, \dots, N,$$

kde  $\{e_i\}$  jsou nezávislé stejně rozdělené veličiny s rozdělením  $N(0, \sigma^2)$ . Rozhodnutí, zda průměrné roční průtoky vykazují lineární trend, můžeme provést pomocí testování hypotéz. V tomto případě stanovíme nulovou hypotézu  $H_0$  a alternativní hypotézu  $A$  následovně:

Nulová hypotéza:  $H_0 : b_1 = 0$

Alternativa:  $A : b_1 \neq 0$

Nulovou hypotézu  $H_0$  zamítáme, jestliže:

$$t = \frac{|\hat{b}_1|}{s_{\hat{b}_1}} > t_{\frac{\alpha}{2}}[N - 2], \quad (3.3)$$

kde  $\hat{b}_1$  je odhad metodou nejmenších čtverců parametru  $b_1$ , tj.

$$\hat{b}_1 = \frac{\sum_{i=1}^N (t_i - \bar{t}) X(t_i)}{\sum_{i=1}^N (t_i - \bar{t})^2},$$

$s_{\hat{b}_1}$  je odhad směrodatné odchylky odhadu  $\hat{b}_1$ , tj.

$$s_{\hat{b}_1} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^N (t_i - \bar{t})^2}},$$

přičemž

$$\hat{\sigma} = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (X_i - (\hat{b}_0 + \hat{b}_1 t_i))^2}.$$

Výsledky testů pro jednotlivé stanice jsou obsaženy v Tabulce 3.3.

Druhý sloupec udává absolutní hodnoty příslušných  $t$ -statistik, viz vzorec (3.3), třetí 2.5% horní kvantily, čtvrtý odpovídající  $p$ -hodnoty. V pátém sloupci nalezneme odhady parametru  $b_1$  a v posledním sloupci odhad jeho směrodatné odchylky.

toky	$t$	$t_{kv}$	$p$	$\hat{b}_1$	$s_{\hat{b}_1}$
<i>BĚLÁ</i>	0.0015	1.9966	0.9988	0.0000	0.0116
<i>BLANICE</i>	0.0305	2.0040	0.9758	-0.0004	0.0117
<i>BRODEČKA</i>	0.1615	1.9966	0.8722	-0.0009	0.0057
<i>ČELADENKA</i>	0.0096	2.0049	0.9924	0.0001	0.0104
<i>DOUBRAVA</i>	0.2651	2.0057	0.7920	0.0062	0.234
<i>JIZERKA</i>	0.2638	1.9883	0.7926	-0.0029	0.0110
<i>KYJOVKA</i>	0.4695	2.0025	0.6405	-0.0020	0.0042
<i>MORAVA</i>	0.0206	2.0017	0.9837	-0.0004	0.0184
<i>MUMLAVA</i>	0.5267	1.9960	0.6001	0.0076	0.0144
<i>ORLICE</i>	0.1889	1.9960	0.8507	0.0059	0.0314
<i>OTAVA</i>	0.1221	1.9908	0.9031	0.0068	0.0068
<i>PORUBKA</i>	0.1390	2.0040	0.8900	-0.0007	0.0053
<i>ROŽNOVBEČVA</i>	0.1741	2.0057	0.8624	0.0006	0.0032
<i>SVRATKA</i>	0.0661	1.9893	0.9475	-0.0008	0.0118
<i>ÚPA</i>	0.2350	1.9960	0.8149	-0.0060	0.0255
<i>VLÁRA</i>	0.2767	2.0066	0.7831	-0.0063	0.0228
<i>VYDRA</i>	0.0093	1.9949	0.9926	-0.0003	0.0315
<i>ZDOBNICE</i>	0.2351	1.9990	0.8149	-0.0058	0.0247

Tabulka 3.3: Testování hypotéz - test existence trendu.

Všechny  $p$ -hodnoty jsou výrazně vyšší než 0.05. Statistický test neprokázal existenci lineárního trendu ani pro jednu stanici.

## Kapitola 4

# Detekce změny průměrného ročního chodu - test pro známý bod změny

Kapitola 4 je věnována nejjednoduššímu dvouvýběrovému testu stacionarity sezónního chování průtokových řad. Sezónní chování řady je určeno vektorem denních průměrných průtoků v jednom kalendářním roce. Základním problémem tedy je, zda jeho vektor středních hodnot zůstává během doby měření stále stejný.

V této kapitole budeme předpokládat, že čas, v kterém došlo ke změně chování časové řady, známe. Jak už bylo řečeno, pokud je změna dost velká, může takový test detekovat změnu i v případě, že k ní došlo v jiném než očekávaném čase  $n$ .

### 4.1 Test změny ve střední hodnotě - dvouvýběrový problém

Nyní popíšeme test pro testování shody středních hodnot dvou výběrů, jejichž prvky jsou náhodné vektory. Jedná se opět o dvouvýběrový problém.

Podobně jako v jednorozměrném případě budeme předpokládat, že oba výběry  $X_1, \dots, X_n$  i  $Y_1, \dots, Y_m$  mají  $p$ -rozměrné normální rozdělení se shodnou varianční matice  $\Sigma$ . Zajímá nás, zda jejich střední hodnoty  $\mu_x$  a  $\mu_y$  jsou shodné.

Testujeme tedy nulovou hypotézu  $H_0$  proti alternativě  $A$ :

$$H_0 : \mu_x = \mu_y \quad A : \mu_x \neq \mu_y.$$

Pokud známe varianční matici  $\Sigma$ , pak testová statistika má tvar:

$$T^2 = \frac{nm}{n+m} (\bar{X} - \bar{Y})^T \Sigma^{-1} (\bar{X} - \bar{Y}). \quad (4.1)$$

Tato testová statistika má za platnosti  $H_0$  rozdělení  $\chi^2$  o  $p$  stupních volnosti, kde  $p$  je dimenze  $\{\mathbf{X}_i\}$  a  $\{\mathbf{Y}_i\}$ . Nulovou hypotézu na hladině  $\alpha$  zamítáme, jestliže

$$T^2 > \chi_\alpha^2[p], \quad (4.2)$$

kde  $\chi_\alpha^2[p]$  je  $\alpha$  100 % horní kvantil  $\chi^2$  rozdělení o  $p$  stupních volnosti.

Pokud matici  $\Sigma$  neznáme, nahrazujeme ji odhadem:

$$\widehat{\Sigma} = \frac{1}{n+m-2} \left\{ \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T + \sum_{i=1}^m (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T \right\}.$$

Nulovou hypotézu  $H_0$  zamítáme, jestliže

$$T^2 > \frac{(n+m-2)p}{n+m-p-1} F_\alpha[p, n+m-p-1],$$

kde  $F_\alpha[p, n+m-p-1]$  je  $\alpha$  100 % horní kvantil  $F$  rozdělení o  $p$  a  $n+m-p-1$  stupních volnosti, viz Anderson (1984).

Pokud  $n$  i  $m$  jsou velké, můžeme díky konzistenci odhadu  $\widehat{\Sigma}$  použít zamítací pravidlo (4.2).

## 4.2 Použití testu pro shodu středních hodnot k detekci změny ročního chodu

Roční chod je v našem případě popsán vektorem o 365 souřadnicích, kde každá složka tohoto vektoru odpovídá dennímu průměrnému průtoku daného kalendářního roku. (V analýze jsme vynechali přestupné 29. únory, takže každý rok je skutečně reprezentován vektorem o 365 souřadnicích.) Základní otázkou je, zda stochastické chování těchto vektorů se během doby měření nezměnilo.

Uvažujme velmi jednoduchý případ porušení stacionárního chování posloupnosti. Předpokládejme, že jsme řadu  $\mathbf{X}_1, \dots, \mathbf{X}_N$  rozdělili na dvě části  $\mathbf{X}_1, \dots, \mathbf{X}_n$  a  $\mathbf{X}_{n+1}, \dots, \mathbf{X}_{n+m}$  ( $N = n + m$ ). Stacionarita může být porušena tak, že první posloupnost  $\mathbf{X}_1, \dots, \mathbf{X}_n$  je stacionární a druhá posloupnost  $\mathbf{X}_{n+1}, \dots, \mathbf{X}_{n+m}$  je také stacionární, ale stochastické chování první a druhé části se vzájemně liší. Poznamenejme, že budeme opět předpokládat, že vektory průměrných denních průtoků  $\mathbf{X}_i = (X_{i1}, \dots, X_{i365})^T$ ,  $i = 1, \dots, N = n + m$ , jsou nezávislé.

Chceme-li detekovat změnu ve střední hodnotě (za předpokladu, že se varianční matice nemění), která by mohla nastat v čase  $n$ , můžeme použít postup z předchozí kapitoly, kde roli prvního výběru hrají průměrné denní průtoky v prvním období  $\mathbf{X}_1, \dots, \mathbf{X}_n$  a roli druhého výběru hrají průměrné denní průtoky v druhém období  $\mathbf{Y}_1 = \mathbf{X}_{n+1}, \dots, \mathbf{Y}_m = \mathbf{X}_{n+m}$ . Testová statistika má pak, viz (4.1), tvar:

$$T^2 = \frac{nm}{n+m} (\bar{\mathbf{X}}^* - \bar{\mathbf{X}}^{**})^T \hat{\Sigma}^{-1} (\bar{\mathbf{X}}^* - \bar{\mathbf{X}}^{**}), \quad (4.3)$$

kde

$$\bar{\mathbf{X}}^* = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, \quad \bar{\mathbf{X}}^{**} = \frac{1}{m} \sum_{i=n+1}^{n+m} \mathbf{X}_i,$$

$$\hat{\Sigma} = \frac{1}{n+m-2} \left\{ \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}^*)(\mathbf{X}_i - \bar{\mathbf{X}}^*)^T + \sum_{i=n+1}^{n+m} (\mathbf{X}_i - \bar{\mathbf{X}}^{**})(\mathbf{X}_i - \bar{\mathbf{X}}^{**})^T \right\}.$$

Nulovou hypotézu zamítáme, platí-li (4.2).

Pokud nevíme, kde by k případné změně mohlo dojít, volíme nejčastěji  $n = N/2$ . V jiném případě můžeme mít představu, že ke změně došlo například blíže konci sledovaného období. V tom případě volíme  $n > N/2$ .

Dvouvýběrový test, který jsme právě popsali, je teoreticky možné použít pro vektory denních průměrných průtoků v po sobě jdoucích kalendářních letech. Jestliže však  $n$  a  $m$  nebudou velké, nebude možné spolehlivě odhadnout kovarianční matici  $\Sigma$ , a tedy provést popsany test o shodnosti dvou vektorů středních hodnot. Kromě toho test, jehož cílem je detekce změny střední hodnoty alespoň jedné souřadnice (změna střední hodnoty jednoho kalendářního dne), bude mít velmi malou sílu, protože počet parametrů (365) je mnohem větší než počet analyzovaných vektorů, který odpovídá počtu let, v kterých byla měření prováděna.

Naším cílem je tudíž porovnávat střední hodnoty dvou vektorů s nižším počtem souřadnic, tedy snížit dimenzi úlohy.



## Kapitola 5

# Redukce dimenze

Redukovat dimenzi úlohy můžeme tak, že transformujeme pozorované vektory pomocí lineární transformace na vektory s menším počtem souřadnic. Pro vektory vzniklé lineární transformací pak budeme testovat shodu jejich středních hodnot. Přesněji řečeno, redukce dimenze spočívá v nahrazení vektorů  $\mathbf{X}_1, \dots, \mathbf{X}_{n+m}$  o  $p$  souřadnicích vektory  $\mathbf{V}_1, \dots, \mathbf{V}_{n+m}$  o  $k$  souřadnicích, kde  $\mathbf{V}_i = \left( \mathbf{c}(1)^T \mathbf{X}_i, \mathbf{c}(2)^T \mathbf{X}_i, \dots, \mathbf{c}(k)^T \mathbf{X}_i \right)^T$ ,  $i = 1, \dots, n + m$ , a  $\mathbf{c}(1), \dots, \mathbf{c}(k)$  jsou vhodně zvolené vektory konstant. To znamená, že studujeme vektory, jejichž souřadnice jsou určitým způsobem zvolené lineární kombinace složek původních vektorů, přičemž se volí  $k \ll p$ . (V našem případě  $p = 365$ .) To znamená, že se zajímáme jen o některé speciální typy změn střední hodnoty sledovaných vektorů. Pokud se nulová hypotéza o shodnosti středních hodnot transformovaných vektorů zamítne, budou i původní vektory mít odlišné vektory středních hodnot. Na druhé straně je zřejmé, že tímto postupem nejsme schopni odhalit jakoukoliv změnu ve vektorech středních hodnot.

Nechť vektory  $\mathbf{X}_1, \dots, \mathbf{X}_n$  jsou nezávislé se střední hodnotou  $\boldsymbol{\mu}_x$  a vektory  $\mathbf{Y}_1 = \mathbf{X}_{n+1}, \dots, \mathbf{Y}_m = \mathbf{X}_{n+m}$  jsou nezávislé se střední hodnotou  $\boldsymbol{\mu}_y$ . Předpokládáme, že varianční matice obou výběrů je shodná a označme ji  $\boldsymbol{\Sigma}$ . Uvažujme matici, jejíž řádky jsou tvořeny vektory  $\mathbf{c}(1)^T, \dots, \mathbf{c}(k)^T$ , pak  $\mathbf{V}_1, \dots, \mathbf{V}_n$  jsou opět nezávislé se střední hodnotou  $\mathbf{C}\boldsymbol{\mu}_x$  a  $\mathbf{V}_{n+1}, \dots, \mathbf{V}_{n+m}$  jsou nezávislé se střední hodnotou  $\mathbf{C}\boldsymbol{\mu}_y$ , přičemž varianční matice obou výběrů je  $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T$ . Odtud plyne, že místo, abychom testovali nulovou hypotézu  $\boldsymbol{\mu}_x = \boldsymbol{\mu}_y$  oproti alternativě  $\boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y$ , testujeme nulovou hypotézu  $\mathbf{C}\boldsymbol{\mu}_x = \mathbf{C}\boldsymbol{\mu}_y$  oproti alternativě  $\mathbf{C}\boldsymbol{\mu}_x \neq \mathbf{C}\boldsymbol{\mu}_y$ . Zřejmě pokud  $\mathbf{C}\boldsymbol{\mu}_x \neq \mathbf{C}\boldsymbol{\mu}_y$ , pak také  $\boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y$ . Na druhé straně ale z rovnosti  $\mathbf{C}\boldsymbol{\mu}_x = \mathbf{C}\boldsymbol{\mu}_y$  neplyne  $\boldsymbol{\mu}_x = \boldsymbol{\mu}_y$ . Záleží tedy na naší intuici, jakým způsobem vybereme  $\mathbf{c}(1), \dots, \mathbf{c}(k)$  tak, abychom případný rozdíl ve středních hodnotách vektorů  $\mathbf{X}_1, \dots, \mathbf{X}_n$  a  $\mathbf{Y}_1 = \mathbf{X}_{n+1}, \dots, \mathbf{Y}_m = \mathbf{X}_{n+m}$  detekovali. Důležité je také vybrat správný počet vektorů. Tato volba je subjektivní a záleží pouze na zkušenosti s podobným typem dat.

## 5.1 Použití dvouvýběrových testů pro průtokové řady

Jak už bylo dříve řečeno, volba času  $n$ , ve kterém řadu vektorů  $V_1, \dots, V_N$  dělíme na dvě části  $V_1, \dots, V_n$  a  $V_{n+1}, \dots, V_{n+m}$ , je subjektivní. Pro lepší porovnání výsledků jsme zvolili čas  $n$  tak, aby dělicím bodem řad byl rok 1997. Vzhledem k tomu, že řady nejsou stejně dlouhé, a začátek i konec sledovaného období je pro každou řadu různý, odpovídá rok 1997 v každé řadě jiné hodnotě  $n$ . Všechny tabulky s výsledky, které jsou uvedené v *Kapitole 5*, odpovídají tomu, že hodnota  $n$  byla zvolena tak, aby uvažovaným bodem změny byl právě rok 1997.

Tabulka 5.1 uvádí, kolik je pro jednotlivé stanice měření do roku 1997 a kolik po něm.

toky	před r. 1997	po roce 1997
<i>BĚLÁ</i>	57	10
<i>BLANICE</i>	45	11
<i>BRODEČKA</i>	57	10
<i>ČELADENKA</i>	45	10
<i>DOUBRAVA</i>	46	9
<i>JIZERKA</i>	75	11
<i>KYJOVKA</i>	47	11
<i>MORAVA</i>	48	11
<i>MUMLAVA</i>	57	11
<i>ORLICE</i>	60	9
<i>OTAVA</i>	65	10
<i>PORUBKA</i>	45	11
<i>ROŽNOV. BEČVA</i>	43	11
<i>SVRATKA</i>	73	10
<i>ÚPA</i>	58	9
<i>VLÁRA</i>	42	11
<i>VYDRA</i>	58	10
<i>ZDOBNICE</i>	53	10

Tabulka 5.1: Počet let před rokem 1997 (včetně) a po něm.



