MODELOVANIE ZÁVISLOSTÍ MEDZI OBJEMOM A KULMINAČNÝM PRIETOKOM POVODŇOVÝCH VĺN POMOCOU KOPULÍ S VYUŽITÍM INFORMÁCIÍ O HISTORICKÝCH UDALOSTIACH V RÁMCI METODIKY MCMC

Ladislav Gaál[1], Ján Szolgay[1], Tomáš Bacigál [2], Silvia Kohnová[1]

Abstrakt

Predložená práca je prípadovou štúdiou, ktorá sa zameriava na odhad návrhových hydrologických veličín na stanici Vltava-Kamýk v Českej republike. Cieľom práce je stanoviť odhad objemu povodňovej vlny na základe pravdepodobnosti podmieneným výskytom 10 000ročného kulminačného prietoku pre vodnú nádrž Orlík.

V prvej časti štúdie sme sa zamerali na lokálnu frekvenčnú analýzu kulminačných prietokov, resp. povodňových vĺn. Nakoľko na danej stanici sa v roku 2002 vyskytla mimoriadna povodňová udalosť, ktorej kulminačný prietok dosiahol viac než dvojnásobok maxima za celé predchádzajúce obdobie (1877–2001), túto udalosť sme považovali za „historickú“ a na jej zahrnutie do radu systematických pozorovaní a následnej frekvenčnej analýzy sme aplikovali metodiku Bayesových MCMC simulácií (Monte Carlo Markovove reťazce). Výsledky MCMC odhadov sme porovnali s lokálnymi štatistickými odhadmi podľa klasických metód. V druhej časti sme sa zamerali na analýzu závislosti medzi kulminačnými prietokmi a povodňovými vlnami prostredníctvom kopúl. Testovali sme viacero rodín kopúl a následne sme odhadli podmienené objemy povodňových vĺn. Získané výsledky návrhových prietokov a objemov sme porovnali s odhadmi používanými v súčasnej inžinierskej praxi.

Kľúčové slová: frekvenčná analýza, maximálne prietoky, objemy povodňových vĺn, kopule, Vltava-Kamýk

1. Úvod

Kopule sú matematickými štruktúrami, ktoré umožňujú modelovať združené rozdelenie pravdepodobnosti dvoch alebo viacerých veličín, medzi ktorými existuje vzájomná závislosť, napr. maximálne prietoky a objemy povodňových vĺn, intenzita a trvanie zrážok atď. Ich výhodou je napr. aj to, že k odhadu návrhových hodnôt združených extrémov nie je potrebné klásť prísne predpoklady na marginálne rozdelenie jednotlivých veličín [7].

Štruktúra závislosti, resp. spoľahlivosť odhadov návrhových hodnôt združených veličín – predovšetkým v oblasti maximálnych hodnôt, t.j. na tzv. “hornom chvoste” – závisí nielen od zvoleného matematického modelu (kopule), ale aj od množstva dostupných údajov pre modelovanie marginálnych rozdelení jednotlivých veličín. Výberové súbory ročných (sezónnych) maxím majú väčšinou obmedzený rozsah, čo je výrazným limitujúcim faktorom vo frekvenčnej analýze. Nedostatku údajov sa však dá vyhnúť, napr. použitím väčšieho množstva simulovaných údajov zo zrážko-odtokových modelov alebo zahrnutím tzv. historických informácií do analýzy. Práve druhá z metodík je aplikovaná v tejto štúdii.

Ciele predloženej práce sú nasledovné: 1) využiť historické informácie vo frekvenčnej analýze marginálnych rozdelení maximálnych prietokov, resp. objemov povodňových vĺn, a to v rámci simulácií MCMC (Monte Carlo Markovove reťazce); 2) odhadnúť združené pravdepodobnosti prietokov a objemov povodňových vĺn pomocou kopúl a testom dobrej zhody posúdiť ich vhodnosť a 3) určiť vybrané kvantily povodňových vĺn za podmienky nastatia 10 000ročného prietoku na stanici Vltava-Kamýk (vodné dielo Orlík) a porovnať ich s odhadmi odborníkov doposiaľ použitými v hydrologickej praxi.

