

Slovenská technická univerzita v Bratislave

Stavebná fakulta

Katedra vodného hospodárstva krajiny

Ing. Veronika Soldánová

**PRÍSPEVOK K VODOHOSPODÁRSKEMU
RIEŠENIU ZÁVLAHOVÝCH NÁDRŽÍ**

Autoreferát dizertačnej práce

Na získanie akademického titulu philosophiae doctor v doktorandskom študijnom programe:

Krajinárstvo

Bratislava 2020

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre vodného hospodárstva krajiny Stavebnej fakulty Slovenskej technickej univerzity
v Bratislave

Predkladateľ: **Ing. Veronika Soldánová**
Katedra vodného hospodárstva krajiny
Stavebná fakulta STU, Radlinského 11,
810 05 Bratislava

Školiteľ: **Prof. Ing. Milan Čistý, PhD.**
Katedra vodného hospodárstva krajiny
Stavebná fakulta STU, Radlinského 11,
810 05 Bratislava

Oponenti: **Prof. Ing. Andrej Šoltész, PhD.**
Prof. Ing. Viliam Bárek, PhD.
Ing. Juraj Bezák, PhD.

Autoreferát bol rozoslaný dňa: 30. 9. 2020

Obhajoba dizertačnej práce sa koná 22. 10. 2020 o 9.30 h online

prof. Ing. Stanislav Unčík, PhD.
dekan Stavebnej fakulty STU v Bratislave

Obsah:

Úvod	3
1 Ciele dizertačnej práce	3
2 Metodika	4
2.1 Použité charakteristiky podobnosti povodí	5
2.2 Výpočet nemeraných prietokov pre potrebu bilancie nádrže	5
2.2.1 Toky bez merania s dvoma a viacerými analógmi	5
2.2.2 Príprava údajov	6
2.2.2.1 Konštrukcia nových premenných (feature engineering)	6
2.2.3 Verzie vstupných údajov	7
2.2.4 Lineárne modely	7
2.2.5 Nelineárne modely	8
2.2.5.1 Trénovanie modelu strojového učenia	8
2.2.5.2 Krížová validácia pre trénovanie modelov strojového učenia	8
2.2.5.3 Ladenie modelu strojového učenia	8
2.2.5.4 Posúdenie fyzikálnej opodstatnenosti modelov	9
2.2.6 Zrážkovo-odtokové modelovanie	9
2.2.7 Hodnotenie kvality modelovania	9
3 Výsledky	10
3.1 Analýza a porovnanie povodí z pohľadu analógie	10
3.1.1 Topografická podobnosť - nadmorské výšky	10
3.1.2 Pedologická a hydrogeologická podobnosť	12
3.1.3 Využitie územia	13
3.1.4 Drenážna sieť	13
3.1.5 Klimatické charakteristiky	14
3.1.6 Topografický index a čísla odtokových kriviek	15
3.1.7 Kalibrovateľnosť	16
3.1.8 Špecifické prietoky	16
3.1.9 Vyhodnotenie podobnosti povodí	17
3.2 Výpočty nemeraných prietokov s dvoma analógmi	17
3.2.1 Zrážkovo-odtokové modelovanie	17
3.2.2 Lineárne a nelineárne štatistické modely	19
3.2.2.1 Ensembleový model pri použití dvoch zdrojových povodí	21
3.2.2.2 Posúdenie fyzikálnej opodstatnenosti modelov	23
3.3 Vodná bilancia nádrže	24
3.3.1 Softvérové riešenie vodohospodárskej funkcie závlahovej nádrže – R balík HydBal	24
3.3.1.1 Výpočet bilancie nádrže	24
3.3.1.2 Postup výpočtu bilancie nádrže v R	25
3.3.1.3 Prechod povodňovej vlny nádržou	26
3.3.1.4 Postup výpočtu povodňovej vlny v R	26
3.3.1.5 Ďalšie funkcie balíka HydBal	26
3.3.2 Výpočet bilancie nádrže Horné Orešany s použitím vypočítaných a meraných prietokov	27
4 Záver	28
5 Literatúra	29
Zoznam publikácií	30

Úvod

Výstavba závlahových nádrží a ich rekonštrukcie sú zásahmi, ktoré výrazne a dlhodobo ovplyvnia štruktúru okolitej krajiny, jej bezpečnosť, využívanie a hospodárstvo, ale taktiež vyžadujú nemalé finančné investície. Z týchto dôvodov je nevyhnutné dôsledné prehodnotenie všetkých konštrukčných parametrov, aspektov a situácií, ktoré môžu v blízkej či vzdialenejšej budúcnosti ovplyvňovať ich existenciu a účel. Toto prehodnocovanie prebieha v réžii vodohospodárskeho manažmentu krajiny. Jednou z jeho úloh je preveriť vplyv náhodnosti, variability a trendu klimatických a hydrologických veličín na návrhové parametre určitého objektu alebo opatrenia, ktorým môže byť výška hrádze protipovodňových stavieb, zásobný objem závlahovej nádrže alebo rôzne parametre závlah. Návrh zavlažovacieho systému vyžaduje vyhodnotenie možnosti zásobnej nádrže poskytnúť vodu na zavlažovanie. Na takéto vyhodnotenie sa používajú simulačné modely a pri ich aplikácii sú nevyhnutným vstupom časové rady klimatických a hydrologických premenných. Táto práca sa venuje situácii, keď niektoré z týchto údajov nie sú k dispozícii.

Témou práce je prispieť k prehodnoteniu závlahových nádrží tvorbou viacerých nástrojov a metodických postupov. Potreba inovovaného prístupu k ich prehodnocovaniu vyplýva napríklad z dokumentu „H2Odnota je voda – Akčný plán na riešenie dôsledkov sucha a nedostatku vody (Uznesenie vlády SR č. 110/2018)“. Jedným z cieľov tohto akčného plánu je „podporiť výskum a modelovanie v oblasti hodnotenia rizika a zraniteľnosti na stanovenie vývoja množstva a kvality vodných zdrojov, požiadaviek na vodu a stanovenie priorít na zásobovanie vodou pri dlhotrvajúcom suchu“. Tento dokument nadväzuje na Zákon č. 364/2004 Z. z. („zákon o vodách“), ktorý je implementáciou Rámcovej smernice o vode (2000/60/ES) a vytvára podmienky na znižovanie nepriaznivých účinkov sucha a nedostatku vody.

Jedným z výstupov práce je inovovaný prístup k posúdeniu zásobného objemu závlahovej nádrže a jej celkovej schopnosti plniť konkrétne požiadavky týkajúce sa potreby vody. Pri návrhu nových vodohospodárskych stavieb je dôležité vyhodnotiť akúkoľvek realistickú kombináciu klimatických a hydrologických situácií, ktoré môžu na konkrétnom mieste v budúcnosti nastať. V praxi často nastáva problém s chýbajúcimi časovými radmi prietokov, ktoré sú veličinou potrebnou pri bilancovaní vody v nádrži, pretože na prevažnej časti malých vodných tokov sa tento údaj nezaznamenáva. V nadväznosti na túto úlohu autorka vypracovala metodiku výpočtu prietokov v nemeráných povodiach s použitím algoritmov strojového učenia, ktorých presnosť je porovnávaná s presnosťou hydrologických modelov.

1 Ciele dizertačnej práce

Cieľom práce je tvorba softvérových nástrojov a metodiky pre vybrané problémy vodohospodárskeho riešenia závlahovej nádrže. V práci navrhnuté riešenia budú predbežne verifikované na pilotnom území, neskôr môžu byť použité v nadväzujúcich prácach.

Jedným z problémov pri riešení vodohospodárskej funkcie závlahovej nádrže je nedostatok meraní na malých tokoch, ktoré takéto nádrže zásobujú. Hlavným cieľom práce je vytvoriť metodiku modelovania prietokov v nemeráných povodiach s použitím analogického prístupu. Tento prístup analyzuje blízke, hydrologicky podobné povodia a využíva ich pre odhad prietokov na nemeranom povodí.

