

Autor diplomové práce: Bc. Waldemar Grešík
Název diplomové práce: Modelování syntetických hydrologických řad v systému stanic
Oponent diplomové práce: Mgr. Ondřej Ledvinka, Ph.D.
Pracoviště oponenta: Český hydrometeorologický ústav, Úsek hydrologie

Kritéria hodnocení diplomové práce:

1. Splnění požadavků zadání:	<i>Hodnocení:</i> velmi dobře (B)
<i>Komentář:</i> Diplomová práce (DP) si kladla za cíl provést rešerši literatury týkající se generování syntetických průtokových řad pro vodohospodářské účely ve vícerozměrném pojetí. Dále byla předmětem aplikace vybraného modelu s využitím reálných měsíčních průtokových dat a také vyhodnocení jeho schopnosti zachování základních statistických charakteristik s důrazem na autokorelaci a křížovou korelaci v soustavě stanic, což je velmi žádoucí. Kromě doporučených dvou metod ze zadání byly krátce diskutovány i metody další, což považuji za pozitivum DP. Za menší prohřešek ovšem považuji fakt, že se autor nesnažil hledat dále, což by mu umožnilo najít i techniky jednodušší. Např. metoda založená na bootstrapu s maximální entropií je velmi dobře publikována a dostupná i na internetu. Vlastně bylo též možné úplně se vyhnout programování a využít softwaru, který se nabízí buď přímo s publikovanými články (na základě dotazu směřovaného autorům), nebo je k dispozici na internetu zdarma (viz např. webové stránky Jose D. Salase). Tímto mohlo být na reálných řadách z Česka otestováno hned několik metod a tyto následně porovnány. Na druhou stranu si ale uvědomuji, že toto by rozsah zadání překračovalo a časové možnosti nebyly autorovi nakloněny (viz dále). Sestavení vlastního modelu (dle zadání) prokázalo schopnosti autora vyvíjet podobný software, což se autor nebál demonstrovat ani poskytnutím kódu ve formě elektronické přílohy. Toho si velmi cením.	
2. Metodika zpracování a logické členění práce:	<i>Hodnocení:</i> uspokojivě (D)
<i>Komentář:</i> Na jednu stranu předpokládám, že postup zpracování byl správný a vycházel z provedené rešerše literatury. Na druhou stranu je tento aspekt těžko hodnotitelný, jelikož několik oddílů přichystaných k popisu tohoto postupu v DP úplně chybí (soudě dle holých nadpisů kapitol), což mě velmi zarmoutilo. Dokonce logické členění DP do kapitol mohlo být promyšlenější. Množství kapitol mohlo být např. vhodně redukováno, o čemž svědčí i jakási roztržitost obsahu DP. Postrádám jak propracovanější diskusi výsledků, tak diskusi se zkušenostmi v zahraničí. Rovněž je znatelné určité zmatení diplomanta v terminologii, kterou je ale jistě těžké si v začátcích vědecké kariéry rychle osvojit. Závěr DP je velice stručný a jistě bylo možné z něj vytěžit více.	
3. Kvalita zpracování výsledků:	<i>Hodnocení:</i> velmi dobře (B)
<i>Komentář:</i> Dle zadání byly výsledky zpracovány tak, že si autor vybral jednodušší metodu ze dvou nabízených (tj. metodu hlavních komponent, PCA), kterou založil na kovarianční matici získané z reálných měsíčních průtokových řad. Aby mohla být metoda správně aplikována, data byla předzpracována (transformována). Pomocí testování hypotéz autor ověřil, jak model určený ke generování syntetických řad zachovává základní statistické charakteristiky, které lze pozorovat u vícerozměrných časových řad. To znamená, že kromě průměru, rozptylu, marginálních distribučních funkcí a autokorelace autora zajímala též tzv. křížová korelace. Na škodu je určité absence podrobnějšího popisu zpracování, a to včetně transformací dat a jejich inverzních forem tak, aby se čtenář mohl dozvědět, jak konkrétně autor dospěl k syntetickým řadám, jejichž vlastnosti následně testuje. Jistě by bylo možné tuto informaci získat ze	

strojového kódu, který je nabízen jako elektronická příloha, ale také jako tištěná Příloha 9. Ovšem pro běžného čtenáře, ačkoliv z oboru, je toto naprosto nepřipustné.