2. Metodika

2.1 Bayesove MCMC simulácie v jednorozmernej frekvenčnej analýze

V [9] prezentovali odhad parametrov distribučnej funkcie, v ktorej sa dajú zohľadniť aj informácie o historických hydrologických udalostiach prostredníctvom funkcie vierohodnosti. Nech $x_1,\ldots,x_i,\ldots,x_s$ označujú prvky systematického pozorovania o dĺžke s a $y_1,\ldots,y_j,\ldots,y_m$ označuje m historických udalostí, ktoré sa vyskytli počas historického obdobia o dĺžke h a nech X_0 je prahová hodnota, pod ktorou leží všetkých h-m (neznámych) udalostí z historického obdobia. Potom

$$\ell(\theta) = \prod_{j=1}^{m} f_X(y_j)\left[F_X(X_0)\right]^{h-m} \prod_{i=1}^{s} f_X(x_i) \qquad (1)$$

je *združenou funkciou vierohodnosti*, kde $f_X(F_X)$ označuje rozdelenie pravdepodobnosti (kumulatívnu distribučnú funkciu, skr. DF) premennej X a θ sú parametre F, resp. f.

Odhady parametrov θ sa získajú *metódou maximálnej vierohodnosti* po logaritmickej transformácii vzťahu (1) a jej následnej maximalizácie. Procedúra odhadu je založená na využití Bayesovej vety [8]. Výhodou Bayesovho prístupu je hlavne to, že výsledkom analýzy je úplné aposteriórne rozdelenie pravdepodobnosti parametrov, vrátane ich intervalov spoľahlivosti alebo ľubovoľných funkcií parametrov θ. *Monte Carlo Markovove reťazce* (MCMC) sú potom iba technickou pomôckou pre aplikáciu Bayesovej vety: metodika MCMC „vzorkuje“ aposteriórne rozdelenie bez nutnosti numerického výpočtu zložitých integrálov, ktoré sa objavia po dosadení vzťahu (1) do Bayesovej vety [8]. Ďalšie detaily o metodike MCMC v lokálnej alebo regionálnej frekvenčnej analýze uvádza napr. [8] alebo [3].

2.2 Združená frekvenčná analýza pomocou kopúl

Združená distribučná funkcia obsahuje informáciu o individuálnych vlastnostiach náhodných vektorov X a Y uloženej v marginálnych DF $F(x)$ a $G(y)$, a zároveň o ich vzťahu. Pre marginálne DF (ďalej len „marginálie“) platí $F(x) = H(x,\infty)$ a $G(y) = H(\infty,y)$, pre $x,y \in \Re$. Pripomeňme, že $H(x,y) = P(X < x \wedge Y < y)$, kde P označuje pravdepodobnosť a $\wedge$ logický súčin. Modelovanie oboch typov informácií združených v jednej funkcii má hlavnú nevýhodu spočívajúcu v obmedzení voľby marginálií na rovnaký typ rozdelenia. Riešením tohto problému je prístup založený na rozklade pravdepodobnostnej štruktúry a oddelenej analýze jednotlivých zložiek. Podľa Sklarovej vety (napr. [4], [7]), združenú DF možno prepísať do tvaru $H(x,y) = C(F(x),G(y))$, kde funkcia C: $[0,1]^2 \rightarrow [0,1]$ sa nazýva kopula. Jej vlastnosti sú

- $C(u,0) = C(0,v) = 0$,
- $C(u,1) = u$, $C(1,v) = v$,
- C je dvojrastúca, t.j. $C(u',v') - C(u',v) - C(u,v') + C(u,v) \geq 0$, pre všetky u, v, u', $v' \in [0,1]$ také, že $u < u'$, $v < v'$.