## 5.2 Nahrazení denních průměrných průtoků průměrnými měsíčními průtoky

Studium průměrných měsíčních průtoků je v hydrologii velmi populární. Nicméně délky měsíců nesledují přesně měsíční cyklus (tj. dobu od novu do novu, která trvá cca 29,5 dne tzv. "synodický měsíc"), ale jsou dány uměle historickým rozhodnutím našich předků. (V gregoriánském kalendáři trvá délka měsíce 28 až 31 dní, odtud náš dnešní "kalendářní měsíc"). Změna se například může projevit uprostřed měsíce nebo se řada změní v určitém období, které nezačíná přesně počátkem měsíce a nekončí s koncem téhož nebo nějakého následujícího měsíce. Proto se změny v ročním chodu v analýze měsíčních průměrných průtoků nemusí projevit nebo se projeví jen tak málo, že použitý statistický test může považovat rozdílnost ve střední hodnotě v prvním a druhém období za nevýznamnou. V článku Jarušková (2013) ukázala, že testy změny založené na průměrných měsíčních průtocích jsou obvykle nejslabší ve srovnání s analýzou založenou na změně koeficientů Fourierovy řady nebo při použití metody hlavních komponent, které jsou popsány v následujících kapitolách.

Nicméně je vhodné si uvědomit, že přechod od denních průměrných průtoků k měsíčním průměrným průtokům odpovídá shora uvedenému popisu redukce dimenze, kde  $\mathbf{c}(1) = \frac{1}{31}(\underbrace{1, \dots, 1}_{31\text{-krát}}, 0, \dots, 0)^T, \dots, \mathbf{c}(12) = \frac{1}{31}(\underbrace{0, \dots, 0}_{334\text{-krát}}, 1, \dots, 1)^T$ , a tedy průměrné měsíční průtoky v  $i$ -tém roce jsou  $\mathbf{c}(1)^T \mathbf{X}_i, \dots, \mathbf{c}(12)^T \mathbf{X}_i$ .

Výhodou detekce změny založené na průměrných měsíčních průtocích je také snadná interpretovatelnost výsledků statistické inference. Pokud zamítneme nulovou hypotézu, pak nám analýza zároveň ukáže, v kterých měsících ke změně došlo.

Tabulka 5.2 udává výsledky testu shody vektoru středních hodnot měsíčních průměrů pro období do roku 1997 a po něm. V prvním sloupci je název toku, v druhém hodnota příslušné testové statistiky. Třetí sloupec udává  $p$ -hodnotu, použijeme-li asymptotické  $\chi^2$  rozdělení. Poznamenejme, že pokud bychom použili přesné rozdělení, byly by odpovídající  $p$ -hodnoty větší.

toky	test.stat.	$p$ -hodnota (asymp.)
<i>BĚLÁ</i>	17.6	0.1279
<i>BLANICE</i>	35.6	0.0004
<i>BRODEČKA</i>	25.0	0.0147
<i>ČELADENKA</i>	25.7	0.0117
<i>DOUBRAVA</i>	24.9	0.0153
<i>JIZERKA</i>	30.0	0.0028
<i>KYJOVKA</i>	13.6	0.3257
<i>MORAVA</i>	23.8	0.0219
<i>MUMLAVA</i>	71.6	0.0000
<i>ORLICE</i>	29.8	0.0030
<i>OTAVA</i>	32.7	0.0011
<i>PORUBKA</i>	16.0	0.1900
<i>ROŽNOV. BEČVA</i>	23.4	0.0247
<i>SVRATKA</i>	9.4	0.6727
<i>ÚPA</i>	27.2	0.0072
<i>VLÁRA</i>	16.1	0.1854
<i>VYDRA</i>	25.2	0.0139
<i>ZDOBNICE</i>	17.6	0.1272

Tabulka 5.2: Výsledky porovnání testu shody vektoru středních hodnot průměrných měsíčních průtoků do r. 1997 a po něm.

### 5.3 Redukce dimenze pomocí aproximace Fourierovou řadou

Myšlenka redukce dimenze pomocí aproximace Fourierovou řadou spočívá v aproximaci vektorů denních průtoků během jednoho roku pomocí periodické funkce. Pro vybrané  $l$  uvažujme nejmenší Fourierovy frekvence  $\{2\pi r/365, r = 1, \dots, l\}$ . Dále pro  $i = 1, \dots, N$ , nahradíme vektory denních měření  $\mathbf{X}_i = (X_{i1}, \dots, X_{i365})^T$  vektory o  $L = (2l + 1)$  složkách  $\mathbf{V}_i = (\hat{m}_i, \hat{A}_i(1), \hat{B}_i(1), \dots, \hat{A}_i(l), \hat{B}_i(l))^T$ ,  $i = 1, \dots, n$ , jejichž souřadnice jsou odhady Fourierových koeficientů  $m_i, A_i(1), B_i(1), \dots, A_i(l), B_i(l)$  ve Fourierově rozvoji, získané metodou nejmenších čtverců:

$$m_{i,j} = m_i + \sum_{r=1}^l A_i(r) \cos \frac{2\pi jr}{365} + B_i(r) \sin \frac{2\pi jr}{365}, \quad j = 1, \dots, 365.$$

Pro odhady parametrů metodou nejmenších čtverců platí:

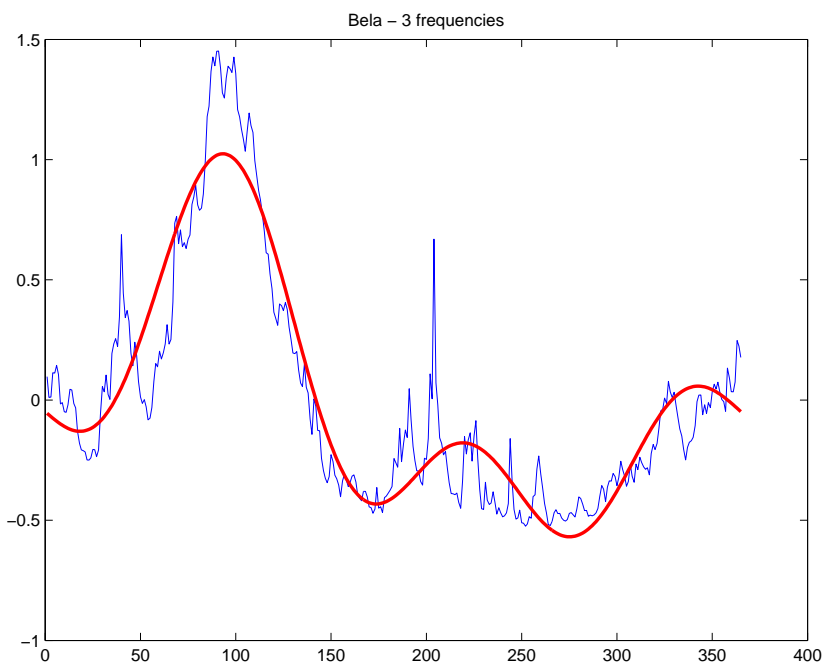
$$\begin{aligned}\hat{m}_i &= (1/365) \sum_{j=1}^{365} X_{i,j}, \\ \hat{A}_i(r) &= (2/365) \sum_{j=1}^{365} X_{i,j} \cos(2\pi jr/365), \\ \hat{B}_i(r) &= (2/365) \sum_{j=1}^{365} X_{i,j} \sin(2\pi jr/365),\end{aligned}$$

$i = 1, \dots, N$  a  $r = 1, \dots, l$ .

Opět se tedy jedná o nahrazení vektorů  $\mathbf{X}_i$  (vektorů denních průměrů) vektory  $\mathbf{V}_i = (\mathbf{c}(1)^T \mathbf{X}_i, \dots, \mathbf{c}(2l+1)^T \mathbf{X}_i)^T$ ,  $i = 1, \dots, N$ , přičemž vektory  $\mathbf{c}(1), \dots, \mathbf{c}(2l+1)$  jsou voleny následovně:  $\mathbf{c}(1) = \frac{1}{365}(1, \dots, 1)^T$ ,  $\mathbf{c}(2) = \frac{1}{365}(\cos(2\pi/365), \dots, \cos(2\pi))^T$ ,  $\mathbf{c}(3) = \frac{1}{365}(\sin(2\pi/365), \dots, \sin(2\pi))^T$ ,  $\dots$ ,  $\mathbf{c}(2l) = \frac{1}{365}(\cos(2\pi l/365), \dots, \cos(2\pi l))^T$ ,  $\mathbf{c}(2l+1) = \frac{1}{365}(\sin(2\pi l/365), \dots, \sin(2\pi l))^T$ . Statistiku a zamítací pravidlo pak použijeme pro výběry  $\{\mathbf{V}_1, \dots, \mathbf{V}_n\}$  a  $\{\mathbf{V}_{n+1}, \dots, \mathbf{V}_{N=n+m}\}$ , kde  $p$  nahradíme  $2l+1$ .

Metoda pro detekci změny funguje dobře, pokud můžeme očekávat, že pro  $i = 1, \dots, N$ , vektory středních hodnot  $E \mathbf{X}_i = (\mu_{i,1}, \dots, \mu_{i,365})^T$  mohou být relativně dobře aproximovány vektory  $(m_{i,1}, \dots, m_{i,365})^T$ . Vzhledem k tomu, že roční cyklus průtoků úzce souvisí s rotací Země kolem Slunce, můžeme předpokládat, že vektory středních hodnot jsou periodickou funkcí indexu času  $j$  ve smyslu, že  $\mu_{i,1}$  a  $\mu_{i,365}$  jsou si blízké. Navíc můžeme předpokládat, že průběh ročního cyklu je pomalu se měnící funkce s malým počtem "vrcholů a údolí". Odtud vyplývá, že modelování ročních chodů pomocí Fourierovy řady je často velmi dobré. Počet  $l$  Fourierových frekvencí je volen subjektivně. Ze zkušenosti s chováním ročního cyklu evropských řek víme, viz Jarušková (2013), že je obvykle dostačující volit  $1 \leq l \leq 3$ . Obrázek ukazuje na příkladu řeky Bělé, jak dobře aproximuje Fourierova řada, kde  $l = 3$ , roční chod.

Tabulka 5.3 udává výsledky testu pro  $l = 3$ . V prvním sloupci je název toku, ve druhém hodnota testové statistiky a ve třetím příslušná  $p$ -hodnota.



Obrázek 5.1: Aproximace ročního chodu pomocí Fourierovy řady s  $l = 3$  pro řeku Bělou.

toky	test. stat.	$p$ -hodnota (asympt.)
<i>BĚLÁ</i>	10.1	0.1203
<i>BLANICE</i>	25.8	0.0002
<i>BRODEČKA</i>	7.2	0.3054
<i>ČELADENKA</i>	13.8	0.0314
<i>DOUBRAVA</i>	12.6	0.0503
<i>JIZERKA</i>	21.0	0.0018
<i>KYJOVKA</i>	6.6	0.3594
<i>MORAVA</i>	17.1	0.0088
<i>MUMLAVA</i>	30.7	0.0000
<i>ORLICE</i>	22.2	0.0011
<i>OTAVA</i>	20.1	0.0027
<i>PORUBKA</i>	7.0	0.3214
<i>ROŽNOV. BEČVA</i>	12.5	0.0515
<i>SVRATKA</i>	7.5	0.2750
<i>ÚPA</i>	15.2	0.0186
<i>VLÁRA</i>	6.3	0.3900
<i>VYDRA</i>	14.5	0.0241
<i>ZDOBNICE</i>	8.6	0.1970

Tabulka 5.3: Výsledky testování pro 3 Fourierovy frekvence (dělicí rok je 1997).

## 5.4 Redukce dimenze metodou hlavních komponent

Metoda hlavních komponent je nejčastěji používanou metodou pro snížení dimenze analyzovaných vektorů. Předpokládejme, že nezávislé vektory  $\mathbf{X}_1, \dots, \mathbf{X}_N$  ( $N = n + m$ ) s nulovou střední hodnotou mají shodnou varianční matici  $\Sigma$ . Předpokládejme, že matice  $\Sigma$  je pozitivně definitní a všechna její vlastní čísla jsou vzájemně různá. Tyto vlastní čísla můžeme uspořádat podle velikosti  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ . Označme  $\mathbf{u}_1, \dots, \mathbf{u}_p$  vlastní vektory odpovídající vlastním číslům  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ . Pak vektory  $\mathbf{V}_i = (\mathbf{u}_1^T \mathbf{X}_i, \dots, \mathbf{u}_k^T \mathbf{X}_i)^T$ ,  $i = 1, \dots, N$ , kde  $k \ll p$ , jsou nezávislé vektory s nulovou střední hodnotou a varianční maticí  $\Sigma_k = \text{diag}(\lambda_1, \dots, \lambda_k)$ . Ve skutečnosti však matici  $\Sigma$  neznáme a musíme ji nahradit odhadem  $\hat{\Sigma}$ , kde  $\hat{\Sigma}$  je výběrová kovarianční matice založená na vektorech pozorování  $\mathbf{X}_1, \dots, \mathbf{X}_{n+m}$ . Vlastní čísla  $\lambda_1, \dots, \lambda_k$  můžeme odhadnout pomocí  $k$  největších vlastních čísel  $\hat{\lambda}_1, \dots, \hat{\lambda}_k$  matice  $\hat{\Sigma}$  a vlastní vektory  $\mathbf{u}_1, \dots, \mathbf{u}_k$  pomocí vlastních vektorů  $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_k$  matice  $\hat{\Sigma}$ , které odpovídají  $k$  největším vlastním číslům  $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ . Vektory  $\mathbf{V}_i = (\mathbf{u}_1^T \mathbf{X}_i, \dots, \mathbf{u}_k^T \mathbf{X}_i)^T$  tak nahradíme vektory  $\mathbf{V}_i = (\hat{\mathbf{u}}_1^T \mathbf{X}_i, \dots, \hat{\mathbf{u}}_k^T \mathbf{X}_i)^T$ ,  $i = 1, \dots, N = n + m$ . Je zřejmé, že tato situace se liší od situace, kterou jsme popsali v úvodu této kapitoly, protože zde pracujeme s lineárními kombinacemi souřadnic pozorovaných vektorů, ale koeficienty v lineární kombinaci jsou opět funkcemi pozorovaných vektorů. Limitní rozdělení statistiky (4.1) je opět  $\chi^2$  rozdělení o  $k$  stupních volnosti, a tedy pro zamítnutí stacionarity můžeme opět použít (4.2).

Při aproximaci pomocí metody hlavních komponent nahrazujeme vektory denních měření  $\mathbf{X}_i$ ,  $i = 1, \dots, n$ , vektory, jejichž složky jsou lineární kombinací složek  $\mathbf{X}_i$  s největšími rozptyly, a hledáme změnu v jejich středních hodnotách. Můžeme se ptát, proč metoda hlavních komponent může dobře fungovat pro detekci změny. Důvod je následující. Jestliže existuje poměrně velká změna ve střední hodnotě některých lineárních kombinací denních měření, potom je jejich výběrový rozptyl velký. Proto s velkou pravděpodobností bude tento vektor zahrnut do souboru  $k$  lineárních kombinací s největšími rozptyly. Pro podrobnější vysvětlení viz Aston a Kirch (2012). Počet  $k$  uvažovaných lineárních kombinací je volen subjektivně. Podle našich zkušeností stačí obvykle vybrat  $6 \leq k \leq 12$ .

Tabulka 5.4 udává výsledky testu pro  $k = 6$ . V prvním sloupci je název toku, ve druhém hodnota testové statistiky a ve třetím příslušná  $p$ -hodnota.

toky	test. stat	$p$ -hodnota (asymp.)
<i>BĚLÁ</i>	11.7	0.0679
<i>BLANICE</i>	17.6	0.0073
<i>BRODEČKA</i>	13.3	0.0388
<i>ČELADENKA</i>	1.6	0.9524
<i>DOUBRAVA</i>	13.8	0.0323
<i>JIZERKA</i>	11.0	0.0876
<i>KYJOVKA</i>	2.6	0.8565
<i>MORAVA</i>	14.0	0.0294
<i>MUMLAVA</i>	21.9	0.0013
<i>ORLICE</i>	11.7	0.0701
<i>OTAVA</i>	14.9	0.0212
<i>PORUBKA</i>	7.8	0.2546
<i>ROŽNOV. BEČVA</i>	0.8	0.9911
<i>SVRATKA</i>	3.5	0.7377
<i>ÚPA</i>	6.8	0.3437
<i>VLÁRA</i>	2.6	0.8624
<i>VYDRA</i>	11.4	0.0761
<i>ZDOBNICE</i>	8.9	0.1781

Tabulka 5.4: Výsledky testování pro metodu hlavních komponent pro  $k = 6$  a rok dělení 1997.