Súčasťou práce je výber povodí vhodných pre tento účel, teda identifikácia povodí s dostatočne podobným hydrologickým správaním ako cieľový tok. Hodnotenie podobnosti možno uskutočniť

na základe charakteristík povodí, ako: topografia, pedologická a hydrogeologická podobnosť, využitie územia, vlastnosti odtokovej siete, klimatické charakteristiky, zrážkovo-odtokové charakteristiky a podobne.

Ďalším cieľom je analýza situácií, ktoré môžu nastať pri požiadavke stanovenia nemeraných prietokov a pre ktoré sú potrebné rôzne metodické prístupy. Napríklad môže byť k dispozícii krátky rad meraní, ktoré sa na toku začali vykonávať v prípravnej fáze konkrétneho plánovaného vodohospodárskeho zámeru. Cieľom je zistiť, aké zlepšenie presnosti modelovaných prietokov je možné dosiahnuť pri použití krátkych, napríklad 1 až 5 ročných radov meraných prietokov.

Ďalšou možnou situáciou je existencia iba jedného blízkeho podobného/analogického povodia. Toto povodie vystupuje ako kalibračné a na ňom vytvorený model je použitý na výpočet neznámych prietokov v cieľovom povodí. Je potrebné zistiť, či je vhodnejšie použitie hydrologického modelovania alebo regresných modelov.

Ak sú k dispozícii 2 a viac vhodných povodí, je situácia komplikovanejšia, ale aj výhodnejšia a v práci je navrhnutý postup, ako využiť údaje z viacerých tzv. zdrojových povodí na určenie neznámych cieľových prietokov v nemeranom povodí.

Posledná fáza a cieľ práce je zameraný na overenie modelovaných prietokov v reálnej vodohospodárskej aplikácii, v bilančnom výpočte nádrže, ktorý zároveň demonštruje použitie vytvoreného softvéru.

Keďže má ísť o pilotnú prácu, dôležité je, aby sa jej výsledky dali použiť pri plošnejšom testovaní navrhovaných metód. Za týmto účelom je navrhnuté softvérové riešenie, spracované formou nadstavby jazyka R, tzv. balíka, pričom ďalším z cieľov je jeho umiestnenie na verejne prístupné úložisko, odkiaľ by bol k dispozícii pre použitie širokej odbornej verejnosti.

2 Metodika

Na výpočet prietokov v nemeranych povodiach (Ungauged catchments) (Blöschl, 2006) je možné použiť zrážkovo-odtokové modely, do ktorých vstupujú zrážky spolu s ďalšími klimatickými veličinami. Možným riešením získania parametrov týchto modelov je využitie kalibrácie na povodiach nachádzajúcich sa v blízkom a podobnom regióne, ktoré majú prietoky merané. Predpokladom podobných odtokových reakcií je podobná klíma, geologické pomery, topografia, vegetácia, využitie územia a pôdy. Takéto podobnosti by vo všeobecnosti mali viesť k podobným odtokovým reakciám.

V hydrológii sa proces prenosu informácií z podobných povodí všeobecne nazýva hydrologická regionalizácia alebo hydrologická analógia (Blöschl, Sivapalan, 1995).

Regionalizačné prístupy možno rozdeliť na tie, ktoré využívajú buď hydrologické modelovanie alebo regresné techniky (v práci sú použité oba uvedené prístupy):

- metódy v prvej skupine prenášajú parametre zrážkovo-odtokového modelu medzi povodiami. Tieto parametre sa potom použijú na generovanie kontinuálneho prietoku;
- druhá skupina metód namiesto prenosu hydrologických parametrov modelu medzi povodiami využíva regresnú závislosť. Tieto modely zvyčajne používajú vzťahy medzi vstupmi (zrážky, teploty, prietoky na podobných povodiach) a výstupmi (prietoky na toku bez meraní).

2.1 Použité charakteristiky podobnosti povodí

Aby bolo možné určiť vhodnosť použitia povodia na (analogický) výpočet prietokov v inom povodí, je potrebné identifikovať jeho hydrologickú podobnosť s cieľovým povodím. Na určovanie tejto podobnosti je možné použiť viacero aspektov podobnosti; v práci boli použité nasledovné:

- topografická podobnosť (nadmorské výšky, sklon, expozícia),
- pedologická a hydrogeologická podobnosť,
- využitie územia,
- drenážna sieť,
- klimatické charakteristiky,
- topografický index (TWI) a čísla odtokových kriviek (CN),
- kalibrovateľnosť.

Topografická podobnosť povodí bola spracovaná na základe analýzy digitálneho výškového modelu (digital elevation model - DEM) s veľkosťou gridu 20x20 metrov. Analýzy DEM prebiehali v jazyku R v prostredí RStudio, odkiaľ boli pomocou balíka RSAGA automatizovane volané príkazy vopred inštalovaného GIS softvéru SAGA.

Celý pracovný postup analýzy DEM bol spracovaný v jednom funkčnom R skripte, ktorý súbory načítaval, predspracoval a postupne na ne aplikoval komplexný rad analýz, ktorých grafické a štatistické výstupy zapisoval na disk. Tento prístup umožňuje automatizáciu topografickej analýzy, čo je vhodné z hľadiska cieľu dizertačnej práce – poskytnúť nástroje pre plošnejšie nasadenie analýzy povodí a nádrží, ktoré z nich zberajú vodu.

Riečna sieť porovnávaných povodí bola generovaná z digitálneho výškového modelu, spolu s ďalšími vlastnosťami – plochou povodia, dĺžkou rozvodnice a dĺžkami jednotlivých rádoz riečnej siete. Z dĺžky rozvodnice a plochy povodí bol vypočítaný index členitosti rozvodnice, z plochy a celkovej dĺžky drenážnej siete jej hustota, vyjadrená dĺžkou drénu v km na 1 km².

Súhrnné vyhodnotenie podobnosti bolo spracované formou penalizácie. V jednotlivých kategóriách charakteristík podobnosti bolo jedným bodom hodnotené povodie, ktoré sa svojimi vlastnosťami popísanými v predošlom texte najviac vymykalo vlastnostiam ostatných povodí. Druhé najmenej podobné povodie získalo pol bodu.

2.2 Výpočet nemeráných prietokov pre potrebu bilancie nádrže

Autorka testovala navrhované metódy tým spôsobom, že vždy jedno povodie bolo považované za nemerané a ostatné sa používali pri výpočte jeho prietokov. Táto obmena cieľového povodia bola vykonaná v záujme dôkladnejšieho preverenia efektívnosti jednotlivých metód. Druhý dôležitý prvok je, že sa pri testovaní predpokladali tri rôzne situácie:

1. Situácia, keď predpokladáme, že sa na cieľovom povodí vykonávalo krátkodobé meranie prietokov a tieto údaje sa použili pri zostavení modelov a následnom určení neznámych prietokov.
2. Výpočet neznámych prietokov bez existencie takýchto meraní pomocou jedného zdrojového povodia (analogu).
3. Výpočet pomocou dvoch, prípadne viacerých zdrojových povodí.

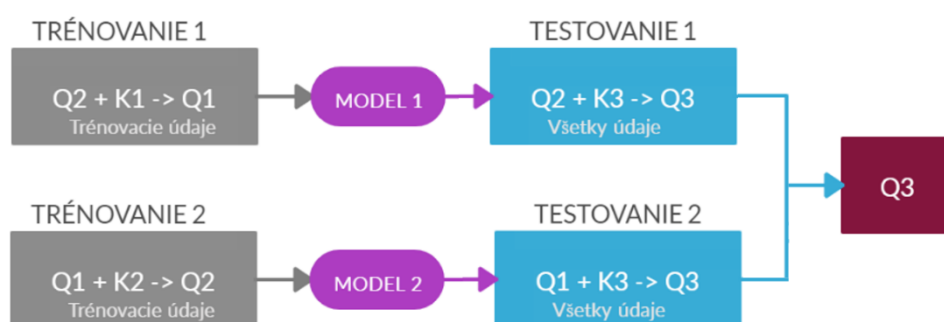
Pre obsírnosť riešenia jednotlivých situácií je v autoreferáte bližšie popísaný len výpočet pomocou dvoch zdrojových povodí, ktorý autorka považuje za originálny prínos práce.

