POROVNANIE REGIONALIZAČNÝCH METÓD NA ODHAD MAXIMÁLNYCH PRIETOKOV NA SLOVENSKU

Ladislav Gaál[1], Silvia Kohnová[1], Ján Szolgay[1]

Abstrakt

Predložená štúdia sa zameriava na porovnanie rozdielnych prístupov k štatistickému modelovaniu návrhových hodnôt maximálnych ročných prietokov na Slovensku, ktorými sú:

1) regionálny model Hoskinga a Wallisa (HW), založený na pevne vymedzených regiónoch,
2) regionálny model vplyvného regiónu (ROI), v ktorom sa regióny definujú samostatne pre každé povodie, resp.
3) lokálny model, ktorý je založený iba na základe pozorovaní v danom mieste.

Do analýzy boli vybrané rady ročných kulminačných prietokov zo 174 malých a stredne veľkých povodí zo Slovenska, v ktorých bolo odvodených 20 fyzickogeografických a klimatických atribútov. Desať regiónov v metóde HW sa vyčlenilo pomocou zhlukovej analýzy na základe 4 charakteristík povodia (plocha povodia, hydrogeologický index povodia, lesnatosť povodia a dlhodobý priemerný úhrn zrážok). Flexibilné regióny v metóde ROI sa vytvárali na základe miery podobnosti povodí, do ktorej vstupovali rôzne kombinácie vyššie uvedených atribútov.

Základom porovnania frekvenčných modelov boli stredná kvadratická chyba a vychýlenie odhadov kvantilov, ktoré sme získali na základe Monte Carlo simulácií. Z výsledkov štúdie vyplýva, že lokálna frekvenčná analýza má pre všetky uvažované doby opakovania výrazne horšie štatistické ukazovatele než regionálne prístupy, avšak na základe doposiaľ vykonaných analýz nie je možné jednoznačne určiť, ktorý z dvoch regionálnych modelov je výhodnejšou voľbou na odhad kvantilov maximálnych prietokov na Slovensku.

Kľúčové slová: regionálna frekvenčná analýza, Hosking-Wallisova metóda, metóda vplyvného regiónu, test homogenity, Monte Carlo simulácie

1. Úvod

Slovenskí hydrológovia a vodohospodári sa počas posledných 10-15 rokov intenzívne venovali metódam odhadu návrhových prietokov v regionálnom kontexte. Metóda Hoskinga a Wallisa (skr. HW) založená na pevne vymedzených regiónoch bola adoptovaná [3] a ďalej vyvinutá, avšak skúsenosti poukázali aj na to, že vyčlenenie homogénnych regiónov (ktoré je nevyhnutnou požiadavkou pre korektnú aplikáciu metódy) nie je triviálnou záležitosťou v orograficky komplexnej krajine akou je Slovensko. V tejto práci sme sa vrátili k metóde HW a prehodnocovali sme ju vo svetle ďalšej alternatívy regionálnej frekvenčnej analýzy, ktorou je metóda vplyvného regiónu (skr. ROI, *region of influence*; napr. [1], [2]). Základnou koncepciou metódy ROI je to, že nie je nutné vymedziť pevné hranice medzi regiónmi, namiesto toho sa regióny definujú flexibilným spôsobom: každé analyzované (cieľové) povodie má svoj vlastný, unikátny región.

Cieľom tejto štúdie bolo: 1) porovnať rôzne alternatívy regionálnej frekvenčnej analýzy na základe metódy ROI a metódy HW (resp. lokálnu frekvenčnú analýzu) objektívnym spôsobom (t.j. na základe viacnásobných Monte Carlo simulácií) a 2) aplikovať najvýhodnejšiu metódu na odhad kvantilov maximálnych ročných prietokov a ich intervalov spoľahlivosti na vybraných povodiach Slovenska.