## Kapitola 6

# Neparametrický odhad střední hodnoty ročního chodu před a po změně a odhad posunu mezi dobami, kde střední roční chody nabývají svého maxima

V předchozím textu jsme ukázali, že použijeme-li pro rozhodování o stacionaritě sezónního chování průtokových řad testování hypotéz, je nulová hypotéza, která tvrdí, že střední roční chod je stacionární, zamítnuta pro více než třetinu analyzovaných řad. Přesněji řečeno, střední hodnota vektoru denních průměrných průtoků z počátku našeho pozorování (tj. do roku 1997) se liší od střední hodnoty vektoru denních průměrných průtoků z konce našeho pozorování (tj. od roku 1997). Přijmeme-li závěry testování hypotéz, že se střední roční chod v průběhu měření změnil, chtěli bychom zjistit, jakým způsobem ke změně došlo. To znamená, že bychom chtěli odhadnout průměrný roční chod před a po změně.

### 6.1 Jádrový odhad regresní funkce

Nyní vysvětlíme, jak se obvykle provádí neparametrický odhad regresní funkce. Popsaný postup se nejčastěji nazývá jádrový odhad regresní funkce, viz Antoch a Vorlíčková (1992).

V kapitolách 4 a 5 jsme označovali písmenem  $n$  počet roků v prvním období, písmenem  $m$  počet roků v druhém období a pomocí  $N = n + m$  celkový počet sledovaných roků. V následu-

jících oddílech, pokud nebude řečeno jinak, bude značit  $n$  počet pozorování prvků v jedné řadě (v naší aplikaci pak v jednom daném roce, tj. konkrétně  $n = 365$ ),  $m$  bude značit odhadovanou funkci a  $N$  počet roků, na základě které se odhad funkce provádí.

Předpokládejme, že v časových okamžicích  $t_1, \dots, t_n \in [0, T]$  pozorujeme posloupnost  $\{X(t_j)\}$ . Veličiny  $\{X(t_j)\}$  splňují vztah:

$$X(t_j) = m(t_j) + e(t_j), \quad j = 1, \dots, n, \quad (6.1)$$

kde  $m(t)$  je nějaká deterministická funkce definovaná na intervalu  $[0, T]$  a  $\{e(t_j)\}$  je posloupnost náhodných veličin s nulovou střední hodnotou. Jedná se tedy o regresní model s regresní funkcí  $m(t)$ .

V modelu (6.1) se regresní funkce nejčastěji odhaduje pomocí jádrového odhadu, tedy neparametricky. Jádrovým odhadem regresní funkce v bodě  $t_0$ , tedy hodnoty  $m(t_0)$ , je:

$$\hat{m}(t_0) = \frac{\sum_{\{j; |t_0 - t_j| \leq h\}} K\left(\frac{t_0 - t_j}{h}\right) X(t_j)}{\sum_{\{j; |t_0 - t_j| \leq h\}} K\left(\frac{t_0 - t_j}{h}\right)}.$$

Je zřejmé, že jádrový odhad je váženým průměrem hodnot  $\{X(t_j)\}$ , kde jsou váhy definovány pomocí jádrové funkce  $K(x)$ , definované na intervalu  $[-1, 1]$ . Jádrová funkce je jakákoliv nezáporná funkce splňující  $\int_{-1}^1 K(x) dx = 1$ .

Velmi často se používá jádrová funkce, která odpovídá hustotě useknutého normálního rozdělení. V takovém případě mluvíme o gaussovském filtru. Dalším často používaným jádrem je Epanechnikovo jádro, kde funkce  $K(x)$  je dána následovně:

$$K(x) = \frac{3}{4}(1 - x^2), \quad x \in [-1, 1].$$

Tuto jádrovou funkci jsme v neparametrickém odhadu střední hodnoty ročního chodu použili i my.

## 6.2 Odhad střední hodnoty ročního chodu

V tomto oddíle popíšeme, jak se provádí jádrový odhad středního ročního chodu za celé sledované období. V následujícím textu budeme prezentovat výsledky jádrového odhadu střední hodnoty ročního chodu pro dvě období, kde dělicím bodem je rok 1997.

Začněme popisem, jak by se pomocí jádrového odhadu odhadovala regresní funkce pro jeden kalendářní rok. Nejprve se interval  $[1, 365]$  převede lineární transformací na interval  $[0, 1]$ , a tedy se předpokládá, že časové okamžiky, v kterých je řada měřena, jsou rovny  $t_j = j/365$ ,  $j = 1, \dots, 365$ , a hodnota  $X(t_j)$ ,  $j = 1, \dots, 365$ , odpovídá průměrnému dennímu



průtoku v  $j$ -tém kalendářním dni. (Tento krok je čistě technický a nebylo by ho třeba provádět.) Důležitá je však volba parametru  $h$ , kterému se anglicky říká "bandwidth", a česky "šířka okna". V našem případě jsme volili  $h = 0.05$ . Abychom mohli odhadnout hodnotu funkce  $m$  i na začátku a konci intervalu  $[0, 1]$ , přidali jsme hodnoty z konce intervalu před bod 0 a hodnoty z počátku za bod 1.

Nyní předpokládejme, že v časových okamžicích  $t_j = j/365$ ,  $j = 1, \dots, 365$  jsou známy hodnoty nejen v jednom kalendářním roce, ale v  $N$  po sobě jdoucích letech. Označme  $X_i(t_j)$ ,  $j = 1, \dots, 365$ ,  $i = 1, \dots, N$ , hodnotu průměrného denního průtoku v kalendářním dnu  $j$  v roce  $i$ . Předpokládejme, že  $X_i(t_j)$  splňují regresní model se stejnou regresní funkcí  $m$ :

$$X_i(t_j) = m(t_j) + e_i(t_j), \quad j = 1, \dots, 365, \quad i = 1, \dots, N,$$

kde  $\{e_i(t_j)\}$  jsou nezávislé, stejně rozdělené a mají nulovou střední hodnotu a rozptyl  $\sigma_e^2$ .

Naším cílem je odhadnout regresní funkci  $m$ . Můžeme postupovat buď tak, že provedeme jádrový odhad regresní funkce pro každý rok zvlášť, a tyto jádrové odhady zprůměrujeme, nebo spočteme průměrné denní průtoky v daných kalendářních dnech za celé sledované období  $i = 1, \dots, N$ , a pak provedeme jádrový odhad funkce  $m$  založený na těchto průměrech. To znamená, že odhadneme  $m(t_0)$  následovně:

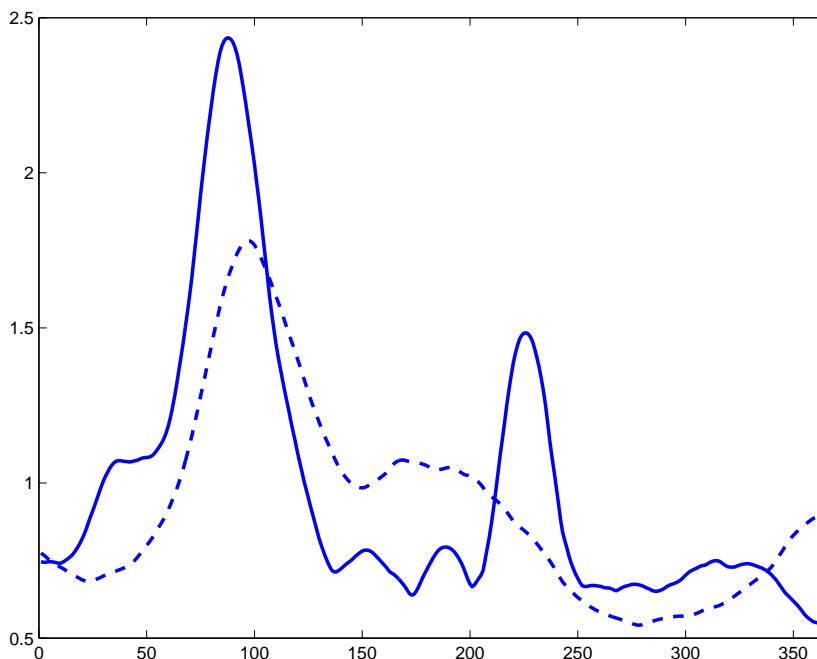
$$\hat{m}(t_0) = \frac{\sum_{\{j; |t_0 - t_j| \leq h\}} K\left(\frac{t_0 - t_j}{h}\right) \bar{X}(t_j)}{\sum_{\{j; |t_0 - t_j| \leq h\}} K\left(\frac{t_0 - t_j}{h}\right)},$$

kde  $\bar{X}(t_j) = (1/N) \sum_{i=1}^N X_i(t_j)$ . Vzhledem k tomu, že jádrový odhad je vážený průměr, budou se odhady při použití obou metod shodovat.

### 6.3 Neparametrický odhad středních hodnot ročního chodu před a po roce 1997

V předchozích kapitolách jsme porovnávali sezónní chování průtoků před rokem 1997 a po něm. V mnoha případech jsme zjistili, že toto chování je statisticky významně odlišné. (Rok 1997 jsme zvolili proto, že při detekování změny se rok 1997 zdál být časovým bodem, ve kterém docházelo ke změně nejčastěji, viz následující kapitoly.)

Z tohoto důvodu jsme rozdělili data denních průtoků pro každou stanici na soubor dat před rokem 1997 a po roce 1997 a způsobem popsaným v předchozím textu jsme odhadli zvlášť průměrný roční chod před rokem 1997 a po něm. Následující obrázky ukazují takto odhadnuté roční chody, a tedy ilustrují, k jakému typu změny došlo. Uvádíme jen obrázky pro ty stanice,



Obrázek 6.1: Odhady středních ročních chodů - Blanice.

v kterých byla testy detekována změna ve střední ročním chodu. Ve všech obrázcích odpovídá čerchovaná čára odhadu středního ročního chodu před rokem 1997 a tučná čára odhadu středního ročního chodu po roce 1997.

## 6.4 Neparametrický odhad polohy maxima středního ročního chodu

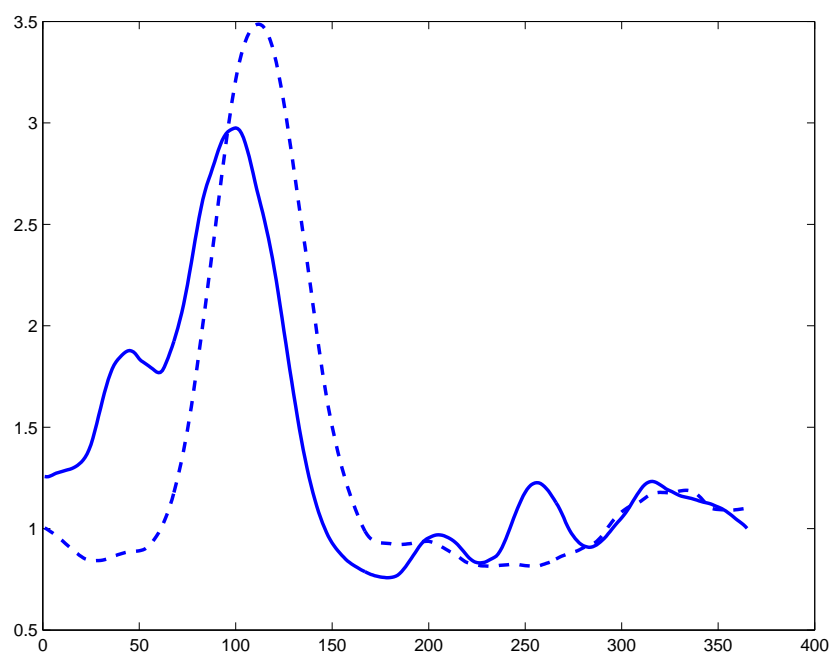
Studiem obrázků 6.1 až 6.8 zjistíme, že jedním z nápadných rysů, kterým se roční chod před rokem 1997 a po roce 1997 liší, je posun času, v kterém dochází k maximálnímu průtoku, směrem k začátku kalendářního roku, viz Horáková a Jarušková (2016). Postup pro neparametrický odhad argumentu maxima regresní funkce navrhli v článcích Gasser a Müller (1984) a Müller (1985).

Předpokládejme opět, že okamžiky pozorování jsou ekvidistantní. V našem případě  $t_j = j/365$ ,  $j = 1, \dots, n = 365$ . Veličina  $X(j/n)$  označuje naměřenou hodnotu v čase  $t_j$ . Návrh Gassera a Müllera (1984) spočívá v nepatrné obměně tvaru jádrového odhadu hodnoty funkce v bodě  $t_0$ . Místo odhadu  $\hat{m}(t_0)$  navrhli použít odhad

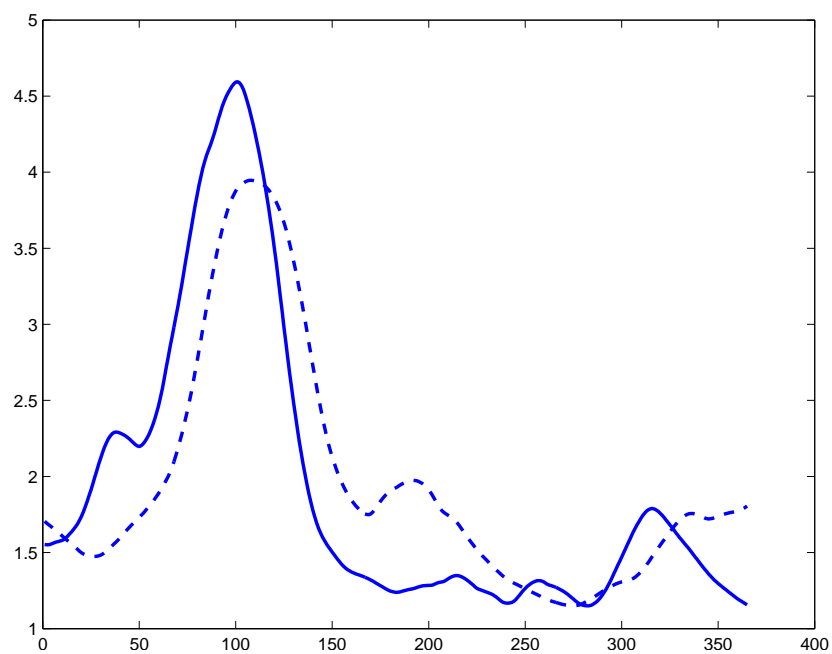
$$\tilde{m}(t_0) = \frac{1}{h} \sum_{\{j; |t_0 - \frac{j}{n}| \leq h\}} \int_{\frac{j-0.5}{n}}^{\frac{j+0.5}{n}} W_0\left(\frac{t_0 - u}{h}\right) du X(j/n).$$

Dále navrhli, aby byla používána funkce:

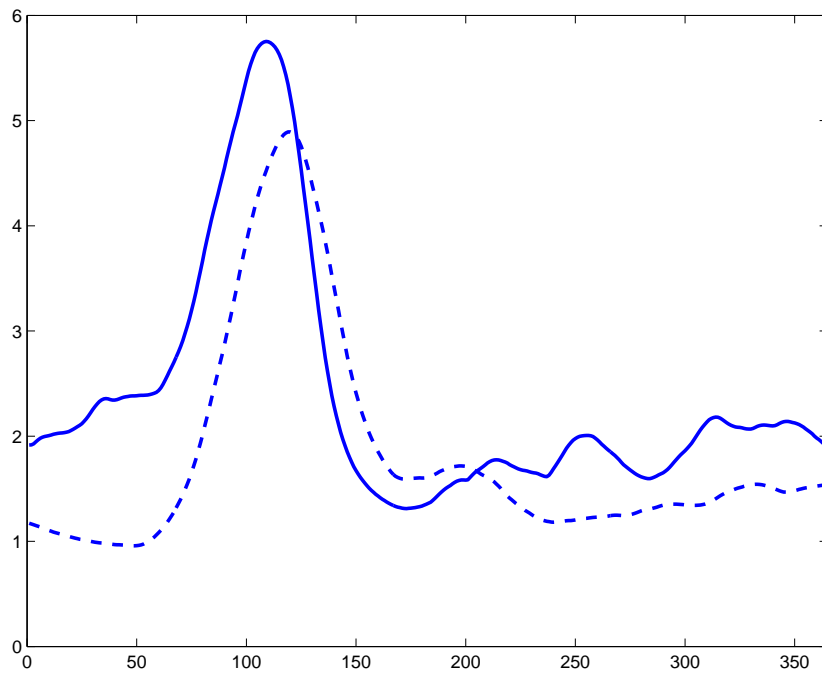
$$W_0(x) = \frac{105}{96}(1 - x^2)^3 \quad \text{pro } x \in [-1, 1].$$



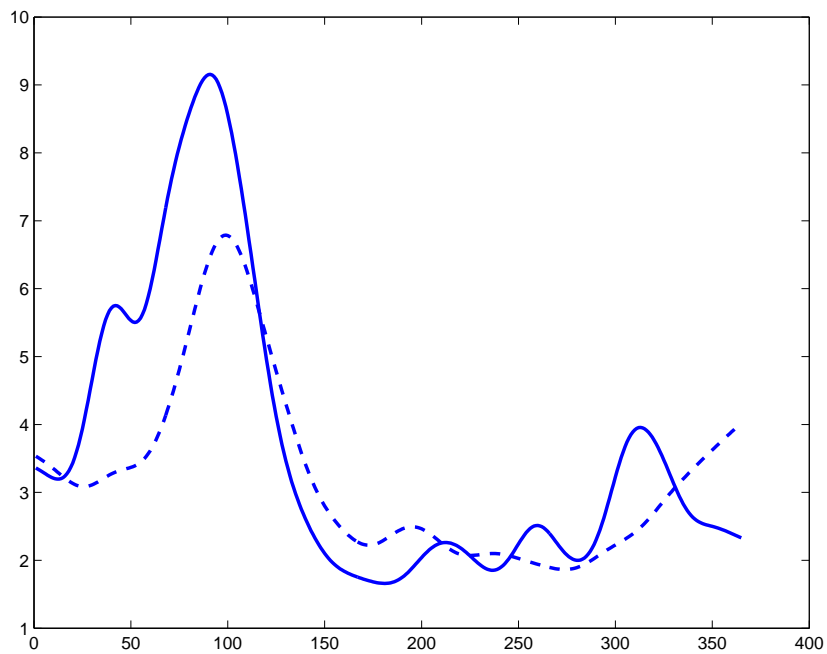
Obrázek 6.2: Odhady středních ročních chodů - Jizerka.