4. Interpretace výsledků, jejich diskuse: *Hodnocení:* dostatečně (E)

Komentář: Interpretace výsledků pomocí zvolených tabelací a uvedení obrázků (ať už přímo v textu, nebo ve formě příloh) plní svůj účel. Na druhou stranu právě diskusní část práce je velice slabá. Při pohledu na množství výsledků, o čemž svědčí i množství příloh, je jasné, že bylo možné vytěžit více. Témat, kterých se DP dotýká je poměrně dost na to, aby mohla být diskusní část rozšířena. Někdy je v textu myšlenka zajímavě započata, ale kýžený dojem z dokončenosti DP se stále nedostavuje. Diplomantovi by jistě prospělo dostat na zpracování mnohem více času. Pevně věřím, že by se efekt dostavil, což naznačují poměrně zajímavé myšlenky o průtocích v měsících, u kterých z nejrůznějších důvodů model selhává. Diskuse by se např. mohla zabývat i možnou nestacionaritou v parametrech, a to zejm. v kontextu hojně studované klimatické změny.

5. Využití literatury a její citace: *Hodnocení:* velmi dobře (B)

Komentář: Chválím zejm. práci se zahraniční literaturou, které je v závěrečné bibliografii uvedeno poměrně hodně. Někdy se však diplomant dopouští tradičních chyb, k nimž náleží neuvádění všech autorů (nyní nemám na mysli případy s velkým množstvím autorů, kdy je nutno přistupovat ke zkracování jejich výčtu), ale také velmi pravděpodobné záměny knih s články (co má např. znamenat svazek a číslo u knihy [32]?). Též není dodržován jednotný citační vzor (např. střídání kurzívy a velkých a malých počátečních písmen v názvech knih). Rovněž se vyskytují neúplné reference, jako např. u publikace [20], kde chybí jak místo vydání, tak nakladatel. Chyby se týkají i citování uvnitř textu, kde se nevhodně střídají číselné odkazy s odkazy ve formě "autor (rok)". V některých pasážích textu je znatelná absence vhodných citací (např. v oddíle 4.5 nebo u popisu Box–Coxovy transformace). Někdy jsem naopak dokonce dospěl k názoru, že knihy, na které se autor odkazuje, uváděné skutečnosti neobsahují. Množství nepřímých citací (tedy zda jsou skutečně nepřímé) svědčí i o tom, že si diplomant nedal příliš práci ani s prohledáváním internetu za účelem nalezení původních článků. Plné texty některých z nich jsem skutečně našel bez nutnosti přístupu k placeným vědeckým databázím, které by ale na půdě vysoké školy neměly činit větší problém.

6. Formální úprava práce, grafická a jazyková úroveň: *Hodnocení:* nedostatečně (F)

Komentář: Na první pohled DP zaujme tím, že je kompletně psaná anglicky. Zpočátku se jazyk zdá být v pořádku, ale po chvíli čtenář přijde na to, že není. V textu je možné nalézt nepovedené formulace nebo překlady, které mají někdy za následek špatné pochopení (které se naopak někdy obrátí v pochopení až po delší době). Typické jsou chyby, které Češi dělají při psaní anglického textu. Mezi ně patří nesprávné používání členů (správně je např. "the Czech Republic" a ne pouhé "Czech Republic"; pokud v odkazech na tabulky či obrázky stojí za "Table" či "Fig./Figure" (skutečně "Figure" a ne "Picture") konkrétní číslo, člen se naopak nepoužívá; před názvy řek se člen vynechává pouze ve výjimečných případech; navíc "River" se v názvech řek píše s velkým počátečním "R"). Nevhodně se střídá britský a americký spelling (např. "modelling" vs. "modeling") a jednotna není dodržena ani při psaní velkých iniciálních písmen (např. "Box–Cox" vs. "Box–cox" (navíc "Cox" se váže ke jménu statistika, takže velké "C" je více než na místě); názvy sekcí/oddílů začínají ve vědecké literatuře slovem "Section" (či zkráceně jen "Sect.") a podobné pravidlo platí i u odkazů na rovnice). "Státní podnik" se do angličtiny překládá jako "State Enterprise" a ne "national organization" (stačí se např. podívat na anglické mutace příslušných webových stránek).