2. Metodika

Dva regionálne prístupy k frekvenčnej analýze maximálnych prietokov, ktoré sme testovali v tejto štúdii (t.j. HW aj ROI) majú veľa spoločného. Obidva prístupy sa zakladajú na metóde indexovej povodne [3], ktorá predpokladá regionálnu homogenitu. Hovoríme, že región alebo regionálny typ je homogénny, ak distribučné funkcie na jednotlivých lokalitách v regióne sú totožné až na škálovací faktor (indexovú hodnotu), ktorý je špecifický pre danú lokalitu (profil). [Poznámka: v tejto práci budeme jednotne používať pojem „región“, aj keď obidva regionálne prístupy v skutočnosti vytvárajú regionálne typy, t.j. množinu povodí „rozhádzané“ v geografickom priestore.]

Metóda HW je založená na homogénnych, pevne vymedzených regiónoch. Regióny sa vyčleňujú väčšinou na základe dobre definovaného súboru charakteristík povodí, najčastejšie (ako aj v tejto štúdii) pomocou zhlukovej analýzy. Homogenita regiónu sa dá preveriť a kvantifikovať prostredníctvom miery homogenity H_1 [3].

Metóda ROI vytvára unikátny (flexibilný) región pre jednotlivé cieľové povodia. V prvom kroku sa definuje miera podobnosti medzi povodiami na základe vopred zvoleného súboru charakteristík (analógia zhlukovej analýzy; používa sa Euklidovská metrika). Flexibilné regióny sa vytvárajú postupným pridávaním najviac podobných povodí k existujúcemu zoskupeniu cieľového povodia na základe dvoch kritérií. Prvou požiadavkou je homogenita regiónu, ktorá sa vyhodnocuje regionálnym testom Lua a Stedingera [5], kým druhou je tzv. pravidlo 5T, ktoré hovorí o tom, že k spoľahlivému odhadu kvantilu s dobou opakovania T je nutné mať údaje v minimálnom rozsahu 5*T staničných rokov [4]. Akonáhle je známy počet podobných povodí pre cieľovú lokalitu, jednotlivým povodiam v regióne sa na základe miery podobnosti priraďujú regionálne váhové koeficienty, ktoré určujú, ako dôležitú úlohu hraje dané povodie pri prenose hydrologickej informácie do cieľovej lokality. Metódu ROI podrobne popisuje napr. [2].

Popri dvoch vyššie uvedených regionálnych prístupov sme do porovnania frekvenčných modelov zaradili aj tradičný spôsob odhadovania návrhových hodnôt, ktorým je *lokálna analýza bez regionálneho prístupu.*

Na odhad bezrozmerných kvantilov v jednotlivých profiloch sme použili zovšeobecnené extremálne rozdelenie (*generalized extreme value*, skr. GEV; pozri napr. [3]).

Základom porovnania frekvenčných modelov boli štatistické charakteristiky strednej kvadratickej chyby a vychýlenia, ktoré sme získali na základe Monte Carlo simulácií s počtom opakovaní 5000 [2].

3. Údaje

Pre analýzy bolo vybraných 174 malých a stredne veľkých povodí (s plochou povodia do 340 km^2) z databázy Slovenského hydrometeorologického ústavu. Dĺžka radov pozorovania ročných kulminačných prietokov sa pohybovala v intervale od 15 do 88 rokov.

Pre každé povodie bolo odvodených približne 20 klimatologických a fyziografických charakteristík povodí. Dodatočné analýzy ukázali, že najvýznamnejšími charakteristikami povodí boli:

- A: plocha povodia [km^2];
- HG: hydrogeologický index povodia [–];
- FA: percentuálny pomer zalesnenia povodia [%];
- q: dlhodobý priemerný špecifický odtok [l.s^{-1}.km^{-2}];
- P: dlhodobý priemerný ročný úhrn zrážok (1931–1980) [mm].