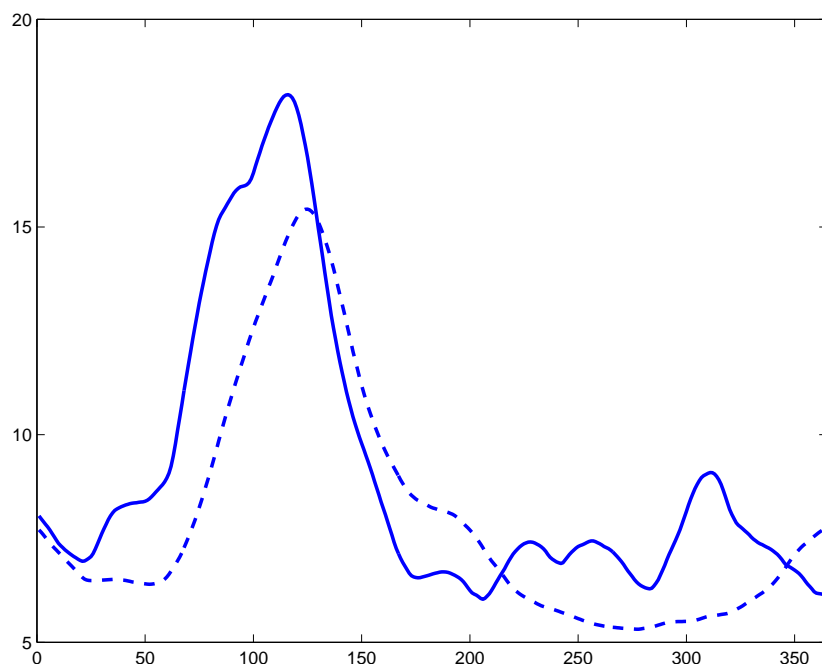
Obrázek 6.3: Odhady středních ročních chodů - Morava.



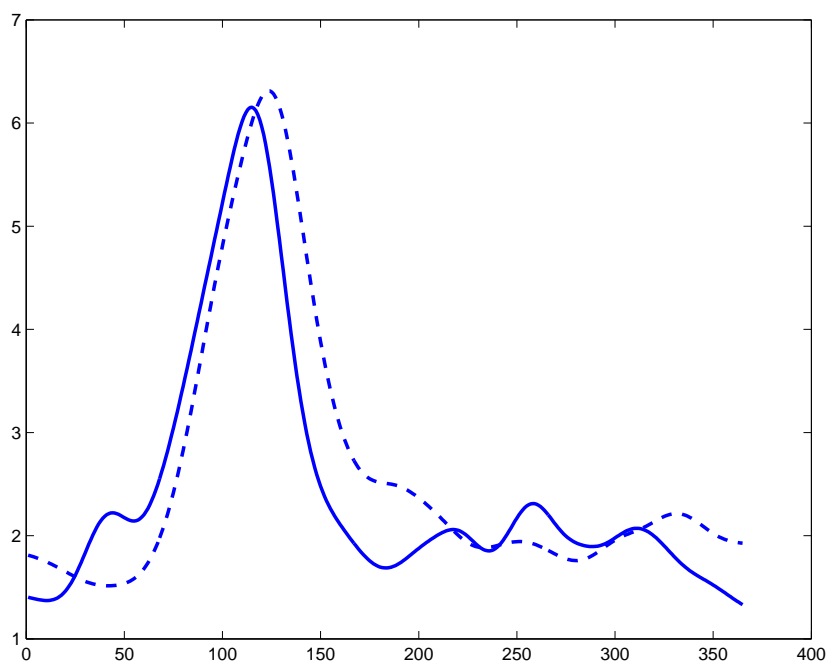
Obrázek 6.4: Odhady středních ročních chodů - Mumlava.



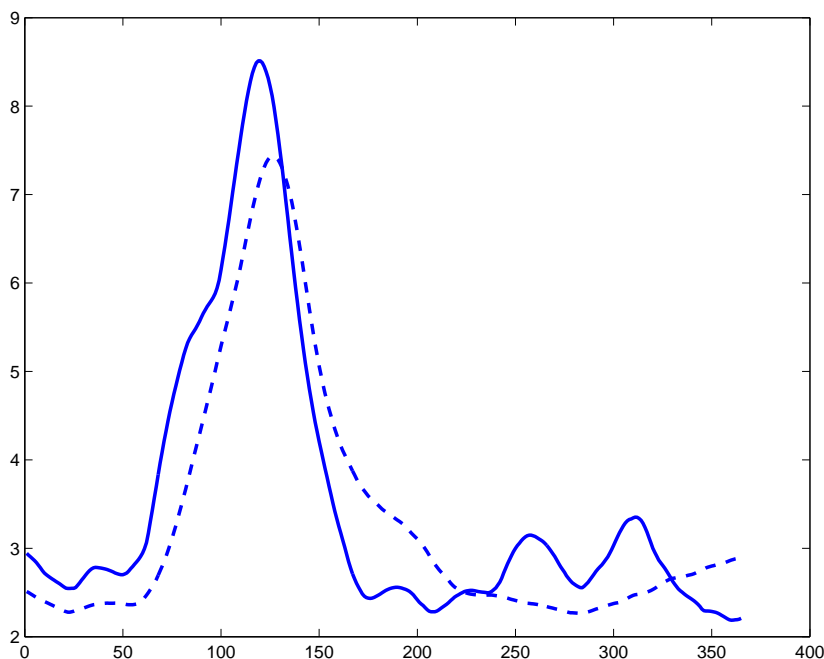
Obrázek 6.5: Odhady středních ročních chodů - Orlice.



Obrázek 6.6: Odhady středních ročních chodů - Otava.



Obrázek 6.7: Odhady středních ročních chodů - Úpa.



Obrázek 6.8: Odhady středních ročních chodů - Vydra.

Konstanta  $h$  opět odpovídá šířce okna. Vzhledem k tomu, že platí

$$\int_{\frac{j-0.5}{n}}^{\frac{j+0.5}{n}} W_0\left(\frac{t-u}{h}\right) du = W_0\left(\frac{t_0 - j/n}{h}\right) \cdot \frac{1}{n} + W_0'\left(\frac{t_0 - \tilde{\theta}_j}{h}\right) \cdot \frac{1}{n^2},$$

kde  $\tilde{\theta}_j \in \left[\frac{j-0.5}{n}, \frac{j+0.5}{n}\right]$ , je zřejmé, že odhad  $\tilde{m}(t_0)$  je velmi podobný odhadu  $\hat{m}(t_0)$ , kde  $W_0(x)$  je jádrová funkce.

Odhad navržený Gasserem a Müllerem (1984) však umožní také odhad první a druhé derivace funkce  $m$ , a odtud pak bodový a intervalový odhad argumentu maxima funkce  $m$ . Pro první a druhou derivaci funkci  $W_0(x)$  platí:

$$W_0'(x) = \frac{105}{16}(-x + 2x^3 - x^5), \quad x \in [-1, 1]$$

a

$$W_0''(x) = \frac{105}{16}(-1 + 6x^2 - 5x^4), \quad x \in [-1, 1].$$

Za odhad první a druhé derivace funkce  $m$  v bodě  $t_0$ , kde  $t_0 = m_0/n$  pro nějaké  $m_0$ , můžeme brát

$$\hat{m}'(t_0) = \frac{1}{h^2} \sum_{j=t_0.n-n.h+0.5}^{t_0.n+n.h-0.5} \int_{\frac{j-0.5}{n}}^{\frac{j+0.5}{n}} W_0'\left(\frac{t_0-x}{h}\right) dx X(j/n)$$

a za odhad druhé derivace

$$\hat{m}''(t_0) = \frac{1}{h^3} \sum_{j=t_0.n-n.h+0.5}^{t_0.n+n.h-0.5} \int_{\frac{j-0.5}{n}}^{\frac{j+0.5}{n}} W_0''\left(\frac{t_0-x}{h}\right) dx X(j/n).$$

Bod maxima funkce  $m$  je pak odhadován jako argument  $t_0$ , kde  $\widehat{m}'(t_0) = 0$ . Takových bodů může být samozřejmě víc. My vybereme takový, aby v něm odhad funkce  $\widehat{m}(t)$  měl největší hodnotu. Zkušební odborníci v oboru jadrových odhadů doporučují, aby se šíře okna volila v závislosti na  $n$ , přičemž by se neměla volit stejná pro odhad průběhu funkce a pro odhad průběhu derivace. Nejčastěji doporučují, aby šíře okna pro odhad průběhu funkce byla zvolena  $h_n = n^{-1/7}$  a pro derivaci  $h_n = n^{-1/9}$ . Ze vztahu  $\widehat{m}'(t_0) = 0$  za předpokladu, že  $\{e(t_j)\}$  jsou stejně rozdělené nezávislé náhodné veličiny s rozptylem  $\sigma_e^2$ , Gasser a Müller odvodili asymptotické chování odhadu argumentu maxima:

$$\widehat{t}_0 \sim N\left(t_0, \frac{\sigma_e^2 \cdot V'}{nh_n^3 [m''(t_0)]^2}\right) \quad (6.2)$$

kde  $\sigma_e^2$  je rozptyl chyb  $\{e(j/n)\}$ ,  $m''(t_0)$  je druhá derivace funkce  $m$  v bodě  $t_0$  a  $V' = \int_{-1}^1 [W_0'(x)]^2 dx = \frac{105}{33}$ . Hodnoty  $\sigma_e^2$  a  $m''(t_0)$  neznáme a musíme je odhadnout.

## 6.5 Neparametrický odhad posunu mezi dobami, kde dochází k maximu středního ročního chodu

V následujícím textu se budeme snažit odhadnout posun mezi dobami, kde dochází k maximu ročního chodu. Předpokládejme, že data v prvním i v druhém období mohou být modelována neparametrickým regresním modelem, přičemž je regresní funkce vyjadřující průměrný roční chod v daném kalendářním roce stejná pro všechny roky v prvním období (před rokem 1997) a obdobně je stejná pro všechny kalendářní roky v druhém období (po roce 1997):

$$X_i(t_j) = m_1(j/365) + e_i(j/365), \quad j = 1, \dots, 365, \quad i = 1, \dots, n,$$

$$X_i(t_j) = m_2(j/365) + e_i(j/365), \quad j = 1, \dots, 365, \quad i = n + 1, \dots, N = n + m,$$

kde  $n$  je nyní počet roků v prvním období a  $m$  je počet roků v druhém období. (Zde jsme se na chvíli vrátili k našemu původnímu značení.) Funkce  $m_1(t)$  a  $m_2(t)$  jsou definované na intervalu  $[0, 1]$ . Postupem popsaným v předchozím textu můžeme odhadnout argumenty maxima funkce  $m_1(t)$  i  $m_2(t)$ . Odhad posunu mezi dobami, kde dochází k maximu ročního chodu, odhadneme jako rozdíl argumentu maxima funkce  $m_1(t)$  a argumentu maxima funkce  $m_2(t)$ . Označme odhad prvního argumentu  $\widehat{t}_0(1)$  a odhad druhého argumentu  $\widehat{t}_0(2)$ . Necht' odhady jejich rozptylů jsou  $s_{\widehat{t}_0(1)}^2$  a  $s_{\widehat{t}_0(2)}^2$ . Díky tomu, že vektory denních průtoků v prvním období a vektory denních průtoků v druhém období považujeme za nezávislé, a díky asymptotické

normalitě odhadů argumentů maxima v jádrových odhadech, můžeme zkonstruovat přibližný 95% interval spolehlivosti následovně:

$$(\hat{t}_0(1) - \hat{t}_0(2) - 2\sqrt{s_{\hat{t}_0(1)}^2 + s_{\hat{t}_0(2)}^2}, \hat{t}_0(1) - \hat{t}_0(2) + 2\sqrt{s_{\hat{t}_0(1)}^2 + s_{\hat{t}_0(2)}^2}).$$

Tabulka 6.1 ukazuje odhad posunu doby, kdy dochází k maximálnímu průtoku, a jeho 95% asymptotický interval spolehlivosti. V prvním sloupci je název toku. V druhém sloupci je dán odhad pořadí dne (v rámci kalendářního roku), pro který byl odhad průměrného ročního chodu nejvyšší pro období do roku 1997. Ve třetím sloupci je tentýž odhad, ale pro období po roce 1997. Ve čtvrtém sloupci je odhad posunu kalendářního dne (rozdíl mezi hodnotou v prvním a druhém sloupci), v kterém docházelo v prvním a druhém období k maximálnímu průtoku, tj. odhad posunu času kulminačního průtoku. V pátém sloupci je odhad směrodatné odchylky odhadu posunu doby kulminačních průtoků. V šestém sloupci je dolní mez 95% intervalu spolehlivosti pro časový posun kulminačních průtoků mezi dvěma obdobími a v sedmém sloupci je obdobně horní mez 95% intervalu spolehlivosti pro tentýž časový posun.



tok	max do 1997	max od 1997	rozdíl	s.odch. rozdílu	dolní mez	horní mez
Bělá	96	92	4	3.9	-3.7	11.7
Blanice	96	88	8	3.7	0.6	15.4
Brodečka	88	85	3	2.3	-1.6	7.6
Čeladenka	103	94	9	3.5	2.0	16.0
Doubrava	86	84	2	2.7	-3.5	7.5
Jizerka	111	96	15	2.8	9.3	20.7
Kyjovka	88	85	3	3.9	-4.9	10.9
Morava	104	99	5	2.9	-0.8	10.8
Mumlava	120	112	8	2.3	3.4	12.6
Orlice	99	92	7	3.8	-0.6	14.6
Otava	125	116	9	3.7	1.7	16.3
Porubka	80	83	-3	5.8	-14.6	8.6
R. Bečva	94	90	4	4.1	-4.2	12.2
Svratka	91	86	5	2.5	-0.1	10.1
Úpa	124	116	8	1.7	4.6	11.4
Vlára	79	83	-4	3.8	-11.5	3.5
Vydra	127	120.2	7	2.3	2.4	11.6
Zdobnice	95	94	1	3.3	-5.7	7.7

Tabulka 6.1: Tabulka odhadů argumentů maxim před a po roce 1997.