Takáto distribučná funkcia (*kopula* C) obsahuje informáciu výhradne o závislosti medzi X a Y, a spája marginálie F a G do združenej distribučnej funkcie H. Existuje niekoľko tried kopúl, z ktorých každá sa vyznačuje špecifickými vlastnosťami [7].

3. Údaje

Štúdia sa zamerala na stanicu Vltava-Kamýk, ktorá na nachádza na hornom toku rieky Vltava, v blízkosti vodného diela Orlík v južných Čechách. Vstupné údaje tvorili rady letných kulminačných prietokov (máj až október) za obdobie 1877–2002, resp. odpovedajúce rady objemov povodňových vĺn, ktoré sme v práci určovali dvoma spôsobmi: 1) *metódou konštantnej dĺžky trvania povodňovej vlny*, t.j. 3 dni trvania vzostupnej vetvy povodňovej vlny, 1 deň trvania kulminácie povodne a 5 dní trvania zostupnej vetvy povodňovej vlny, resp. 2) tzv. *subjektívnou metódou*, t.j. vizuálnou separáciou každej povodňovej vlny.

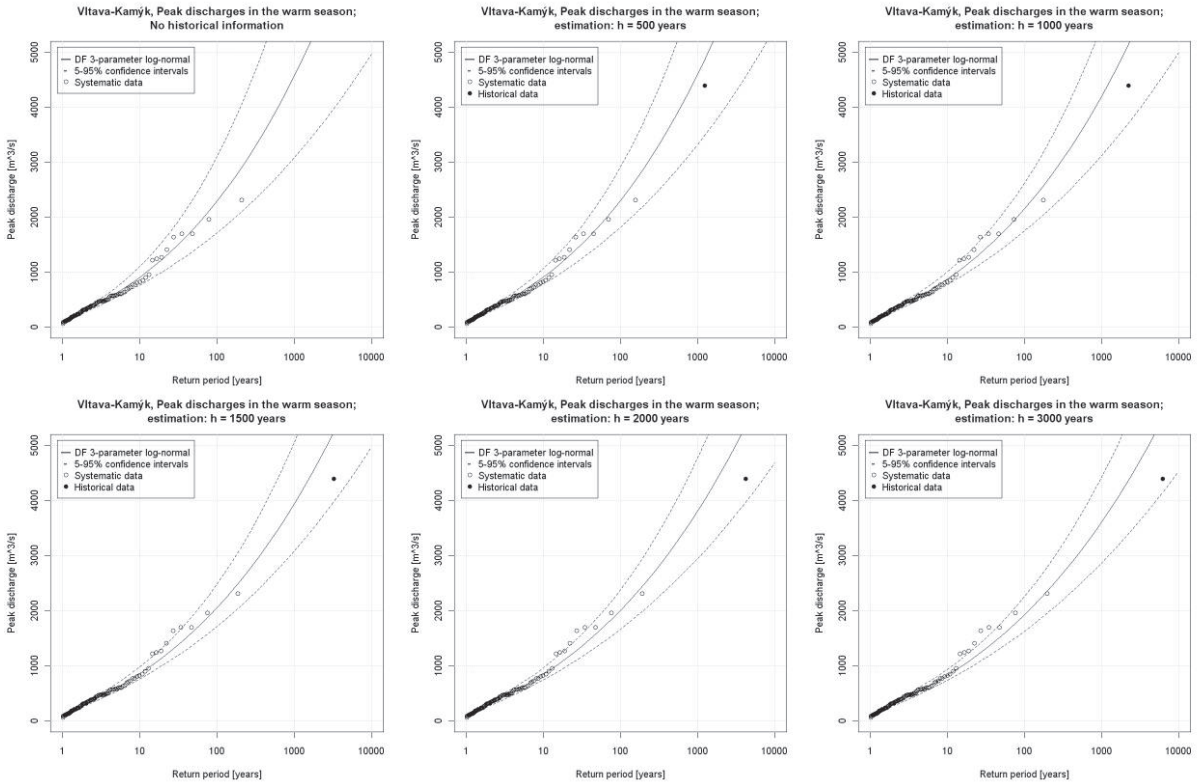
Rad kulminačných prietokov je výnimočný tým, že zahŕňa aj mimoriadnu povodeň z roku 2002: hodnota tohto maxima (4390 $m^3.s^{-1}$) takmer dvojnásobne prevyšuje maximum z ostatnej časti radu (2309 $m^3.s^{-1}$) z obdobia 1877–2001. V dôsledku tohto faktu sme udalosť z roku 2002 definovali ako „historickú“ (aj napriek tomu, že sa vyskytla v súčasnosti) s cieľom ilustrovať výhodné aspekty zahrnutia informácie o mimoriadnych historických udalostiach do lokálnej frekvenčnej analýzy.