4. Výsledky

V rámci zhlukovej analýzy sme testovali takmer 70 rôznych kombinácií charakteristík, pričom sme sa snažili dodržať požiadavku nízkej hodnoty korelačných koeficientov medzi vybranými atribútmi; na druhej strane, výstupy zo zhlukovej analýzy, t.j. vytvorené regióny museli spĺňať požiadavku regionálnej homogenity. Ukázalo sa, že najvýhodnejšou kombináciou charakteristík povodí je štvorica *A*, *HG*, *FA* and *P*, na základe ktorých sa vyčlenilo 10 homogénnych regiónov. Počet povodí zahrnutých do jednotlivých regiónov bolo v rozmedzí 10 až 33, pričom každý z regiónov mal prijateľnú mieru regionálnej heterogenity $H_1 \leq 2$ [3].

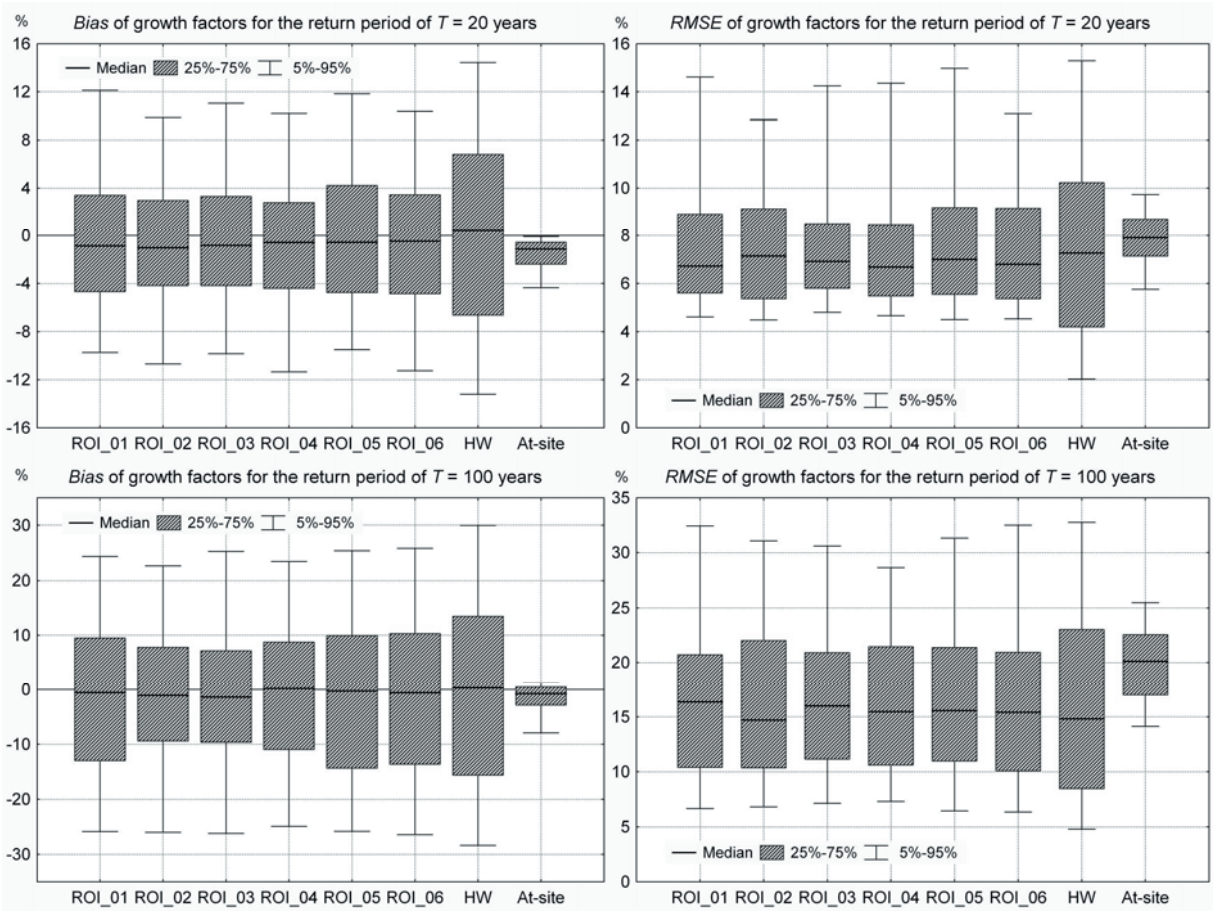
Kombinácie atribútov povodí pre definíciu miery podobnosti povodí v metodike ROI (t.j. zjednodušene: schémy metódy ROI) sme vytvárali na základe štyroch preferovaných atribútov zo zhlukovej analýzy, ku ktorým sme ešte pridali dlhodobý priemerný špecifický odtok *q*. Definovali sme spolu 6 rôznych schém ROI, prehľad ktorých prezentujeme v Tab. 1.