## 6.6 Model s odlišným středním chodem pro různé roky

Model, který uvažuje stejnou regresní funkci ve všech letech, není patrně úplně realistický. Z tohoto důvodu jsme ještě uvažovali model, kde roční chody jsou náhodné funkce a polohy jejich maxima jsou tudíž náhodné veličiny. O středních hodnotách poloh maxim jsme předpokládali, že jsou v prvním období (do roku 1997) stejné a rovněž jsou stejné v druhém období (po roce 1997). O rozptylech poloh maxim  $\sigma_j^2$  jsme předpokládali, že je možné je odhadnout pomocí (6.2). Střední hodnoty umístění maxim v prvním i druhém období jsme odhadovali jako vážené průměry s vahami úměrnými  $1/\hat{\sigma}_j^2$ .

Za velmi obecných předpokladů mají vážené průměry asymptoticky normální rozdělení, a je tudíž možno zkonstruovat 95% asymptotický interval spolehlivosti. Další použitelnou metodou pro odhad intervalu spolehlivosti je bootstrap. Bootstrapový interval spolehlivosti jsme

konstruovali následovně. Nejprve jsme pro každý rok neparametricky odhadli polohu maxima. Tyto odhady jsme rozdělili na dva soubory - odhadnuté polohy maxim pro první období a odhadnuté polohy maxim pro druhé období. Z obou souborů jsme vytvářeli bootstrapové výběry (výběry s opakováním z naměřených dat) a počítali hodnotu odhadu střední hodnoty pro první období i pro druhé období a rozdíl mezi těmito odhady. Postup jsme mnohokrát opakovali. Jako dolní mez 95% bootstrapového intervalu spolehlivosti jsme vzali 2.5% empirický kvantil a jako horní mez 97.5% empirický kvantil bootstrapových odhadů rozdílu středních hodnot. Více o metodě bootstrap viz Efron a Tibshirani (1993).

Tabulka 6.2 uvádí dolní a horní meze 95% intervalu spolehlivosti pro posun doby (ve dnech), ve kterých odhadujeme, že dochází k jarnímu kulminačnímu průtoku. V druhém a třetím sloupci je dolní a horní mez 95% asymptotického intervalu spolehlivosti pro posun doby (ve dnech), ve kterém dochází k jarnímu kulminačnímu průtoku. Ve čtvrtém a pátém sloupci jsou opět dolní a horní meze 95% intervalu spolehlivosti pro posun získané metodou bootstrap. Je vidět, že intervaly jsou velmi široké. Odhad je tudíž velmi nepřesný.

tok	d. mez (asym)	h. mez (asym)	d. mez (bootst)	h. mez (bootst)
Bělá	-8.6	41.7	-5.5	41.5
Blanice	-3.2	35.8	-0.7	36.0
Brodečka	-7.6	39.7	-4.9	39.6
Čeladenka	-3.7	26.9	-1.8	26.7
Doubrava	-16.5	28.7	-13.2	28.1
Jizerka	8.8	47.1	11.2	47.0
Kyjovka	-2.3	43.6	-0.4	42.8
Morava	-7.0	33.9	-3.7	35.1
Mumlava	-1.2	40.2	3.3	40.8
Orlice	-15.8	30.0	-11.9	28.8
Otava	-8.5	52.4	-3.7	53.5
Porubka	17.7	58.9	19.1	57.9
R. Bečva	-4.6	25.6	-3.7	25.1
Svratka	-2.7	37.2	-0.8	36.6
Úpa	-4.7	8.9	-4.9	8.2
Vlára	-21.8	21.6	-20.2	20.5
Vydra	-6.5	57.5	-2.7	55.9
Zdobnice	-11.9	38.9	-8.3	39.6

Tabulka 6.2: 95% intervaly spolehlivosti pro posun doby jarní kulminace ve dnech v období do roku 1997 a po něm.



## Kapitola 7

# Odhad střední hodnoty ročního chodu před a po změně pomocí periodické funkce a odhad posunu mezi dobami, kde roční chody nabývají svého maxima

### 7.1 Odhad střední hodnoty ročního chodu a odhad argumentu maxima střední hodnoty ročního chodu pomocí periodické funkce

V kapitole 5 jsme se kromě jiného zabývali i postupy, ve kterých se regresní funkce modeluje pomocí periodické funkce, to znamená pomocí lineární kombinace sinů a kosinů s několika málo nejmenšími Fourierovými frekvencemi. V mnoha případech postačovalo uvažovat pouze tři Fourierovy frekvence. Tento model budeme používat také v této kapitole. Stejně jako v předchozím textu  $\{X_i(t_j)\}$  odpovídají denním průměrným průtokům měřeným v roce  $i$  v kalendářním dni  $j$ , tj.  $t_j = j/365$ .

Předpokládejme opět regresní model ve tvaru

$$X_i(t_j) = m(t_j) + e_i(t_j), \quad j = 1, \dots, 365, \quad i = 1, \dots, N,$$

kde vektory  $\{(e_i(1/365), \dots, e_i(1)), i = 1, \dots, N\}$  jsou nezávislé, stejně rozdělené, a mají nulovou střední hodnotu a varianční matici  $\Sigma$ . (Všimněme si, že v tomto modelu mohou být průměrné průtoky v jednom kalendářním roce závislé veličiny. Na druhé straně ale je vari-

anční matice vektoru průměrných denních průtoků stejná pro všechny analyzované roky.) Nyní ovšem předpokládáme, že regresní funkce  $m(t)$ ,  $t \in [0, 1]$ , je dána parametricky jako určitá periodická funkce:

$$m(t) = \mu + A_1 \cos 2\pi t + B_1 \sin 2\pi t + A_2 \cos 4\pi t + B_2 \sin 4\pi t + A_3 \cos 6\pi t + B_3 \sin 6\pi t. \quad (7.1)$$

Koeficienty  $\mu$ ,  $A_1$ ,  $B_1$ ,  $\dots$ ,  $B_3$  představují regresní parametry. Předpokládejme, že funkce  $m$  nabývá na intervalu  $[0, 1]$  pouze jednoho maxima, a to ve vnitřním bodě  $[0, 1]$ , který označíme  $t_0$ . Dále předpokládejme, že derivace funkce  $m$  (označme ji  $g = m'$ ) nabývá v bodě  $t_0$  hodnoty nula, tj.  $g(t_0) = m'(t_0) = 0$ .

Vzhledem k tomu, že odhady regresních parametrů na základě dat pouze z jednoho kalendářního roku jsou lineárními funkcemi těchto dat, můžeme parametry shora uvedeného modelu odhadnout následovně:

$$\begin{aligned} \hat{\mu} &= (1/365) \sum_{j=1}^{365} \bar{X}(j/365), \\ \hat{A}_1 &= (2/365) \sum_{j=1}^{365} \bar{X}(j/365) \cos(2\pi j/365), \\ &\vdots \\ \hat{A}_3 &= (2/365) \sum_{j=1}^{365} \bar{X}(j/365) \cos(6\pi j/365), \\ \hat{B}_3 &= (2/365) \sum_{j=1}^{365} \bar{X}(j/365) \sin(6\pi j/365). \end{aligned}$$

Matici  $\Sigma$  odhadneme pomocí reziduálních vektorů  $\{r_i = (r_{i1}, \dots, r_{i365})^T, i = 1, \dots, N\}$ , kde

$$r_{ij} = X_i(t_j) - \hat{\mu} - \hat{A}_1 \cos(2\pi j/365) - \dots - \hat{B}_3 \sin(6\pi j/365), \quad i = 1, \dots, N, \quad j = 1, \dots, 365.$$

Následovně  $\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N r_i r_i^T$ . Varianční matice vektoru  $(\bar{X}(1/365), \dots, \bar{X}(1))^T$  je pak  $\hat{\Sigma}/N$ . Odtud plyne, že odhad varianční matice vektoru  $(\hat{\mu}, \hat{A}_1, \hat{B}_1, \dots, \hat{A}_3, \hat{B}_3)^T$  může být spočten takto:  $(\frac{2}{365})^2 \mathbf{B}(\hat{\Sigma}/N)\mathbf{B}^T$ , kde

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \cos \frac{2\pi}{365} & \cos \frac{4\pi}{365} & \dots & \cos 2\pi \\ \sin \frac{2\pi}{365} & \sin \frac{4\pi}{365} & \dots & \sin 2\pi \\ \cos \frac{6\pi}{365} & \cos \frac{12\pi}{365} & \dots & \cos 6\pi \end{pmatrix}.$$

Pro funkci  $g$ , tj. derivaci funkce  $m$ , platí:  $g(t) = -2\pi A_1 \sin(2\pi t) + 2\pi B_1 \cos(2\pi t) - 4\pi A_2 \sin(4\pi t) + 4\pi B_2 \cos(4\pi t) - 6\pi A_3 \sin(6\pi t) + 6\pi B_3 \cos(6\pi t)$ . Vztah  $g(t_0, A_1, \dots, B_3)$  definuje implicitně funkci  $t_0 = G(A_1, \dots, B_3)$ , přičemž platí

$$\frac{\partial G}{\partial A_1} = -\frac{\frac{\partial g}{\partial A_1}}{\frac{\partial g}{\partial t_0}}, \dots, \frac{\partial G}{\partial B_3} = -\frac{\frac{\partial g}{\partial B_3}}{\frac{\partial g}{\partial t_0}}.$$

Definujeme odhad  $\hat{t}_0$  jako řešení rovnice  $g(\hat{t}_0, \hat{A}_1, \dots, \hat{B}_3) = 0$  nebo  $\hat{t}_0 = G(\hat{A}_1, \dots, \hat{B}_3)$ .

Pro velký počet roků  $N$  má vektor  $(\hat{A}_1, \dots, \hat{B}_3)^T$  asymptoticky normální rozdělení

$$\sqrt{N} \begin{pmatrix} \hat{A}_1 - A_1 \\ \vdots \\ \hat{B}_3 - B_3 \end{pmatrix} \sim \mathbf{N}(\mathbf{0}, \Sigma_{CS}),$$

kde  $\Sigma_{CS} = (\frac{2}{365})^2 \mathbf{B}_{(-1)} \Sigma \mathbf{B}_{(-1)}^T$  a  $\mathbf{B}_{(-1)}$  je matice  $\mathbf{B}$  po vynechání prvního řádku. Použitím  $\Delta$ -metody lze ukázat, že  $\sqrt{N}(\hat{t}_0 - t_0)$  má asymptoticky normální rozdělení s rozptylem  $\sigma_{\hat{t}_0}^2 = (\frac{\partial G}{\partial A_1}, \dots, \frac{\partial G}{\partial B_3}) \Sigma_{CS} (\frac{\partial G}{\partial A_1}, \dots, \frac{\partial G}{\partial B_3})^T$ . Ve skutečnosti rozptyl  $\sigma_{\hat{t}_0}^2$  neznáme a musíme ho nahradit konzistentním odhadem

$$\hat{\sigma}_{\hat{t}_0}^2 = \left( \frac{\partial G}{\partial A_1}(\hat{A}_1, \dots, \hat{B}_3), \dots, \frac{\partial G}{\partial B_3}(\hat{A}_1, \dots, \hat{B}_3) \right) \hat{\Sigma}_{CS} \left( \frac{\partial G}{\partial A_1}(\hat{A}_1, \dots, \hat{B}_3), \dots, \frac{\partial G}{\partial B_3}(\hat{A}_1, \dots, \hat{B}_3) \right)^T,$$

$$\text{kde } \hat{\Sigma}_{CS} = (\frac{2}{365})^2 \mathbf{B}_{(-1)} \hat{\Sigma} \mathbf{B}_{(-1)}^T.$$

## 7.2 Odhad posunu mezi dobami, kde dochází k maximu ročního chodu, v regresním modelu s periodickou regresní funkcí

V následujícím textu se budeme opět snažit odhadnout posun mezi dobami, kde střední roční chody nabývají svého maxima. Nyní ovšem předpokládejme, že data v prvním i v druhém období mohou být modelována regresním modelem s periodickou regresní funkcí, přičemž je regresní funkce vyjadřující průměrný roční chod v daném kalendářním roce stejná pro všechny kalendářní roky v prvním období (před rokem 1997) a obdobně je stejná pro všechny roky v druhém období (po roce 1997):

$$X_i(t_j) = m_1(j/365) + e_i(j/365), \quad j = 1, \dots, 365, \quad i = 1, \dots, n,$$

$$X_i(t_j) = m_2(j/365) + e_i(j/365), \quad j = 1, \dots, 365, \quad i = n+1, \dots, N = n+m,$$

kde  $n$  je nyní počet roků v prvním období,  $m$  je počet roků v druhém období a  $N = n + m$  je celkový počet všech roků. (Zde jsme se opět na chvíli vrátili k našemu původnímu značení.)

Funkce  $m_1(t)$  a  $m_2(t)$  jsou definované na intervalu  $[0, 1]$  předpisem (7.1).

Postupem popsaném v předchozím textu můžeme odhadnout argumenty maxima funkce  $m_1(t)$  i  $m_2(t)$ . Odhad posunu mezi dobami, kde dochází k maximu ročního chodu, odhadneme jako rozdíl argumentu maxima funkce  $m_1(t)$  a argumentu maxima funkce  $m_2(t)$ . Označme odhady těchto argumentů maxim jako  $\hat{t}_0(1)$  a  $\hat{t}_0(2)$ . Dále označme odhady jejich rozptylů  $s_{\hat{t}_0(1)}^2$  a  $s_{\hat{t}_0(2)}^2$ . Díky tomu, že vektory denních průtoků v prvním období a vektory denních průtoků v druhém období považujeme za nezávislé, a díky asymptotické normalitě odhadů argumentů maxima, kterou jsme dokázali pomocí  $\Delta$  metody, můžeme zkonstruovat přibližný 95% interval spolehlivosti následovně:

$$\left( \hat{t}_0(1) - \hat{t}_0(2) - 2\sqrt{s_{\hat{t}_0(1)}^2 + s_{\hat{t}_0(2)}^2}, \hat{t}_0(1) - \hat{t}_0(2) + 2\sqrt{s_{\hat{t}_0(1)}^2 + s_{\hat{t}_0(2)}^2} \right).$$

V Tabulce 7.1 jsou uvedeny údaje, které byly získány použitím shora uvedeného modelu. V prvním sloupci je název toku. V druhém sloupci je dán odhad pořadí dne (v rámci kalendářního roku), pro který byl odhad průměrného ročního chodu nejvyšší pro období do roku 1997. Ve třetím sloupci je tentýž odhad, ale pro období po roce 1997. Ve čtvrtém sloupci je odhad posunu kalendářního dne (rozdíl mezi hodnotou v prvním a druhém sloupci), v kterém docházelo v prvním a druhém období k maximálnímu průtoku, tj. odhad posunu času kulminačního průtoku. V pátém sloupci je odhad směrodatné odchylky odhadu posunu doby kulminačních průtoků. V šestém sloupci je dolní mez 95% intervalu spolehlivosti pro časový posun kulminačních průtoků mezi dvěma obdobími a v sedmém sloupci je obdobně horní mez 95% intervalu spolehlivosti pro tentýž časový posun.

Všimněme si, že spočtené intervaly spolehlivosti pro odhad posunu doby, v které dochází k největšímu průtoku, jsou relativně dlouhé. To je způsobeno nejen náhodným chováním ročních chodů v jednotlivých letech, ale i chybou modelu. Roční chody nejsou jistě přesně dány lineární kombinací sinů a kosinů odpovídajícím třem nejmenším Fourierovým frekvencím.

Na druhou stranu si všimněme, že odhad posunu je pro všechny stanice kladný, což vypovídá o tom, že jarní kulminace opravdu začíná v posledních letech dříve. To jistě ještě podtrhuje fakt, že pro 11 z 18 studovaných stanic neobsahuje 95% interval spolehlivosti nulu. Pokud bychom tedy na hladině významnosti 0.05 testovali, zda dochází k posunu jarních kulminačních průtoků při použití shora popsaného modelu, pak bychom pro 11 z 18 toků nulovou hypotézu, která tvrdí, že k posunu nedochází, zamítli.



tok	max do 1997	max od 1997	rozdíl	sm. odch. rozdílu	dolní mez	horní mez
Bělá	95.4	85.6	9.9	6.3	-2.8	22.5
Blanice	101.3	86.7	14.6	5.0	4.7	24.5
Brodečka	87.5	84.2	3.2	1.0	1.2	5.2
Čeladenka	105.0	90.5	14.5	6.0	2.5	26.6
Doubrava	83.5	73.4	10.1	7.3	-4.5	24.7
Jizerka	111.8	92.2	19.6	7.6	4.4	34.7
Kyjovka	90.2	85.8	4.4	11.8	-19.3	28.1
Modrava	108.3	93.5	14.7	5.9	3.0	26,4
Mumlava	117.7	102.8	14.9	4.4	6.1	23.8
Orlice	100.2	82.8	17.4	6.3	4.8	29.9
Otava	123.1	106.0	17.0	5.8	5.4	28.7
Porubka	84.0	79.9	4.1	8.4	-12.7	20.9
R. Bečva	94.4	80.7	13.8	6.4	1.0	26.5
Svratka	89.5	75.9	13.6	7.4	-1,1	28.3
Úpa	121.0	109.1	11.9	5.5	0.8	23.0
Vydra	125.1	114.2	10.9	5.1	0.7	21.1
Vlára	73.3	70.2	3.1	7.6	-12.2	18.3
Zdobnice	97.5	86.9	10.6	6.5	-2.4	23.5

Tabulka 7.1: Odhady doby (ve dnech), kdy dochází k jarní kulminaci.