4. Výsledky

4.1 Frekvenčná analýza maximálnych prietokov, resp. povodňových vĺn

Návrhové hodnoty analyzovaných veličín, odhadovaných na základe vzťahu (1) závisia predovšetkým od voľby teoretickej DF $F(x)$, od prahovej hodnoty X_0 a od dĺžky historického obdobia h. Z tohto dôvodu sme vykonali niekoľko sérií citlivostných analýz, v ktorých sme skúmali efekt jednotlivých faktorov na finálne odhady návrhových hodnôt a ich neurčitosti.

V analýze kulminačných prietokov, resp. povodňových vĺn sme skúmali vhodnosť nasledujúcich DF ([6], [5]): 4-parametrické Johnsonovo rozdelenie (JOH), 2-parametrické log-normálne rozdelenie (LN2), resp. 3-parametrické rozdelenia, a to zovšeobecnené extremálne (GEV), zovšeobecnené logistické (GLO), log-normálne (LN3), Pearsonovo rozdelenie III. typu (P3) a logaritmické Pearsonovo rozdelenie III. typu (LP3). Dobrú zhodu pozorovaných údajov s preloženými teoretickými DF sme posudzovali subjektívnym spôsobom, a to na základe toho, či 90 % pozorovaných údajov znázornených pomocou empirického vzťahu pre polohu kvantilov [1] ležia v 90percentnom intervale spoľahlivosti (5 % – 95 %) určeným simuláciou MCMC [10].



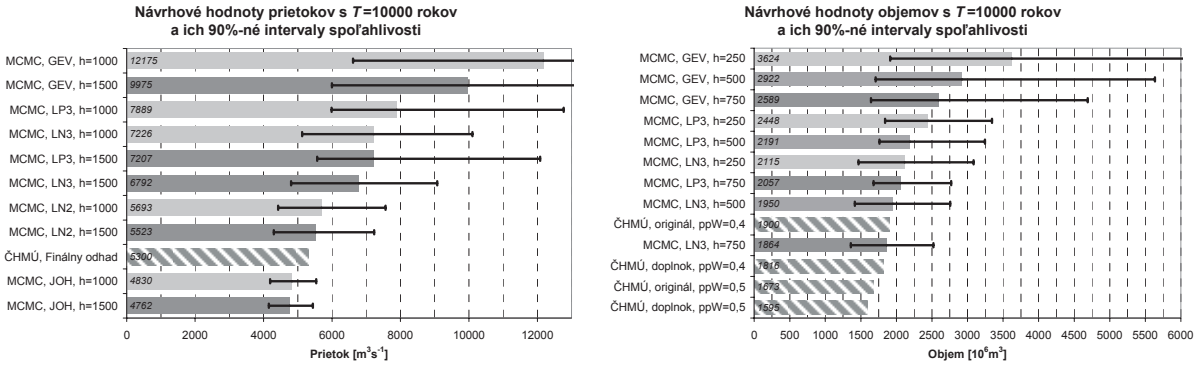
Obr. 1: Letné kulminačné prietoky preložené 3-parametrickým log-normálnym rozdelením.