2.2.1 Toky bez merania s dvoma a viacerými analógmi

Vylepšenou metódou výpočtu prietokov v nemeráných povodiach na základe analógie, je použitie dvoch alebo viacerých analógových povodí namiesto jedného. Vhodným nástrojom pri takýchto podmienkach (údajových zdrojoch) sú štatistické modely a modely strojového učenia (nazývané aj dátovo riadené modely). Model je kalibrovaný na základe prietokov jedného analógového povodia a klimatických údajoch druhého povodia, pričom cieľom tejto kalibrácie je výpočet prietokov

v druhom z analogických povodí. Takýto model už potrebuje pre výpočet cieľového povodia iba klimatické údaje tohto cieľového povodia (v podstate v tejto štruktúre prichádza cieľové povodie na miesto druhého analógového povodia z kalibračnej fázy). Takto vytvorený model je preto možné použiť na výpočet prietokov v nameranom povodí, pričom do neho vstupujú údaje o jeho klíme a použijú sa prietoky z „prvého“ analógu. V tejto práci je skúmaný potenciál tejto metódy iba s dvoma analógmi; ak by sa tento prístup ukázal sľubný, je metódu možné vylepšiť použitím viacerých „prvých“ analógov. Kľúčovým krokom je výber vhodných analógov, ktorý bude riešený v kapitole 3.1.

Ďalším rozšírením tejto metódy je možnosť kombinácie analógových povodí. Tento variant používa na predikciu cieľového prietoku kombináciu analógových povodí vo vymenenom poradí. Podstatou výpočtu je tvorba dvoch modelov, pričom prvý je trénovaný na prietokoch druhého analogického povodia a klíme z prvého analogického povodia a druhý opozitne; do jeho tréningu vstupuje prietok z druhého analógu a klíma z prvého. Výsledný prietok je získaný jednoduchým ensemblovým modelom, ktorý bol v tejto práci vypočítaný ako aritmetický priemer prietokov predpovedaných týmito dvoma modelmi, aj keď sú k dispozícii aj sofistikovanejšie varianty. Schéma kombinovaného prístupu je znázornená na obrázku 2.1.



Obrázok 2.1: Schéma výpočtu prietokov s dvoma analógovými povodiami – kombinovaný variant, Q1, Q2 – merané prietoky analógových povodí, K1, K2 – klimatické veličiny popisujúce miestnu klímu v analógových

2.2.2 Príprava údajov

Aby bol celý postup reprodukovateľný a alternatívy výpočtov medzi sebou porovnateľné, boli definované rady dátumov, ktoré boli pre každé povodie a výpočtový algoritmus totožné a slúžili na rozdelenie pozorovaní na tréningovú a testovaciu množinu. Sada údajov každého z povodí obsahovala spolu 13729 pozorovaní, pričom rozdelenie na tréningové a testovacie údaje bolo vykonané v pomere 70:30 (tréningovanie:testovanie). Dátumový rozsah použitých údajov bol od 30.01.1980 do 31.08.2017.

2.2.2.1 Konštrukcia nových premenných (feature engineering)

Lepšie výsledky v presnosti simulovaných prietokov možno dosiahnuť rozšírením súboru vstupných údajov o nové premenné odvodené z pôvodných (základných) údajov na základe hydrologických znalostí alebo rôznymi transformáciami pôvodných údajov. V „machine learning“ terminológii sa pre túto časť modelovania používa termín „feature engineering“ a jeho aplikácia je v kontexte hydrologickej aplikácie strojového učenia ďalším prínosom tejto práce, keďže jeho použitie v hydrologickej komunite zatiaľ nie je úplne bežné.

Premennými vzniknutými v tomto procese sú:

- **potenciálna evapotranspirácia,**
- **premenná zohľadňujúca kumuláciu a topenie snehu,**
- **posunuté premenné,**
- **kumulované premenné.**

Evapotranspirácia úzko súvisí s teplotou a ďalšími klimatickými veličinami, avšak transformácia týchto údajov do evapotranspirácie a zahrnutie tohto údaje do výpočtu samostatne, môže algoritmom pomôcť dosiahnuť lepšie výsledky.

V zimných mesiacoch nemožno považovať zrážku za iniciátora odtoku – ak padla ako sneh, odtok sa tvorí až po jeho roztopení. Pre oblasti bez monitorovania hodnôt topenia snehu, je možný ich dopočet viacerými, na údaje rôzne náročnými spôsobmi (v práci použitá metóda Walter (2005) implementovaná v balíku EcoHydrology (Fuka et al., 2018)). Premennú, ktorá v tomto kroku vznikla predstavuje časový rad s denným krokom združenej premennej - súčtom kvapalných zrážok a roztopeného snehu v danom dni. Tieto časové rady sa pre jednoduchosť v práci nazývajú ako „sneh“, aj keď nejde o vyslovene snehový údaj.

Posunuté premenné zohľadňujú časovo závislé veličiny nielen z aktuálneho dňa, ale aj z tých predchádzajúcich, pretože na odtok z povodia nemá vplyv napríklad iba dážď z daného dňa, ale aj zrážky z predchádzajúcich dní. Vo väčšine výpočtov sa preto predpoveď prietoku v nasledujúcom časovom kroku nepočíta iba na základe súčasných hodnôt klimatických premenných, ale aj ich hodnôt 1,2,3 a 4 dni dozadu.

Tiež možno predpokladať, že po obdobiach s väčším množstvom zrážok, keď sa povodie nasýti vodou, bude okamžitý odtok a prietoky vyššie ako po suchších obdobiach. Pre každý deň bola preto skonštruovaná kumulovaná premenná s nasledovnými časovými rozsahmi:

- kumulované hodnoty zrážok a topenia snehu za 5. až 30. deň,
- kumulované hodnoty evapotranspirácie za 5 až 30 deň,
- kumulované hodnoty priemerných denných teplôt za 1. až 3. deň a za 4. až 15. deň.

Výsledný súbor s údajmi pripravenými na tréning modelov obsahoval spolu 22 vysvetľujúcich premenných, tzv. features, z čoho 3 boli základné, dve dopočítané a ďalších 17 vzniklo v procese tvorby nových premenných.

2.2.3 Verzie vstupných údajov

Údaje, ktorých tvorba bola popísaná v predchádzajúcom texte, boli základom pre 3 rôzne kombinácie údajov vstupujúcich do výpočtov:

1. základné údaje, t.j. prietoky Q [mm.km^{-2}] (prepočítané na jednotku plochy podľa rozlôh jednotlivých povodí), priemerné teploty T_{ave} [$^{\circ}\text{C}$] a zrážky Z [mm],
2. zrážkové údaje + kumulované + posunuté premenné,
3. zrážkovo/snehové údaje + kumulované + posunuté premenné.

Hlavným cieľom testovania rôznych vstupov bolo preveriť, na koľko presný môže byť výpočet analógie so základnými údajmi a následne porovnať, aký veľký vplyv môže mať doplnenie premennej topenia snehu namiesto zrážok a zároveň preskúmať, aký potenciál na zvýšenie presnosti modelu má zahrnutie novovytvorených údajov.

2.2.4 Lineárne modely

Najpoužívanejším modelom pre regresnú analýzu je lineárna regresia, v prípade viacerých vysvetľujúcich premenných - viacnásobná lineárna regresia, ktorá je najbežnejšou regresnou metódou používajúcou sa na nájdenie príslušných koeficientov výslednej lineárnej závislosti metódou najmenších štvorcov. Základná rovnica je:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \dots + \epsilon \quad (2.1)$$

kde, Y je závislá premenná, X_i sú vysvetľujúce premenné, β_0 je konštantný člen, β_i je koeficient pre každú vysvetľujúcu premennú a ϵ je chyba merania.