Tab. 1: Schémy metódy ROI použité v tejto štúdii, t.j. rôzne kombinácie vstupných atribútov povodí, na základe ktorých sme definovali podobnosť povodí v metóde ROI. Symbol ✓ (–) znamená, že daný atribút je (nie je) zahrnutý do príslušnej schémy označenej skratkou ROI_01 ... ROI_06.

Použité schémy metódy ROI	*A*	*HG*	*FA*	*q*	*P*
ROI_01	✓	✓	✓	✓	–
ROI_02	✓	✓	✓	–	–
ROI_03	✓	✓	–	✓	–
ROI_04	✓	–	✓	✓	–
ROI_05	–	✓	✓	✓	–
ROI_06	–	✓	✓	–	✓

Tab. 2: Priemerné hodnoty vychýlenia a strednej kvadratickej chyby jednotlivých frekvenčných modelov v % pre rôzne významné doby opakovania *T* (v rokoch). Najnižšie hodnoty štatistických charakteristík (v absolútnom zmysle) pre danú dobu opakovania sú zvýraznené tučným písmom. „At-site“ označuje lokálny odhad.

T [roky]	Priemerné vychýlenie [%]							
	ROI_01	**ROI_02**	**ROI_03**	**ROI_04**	**ROI_05**	**ROI_06**	**HW**	**At-site**
10	-0,177	-0,353	-0,528	-0,278	**-0,135**	-0,280	0,693	-1,016
20	-0,290	-0,712	-0,627	-0,517	**-0,052**	-0,386	0,235	-1,579
50	-0,651	-1,048	-0,851	-0,692	-0,542	-0,354	**-0,210**	-1,843
100	-1,040	-1,120	-1,020	-0,877	-1,183	-1,258	**-0,378**	-1,621
200	-1,015	-1,098	-1,180	-0,591	-1,390	-1,297	**-0,370**	-0,981
T **[roky]**	**Priemerná stredná kvadratická chyba [%]**							
	ROI_01	**ROI_02**	**ROI_03**	**ROI_04**	**ROI_05**	**ROI_06**	**HW**	**At-site**
10	4,233	4,303	4,135	**4,063**	4,233	4,200	4,097	4,339
20	7,792	7,641	7,665	**7,506**	7,842	7,671	7,608	7,896
50	13,047	12,649	12,878	**12,587**	12,740	12,880	12,628	14,337
100	16,806	**16,507**	16,783	16,554	16,540	16,597	16,536	20,013
200	20,992	20,783	21,094	20,968	20,542	20,745	**20,532**	26,457



Obr. 1: „Krabicové grafy“ vychýlenia (*Bias*) a strednej kvadratickej chyby (*RMSE*) jednotlivých frekvenčných modelov (v %) pre doby opakovania T = 20 a 100 rokov.

Súhrnné výsledky Monte Carlo simulácií uvádzame vo forme tabuľky (Tab. 2) a tzv. krabicových grafov (Obr. 1). Tab. 2 prezentuje bodové charakteristiky frekvenčných modelov: priemerné vychýlenie a priemernú strednú kvadratickú chybu simulovaných čiar prekročenia (pre detaily pozri napr. [1]). Krabicové grafy ponúkajú ešte širší pohľad na analyzované štatistické modely v transparentnom prevedení. Popri ďalšej bodovej charakteristike výsledkov (medián – bodkovaná čiara v strede obdĺžnika) sa prostredníctvom krabicových grafov prezentujú aj rozptylové charakteristiky simulácií: interkvartilové rozpätie (obdĺžnik) a 90% rozpätie hodnôt (tzv. antény).