Citlivostná analýza vzhľadom na dĺžku historického obdobia h. Pri konštantnej prahovej hodnote X_0 = 4000 $m^3.s^{-1}$ sme uvažovali 5 rôznych hodnôt h od 500 po 3000 rokov. Grafy na Obr. 1 znázorňujú výsledky citlivostnej analýzy pre DF LN3, pričom v pravom hornom rohu

sa nachádza aj prípad bez zahrnutia historickej informácie. Môžeme konštatovať, že zahrnutie historickej informácie evidentne zužuje neurčitosť odhadov kvantilov v porovnaní so situáciou, kedy sa analyzujú iba systematické údaje. Ďalej sme usúdili, že najprijateľnejším odhadom dĺžky historického obdobia *h* je 1000, prípadne 1500 rokov.

Citlivostná analýza vzhľadom na prahovú hodnotu X_0. Pre konštantnú hodnotu dĺžky historického obdobia h = 1000 (aj 1500) rokov sme si zvolili rôzne prahové hodnoty v rozmedzí od absolútneho maxima až po sekundárne maximum. Z analýzy sme usúdili (tu bez ilustrácií), že hraničná hodnota X_0 = 4000 $m^3.s^{-1}$, ktorá je iba o niečo nižšia než hodnota historickej kulminácie, je adekvátna pre danú analýzu MCMC. Pri nižších hodnotách X_0 síce pozorujeme zužovanie intervalu spoľahlivosti návrhových hodnôt, avšak tento kladný aspekt je výrazne znehodnotený záporným dôsledkom, ktorým je fakt, že extrémy v strednej časti spektra (pre doby opakovania T = 10 až 100 rokov) sa postupne dostávajú von z konfidenčného pásu.

Po sériách citlivostných analýz sme si zúžili množinu prijateľných DF, resp. možných kombinácií hodnôt h a X_0. Obr. 2 (vľavo) prezentuje návrhové hodnoty letných kulminačných prietokov s dobou opakovania T = 10 000 rokov pre 5 najpriaznivejších DF (GEV, LP3, LN3, LN2 a JOH), resp. dve hodnoty h (1000 a 1500 rokov) a X_0 = 4000 $m^3.s^{-1}$. Naše odhady na základe metodiky MCMC sú ďalej porovnané aj s odhadom odborníkov ČHMÚ [2]. Z Obr. 2 je zrejmé, že spektrum odhadov $Q_{10\,000}$ je relatívne široké, aj keď sa obmedzíme iba na konkrétnu hodnotu h. Odhad ČHMÚ sa nachádza v spodnej časti tohto spektra.



Obr. 2: Porovnanie návrhových hodnôt kulminačných prietokov (vľavo), resp. objemov povodňových vĺn (vpravo) s dobou opakovania T = 10 000 rokov, resp. ich 90% intervalom spoľahlivosti s podobnými odhadmi ČHMÚ [2].

V prípade objemov povodňových vĺn sme postupovali analogicky: vykonali sme série citlivostných analýz tak pre objemy odpovedajúce konštantnej dĺžke povodňovej vlny ako aj objemy určené subjektívnou metódou. Udalosť z roku 2002 sme taktiež považovali za „historickú“, avšak odhad dĺžky historickej periódy sme mali o niečo menší: uvažovali sme hodnoty v rozmedzí h = 250 až 750 rokov. Porovnanie návrhových hodnôt objemov povodňových vĺn s T = 10 000 rokov s odhadmi ČHMÚ sa nachádza na Obr. 2 (vpravo).