Hlavnou podmienkou lineárnej regresie je to, že vysvetľujúce premenné X_i nesmú byť vzájomne príliš korelujúce. V prípade, že túto podmienku nemožno splniť, je možné použiť vhodnejšie, takzvané regularizované algoritmy, ktoré jednak hľadajú riešenie horeuvedenej rovnice, ale snažia sa riešenie „regularizovať“ tak, aby malo isté vlastnosti. Pri použití metódy Least Absolute Selection And Shrinkage Operator (LASSO) sa regularizácia vykonáva v tom zmysle, aby sa čo najviac koeficientov premenných vynulovalo. Vznikne tak jednoduchší model s menej premennými, v ktorom sa eliminuje multikolinearita a ktorý zvyčajne lepšie generalizuje. Algoritmus hľadá nadbytočné vysvetľujúce premenné a ich koeficienty nastaví na nulu (Roth, Fisher, 2008).

Ďalšia z použitých lineárnych metód je tzv. „best subsets“ regresia, kde najlepšou podmnožinou (subsetom) sa myslí taká podmnožina vysvetľujúcich premenných, ktorých použitie v regresnej rovnici vedie k najlepšiemu výsledku. Použitý bol balík (nastavba) programu R „leaps“ (Lumley, 2017).

2.2.5 Nelineárne modely

Medzi výhody nelineárnych modelov patrí často vyšší stupeň presnosti výsledkov; vo vedeckej literatúre sa preukázala lepšia presnosť najmä pri krátkodobých predpovediach prietokov (Yaseen et al., 2015). Nasleduje stručný popis použitých modelov.

Algoritmus Random Forest (RF) (Breiman, 2001) je „les“ tvorený súborom stromov, ktoré môžu byť v závislosti od riešeného problému buď klasifikačné alebo regresné. Je to vlastne ensemblová metóda. Výsledná hodnota je priemerom hodnôt, ktoré vypočíta každý strom. Stromy sú na riešenie tréňované pomocou tréňovacích údajov – toto je terminológia strojového učenia pre kalibráciu. Užívateľ si volí, alebo hľadá pri procese ladenia modelu optimálny počet stromov v lese. Každý strom je tréňovaný pomocou inej vzorky údajov (bootstrapu), čo vedie k získaniu rôznych stromov.

XGBoost je skratka výrazu Extreme Gradient Boosting a je to algoritmus založený na tzv. boostovaní (posilňovaní – zlepšovaní riešenia). Je ensemblovou metódou zvyčajne využívajúcou plytké regresné stromy – podobne ako Random Forest. XGBoost vyhodnotí presnosť modelu v predchádzajúcom štádiu a potom vyvinie ďalší model, ktorý sa sústreďuje hlavne na v predchádzajúcom modeli nesprávne vypočítané vzorky. Takéto „zosilňovanie“ pokračuje, kým sa nedosiahne požadovaná úroveň presnosti. XGBoost vyžaduje ladenie pomerne veľa parametrov, zatiaľ čo RF je takmer bez ladenia. V tejto práci bol použitý balík pre softvér R catboost (Dorogush, Ershov, Gulin, 2018).

Ďalší použitý model využíva framework H2O (LeDell et al., 2019). H2O je open source platforma, charakteristická škálovateľnosťou pre rôzne veľké modely a podporuje najpoužívanejšie štatistické algoritmy a algoritmy strojového učenia vrátane XGBoost, algoritmov hlbokého učenia (Deep Learning), náhodných stromov a ďalších. H2O má špičkovú funkčnosť AutoML, ktorá automaticky prechádza všetkými algoritmi a ich hyperparametrami, aby vytvorila rebríček najlepších modelov a následne ensemblový model typu stacking vyberie najlepšie modely a pomocou ich výsledkov, t.j., nie pomocou pôvodných údajov, natrénuje základným modelom výsledok.

2.2.5.1 Tréňovanie modelu strojového učenia

Model bol vytváraný pomocou tzv. tréningových údajov (v hydrológii sa používa pojem kalibračné údaje) a overený pomocou nezávislých testovacích údajov, aby sa vyhodnotila očakávaná presnosť modelu. Pri tréňovacích aj testovacích údajoch musíme poznať okrem napr. klimatických vstupov aj cieľovú / počítanú premennú – napríklad prietok. Až pri reálnom použití modelu cieľovú premennú nepoznáme – pre jej výpočet bol model zostavený. Výsledkom je model, alebo parametre modelu a predikovaná premenná získaná pri najmenšej dosiahnutej chybovosti v testovacej časti tohto procesu.

2.2.5.2 Krížová validácia pre tréňovanie modelov strojového učenia

V tomto prístupe boli dostupné údaje rozdelené na tréňovací a testovací súbor, pričom pri tréňovaní sa používala krížová validácia. Tréňovací súbor sa pri tomto type validácie náhodne rozdelí napr. na 10 rovnako veľkých častí (10 foldová validácia). V prvom cykle bude prvých deväť oddielov použitých ako tréňovací súbor a desiaty bude použitý ako testovací. V ďalšom cykle bude na testovanie použitý deviaty oddiel a všetky zvyšné na tréňovanie modelu, atď. Tento postup sa opakuje, pokiaľ sa na testovanie nepoužije každý z 10 počiatočných foldov. Výsledná presnosť modelovania je priemerom presností dosiahnutých počas všetkých prebehnutých cyklov (Solomatine, 2006).

2.2.5.3 Ladenie modelu strojového učenia

Každý z modelov strojového učenia má svoje špecifiká. Obsahuje rôzne parametre, ktoré je potrebné nastaviť podľa konkrétnej úlohy. V procese ladenia modelov je vyžadovaná znalosť použitých algoritmov a účel ich parametrov.

2.2.5.4 Posúdenie fyzikálnej opodstatnenosti modelov

Aj keď sú modely strojového učenia do určitej miery čiernymi skrinkami, existuje v súčasnosti mnoho nástrojov pomocou ktorých možno posúdiť zmysluplnosť vyvinutých modelov. Jedným zo základných nástrojov pre tento účel je vyhodnotenie dôležitosti jednotlivých nezávislých premenných (variable importance). Dôležitosť vstupných premenných, je mierou ich relatívnej dôležitosti pri výpočte prietokov. Jej to pomerné číslo, ktoré buď model zistí a vyhodnotí v rámci výpočtu alebo sa dá zisťovať rôznymi inými technikami pre interpretáciu modelov vyvinutých v nedávnom období v rámci strojového učenia. Umožňuje posúdiť fyzikálnu opodstatnenosť modelu na základe jeho hydrologického posúdenia. Vieme teda predpokladať, ktorá premenná by mala byť pri výpočte neznámych prietokov dôležitejšia a ktorá menej. Údaj o dôležitosti premenných poskytuje preto určitú možnosť interpretácie modelu.

2.2.6 Zrážkovo-odtokové modelovanie

Výpočet pomocou hydrologického modelu prebiehal s použitím koncepčného hydrologického modelu TUV (Viglione, Parajka, 2014), ktorého štruktúra vychádza zo švédskeho HBV modelu (Bergström, 1995). TUV model pracuje v dennom časovom kroku a má 15 kalibračných parametrov, ktoré umožňujú kontrolovať a upravovať vlastnosti simulovaných javov a stavy zásob vody. Ako vstupné údaje model vyžaduje: úhrny zrážok na povodí, priemerné teploty vzduchu na povodí, priemernú potenciálnu evapotranspiráciu na povodí.

Analogický výpočet prietokov pomocou TUV modelu spočíva v kalibrácii jeho 15 parametrov na analógovom (zdrojovom) povodí. Kalibrované parametre hydrologického modelu možno uskutočniť napríklad pomocou genetického algoritmu (GA), v práci bol použitý variant GA, ktorý sa nazýva rgenoud (Mebane et al., 2011).