## Kapitola 8

# Testování existence bodu změny použitím metod analýzy bodu změny

### 8.1 Testová statistika maximálního typu

Metody bodu změny se používají ke zjištění, zda pozorovaná posloupnost  $\mathbf{X}_1, \dots, \mathbf{X}_N$  je stacionární nebo zda ji lze rozdělit na dvě části  $\mathbf{X}_1, \dots, \mathbf{X}_n$  a  $\mathbf{X}_{n+1}, \dots, \mathbf{X}_{N=n+m}$ , které jsou stacionární, ale jejich rozdělení je vzájemně odlišné. Na rozdíl od Kapitoly 4 se však předpokládá, že čas, ve kterém ke změně může dojít, je předem neznámý. (Stejně jako v předchozím budeme předpokládat, že  $\mathbf{X}_1, \dots, \mathbf{X}_N$  jsou nezávislé.)

V rámci matematické statistiky se opět rozhoduje pomocí testování hypotéz. Proto je třeba nejprve nalézt vhodnou testovou statistiku. Jedním ze způsobů, jak se testová statistika hledá, je následující úvaha.

Předpokládejme nejprve, že pokud ke změně dochází, pak k ní dojde v čase  $k$ , který je předem znám. V takovém případě se posloupnost  $\mathbf{X}_1, \dots, \mathbf{X}_N$  rozdělí na dvě posloupnosti  $\mathbf{X}_1, \dots, \mathbf{X}_k$  a  $\mathbf{X}_{k+1}, \dots, \mathbf{X}_N$ . Jestliže se rozdělení  $\mathbf{X}_1, \dots, \mathbf{X}_k$  a  $\mathbf{X}_{k+1}, \dots, \mathbf{X}_N$  liší pouze v hodnotě nějakého parametru rozdělení  $\theta$ , a jinak je stejné, pak chceme testovat nulovou hypotézu, která tvrdí, že hodnota parametru v první i druhé části je stejná (a tedy  $\mathbf{X}_1, \dots, \mathbf{X}_N$  je stacionární) proti alternativě, která tvrdí, že parametr odpovídající první a druhé části má různé hodnoty. V takovém případě se jedná o dvouvýběrový parametrický test. Pro speciální případ testování shody dvou středních hodnot výběrů z vícerozměrného normálního rozdělení byl postup popsán v sekci 4.1.

Označme obecně testovou statistiku pro testování shora popsaného dvouvýběrového parametrického testu jako  $U(k)$ . Nulová hypotéza se zamítá, pokud  $U(k)$  je příliš velké. To znamená, že překročí nějakou kritickou hodnotu.

Vraťme se však k situaci, že čas  $k$  neznáme. V takovém případě můžeme spočítat  $U(k)$  pro každou hodnotu  $k$  a dostaneme tak posloupnost testových statistik  $U(1), \dots, U(N-1)$ . Je přirozené zamítnout nulovou hypotézu tehdy, jestliže alespoň jedna ze statistik  $U(1), \dots, U(N-1)$  je velká, což znamená, že nás zajímá

$$\max_{1 \leq k \leq N-1} U(k). \quad (8.1)$$

Tímto způsobem získáme testovou statistiku takzvaného maximálního typu.

Někdy předpokládáme, že ke změně nemohlo dojít v prvních a posledních  $[\beta N]$  časových okamžicích (kde  $0 < \beta < 1$  je malé). V takovém případě použijeme statistiku:

$$\max_{[\beta N] \leq k \leq N - [\beta N]} U(k). \quad (8.2)$$

K nalezení pravidla pro zamítnutí nulové hypotézy musíme znát chování testové statistiky za platnosti nulové hypotézy. Najít přesné rozdělení testové statistiky maximálního typu je však velmi obtížné. Z tohoto důvodu se snažíme najít kritické hodnoty alespoň přibližně.

K nalezení přibližných kritických hodnot slouží tři základní metody, viz Antoch et al. (2002):

- Metoda založená na Bonferroniho nerovnosti
- Metoda asymptotického rozdělení
- Metoda založená na permutačním principu

### Metoda založená na Bonferroniho nerovnosti

Pro libovolný systém jevů  $A_1, \dots, A_m$  platí

$$P\left(\bigcup_{k=1}^m A_k\right) \leq \sum_{k=1}^m P(A_k). \quad (8.3)$$

Nyní pro libovolné reálné  $x$  definujme jevy  $A_k = \{U(k) > x\}$ ,  $k = 1, \dots, N$ . Pak

$$P\left(\max_{1 \leq k \leq N-1} U(k) > x\right) = P\left(\bigcup_{k=1}^{N-1} \{U(k) > x\}\right) \leq \sum_{k=1}^{N-1} P(U(k) > x) \quad (8.4)$$

Předpokládejme, že všechny statistiky  $U(k)$  mají stejné rozdělení. Zvolíme-li nyní  $x$  rovné  $\alpha/(N-1) \cdot 100\%$  hornímu kvantilu  $u_{\alpha/(N-1)}$  statistiky  $U(k)$ , tedy  $P(U(k) > u_{\alpha/(N-1)}) = \alpha/(N-1)$ , pak

$$P\left(\max_{1 \leq k \leq N-1} U(k) > u_{\alpha/(N-1)}\right) \leq \sum_{k=1}^{N-1} P(U(k) > u_{\alpha/(N-1)}) = \alpha,$$

a tedy  $u_{\alpha/(N-1)}$  je konzervativní kritická hodnota. (Skutečná kritická hodnota je menší nebo rovná této hodnotě.)

Použití kritických hodnot získaných z Bonferroniho nerovnosti se doporučuje, pokud  $N$  je malé (zhruba řečeno  $N \leq 100$ ).

### **Metoda asymptotického rozdělení**

V tomto případě se snažíme zjistit limitní chování statistik (8.1) a (8.2) pro  $N$  blízcí se nekonečnu. Statistika (8.1) obvykle konverguje k nekonečnu, ale je možné zjistit, jak rychlá je konvergence. Statistika (8.2) má limitní rozdělení. Problém spočívá v tom, že přesné rozdělení limitní náhodné veličiny je opět natolik složité, že není možné ho využít pro zjištění asymptotických kritických hodnot. Nicméně velmi často můžeme použít přibližné chování limitní náhodné veličiny pro velké hodnoty argumentů, tj. takzvané chování chvostu náhodné veličiny.

### **Metoda založená na permutačním principu**

Pro získání přibližných kritických hodnot, postupujeme následovně. Posloupnost dat mnohokrát po sobě náhodně zpermutujeme (například 1 000 000 krát). Z každé získané permutace spočteme hodnotu statistiky (8.1), respektive (8.2). Ze získané množiny hodnot testové statistiky spočteme  $\alpha$  100% empirický horní kvantil. Ten pak slouží jako odpovídající kritická hodnota. Antoch a Hušková (2001) dokázali, že pro velké rozsahy výběru  $N$  pro případ detekce změny ve střední hodnotě náhodného vektoru se takto získaná přibližná kritická hodnota blíží ke správné kritické hodnotě, a to jak za platnosti nulové hypotézy, tak i alternativy.

Permutační testy jsou testy, které jsou velmi blízké postupům založených na metodě bootstrap. Na rozdíl od bootstrapu se v permutačních testech provádějí výběry bez opakování.

Výhodou postupů založených na statistikách maximálního typu je, že v případě zamítnutí nulové hypotézy ihned získáváme odhad času, v kterém ke změně došlo. Tímto odhadem je argument, v kterém posloupnost testových statistik  $\{U(k)\}$  nabývá svého maxima.

## **8.2 Test shody střední hodnoty posloupnosti náhodných vektorů**

Problematikou detekce bodu změny střední hodnoty náhodných vektorů a jejím použitím v klimatologii se zabývali například Horváth et al. (1999) nebo Jarušková (2010).

Předpokládejme, že pozorujeme posloupnost náhodných vektorů  $\mathbf{X}_1, \dots, \mathbf{X}_N$ . Pro jednoduchost předpokládejme, že všechny vektory jsou nezávislé a mají  $p$ -rozměrné normální rozdělení se stejnou varianční-kovarianční maticí  $\Sigma$ . Střední hodnota  $E \mathbf{X}_i = \boldsymbol{\mu}_i, i = 1, \dots, N$ .

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_N, \quad i = 1, \dots, N; \quad (8.5)$$

$A : \exists k^* \in \{1, \dots, n-1\}$  takové, že

$$\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_{k^*}, \quad (8.6)$$

$$\boldsymbol{\mu}_{k^*+1} = \cdots = \boldsymbol{\mu}_N, \quad \text{ale } \boldsymbol{\mu}_{k^*} \neq \boldsymbol{\mu}_{k^*+1}.$$

Použijeme-li postup popsáný v oddíle 8.1, bude mít testová statistika tvar:

$$U = \max_{1 \leq k \leq N-1} \frac{k(N-k)}{N} (\bar{\mathbf{X}}_k^* - \bar{\mathbf{X}}_k^{**})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{X}}_k^* - \bar{\mathbf{X}}_k^{**}), \quad (8.7)$$

respektive

$$U(\beta) = \max_{[\beta n] \leq k \leq N - [\beta n]} \frac{k(N-k)}{N} (\bar{\mathbf{X}}_k^* - \bar{\mathbf{X}}_k^{**})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\bar{\mathbf{X}}_k^* - \bar{\mathbf{X}}_k^{**}), \quad (8.8)$$

kde

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

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

Použijeme-li Bonferroniho nerovnost, pak nulovou hypotézu zamítáme na hladině významnosti  $\alpha$ , jestliže  $U > \chi_{\alpha/(N-1)}^2(p)$ , kde  $\chi_{\alpha/(N-1)}^2(p)$  je  $\alpha/(N-1) \cdot 100\%$  horní kvantil  $\chi^2$ -rozdělení o  $p$  stupních volnosti. Připomeňme, že uvedený postup by byl zcela správný, pokud bychom matici  $\boldsymbol{\Sigma}$  znali. Pokud matici  $\boldsymbol{\Sigma}$  nahradíme maticí  $\widehat{\boldsymbol{\Sigma}}$ , mají jednotlivé statistiky rozdělení  $\chi^2$  jen přibližně. Nicméně chyba, kterou se takto dopustíme, je relativně malá.

Pokud  $N$  je velké, můžeme použít asymptotické chování (8.7) a (8.8). Csörgő M. and Horváth L. (1997) dokázali, že

$$P\left(U > 2 \log \log N + p \log \log \log N - 2 \log \Gamma\left(\frac{p}{2}\right) + 2t\right) \rightarrow 1 - \exp\{-2e^{-t}\} \quad \text{as } N \rightarrow \infty, \quad (8.9)$$

$$U(\beta) \xrightarrow{\mathcal{D}} \max_{\beta \leq t \leq 1-\beta} \frac{B_1^2(t) + \cdots + B_p^2(t)}{t(1-t)} \quad \text{as } N \rightarrow \infty, \quad (8.10)$$

kde  $\{B_i(t), 0 \leq t \leq 1\}$ ,  $i = 1, \dots, p$  jsou nezávislé Brownovy můstky.

Pro velká  $u$  platí

$$P\left(\max_{\beta \leq t \leq 1-\beta} \frac{B_1^2(t) + \cdots + B_p^2(t)}{t(1-t)} \geq u^2\right) \sim \left(-\log \frac{\beta}{1-\beta}\right) \frac{1}{2^{(N/2)-1}} \frac{\sqrt{2\pi}}{\Gamma(n/2)} u^{N+1} (1-\Phi(u)), \quad (8.11)$$

kde  $\Phi(\cdot)$  je distribuční funkce standardního normálního rozdělení.

Limitní chování (8.9) a (8.10) spolu s (8.11) můžeme použít k nalezení přibližných kritických hodnot testu, pokud je  $N$  velké.

Poslední možností je použít permutační princip, který byl popsán v sekci 8.1, viz Hušková (1998).

Je zřejmé, že shora uvedený test je možné použít nikoliv pouze pro původní vektory, jejichž složky jsou denní průměrné průtoky, ale také pro libovolnou posloupnost náhodných vektorů, které vznikly redukcí dimenze původních vektorů, to je například pro posloupnost průměrných měsíčních průtoků, Fourierovy koeficienty a hlavní komponenty, viz Berkes et al. (2009) nebo Aston a Kirch (2012).

Vhledem k tomu, že analyzované řady ročních průtoků jsou velmi krátké, jsou metody pro detekci bodu změny použitelné jen s obtížemi. Ze všech shora popsaných postupů pro rozhodnutí, zda se sezónní chování řady nezměnilo, je asi nejvhodnější odhad  $p$ –hodnoty pomocí permutačního principu. Tabulka 8.1 uvádí přibližné  $p$ –hodnoty odpovídajících testů pro stanice, pro které alespoň jednou byla nulová hypotéza zamítnuta na hladině významnosti  $\alpha = 0.1$ . Poznamenejme, že odhady  $p$ –hodnot v Tabulce 8.1 nejsou příliš přesné. Na druhé straně se seznam stanic, které metoda detekce bodu změny identifikovala jako řady, kde mohlo dojít ke změně sezónního chování, poměrně dobře shoduje se seznamem stanic, kde jsme detekovali změnu dělením řady na dvě části, přičemž dělícím bodem byl rok 1997.

tok	metoda hlav. komp.	metoda aprox. Four. řadou	metoda měs. průměrů
Blanice	0.03	0.005	0.1
Čeladenka	nezamítáme $H_0$	nezamítáme $H_0$	0.1
Doubrava	0.08	nezamítáme $H_0$	nezamítáme $H_0$
Jizerka	0.1	0.1	0.1
Morava	0.01	nezamítáme $H_0$	0.1
Mumlava	0.08	0.03	0.01
Orlice	nezamítáme $H_0$	nezamítáme $H_0$	0.1
Otava	0.04	0.05	0.09
Vydra	0.08	0.1	0.1

Tabulka 8.1: Přibližné  $p$ –hodnoty získané pomocí permutačního principu.

Jak už bylo řečeno, získáme při použití shora uvedeného postupu také odhad bodu, v kterém (nebo v blízkosti kterého) došlo ke změně. V případě analyzovaných řad pomocí metod analýzy bodu změny byl nejvěrohodnějším bodem změny rok 1997.





## Kapitola 9

# Detekce více změn

Předpokládejme opět, že v ekvidistantních časových okamžicích  $i = 1, \dots, N$  pozorujeme posloupnost  $p$ -rozměrných vektorů  $\{\mathbf{X}_i\}$ . Zajímá nás opět stacionarita této posloupnosti. Speciálně se zajímáme o to, zda vektor středních hodnot této posloupnosti  $\boldsymbol{\mu}_i = E \mathbf{X}_i$  zůstává během doby měření stále stejný, tj.  $\boldsymbol{\mu}_i = \boldsymbol{\mu}$  pro  $i = 1, \dots, N$ .

V předchozích kapitolách jsme zkoumali situaci, kdy se vektor středních hodnot v nějakém neznámém časovém okamžiku náhle změní. Může se ovšem stát, že takovýto změn může nastat víc. Pokud jsou změny dost velké a jsou od sebe dostatečně vzdálené, pak testová statistika maximálního typu navržená pro testování nejvýše jedné změny, nabude velké hodnoty i v případě více změn, a tedy nulová hypotéza o neměnnosti vektoru středních hodnot se zamítne. Pokud však změny nejsou dostatečně velké, může se stát, že stacionaritu nezamítneme, ačkoliv řada není stacionární. To znamená, že postup založený na testové statistice navržené pro detekci jedné změny ztrácí sílu. To znamená, že v některých případech nezamítneme stacionaritu, ačkoliv řada není stacionární.

Pokud bychom předem tušili, ke kolika změnám teoreticky může ve vektoru středních hodnot dojít, mohli bychom navrhnout testovou statistiku, která by měla větší sílu, a tedy by odhalila i menší změny ve vektoru středních hodnot.