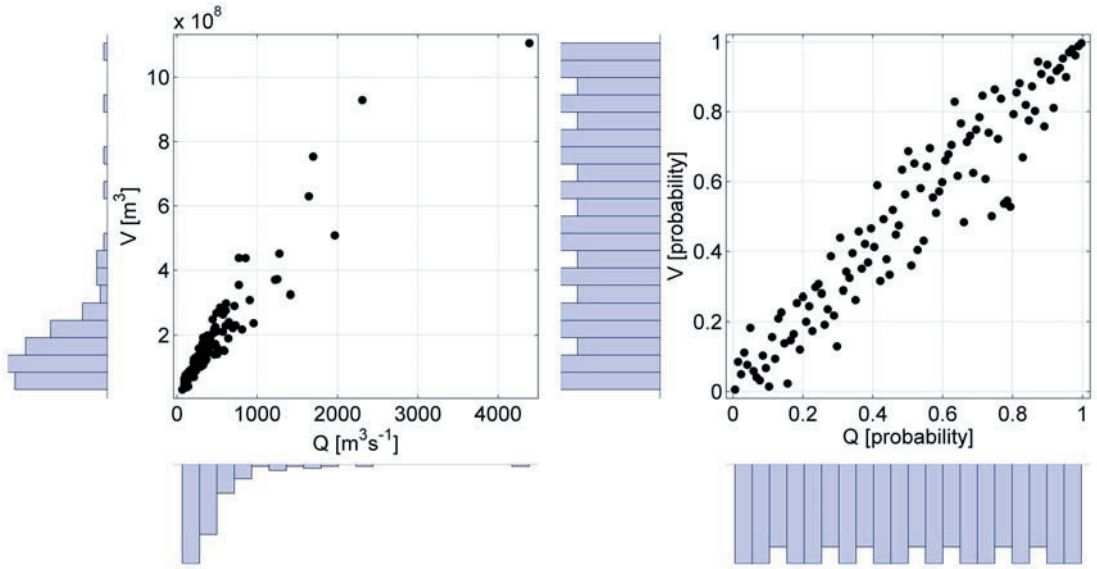
4.2 Združená frekvenčná analýza dvojíc maximálnych prietokov a objemov vĺn

Naším cieľom v združenej frekvenčnej analýze bolo vypočítať odhad objemu povodňovej vlny za podmienenej pravdepodobnosti výskytu kulminačného prietoku $p = 10^{-4}$, t.j. pre dobu opakovania T = 10 000 rokov. K tomuto účelu sme si zvolili matematický aparát kopúl.

Parciálne výsledky jednorozmernej frekvenčnej analýzy pomocou simulácií MCMC, resp. procedúra fitovania kopúl ponúkajú rôzne alternatívne voľby parametrov a metód k získaniu finálnych odhadov návrhových hodnôt. Výsledky výpočtov sumarizované v nasledujúcich častiach štúdia sú triedené podľa:

- *spôsobu výpočtu zodpovedajúcich objemov povodňových vĺn*: okno konštantnej (3 + 1 + 5 dní) a premenlivej (subjektívnej) dĺžky;
- *predpokladanej doby opakovania extrémov*: T = 500 rokov pre objemy povodňových vĺn, resp. 1000 alebo 1500 rokov pre kulminačné prietoky;
- *parametrickej triedy marginálií*: LN3, GEV, LP3, JOH a LN2;
- *rodiny kopúl*: Gumbel, Clayton, Frank, Normal (Gauss), Plackett, t.j. najpoužívanejšie jednoparametrické rodiny kopúl;
- *metódy fitovania kopule*: parametrická (metóda maximalizácie pseudo-vierohodnostnej funkcie) a neparametrická (dve metódy, založené na mierach asociácie Kendallovo τ či Spearmanovo ρ; [7]);
- *pravdepodobnostnej hodnoty kvantilu*: medián (0,5) a 0,6 kvantil.

Na Obr. 3 (vľavo) je zobrazený graf hodnôt maximálneho prietoku Q a k nim vypočítaných hodnôt objemu vlny V (metóda konštantnej dĺžky povodňovej vlny), pričom na Obr. 3 (vpravo) sú rovnaké veličiny po transformácii zodpovedajúcou marginálnou DF na dvojice pravdepodobností.