Kalibrácia (hľadanie optimálnych hodnôt parametrov hydrologického modelu) bola vykonaná prostredníctvom genetického algoritmu na údajoch o prietoku a klíme z analógového (zdrojového) povodia (zrážkach, priemernej teplote a potenciálnej evapotranspirácii). Najdôležitejšími parametrami genetického algoritmu boli veľkosť populácie, počet generácií a hraničné hodnoty pre 15 kalibrovaných parametrov modelu TUV. Súbor najlepších získaných parametrov sa použil pri modelovaní prietokov v „nemeranom“ povodí. Ďalšími údajmi vo výpočte nemeráných prietokov boli klimatické údaje tohto povodia. Takto vypočítané denné prietoky boli prevedené na mesačné, na ktorých bola následne zisťovaná ich zhoda s meranými mesačnými prietokmi, keďže pre účely testovania boli merané údaje známe.

2.2.7 Hodnotenie kvality modelovania

Na hodnotenie kvality modelovania bol v práci použitý R balík hydroGOF (Zambrano-Bigiarini, 2017), ktorý obsahuje rozsiahly súbor štatistických ukazovateľov a umožňuje ich jednoduché použitie. Ako hlavný ukazovateľ zhody modelovaných a meraných prietokov bola v hydrológii často používaná štatistika - Nash-Sutcliffe Efficiency (NSE). NSE určuje mieru zhody denných simulovaných a meraných prietokov, pričom dokonalá zhoda je vyjadrená číslom 1. Ďalšími použitými ukazovateľmi boli:

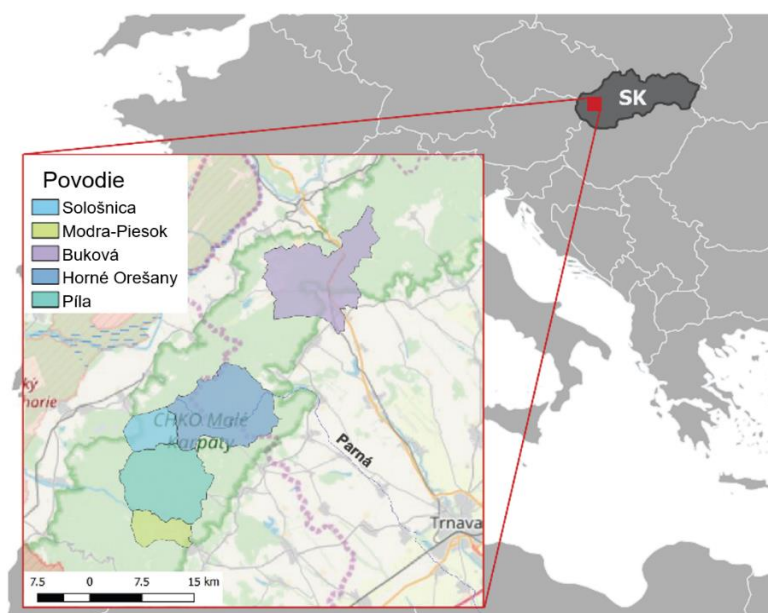
- PBIAS - priemerná percentuálna odchýlka medzi pozorovaným a simulovaným prietokom,
- pearsonov korelačný koeficient (r) popisuje úroveň kolinearity medzi simulovanými a nameranými údajmi,
- stredná kvadratická chyba (RMSE) určuje veľkosť chyby v jednotkách skúmaných premenných,
- stredná a stredná absolútna chyba (ME, MAE),
- CP - koeficient perzistencie ,
- VE - objemová účinnosť (volumetric efficiency).

3 Výsledky

3.1 Analýza a porovnanie povodí z pohľadu analógie

Metodika analogického výpočtu prietokov v nemeraných povodiach bola testovaná na piatich povodiach nachádzajúcich sa v Malých Karpatoch, na západnom Slovensku. Išlo o povodie prislúchajúce k vodomernej stanici Horné Orešany, nachádzajúcej sa na malom horskom toku Parná (veľkosť povodia 37,33 km²), povodie stanice Buková, na toku Trnávka (42,96 km²), povodie stanice Modra-Piesok (9,38 km²), nachádzajúcej sa na Vištuckom potoku, ďalej povodie Píla (32,9 km²) na toku Gidra a Sološnica (10,47 km²) na Sološnickom potoku.

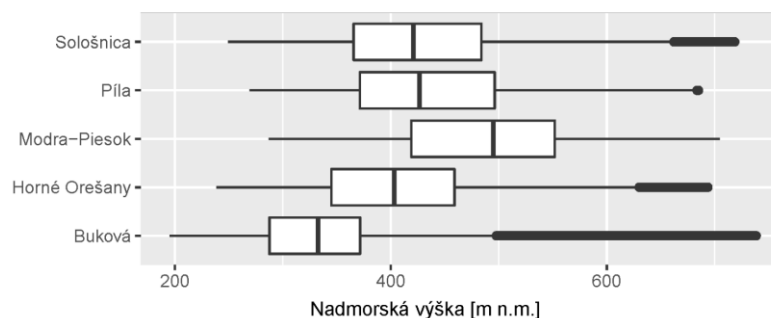
Keďže ide o susedné povodia existuje predpoklad, že sa budú vyznačovať podobnými črtami a to podobnou klímou, topografiou, geológiou, pôdnym pokryvom, využitím pôdy, genézou odtoku, atď. To je prvý zohľadňovaný aspekt v tejto práci – vyberané sú blízke povodia k povodiu, na ktorom sú určované neznáme prietoky. Autorka odporúča vybrať čo najviac okolitých povodí a následne ich prefiltrovať vyhodnotením uvedených vlastností a zvoliť pre konečný výpočet tie najvhodnejšie (najpodobnejšie cieľovému povodiu). Lokalizácia riešených povodí je znázornená na obrázku 3.1.



Obrázok 3.1: Poloha vybraných povodí

3.1.1 Topografická podobnosť - nadmorské výšky

Najmenšiu, avšak aj najväčšiu nadmorskú výšku možno pozorovať v povodí Buková, pričom minimálna hodnota nadmorskej výšky je 195,31 m n. m. a maximálna 738,35 m n. m. Medián tohto povodia sa oproti ostatným povodiám pohybuje vo výrazne nižších číslach, a nachádza sa tu aj najväčší počet odľahlých hodnôt, čo možno dobre vidieť na porovnaní krabicových grafov nadmorských výšok v jednotlivých povodiach na obrázku 3.2.



Obrázok 3.2: Krabicové grafy nadmorských výšok riešených povodí

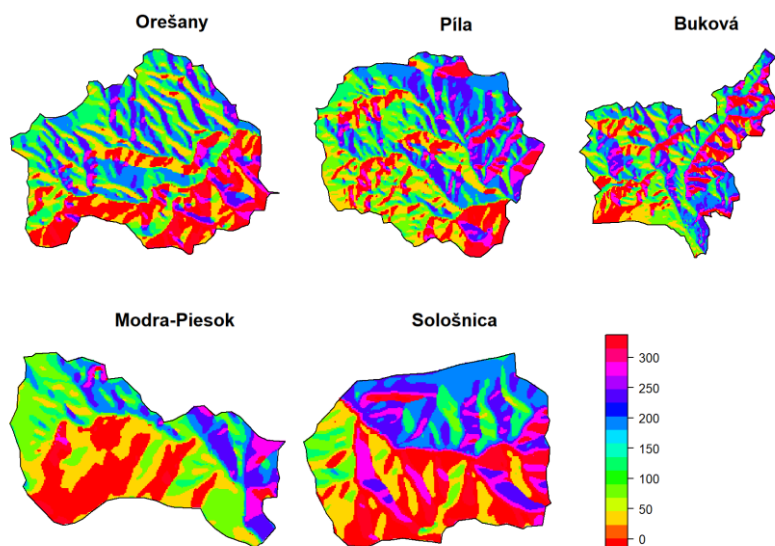
Pri povodí Modra-Piesok možno v porovnaní s ostatnými povodiami pozorovať priestorovo odlišné usporiadanie terénu. Ako jediné z povodí ma záporný koeficient šikmosti, čo znamená, že sa v povodí nachádza viac veľkých nadmorských výšok, ako tých menších. Na základe uvedeného popisu možno predpokladať, že sa v povodí budú vyskytovať strmšie svahy, čoho následkom bude aj väčší povrchový odtok.