Viděli jsme, že v případě jedné změny byla testová statistika maximálního typu odvozena z dvouvýběrových testových statistik. Speciálně testová statistika byla maximum posloupnosti dvouvýběrových testových statistik, kde maximum bylo uvažováno přes všechny časové okamžiky, v kterých by se mohla řada změnit. Pokud předpokládáme, že v řadě došlo k více než jedné změně, je základním stavebním kamenem testová statistika pro testování shody vektorů středních hodnot více souborů. Postupy, které se zabývají těmito typy problémů, patří do části matematické statistiky známé pod názvem Analýza rozptylu, viz Anděl (1985).

## 9.1 Detekce $d$ bodů změny v případě, kdy je známo, v kterých časových okamžicích by se řada mohla měnit

Předpokládejme, že měřená řada se rozpadá na  $d+1$  úseků  $\mathbf{X}_1, \dots, \mathbf{X}_{n_1}, \dots, \mathbf{X}_{n_d+1}, \dots, \mathbf{X}_N$ , uvnitř kterých zůstává vektor středních hodnot stejný. Zajímá nás, zda se vektor středních hodnot pro některé dva úseky liší.

Podobně jako v předchozím textu rozhodnutí můžeme založit na testování statistických hypotéz. Nulovou hypotézu

$$H_0 : \mathbf{X}_i = \boldsymbol{\mu} + \mathbf{e}_i, \quad i = 1, \dots, N, \quad (9.1)$$

testujeme proti alternativě.

$$\begin{aligned} A : \mathbf{X}_i &= \boldsymbol{\mu}_1 + \mathbf{e}_i, & i &= 1, \dots, n_1, \\ \mathbf{X}_i &= \boldsymbol{\mu}_2 + \mathbf{e}_i, & i &= n_1 + 1, \dots, n_2, \\ & \vdots \\ \mathbf{X}_i &= \boldsymbol{\mu}_{d+1} + \mathbf{e}_i, & i &= n_d + 1, \dots, N, \\ \boldsymbol{\mu}_i &\neq \boldsymbol{\mu}_{i+1} \text{ alespoň pro jedno } i, & 1 \leq i \leq d. \end{aligned} \quad (9.2)$$

Vektory  $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{d+1}$  jsou neznámé. Náhodné vektory  $\{\mathbf{e}_i\}$  jsou nezávislé stejně rozdělené s nulovým vektorem středních hodnot a varianční maticí  $\boldsymbol{\Sigma}$ .

Jestliže jsou všechny délky úseků  $n_1, n_2 - n_1, \dots, N - n_d$  známé, jedná se o problém analýzy rozptylu ve vícerozměrném případě, kde se testuje, zda je vektor středních hodnot  $d+1$  souborů stejný. Testová statistika má tvar:

$$\chi^2(n_1, \dots, n_d) = \sum_{k=1}^{d+1} (\bar{\mathbf{X}}_{k\cdot} - \bar{\mathbf{X}}_{\cdot\cdot})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}}_{k\cdot} - \bar{\mathbf{X}}_{\cdot\cdot}), \quad (9.3)$$

kde  $p$ -rozměrný vektor  $\bar{\mathbf{X}}_{k\cdot} = (\bar{X}_{k\cdot}(1), \dots, \bar{X}_{k\cdot}(p))$  odpovídá vektoru průměrů přes  $k$ -tý úsek, jehož složky jsou  $\bar{X}_{k\cdot}(j) = \sum_{i=n_{k-1}+1}^{n_k} X_i(j) / (n_k - n_{k-1})$ ,  $j = 1, \dots, p$ . Vektor průměrů přes všechny časové okamžiky jsme označili  $\bar{\mathbf{X}} = (\bar{X}(1), \dots, \bar{X}(p))$  se složkami  $\bar{X}(j) = \sum_{i=1}^N X_i(j)$ ,  $j = 1, \dots, p$ . Pokud matice  $\boldsymbol{\Sigma}$  není známá, nahrazujeme ji konzistentním odhadem.

Pokud jsou všechny úseky velké, má testová statistika (9.3) rozdělení  $\chi^2$  o  $d \times p$  stupňů volnosti, viz Anderson (1984).

## 9.2 Detekce $d$ bodů změny v případě, kdy není známo, v kterých časových okamžicích by se řada mohla měnit

Ve většině případů netušíme, kde by se řada mohla změnit. V tom případě testujeme nulovou hypotézu

$$H_0 : \mathbf{X}_i = \boldsymbol{\mu} + \mathbf{e}_i, \quad i = 1, \dots, N, \quad (9.4)$$

proti alternativě

$$\begin{aligned} A_d : \quad & \text{existují } 1 \leq n_1 < n_2 < \dots < n_d < n \text{ takové, že} \\ & \mathbf{X}_i = \boldsymbol{\mu}_1 + \mathbf{e}_i, \quad i = 1, \dots, n_1, \\ & \mathbf{X}_i = \boldsymbol{\mu}_2 + \mathbf{e}_i, \quad i = n_1 + 1, \dots, n_2, \\ & \vdots \\ & \mathbf{X}_i = \boldsymbol{\mu}_{d+1} + \mathbf{e}_i, \quad i = n_d + 1, \dots, N, \\ & \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i+1} \text{ alespoň pro jedno } i, 1 \leq i \leq d. \end{aligned} \quad (9.5)$$

Kdybychom při výběru testové statistiky postupovali stejně jako v případě jedné změny, definovali bychom testovou statistiku jako maximum testových statistik (9.3) přes všechna možná dělení  $n_1 < n_2 < \dots < n_d$ .

Musíme si však uvědomit, že pro stanovení hodnoty testové statistiky je třeba odhadnout vektor středních hodnot pro každý úsek. Aby tento odhad byl kvalitní, je třeba, aby v každém úseku bylo dostatečné množství dat. Z tohoto důvodu místo statistiky maximálního typu, kde se počítá maximum pro všechna dělení časového úseku  $i = 1, \dots, N$ , uvažují se jen ta dělení, pro která každý úsek obsahuje alespoň  $\epsilon N$  dat, přičemž  $0 < \epsilon < 1$  je nějaká malé kladné číslo. Testová statistika má v takovém případě tvar:

$$\chi^2 = \max_{\substack{1 \leq n_1 < n_2 < \dots < n_d < N \\ n_1 \geq \lceil \epsilon N \rceil, n_2 - n_1 \geq \lceil \epsilon N \rceil, \dots, N - n_d \geq \lceil \epsilon N \rceil}} \chi^2(n_1, \dots, n_d) \quad (9.6)$$

Testová statistika je pak zaměřená jen na detekci takových změn, které jsou od sebe  $\epsilon N$  časových okamžiků vzdálené.

## 9.3 Dvě změny - Dunaj

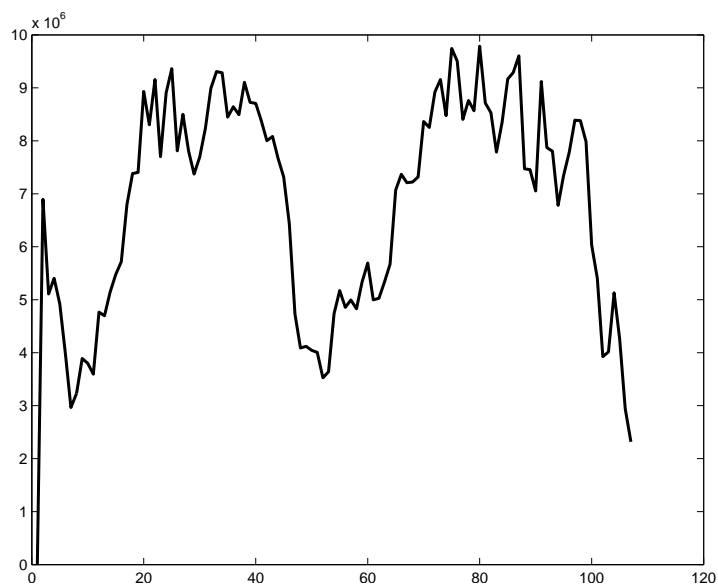
Průtokové řady, které jsme studovali až doposud, jsou příliš krátké, aby v nich bylo možno detekovat více změn. Proto se v tomto oddíle budeme zabývat průtokovou řadou, která byla měřena více než sto let. Jedná se o průtokovou řadu Dunaje, pro kterou jsme v době její analýzy

měli dostupné průměry denních průtoků z období 1901 - 2009. Připomeňme, že zkoumáním průtoků Dunaje se věnovali také Pekarova a Pekar (2006).

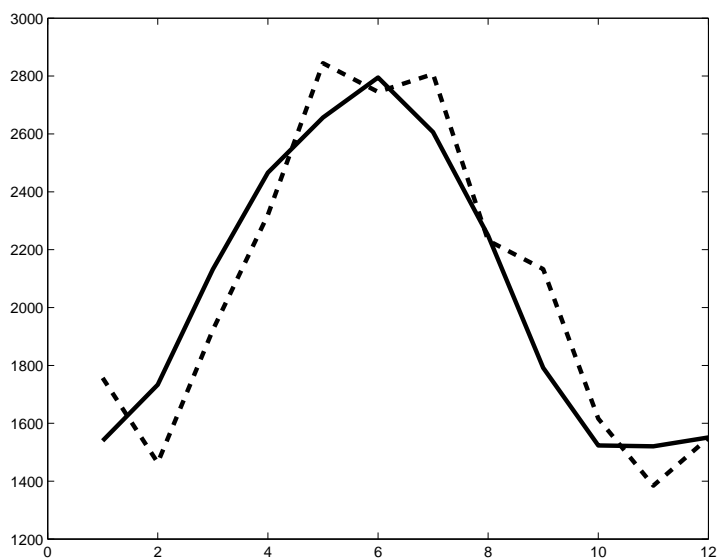
Stejně jako v předchozím textu nás zajímá stacionarita sezónního chování. Změnu (změny) sezónního chování zde budeme detekovat pomocí měsíčních průměrných průtoků. Jak už bylo řečeno, testy založené na měsíčních průměrech bývají často nejslabší, na druhé straně jsou dobře interpretovatelné. Z obrázku průměru měsíčních průměrných průtoků v prvním a druhém období vidíme ihned, k jakému typu změny dochází.

Nejprve použijme test pro detekci jedné změny, popsany v kapitole 8. Vykreslíme-li si posloupnost statistik (9.1), pak vidíme, že velké hodnoty nabývá pro pořadí 20 - 40 (roky 1921 - 1941) a pro pořadí 70 - 90 (roky 1971 - 1991). Znamená to, že v obou těchto obdobích se sezónnost výrazně měnila.

Použijeme-li pro detekci dvou změn test popsany v předchozím oddíle s hodnotou  $\epsilon = 0.15$ , pak testová statistika nabývá hodnoty  $\chi^2 = 48.1$ . Pomocí simulací jsme odhadli odpovídající  $p$  hodnotu, která byla přibližně rovna 0.1. Dále jsme odhadli, že se sezónnost nejvýrazněji změnila kolem roku 1923 (pořadí 22) a roku 1974 (pořadí 73).

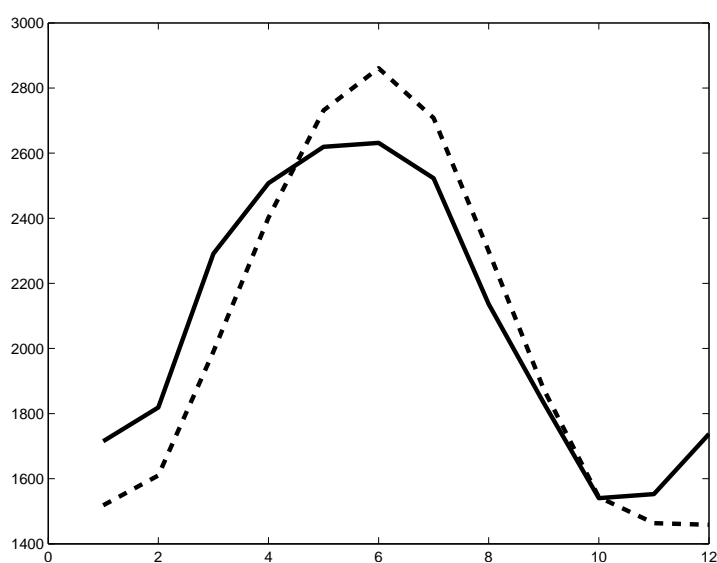


Obrázek 9.1: Průběh testových statistik pro detekci změny vektoru středních hodnot průměrných měsíčních průtoků Dunaje.



Obrázek 9.2: Průměry měsíčních průměrných průtoků do roku 1923 (čerchovaná čára) a od roku 1923 (plná čára).

Podívejme se nyní na obrázky, které nám ilustrují, o jakou změnu se jednalo. V obou případech dochází v druhém období k nárůstu střední hodnoty průměrných měsíčních průtoků pro měsíce únor - duben. Tato změna je po roce 1974 ještě spojena s posunem jarní kulminace směrem k počátku kalendářního roku. Jedná se o stejný jev, který jsme pozorovali také u českých řek. Pro Dunaj však není zcela jasné, zda důvodem tohoto posunu je skutečnost, že srážky již nejsou zadržovány ve sněhové pokrývce během delší doby, jak tomu bylo v první polovině dvacátého století, nebo zda se jedná o umělé zásahy lidí, kteří v zimním a časném jarním období vypouštějí nádrže a přehrady tak, aby byly schopny zadržet vysoké jarní průtoky způsobené táním sněhu.



Obrázek 9.3: Průměry měsíčních průměrných průtoků do roku 1974 (čerchovaná čára) a od roku 1974 (plná čára).

# Kapitola 10

## Odhad společného bodu změny

### 10.1 Metoda pro odhad společného bodu změny

V kapitole 5 jsme pro testování shody dvou úseků vícerozměrné časové řady uvažovali dělicí bod odpovídající roku 1997. To mělo následující důvod. V případě, že jsme studovali každou stanicí odděleně a používali jsme testovou statistiku maximálního typu pro detekci změny, pak posloupnost testových statistik, jejichž maximum hledáme, nabývala velmi často svého maxima buď v bodě odpovídající roku 1997 nebo blízko něho.

V této kapitole budeme předpokládat, že se všechny studované řady mění ve stejný časový okamžik  $k^*$  (případně se některé nemusí měnit vůbec). Cílem je společný bod změny  $k^*$  odhadnout. Nejčastěji se bod změny odhaduje následovně:

$$\hat{k}^* = \operatorname{argmax}_{1 \leq k \leq N_1} \sum_{l=1}^R U^{(l)}(k), \quad (10.1)$$

kde  $\{U^{(l)}(k), k = 1, \dots, N_1\}$  je posloupnost testových statistik odpovídajících  $l$ -té stanici, viz sekce 8.2.

Pro model, který odpovídá situaci, kde dochází ke změně vektoru středních hodnot posloupnosti náhodných vektorů  $\mathbf{X}_1^{(l)}, \dots, \mathbf{X}_N^{(l)}$ :

$$\begin{aligned} \mathbf{X}_i^{(l)} &= \boldsymbol{\mu}_1^{(l)} + \mathbf{e}_i^{(l)}, & i = 1, \dots, k^*, \\ \mathbf{X}_i^{(l)} &= \boldsymbol{\mu}_2^{(l)} + \mathbf{e}_i^{(l)}, & i = k^* + 1, \dots, N, \\ \boldsymbol{\mu}_1 &\neq \boldsymbol{\mu}_2, \end{aligned}$$

mají statistiky  $\{U^{(l)}(k)\}$  za předpokladu, že vektory  $\{\mathbf{e}_i^{(l)}\}$  jsou nezávislé stejně rozdělené řídicí se vícerozměrným normálním rozdělením s nulovým vektorem středních hodnot a variančně-kovarianční maticí  $\boldsymbol{\Sigma}^{(l)}$ , tvar:

$$U^{(l)}(k) = \frac{k(N-k)}{N} (\overline{\mathbf{X}}_k^{*(l)} - \overline{\mathbf{X}}_k^{**l})^T (\widehat{\boldsymbol{\Sigma}}^{(l)})^{-1} (\overline{\mathbf{X}}_k^{*(l)} - \overline{\mathbf{X}}_k^{**l}),$$

kde vektor  $\overline{\mathbf{X}}_k^{*(l)}$  je vektor, jehož složky jsou průměry odpovídajících složek vektorů  $\mathbf{X}_1^{(l)}, \dots, \mathbf{X}_k^{(l)}$ , a dále vektor  $\overline{\mathbf{X}}_k^{**^{(l)}}$  je vektor, jehož složky jsou průměry odpovídajících složek vektorů  $\mathbf{X}_{k+1}^{(l)}, \dots, \mathbf{X}_N^{(l)}$ , a  $\widehat{\Sigma}^{(l)}$  je odhad variančně-kovarianční matice. Index  $l$  pak označuje, že pracujeme s  $l$ -tou stanicí.