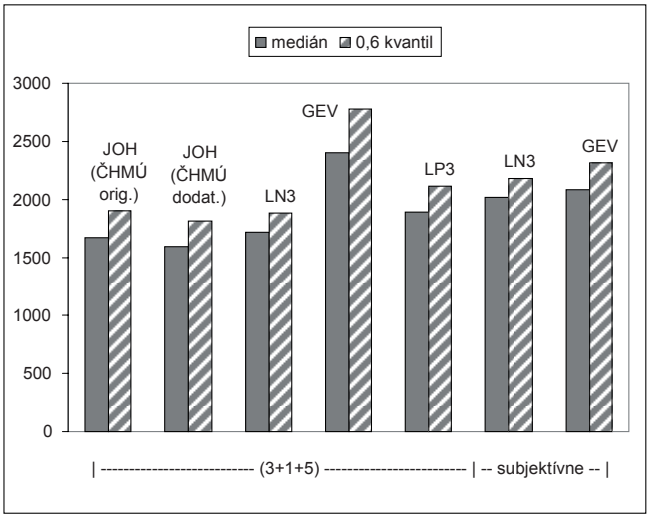
Obr. 3: Pozorované dvojice kulminačných prietokov a odpovedajúcich objemov povodňových vĺn (vľavo) a ich marginálne pravdepodobnosti (vpravo) pri konštrukcii objemov na základe metódy konštantnej dĺžky povodňovej vlny.

Tab. 1: Maximálne hodnoty vierohodnostnej funkcie v procese fitovania a p-hodnoty GoF testov (*goodness-of-fit, testov dobrej zhody*) pre kopule s parametrom odhadnutým troma rôznymi spôsobmi.

Metóda odhadu \ Kopula	Gumbel	Clayton	Frank	Normal	Plackett
okno konštantnej dĺžky (3+1+5)					
Max. vierohodnosť	126,4	96,5	128,1	**135,8**	125,7
GoF test, max. vierohodnosť	0,0235	0,0005	0,0904	**0,1054**	0,0305
GoF test, Kendallovo τ	0,0664	0,0005	0,1004	0,1723	**0,2063**
GoF test, Spearmanovo ρ	——	0,0025	0,1204	0,1773	**0,4151**
okno subjektívnej dĺžky					
Max. vierohodnosť	73,7	59,2	66,7	**77,3**	68,6
GoF test, max. vierohodnosť	0,0165	0,0035	0,0205	**0,1983**	0,0624
GoF test, Kendallovo τ	0,0145	0,0025	0,0275	**0,1613**	0,0884
GoF test, Spearmanovo ρ	0,0125	0,0065	0,0285	**0,1304**	0,1216

Výber najlepšieho odhadu kopule podľa hodnoty vierohodnostnej funkcie a *p*-hodnoty testu dobrej zhody dokumentuje Tab. 1. Väčšie hodnoty znamenajú vyššiu preferenciu (najväčšie sú zvýraznené), pričom stabilne najvyššie preferencie dosiahla *normálna* kopula.

Na Obr. 4 je vizuálne porovnanie návrhových hodnôt podmienených objemov pre 10 000ročný kulminačný prietok podľa metódy výpočtu objemov, aj podľa použitých marginálií. Výška prvého stĺpca zodpovedá mediánu (0,5 kvantilu) a výška druhého stĺpca 0,6 kvantilu. Odhady podľa ČHMÚ sú 1673 mil. m^3, resp. 1900 mil. m^3 z originálnej štúdie a 1595 mil. m^3, resp. 1816 mil. m^3 z jej dodatku [2]. Z uvedených výsledkov je zjavná najmä variabilita medzi metódami, zvýraznená požiadavkou poznať rozdelenie pravdepodobnosti náhodných veličín za hranicou, kde nie sú dostupné pozorovania.



Obr. 4: Porovnanie podmienených objemov letných povodňových vĺn podľa metodiky definície vlny a použitých marginálií s výsledkom analýzy ČHMÚ [2].