Sklon

Z hľadiska sklonu sú si najpodobnejšími povodiami Píla a Buková, s priemernými sklonmi $9,9^\circ$ (Píla) a $9,47^\circ$ (Buková). Povodím s najmenšou priemernou sklonitosťou, $8,92^\circ$, je povodie Modra-Piesok, ktorého distribúcia hodnôt sklonitosti zároveň vykazuje najmenšiu smerodajnú odchýlku v súbore a to 4,46. Naopak najväčšiu smerodajnú odchýlku má povodie Sološnica (5,25), v ktorom sa nachádzajú aj najstrmšie svahy so sklonom až $33,83^\circ$ stupňa. V tomto ohľade je toto povodie najpodobnejšie Horným Orešanom, ktorých najstrmšie svahy sa pohybujú v hodnotách okolo 29° . Kompletný súhrn porovnania sklonov v riešených povodiach je uvedený v tabuľke 3.1.

Expozícia

Expozícia povodí je znázornená na obrázku 3.3. V porovnávaných povodiach je najviac zastúpenou orientáciou severná a severovýchodná, čo podmieňuje pomalšie topenie snehu v jarných mesiacoch a zamrzanie tokov v zime.



Obrázok 3.3: GIS analýza expozície riešených povodí

Tabuľka 3.1 obsahuje zhrnutie štatistických charakteristík predošlých analýz nadmorských výšok, sklonitosti a expozície povodí.

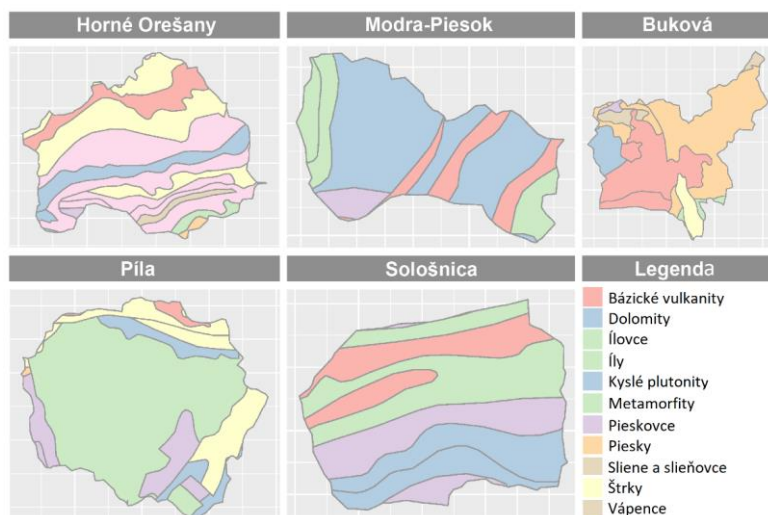
Tabuľka 3.1: Štatistické charakteristiky popisujúce topografické vlastnosti povodí

Charakteristika		Povodie				
		Horné Orešany	Píla	Buková	Modra-Piesok	Sološnica
Nadmorská výška	max	693.65	684.92	738.35	704.64	718.52
	min	238.26	269.05	195.31	286.94	249.44
	mean	406.97	434.25	332.47	488.21	426.67
	sd	82.41	79.50	72.40	85.94	83.21
	skew	0.44	0.34	0.91	-0.13	0.40
Sklon	max	29.02	27.66	27.17	25.55	33.83
	mean	11.12	9.91	9.47	8.92	15.18
	sd	5.25	4.76	5.71	4.46	5.25
	skew	0.22	0.31	0.46	0.72	-0.39
Expozícia	mean	134.23	145.61	149.56	102.18	152.87
	sd	96.16	91.79	94.55	87.72	108.71
	skew	0.27	0.14	0.05	0.69	-0.04

3.1.2 Pedologická a hydrogeologická podobnosť

Horninové zloženie riešených povodí je značne diferencované. Povodie Píla má prevládajúce zastúpenie kyslých plutonitov (74.60%) a len malé zastúpenie ďalších horninových prvkov. Povodie Orešany má najrozmanitejšiu horninovú skladbu a vyskytuje sa v ňom až 8 typov hornín, pričom najviac zastúpené sú vápence a najmenej kyslé plutonity. Najmenej rôznorodé sú povodia Modra-Piesok a Sološnica; obsahujú len 4 typy hornín. Vzájomnú podobnosť v horninovom zložení možno pozorovať medzi povodiami Píla a Modra-Piesok, nakoľko prevládajúcou horninou v oboch sú spomínané kyslé plutonity. Povodie Buková ako jediné obsahuje v menšom množstve aj štrky a íly, čo svedčí o rôznorodosti horninového zloženia tohto povodia; najväčšie zastúpenie v ňom majú piesky a dolomity. GIS analýza horninového zloženia všetkých povodí je na obrázku 3.4.

Pri určovaní podobnosti povodí na základe zastúpenia zvodnencov sa zistilo, že vo všetkých povodiach okrem jedného má jednoznačne najvyššie zastúpenie typ „oblasti s takmer žiadnymi množstvami podzemných vôd“. Výnimku predstavuje Buková, u ktorej je táto kategória zastúpená najmenšou výmerou. Druhým najčastejšie zastúpeným typom zvodnencov sú „Priestorovo obmedzené alebo nespojité hydrogeologicky vysoko produktívne zvodnence, alebo rozsiahle a stredne produktívne zvodnence“. Najmenej zastúpeným typom sú „menšie zvodnence s obmedzenými množstvami podzemných vôd miestneho významu“. Podľa výmery jednotlivých zvodnencov možno ako najpodobnejšie povodia označiť dvojicu Horné Orešany a Pílu a ako druhý pár Modru-Piesok a Sološnicu.



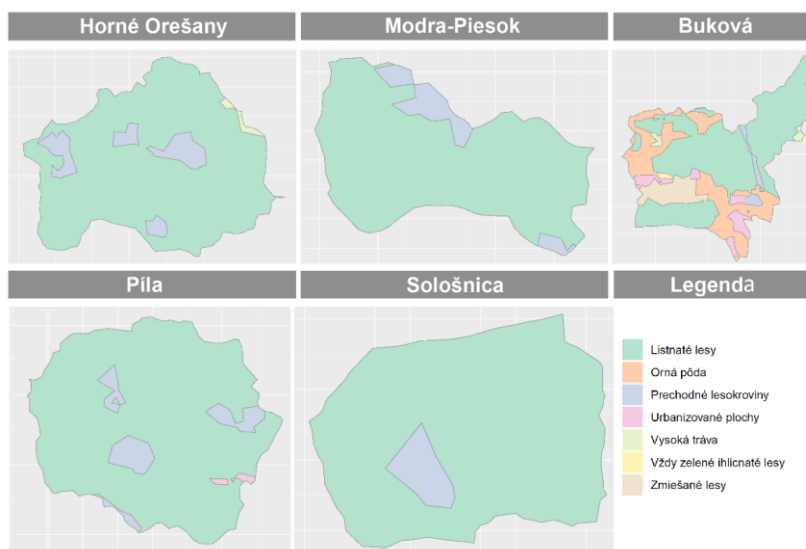
Obrázok 3.4: Horninové zloženie povodí

Pedologické zloženie riešených povodí pozostáva výlučne z hlinitej, piesočnato-hlinitej a hlinito-piesočnatej pôdy. Hlinitá pôda je prevládajúcim pôdnym druhom pri troch povodiach a to Horných Orešanoch, Bukovej a Sološnici, pričom Sološnica je ňou pokrytá homogénne celá. Horné Orešany a Buková majú v menšej miere zastúpenú piesočnato-hlinitú pôdu, čo ich robí spomedzi všetkých povodí navzájom najpodobnejšími. Píla má ako jediná zastúpené všetky 3 pôdne druhy; najviac zastúpená je piesočnato-hlinitá pôda a najmenej hlinitá. Modra-Piesok má ako jediná prevládajúci pôdny druh hlinito-piesočnatú pôdu.