Celková formální úprava DP opravdu strádá, což si vysvětlují nadměrným spěchem studenta (ale patrně i jeho školitele). Přestože se řešené téma týká statistické/stochastické hydrologie a rovnice jsou tedy nedílnou součástí, diplomant si nedal pozor při jejich číslování, které je ve výsledku nesprávné a zmatečné. To způsobuje snížení čtenářovy orientace v textu. Číslování

rovníc dokonce nerespektuje čísla kapitol (zde se omlouvám, pokud to původně nebylo zamýšlené) či je jiným způsobem zpřeházené. Toto se týká i samotného číslování kapitol/oddílů. Přitom některé oddíly, na které se text odkazuje, dokonce chybí (např. 4.5.2, 5.3.4 nebo 5.6.1), nemluvě o naprosté absenci textu pod některými nápisy, které autor pouze naznačil, ale započatou myšlenku u nich již nedokončil. Toto se zákonitě projevuje i v podobě obsahu, dle kterého je hned několik (pod)kapitol na jedné straně. Někdy jsou názvy tabulek uvedeny na jiné straně, než je tabulka samotná, což vůbec nenapomáhá celkovému dojmu z předložené DP. Samostatnou část kritiky formální úpravy by mohla tvořit matematická notace. DP není natolik objemná, aby nebylo možné volit odlišné značení pro různé matematické objekty. Někde je např. písmenem "c" značena nezávislá komponenta, jinde naopak koeficient variace apod. Značení dokonce není jednotné. Přitom právě nezávislé komponenty jsou na jednom místě práce značeny jako "ic" a na jiném místě jako "c". To si vysvětluji opisováním rovnic z různých zdrojů, jako tomu bude určitě i v případě první derivace CDF na str. 11, kterou má jistě být "p(x)" a ne "f(x)". Seznam matematických značek by mohl být doplněn např. o písmeno "k", které značí tzv. "lag" (neboli zpoždění), o písmeno "t" (tedy časový index) nebo o písmeno "z" (viz autoregresní model). Na druhou stranu jsou některé značky, se kterými jsem se v textu nesetkal, v seznamu nadbytečné (".").

K formálnímu vzhledu práce bych měl na závěr několik doporučení. Pokud se text neodkazuje na vloženou tabulku (obrázek), je taková tabulka (obrázek) zbytečná (např. tabulky 6.2 a 6.4). To se týká i příloh, ke kterým v textu nenacházím žádné další vysvětlení. A nakonec, pokud se v titulku nějakého obrázku objeví slovo "mapa", měla na obrázku mapa opravdu být. Mapy totiž kromě reprezentace geografických prvků obsahují minimálně (pokud možno úplnou) legendu a měřítko, což Příloha 1 nesplňuje.

Z hlediska formální úpravy tedy jednoznačně doporučuji komisi umožnit autorovi opravu DP, pokud ta bude nakonec obhájena!!!

7. Závěry práce a jejich formulace:

Hodnocení: dostatečně (E)

Komentář: Jak bylo zmíněno v bodu (2), závěr práce je velmi strohý, což je velice překvapivé vzhledem k tomu, jak je zpracována literární rešerše či dokonce početné přílohy. Právě při pohledu na ně se domnívám, že, kromě závěru, dokonce diskuse výsledků mohla být bohatší. Takto se nemohu zbavit pocitu (a pevně doufám, že nejde jen o můj názor), že je DP jaksi nedotažená ke svému kýženému konci a stále něco chybí. Přesto jsem se rozhodl práci doporučit k obhajobě. Předpokládám tímto, že rozhodující slovo budou mít odpovědi diplomanta na mé otázky.

8. Otázky k obhajobě a případné další připomínky k práci:

- 1) Co si diplomant představuje pod pojmem "deterministická část (autoregresního) modelu"?
- 2) Co si diplomant představuje pod pojmem "parciální autokorelační funkce (PACF)"?
- 3) Je autoregresní koeficient totéž, co autokorelační koeficient?
- 4) Může autoregresní koeficient ρ v reprezentaci uváděné pro AR(1) na str. 14 nabývat hodnot větších než 1? Co se děje v případě, kdy je $\rho = 1$? A co v případě, kdy $-1 < \rho < 0$?
- 5) Mohl by diplomant ve stručnosti vlastními slovy popsat, jak se vypočítávají parametry tříparametrického logaritmicko-normálního rozdělení? Ve vzorečkách na str. 31 navíc vidím dvě různá "e". Jde vždy o Eulerovo číslo, nebo také o nějakou jinou konstantu?
- 6) Jaký význam mají rovnice 1 a 2 na str. 34? Mám dojem, že jde opět o Box–Coxovu transformaci. A mohl by diplomant stručně zopakovat proceduru nastíněnou v oddíle 6? Co tedy probíhalo nejprve? Standardizace na rozdělení s nulovým průměrem a jednotkovým rozptylem? A jak tedy byla nakonec normalizace provedena? Něco mi napovídá, že právě Box–Coxovou transformací. Je to pravda?
- 7) Mohl by mne diplomant odkázat na nějakou publikaci, kde byla otázka kvality získaných dat již řešena, jak ostatně popisuje v oddíle 6.3.1? A odkud vlastně data pochází? Jaké byly

délky těchto řad průtoků? Byly všechny tyto řady stejně dlouhé? A byly úplné, nebo obsahovaly nějaké chybějící hodnoty?