Poznamenejme, že za předpokladu, že jsou řady pro různé stanice nezávislé, je odhad bodu změny (10.1) maximálně věrohodným odhadem.

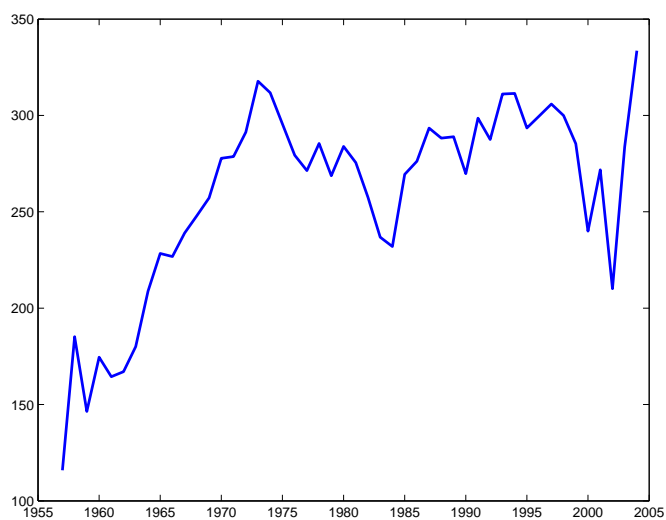
## 10.2 Odhad společného bodu změny pro malé české toky

Je zřejmé, že shora popsáný postup pro odhad společného bodu změny vektoru středních hodnot je možný použít například pro vektory měsíčních průměrů nebo vektory Fourierových koeficientů, a tak podobně.

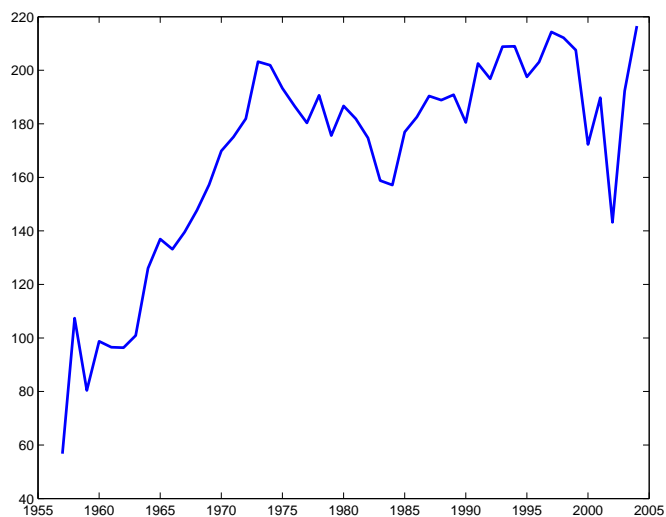
Pro studované průtokové řady malých českých řek, které byly popsány v kapitole 2, jsme použili shora definovaný způsob odhadu  $\widehat{k}^*$  v případě, že se vyšetřuje stabilita sezónního chování pomocí vektorů průměrných měsíčních průtoků. Obrázek 10.1 ukazuje průběh součtu posloupností testových statistik  $\{U^{(l)}(k)\}$  pro všech  $l = 1, \dots, R = 18$  řek. V tomto případě je odhadnut bod změny rokem 1973. (Poslední hodnotu neuvažujeme.)

Obrázek 10.2 ukazuje průběh součtu posloupností testových statistik  $\{U^{(l)}(k)\}$ , kde je součet tvořen jen těmi  $R = 12$  řadami, pro které byla nulová hypotéza, která tvrdí, že posloupnost vektorů středních hodnot je stále stejná, zamítnuta. V tomto případě je odhadnut bod změny rokem 1997. (Poslední hodnotu neuvažujeme.)





Obrázek 10.1: Součet posloupností testových statistik pro detekci změny vektoru středních hodnot pro všech 18 analyzovaných průtokových řad.



Obrázek 10.2: Součet posloupností testových statistik pro detekci změny vektoru středních hodnot pro 12 průtokových řad, pro které byla hypotéza o stacionaritě sezónního chování zamítnuta.



# Kapitola 11

## Závěr

Cílem předkládané práce bylo navrhnout statistické metody pro detekci změny sezónního chování časových řad. Jedná se o metody, které kombinují postupy matematické statistiky známé pod názvem Detekce bodu změny, viz Antoch et al. (2002), s postupy navrženými pro zpracování funkcionálních dat, viz Ramsay a Silverman (2005). Metody byly s výjimkou kapitoly o vícenásobných změnách (zde byla použita průtoková řada Dunaje v Bratislavě) použity pro průtokové řady 18 malých českých toků. Pro aplikaci metod byla vytvořena makra v rámci softwaru Matlab.

Analyzované průtokové řady byly vybrány Doc. Satrapou tak, aby průtokové řady v určených profilech nebyly ovlivněny lidskou činností. Jinak řečeno, data pocházela z měřících stanic, nad kterými se nenacházejí žádné vodní nádrže, kde by se s průtoky uměle manipulovalo. Jednalo se tedy o měřící stanice na řekách blízko jejich pramene, kde jsou průměrné denní průtoky obvykle velice nízké. Takováto data jsou pro statistické zpracování pomocí metod detekce bodu změny jen velmi málo vhodná. Řady jsou totiž poměrně krátké a mají velkou variabilitu. Jedna z hlavních otázek, na kterou se snažila dát práce odpověď, byla, zda je možné navržené postupy použít i na takovéto řady. Ukázalo se, že je lze použít vcelku s úspěchem. Stejně dobře lze použít i metody k odhadu středních ročních chodů a bodových odhadů poloh jejich maxim. Navržené metody pro intervalové odhady (intervaly spolehlivosti) posunu mezi polohou maxima před a po změně nedávaly smysluplné výsledky.

Do práce byla také zařazena kapitola týkající se detekce a odhadu vícenásobných změn. Tato část navazuje na práci Antoch a Jarušková (2013) a ukazuje, jak lze metody navržené těmito autory zobecnit pro detekci více změn ve střední hodnotě náhodných vektorů. Metoda detekce vícenásobných změn ve střední hodnotě vektoru je nová, nebyla nikdy publikovaná. Použití pro detekci změn v sezónním chování průtoků Dunaje ukázalo, že se dá úspěšně použít.

Poslední metoda, která byla do práce zařazena, umožňuje odhad spočteného bodu změny

v případě, že sledujeme více časových řad současně, a kterou jsme aplikovali jednak na všechny analyzované malé české řeky, jednak na řeky, pro které bylo předem ukázáno, že v nich dochází ke změně sezónního chování. Hlavní přínos práce spočívá v tom, že podává rozsáhlý přehled metod vhodných k detekci a popisu změn v sezónním chování průtokových řad. Aplikace metod pak ukázala, které metody jsou vhodné i pro relativně krátké řady a naopak, které nedávají smysluplné výsledky. Především se ukázalo, že pro řady kratší než 100 let není příliš vhodné použití asymptotických postupů (tj. přibližných asymptotických kritických hodnot a asymptotických intervalů spolehlivosti). V takovém případě je vhodnější použít tzv. "resampling" techniky, to znamená metody založené na bootstrapu nebo permutačním principu. Na téma dizertace vzniklo několik článků, z nichž citujeme Horáková a Jarušková (2016), Horáková a Jarušková (2015), Jarušková et al. (2015), Horáková et al. (2012).

## **11.1 Poděkování**

Poděkování patří grantům SGS v letech 2012 - 2016, které vypisuje ČVUT na podporu výzkumné práce studentů a doktorandů, a grantu GAČR 15-09663S, jehož spoluřešitelkou byla vedoucí práce Prof. RNDr. Daniela Jarušková, CSc.

## 11.2 Reference

- ANDĚL J.: *Matematická statistika*, SNTL, 1985.
- ANDERSON T. W.: *An introduction to multivariate statistical analysis*, John Wiley and Sons, 1984.
- ANTOCH J., JARUŠKOVÁ D.: *Testing for multiple change-points*, Computational Statistics 28, 2013, 2161 – 2184.
- ANTOCH J., VORLÍČKOVÁ D.: *Vybrané metody statistické analýzy dat*, Academia, 1992.
- ANTOCH J., HUŠKOVÁ M.: *Permutation tests for a change point analysis*. Statist. Probab. Letters 53, 2001, 37 – 46.
- ANTOCH J., HUŠKOVÁ M., JARUŠKOVÁ D.: *Off-line statistical process control*. in Multivariate Total Quality Control, eds. C. Lauro, J. Antoch, V. E. Vinzi, G. Saporta, 2002, Physica-Verlag, Springer Verlag Company, Heidelberg.
- ASTON J.A.D., KIRCH C.: *Detecting and estimating changes in dependent functional data*. J. of Mult. Anal. 109, 2012, 204 – 220.
- BERKES I., GABRYS R., HORVÁTH L., KOKOSZKA P.: *Detecting changes in the mean of functional observations*. J.R. Stat. Soc. Ser. B stat. Methodol. 71, 2009, 927 – 946.
- CSÖRGŐ M., HORVÁTH L.: *Limit Theorems in Change Point Analysis*, 1997, J. Wiley, New York.
- DAI A., QIAN T., TRENBERTH K. E. MILLIMAN J. D.: *Changes in continental freshwater discharge from 1948 to 2004.*, 2009, J. Climate 22, 2773 – 2792.
- EFRON B., TIBSHIRANI R.J.: *An introduction to the bootstrap.*, 1993, Chapman and Hall.
- GASSER T., MÜLLER H-G.: *Estimating regression functions and their derivatives by the kernel method*. Scand. J. Statist. 11, 1984, 171 – 185.
- HANEL M., VIZINA A., MARTÍNKOVÁ M.: *Changes of draught characteristics in small Czech and Slovak catchments projected by the CMIP5 GCM ensemble*. Hydrology in changing world: Environmental and human dimensions. Book Series: IAHS, 2014, Vol. 363, 78 – 83.
- HANEL M., MRKVIČKOVÁ M., MÁČA P.: *Evaluation of simple statistical downscaling methods for monthly regional climate model simulations with respect to the estimated changes in runoff in the Czech Republic*, Water Resources Management, 2013, vol. 27, 5261 – 5279.

- HORÁKOVÁ H., JARUŠKOVÁ D., SATRAPA L.: *Detekce změny ročního chodu průtokových řad*. In: Voda a krajina 2012 - sborník příspěvků. Praha: Vydavatelství ČVUT v Praze, 2012, s. 83-90. ISBN 978-80-01-05107-8
- HORÁKOVÁ, H.; JARUŠKOVÁ, D.: *Estimation of spring peak timing shift*. In: APLIMAT 2016 - 15th Conference on Applied Mathematics 2016, Proceedings. Bratislava: Slovak University of Technology in Bratislava, 2016, pp. 489-497. ISBN 9788022745314.
- HORÁKOVÁ H., JARUŠKOVÁ D.: *Metody detekce více změn v sezónním chování průtokových řad*. In: Voda a krajina 2016 - sborník příspěvků. Praha: Vydavatelství ČVUT v Praze, 2015, s. 69-76. ISBN 978-80-01-06024-7
- HORÁKOVÁ H., JARUŠKOVÁ D.: *Odhad změny ročních maxim průtokových řad ve dvouvýběrovém případě*. In: Voda a krajina 2015 - sborník příspěvků. Praha: Vydavatelství ČVUT v Praze, 2015, s. 15-24. ISBN 978-80-01-05794-0
- HORVÁTH L., KOKOSZKA P., STEINEBACH J.: *Testing for changes in multivariate dependent observations with an application to temperature changes*. Journal of Multivariate Analysis 68, 1999, 96 – 119.
- IRMAK S., KABENGE I., SKAGGS K. E., MUTIIBWA D.: *Trend and magnitude of changes in climate variables and reference evapotranspiration over 116-yr period in the Platte River Basin, central Nebraska-USA*. Journal of Hydrology 420–421, 2012, 228 – 244.
- JANDHYALA V., FOTOPOULOS S., MAC NEIL I., LIU P.: *Inference for single and multiple change-points*. Journal of Time Series 34 (2012), 423-446.
- JARUŠKOVÁ D.: *Asymptotic behavior of a test statistic for detection of change in mean of vectors*. JSPI 140, 2010, 616-625.
- JARUŠKOVÁ D.: *Statistical methods for detecting changes in mean annual cycle and their application to several runoff series of European rivers*. In: Proceedings of ICWRER 2013, DOI 10.5675/ICWRER 2013, 329-345.
- JARUŠKOVÁ D., HORÁKOVÁ H., SATRAPA: *Detection of nonstationarities of several small Czech rivers by statistical methods*, Civil Engineering Journal 1, 2015, DOI: 10.14311/CEJ.2015.01.0005
- KLIMENT Z., MATOUŠKOVÁ M., LEDVINKA O.: *Trend analysis of rainfall-runoff regimes in selected headwater areas of the Czech Republic*, Journal of Hydrology and Hydromechanics 59, 2011, 36 – 50.

- LEDVINKA O., LAMAČOVÁ A.: *Detection of field significant long-term monotonic trends in spring yields*. Stochastic Environmental Research and Risk Assessment 9, 2015, 1463 – 1484.
- MILLY P. C. D., DUNNE K. A.: *Global pattern of trends in streamflow and water availability in a changing climate*. Nature 438, 2005, 347 – 350.
- MÜLLER H-G.: *Kernel estimators of zeros and locations and size of extrema of regression functions*, Scand. J. Statist. 12, 1985, 221-232.
- NOHARA D., KITO A., HOSAKA M., OKI T.: *Impact of climate change on river discharge projected by multimodel ensemble*. J. Hydrometeor. 7, 2006, 1076 – 1089.
- PEKAROVA P., PEKAR J.: *Long-term discharge prediction for the Turnu Severin station (the Danube) using a linear autoregressive model*. Hydrological Processes 20, 1217-1228 (2006). DOI: 10.1002/hyp.5939
- PEKAROVA P., MIKLANEK P., PEKAR J.: *Spatial and temporal runoff oscillation analysis of the main rivers of the world during the 19th - 20th centuries*. Journal of Hydrology 274, 2003, 62 – 79.
- RAMSAY J. O., SILVERMAN B. W.: *Functional Data Analysis.*, 2005, Springer, New York.
- SATRAPA, L., ZUKAL, M.: *Využití vodních toků v měnícím se klimatu a hydrologickém režimu*. In: Young Scientist 2012. Košice: Stavebná fakulta TÚ, díl 5., čl. č. 25., 163-170. ISBN 978-80-553-0904-0.
- STEWART I. T., CAYAN D. R., DETTINGER M. D.: *Changes toward Earlier Streamflow Timing across Western North America*. J. Climate 18, 2005, 1136 – 1155.
- VILLARINI G., SMITH J.A., SERINALDI F., NTELEKOS A.A.: *Analyses of seasonal and annual maximum daily discharge records for central Europe*. Journal of Hydrology 399(3):299-312, March 2011, DOI: 10.1016/j.jhydrol.2011.01.007
- WETHERALD R.T.: *Changes of time mean state and variability of hydrology in response to a doubling and quadrupling of CO<sub>2</sub>*. Climatic Change 102, 2010, 651 – 670.
- YE H., YANG D., ZHANG T., ZHANG X., LADOCHY S., ELLISON M.: *The Impact of climatic conditions on seasonal river discharges in Siberia*. J. Hydrometeor. 5, 2004, 286 – 295.
- National Geographic Magazine*, April 2017, české vydání.

#### INTERNETOVÉ ZDROJE:

<http://geologie.vsb.cz/jelinek/tc-global-oteplovani.htm>

<http://www.ovodarenstvi.cz/clanky/prutoky-nekterych-rek-v-cesku-se-pohybují-na-historickem-minimu>

[https://cs.wikipedia.org/wiki/Kjótský\\_protokol](https://cs.wikipedia.org/wiki/Kjótský_protokol)

[https://cs.wikipedia.org/wiki/Montrealský\\_protokol](https://cs.wikipedia.org/wiki/Montrealský_protokol)

[https://cs.wikipedia.org/wiki/Pařížská\\_dohoda](https://cs.wikipedia.org/wiki/Pařížská_dohoda)

[https://cs.wikipedia.org/wiki/Rámcová\\_úmluva\\_OSN\\_o\\_změnách\\_klimatu](https://cs.wikipedia.org/wiki/Rámcová_úmluva_OSN_o_změnách_klimatu)

[https://cs.wikipedia.org/wiki/Skleníkový\\_efekt](https://cs.wikipedia.org/wiki/Skleníkový_efekt)

[https://cs.wikipedia.org/wiki/Skleníkové\\_plyny](https://cs.wikipedia.org/wiki/Skleníkové_plyny)