Tabuľka obsahujúca zhrnutie charakteristík povodí, ktoré boli získané spracovaním GIS vrstiev vo formáte *.shp*, sa nachádza v úplnej verzii dizertačnej práce.

3.1.3 Využitie územia

V každom z riešených povodí majú jednoznačne najväčší podiel listnaté lesy. Pri pohľade na obrázok 3.5 je vidieť takmer dokonalú podobnosť povodí Horné Orešany, Modra-Piesok, Píla a Sološnica, ktoré sú pokryté z prevažnej väčšiny listnatými lesmi a prechodnými lesokrovinami. V povodí Buková sa ako v jedinom zo súboru vyskytujú aj väčšie výmery ostatných typov využitia územia, z čoho najpodstatnejším je orná pôda a urbanizované plochy. Zároveň je v tomto povodí najmenej listnatých lesov. Nakoľko sú to plochy náchylné na povrchový odtok, čím výrazne ovplyvňujú hydrologickú odozvu povodia, je pravdepodobné, že budú mať zásadný vplyv aj na využiteľnosť tohto povodia v analogických výpočtoch.



Obrázok 3.5: Využitie územia riešených povodí

Percentuálne zastúpenie jednotlivých plôch je uvedené v tabuľke 3.2.

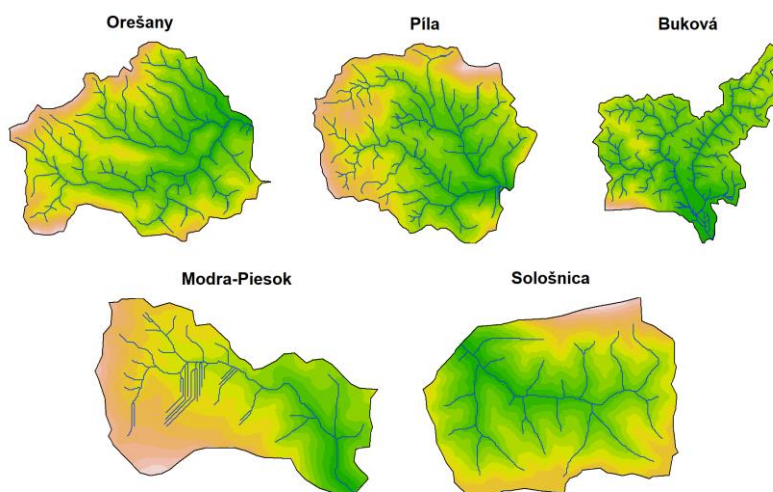
Tabuľka 3.2: Percentuálne zastúpenie plôch využitia územia v riešených povodiach

Charakteristika		Povodie				
		Horné Orešany	Píla	Buková	Modra-Piesok	Sološnica
Využitie územia	Listnaté lesy	88.91	91.17	51.17	89.13	91.36
	Orná pôda	-	-	20.24	-	-
	Prechodné lesokroviny	9.25	7.13	5.33	10.87	8.64
	Urbanizované plochy	-	1.70	10.17	-	-
	Vysoká tráva	1.84	-	2.16	-	-
	Vždy zelené ihličnaté lesy	-	-	1.14	-	-
	Zmiešané lesy	-	-	9.80	-	-

3.1.4 Drenážna sieť

Rovnaké hodnoty hustoty drenážnej siete (0,13) majú povodia Píla a Modra-Piesok. Podľa grafického znázornenia drenážnych sietí na obrázku 3.6 možno konštatovať, že usporiadanie drenážnej siete Píly, je podobnejšie skôr povodiu Horné Orešany, ktoré dosahuje hustotu 0,09 a Bukovej, s hustotou 0,16.

Index členitosti rozvodnice označuje za najpodobnejšie povodia Horné Orešany a Bukovú, Pílu a Sološnicu, pričom Modra-Piesok je s hodnotou indexu 1,71 tvarovo najodlišnejším povodím z tejto skupiny.



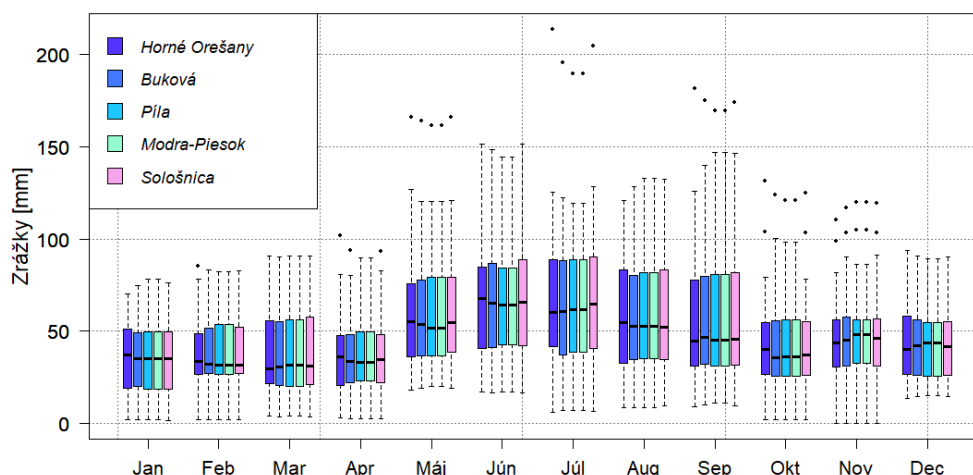
Obrázok 3.6: Drenážne siete povodí

Tabuľka 3.3: Porovnanie vlastností drenážnej siete riešených povodí

Charakteristiky (v km a km ²)	Povodie				
	Horné Orešany	Píla	Buková	Modra-Piesok	Sološnica
Dĺžka rozvodnice	29.00	24.55	15.00	39.65	13.14
Plocha povodia	37.33	32.90	9.38	42.96	10.46
Dĺžka drenážnej siete	3.18	4.28	1.52	5.72	1.00
Hustota drenážnej siete I. rádu	0.046	0.070	0.088	0.070	0.051
Hustota drenážnej siete II. rádu	0.008	0.015	0.015	0.012	0.012
Hustota drenážnej siete III. rádu	0.005	0.008	0.009	0.008	0.004
Hustota drenážnej siete	0.09	0.13	0.16	0.13	0.10
Index členitosti rozvodnice	1.34	1.21	1.38	1.71	1.15

3.1.5 Klimatické charakteristiky

Na hodnotenie klimatickej podobnosti môže byť z klimatických vlastností vybraná korelácia medzi povodiami v nasledujúcich veličinách: priemerné mesačné zrážky, potenciálna evapotranspirácia a teplota, priemerné ročné zrážkové úhrny, priemerná ročná potenciálna evapotranspirácia a teplota. Skutočnosťou však je, že ak sa nachádzajú porovnávané povodia v tesnej blízkosti, nebudú klimatické rozdiely medzi nimi veľmi významné. Túto skutočnosť možno potvrdiť vizuálnym porovnaním krabicových grafov priemerných mesačných zrážok (obrázok 3.7) jednotlivých povodí.

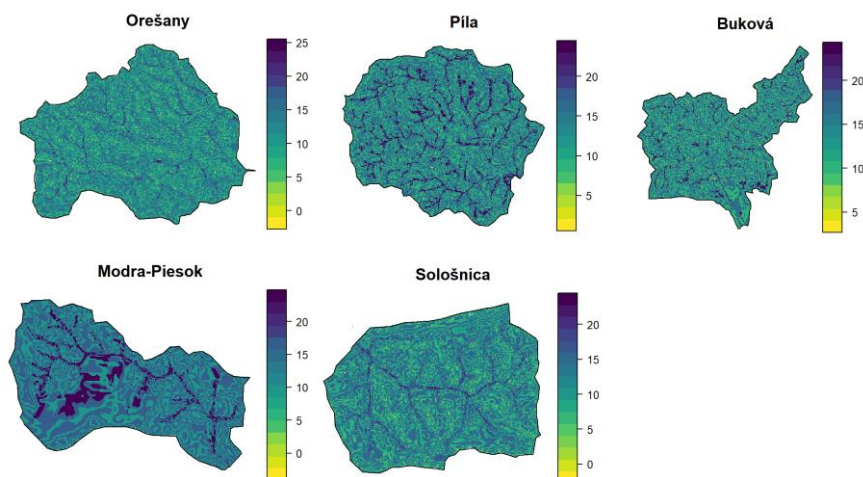


Obrázok 3.7: Priemerné mesačné úhrny zrážok porovnávaných povodí

Numericky vyjadrená podobnosť klímy prostredníctvom korelačného koeficientu toto tvrdenie potvrdzuje. Pri vzájomnej korelácii zrážok z povodí dosahoval najnižší korelačný koeficient hodnotu 0.985 pri povodiach Píla a Buková. Pri všetkých ostatných vzájomných koreláciách zrážok a priemerných teplôt to boli hodnoty väčšie ako 0.991.