Kým priemerné hodnoty vychýlenia (horná časť Tab. 2) neposkytujú relevantnú informáciu o rozdieloch vo frekvenčných modeloch, z krabicových grafov je možné videť, že regionálny model HW sa vyznačuje najmenej priaznivejším správaním (najširšie rozpätie medzi kvartilmi 25–75 % aj 5–95 %). Treba si poznamenať, že úzke 50% aj 90% intervaly rozpätia hodnôt pre model lokálneho odhadu odrážajú iba skutočnosť, že v tejto metóde nedochádza k prenosu regionálnej informácie.

Na základe priemerných hodnôt strednej kvadratickej chyby (dolná časť Tab. 2) môžeme konštatovať, že model *ROI_04* má pre kratšie doby opakovania najnižšie ukazovatele spomedzi všetkých modelov. Tento fakt potvrdzuje aj krabicový graf pre T = 20 rokov (Obr. 1, vpravo hore), kde je tiež vidieť, že modely ROI sú výhodnejšími v porovnaní s regionálnym modelom HW a neporovnateľne lepšie voči lokálnemu odhadu. Akonáhle sa však približujeme k vyšším dobám opakovania, rozdiely medzi modelmi ROI a HW sa postupne eliminujú (ba prevezme dominanciu model HW čo sa týka priemernej strednej kvadratickej chyby pre T = 200 rokov, Tab. 2).

5. Záver

Z predloženej analýzy vyplýva, že regionálne prístupy sa ukazujú jednoznačne v lepšom svetle než lokálna frekvenčná analýza, a to pre všetky uvažované doby opakovania *T*. Na základe doterajších analýz však nie je možné rozhodnúť, ktorý z regionálnych prístupov dáva prijateľnejšie výsledky, lebo 1) metóda ROI sa zdá byť výhodnejšia pre kratšie doby opakovania (zrejme kvôli tomu, že veľkosť regionálnych typov je prispôsobená cieľovej dobe opakovania, pričom odhady kvantilov na základe metódy HW môžu byť skreslené v dôsledku redundantnej informácie) a 2) metóda HW má vo väčšine prípadov lepšie štatistické ukazovatele (medián a priemer strednej kvadratickej chyby) pre dlhšie doby opakovania.

V blízkej budúcnosti preto plánujeme rozšírenie analýzy v niekoľkých smeroch, napr. zahrnutím ďalších kombinácií charakteristík povodí (t.j. definíciou ďalších alternatív schém ROI) alebo zavedením relatívnych váhových koeficientov pre charakteristiky vstupujúce do vzťahu pre mieru podobnosti.

Poďakovanie

Táto práca bola podporovaná grantmi VEGA 1/0496/08 a 1/0103/10. Článok vznikol tiež vďaka podpore v rámci OP Výskum a vývoj pre projekt „Centrum excelentnosti protipovodňovej ochrany územia“ ITMS 26240120004 spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.

Literatúra

[1] BURN, D. H., 1990. Evaluation of regional flood frequency analysis with a region of influence approach. Water Resources Research 26 (10), 2257–2265.

[2] GAÁL, L. – KYSELÝ, J., 2009. Comparison of region-of-influence methods for estimating high quantiles of precipitation in a dense dataset in the Czech Republic. Hydrology and Earth System Sciences 13, 2203–2219.

[3] HOSKING, J. R. M. – WALLIS J. R., 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, 224 pp.

[4] JAKOB, D. – REED, D. W. – Robson, A. J., 1999. Selecting a pooling-group: In: *Flood Estimation Handbook*, Vol. 3. Institute of Hydrology, Wallingford, UK.

[5] LU, L. H. – STEDINGER, J. R., 1992. Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. Journal of Hydrology 138 (1-2), 223–245.

Kontakty na autora

1 Katedra vodného hospodárstva krajiny, Stavebná fakulta Slovenskej technickej univerzity v Bratislave; e-mail: ladislav.gaal@stuba.sk