8) V práci je uvedeno, že pro statistické testování (průměrů, rozptylů, autokorelací apod.) bylo modelem vygenerováno 1000 samlů. Jak tyto samplly byly dlouhé? Byla jejich délka stejná jako u historických řad?

9) Bylo by možné rozvést testování rovnosti dvou korelačních koeficientů? Popis metody v práci je poněkud vágní. Předpokládáte předem např., že řady mají odlišné délky (soudě dle dvou "n" ve vzorci? Mohl by též diplomant sdělit, kde přesně v citované publikaci je uvedena tato zmiňovaná procedura?

10) V práci nebyla hodnocena tzv. zpožděná křížová korelace (tj. "lagged cross-correlation"). Čím tedy diplomant vysvětluje časovou korelaci v tabulkách obsahujících pouze křížové korelace (mám na mysli tabulky 7.8 a 7.9 v Příloze 5)?

11) Proč autor zvolil odhad rozptylu s multiplikativním členem $1/n$ namísto s v české hydrologii vžitým členem $1/(n - 1)$? Narážím na to proto, že jsem si všimnul, že tento multiplikativní člen není u odhadu koeficientu šikmosti (na str. 10) uveden vůbec. Proč?

12) Je poslední člen rovnice 4.3 (na str. 12) opravdu správně? Domnívám se totiž, že toto latí pouze pro vektor, kdy jsou střední hodnoty jeho jednotlivých členů odečteny (tj. pro tzv. centrováný vektor).

13) Jak konkrétně byl počítán autokorelační koeficient 1. řádu mezi řadami listopadů a předchozími řadami prosinců (viz oddíl 7.4.1)?

14) Co je "N" v matici 4.5 na str. 12? V seznamu matematických značek není. Nebylo "n" zaměněno za "N"?

15) Jsou obrázky 7.1 a 7.2 odněkud převzaty? Mohl byste uvést zdroj?

A na závěr několik doporučení, která se nevešla do jiných kolonek:

1) Praktická část DP je spíše o PCA. Proto bych ji v teoretické části rozvedl více.

2) Pokud už bylo nutné vkládat do DP obrázky 7.1 a 7.2, určitě by si zasloužily detailnější popis. Alternativně by bylo možné čtenáře odkázat na oddíly, které o nich pojednávají podrobněji.

3) Předpokládám, že definice PDF na str. 11 je myšlena pro diskrétní veličinu. U spojitě totiž rovnost $x = a$ nemá smysl a je tedy nutné přistoupit k intervalům. V diskrétním pojetí však ovšem nelze hovořit o derivacích, nýbrž o diferencích.

4) Na str. 26 bych byl opatrnější s tvrzením o špičatosti a její standardizaci. Některé vzorce platí totiž jen pro centrované veličiny. Taktéž uváděné literární zdroje jsou jaksi promíchané.

5) V rámci celé DP bych doporučoval používat jednotnou terminologii (viz např. "normal distribution vs. "Gaussian distribution"; u rozdělení v anglickém textu bych také doporučoval používat správně členy).

6) Příště bych psal "the University of Helsinki" a ne "University of Helsinki" (viz str. 20).

Celkové hodnocení diplomové práce*:

Práci doporučuji k obhajobě: ANO

Návrh hodnocení: USPOKOJIVĚ (D)

*** ČVUT v Praze v souladu s ustanovením § 47b zákona č. 111/1998 Sb., o vysokých školách v platném znění, nevýdělečně zveřejňuje závěrečné práce včetně posudků a záznamu o průběhu a výsledku obhajoby. Odevzdáním posudku oponent souhlasí s jeho zveřejněním.*

V Praze dne 6. 2. 2018

.....
Podpis oponenta

(*) Celkové hodnocení diplomové práce nemusí být průměrem výše uvedených hodnocení jednotlivých částí. Váhu dílčích kritérií určuje oponent.

(**) Informace ke zveřejnění Vámi vypravovaného posudku.