5. Záver

Napriek tomu, že frekvenčná analýza systematických a historických údajov prostredníctvom MCMC simulácií vyžaduje subjektívnu voľbu niekoľkých parametrov, táto metodika je transparentná: hypotézy sa dajú explicitne formulovať vo funkcii vierohodnosti, zvolené parametre majú hydrologický význam a ich vplyv na výsledky sa dá overiť na základe citlivostných analýz. Ukázalo sa, že najvýznamnejší vplyv na výsledné intervaly spoľahlivosti odhadov kvantilov majú dva parametre: dĺžka historického obdobia h a hodnota X_0, ktorá definuje hranicu pre zahrnutie „neznámych“ hydrologických udalostí z historického obdobia.

Ako sme predpokladali, rôzne marginálne distribučné funkcie, resp. rodiny kopúl vedú k rozdielnym odhadom návrhových hodnôt objemov povodňových vĺn podmienených návrhovou hodnotou maximálneho prietoku. Na druhej strane môžeme konštatovať, že odhady ČHMÚ sú konzistentné našimi výsledkami aj v prípade ich nezávislého posúdenia pomocou kopúl.

Ďalším zistením štúdie je, že aj keď rozšírenie systematických pozorovaní o jedinú historickú udalosť prináša jednoznačné zlepšenie odhadov v jednorozmernej frekvenčnej analýze, tento krok je takmer nedetekovateľný v združenej frekvenčnej analýze založenej na kopuliach (t.j. návrhové hodnoty objemov povodňových vĺn sú takmer nezávislé od zvolenej hodnoty h; bez ilustrácií). Toto zistenie nie je prekvapením ak si vezmeme do úvahy, že aplikácia kopúl je sprevádzaná nevyhnutnou transformáciou pozorovaných dvojíc údajov do intervalu [0;1] x [0;1], čo spôsobuje citeľnú “stratu informácie“, ktorá je zahrnutá v magnitúde pridanej historickej udalosti.

Poďakovanie

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe zmluvy č. APVV-0443-07 a grantom VEGA 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] BAYLISS, A. C. – REED, D. W., 2001. The use of historical data in flood frequency estimation. Report to MAFF. CEH Wallingford, 87 pp.

[2] BOHÁČ M. – KULASOVÁ, B., 2005. Hydrologická studie pro VD Orlík: Průběhy teoretických povodňových vln s kulminačním průtokem s pravděpodobnosti překročení pQ = 0.0001 s podmíněnými pravděpodobnostmi překročení objemu. Český hydrometeorologický ústav., 14 s.

[3] GAUME, E. – GAÁL, L. – VIGLIONE, A. – SZOLGAY, J. – KOHNOVÁ, S. – BLÖSCHL., G., 2010. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events on ungauged sites. Journal of Hydrology [in press], doi: 10.1016 / j. jhydrol. 2010.01.08.

[4] GENEST, C. – FAVRE, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12 (4), 347-368.

[5] HILL, I. D. – HILL, R. – HOLDER, R. L., 1976. Algorithm AS 99: Fitting Johnson curves by moments. Applied Statistics 25 (2), 180-189.

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

[7] NELSEN, R. B., 2006. An Introduction to Copulas. Springer-Verlag, New York, 2nd edition.

[8] REIS, D. S. – STEDINGER, J. R., 2005. Bayesian MCMC flood frequency analysis with historical information. Journal of Hydrology 313: 97-116.

[9] STEDINGER, J. R. – COHN, T. A., 1986. Flood frequency analysis with historical and paleoflood information. Water Resources Research 22(5): 785-793.

[10] VIGLIONE, A., 2008. The nsRFA Package (Non-supervised Regional Frequency Analysis). Version 0.6-2. Last modified 19-09-2008. Available at: http://cran.r-project.org/ web/packages/ nsRFA/index.html.

Kontakty na autora

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

2 Katedra matematiky a deskriptívnej geometrie, Stavebná fakulta Slovenskej technickej univerzity v Bratislave.