3.1.6 Topografický index a čísla odtokových kriviek

Index topografickej vlhkosti (TWI) je najčastejšie používaný hydrologicky založený topografický index, ktorý opisuje tendenciu bunky akumulovať vodu (bunky DEM). Hodnotené povodia vykazujú veľmi podobnú distribúciu hodnôt TWI. S mierne vyššou priemernou hodnotou TWI možno vyčleniť povodie Modra-Piesok. Ostatné povodia nevykazujú väčšie rozdielnosti a preto ich možno na základe hodnotenia indexu topografickej vlhkosti považovať za podobné.



Obrázok 3.8: Index topografickej vlhkosti

Číslo CN krivky určuje schopnosti pôdy infiltrovať povrchový odtok vzhľadom na využívanie pôdy, pôdny pokryv a predchádzajúci stav pôdnej vlhkosti (Amutha, Porchelvan, 2009) a pohybuje sa v rozpätí hodnôt 0 až 100, pričom 0 znamená, že všetka zrážková voda infiltrovala a povrchový odtok neexistuje. Povodia Horné Orešany, Píla a Sološnica majú veľmi podobné odtokové vlastnosti. Priemerná CN hodnota týchto povodí je 72,24 pre Horné Orešany, 72,58 pre Pílu a 72,77 pre Sološnicu. Najvyššiu priemernú CN hodnotu (73,43) dosahuje Buková, v ktorej sa vyskytujú vyššie CN hodnoty hlavne v miestach urbanizovaných plôch a ornej pôdy. Najvyššia CN hodnota v tomto povodí je 91. Najnižšia priemerná CN hodnota sa nachádza v povodí Modra-Piesok, ktoré je ňou pokryté homogénne. Pri podrobnejších analýzach by bolo vhodné zistiť CN krivky vlastným výpočtom, čo sa vzhľadom na prácnosť v tejto práci nerobilo.

3.1.7 Kalibrovateľnosť

Dôležitým ukazovateľom potenciálnej vhodnosti povodia pre analogické výpočty prietokov je jeho kalibrovateľnosť, t.j. schopnosť povodia byť dostatočne dobrým prediktorom pre predpovedanie vlastných budúcich prietokov. Kalibrovateľnosť bola zisťovaná hydrologickým modelom TUW. Každé z povodí bolo hraničným dátumom 1.1.2005 rozdelené na tréningové (do tohoto dátumu; spolu 9103 pozorovaní) a testovacie údaje (od tohoto dátumu; spolu 4626 pozorovaní). Na prietokoch a klimatických údajoch spadajúcich do tréningového obdobia bol kalibráciou genetickým algoritmom vytvorený najlepší možný model predikujúci budúce (testovacie) prietoky.

Hodnotou ukazovateľa Nash-Sutcliffe (NSE) bola pre všetkých 5 povodí vyjadrená miera presnosti konečného modelu. Najlepšie dosiahnuté NSE, ktoré v tomto prípade predstavujú kalibrovateľnosť povodí, sú uvedené v tabuľke 3.4.

Tabuľka 3.4: Kalibrovateľnosť povodí

Povodie	Nash-Sutcliffe (NSE)
Horné Orešany	0,617
Píla	0,681
Buková	0,426
Modra-Piesok	0,683
Sološnica	0,317

Najvyššie hodnoty NSE (0,681 a 0,683) boli dosiahnuté v povodiach Píla a Modra-Piesok. Tento fakt značí, že v týchto povodiach nenastali v priebehu pozorovaní významnejšie zmeny a že ich hydrologický režim je stabilný a teda sú vhodnými kandidátmi pre použitie v analogickom výpočte prietokov iného povodia. Najnižšiu hodnotu kalibrovateľnosti dosiahli povodia Sološnica (0,317) a Buková (0,426), z čoho je možné predpokladať, že aj v prípade, ak by boli tieto povodia na záver vyhodnotené ako vysoko podobné ostatným povodiám, nebudú vhodnými pre hydrologickú analógiu.

3.1.8 Špecifické prietoky

Porovnávanie prietokov v reálnej aplikácii navrhutej metódy použiť nemožno, pre tento účel však môžeme prostredníctvom nich overiť, ktoré zo skúmaných atribútov podobnosti povodí správne identifikovali hydrologicky najpodobnejšie povodia.

Vzájomný vzťah špecifických prietokov povodí možno vyjadriť korelačnou maticou (obrázok 3.9). Na tejto matici je ku každému povodiu doplnený súčet korelačných koeficientov.



Obrázok 3.9: Korelačná matica špecifických prietokov riešených povodí

Povodie Buková má tento súčet najmenší, z čoho možno usúdiť, že čo sa týka prietokového režimu, je ostatným povodiám najmenej podobná. Takéto povodia môže byť vhodnejšie pre výpočet prietokov v nemeraných povodiach nepoužívať, pretože nízka korelácia môže znamenať, že v povodí

existujú príčiny ovplyvňujúce odtok, ktoré na ostatných povodiach, alebo v danej oblasti vo všeobecnosti nie sú.

Z korelácií jednotlivých prietokov vyplýva, že napríklad pre analogický výpočet prietokov v povodí Horné Orešany sú najvhodnejšími povodiami Píla a Sološnica. Sološnica však pri výpočte kalibrovateľnosti povodí dosiahla zo všetkých povodí najmenšiu hodnotu NSE a to 0,317, z čoho vyplýva, že aj na túto charakteristiku je potrebné nahliadať komplexnejšie, pretože môže byť zavádzajúca.

3.1.9 Vyhodnotenie podobnosti povodí

V jednotlivých kategóriách charakteristík podobnosti bolo jedným penalizačným bodom hodnotené povodie, ktoré sa svojimi vlastnosťami popísanými v predošlom texte najviac vymykalo vlastnostiam ostatných povodí. Druhé najmenej podobné povodie získalo pol bodu. Výsledky penalizácie jednotlivých povodí, spolu s ich celkovým súčtom sú uvedené v tabuľke 3.5.

Tabuľka 3.5: Vyhodnotenie podobnosti povodí penalizáciou

Charakteristika	Povodie				
	Horné-Orešany	Píla	Buková	Modra-Piesok	Sološnica
Topografia	-	-	1	-	0.5
Pedologické a hydrogeologické pomery	-	-	0.5		1
Využitie územia	-	-	1	-	-
Drenážna sieť	-	-	-	1	0.5
Klimatické podmienky	-	-	-	-	-
Topografický index a čísla odtokových kriviek	-	-	1	0.5	-
Kalibrovateľnosť	-	-	0.5	-	1
Súhrnná korelácia prietokov	2.82	2.64	2.23	2.3	2.47
Súčet :	0	0	4	1.5	3

Najviac penalizovaným a teda najodlišnejším povodím je Buková. Ako druhé najmenej podobné povodie vyšla z hodnotenia Sološnica, ktorá získala po jednom bode v kategóriách pedologické a hydrogeologické pomery a kalibrovateľnosť, pričom práve kalibrovateľnosť je jedným z najdôležitejších ukazovateľov vhodnosti povodia pre výpočet prietokov analógiou. Pri porovnaní klimatických podmienok nebolo bodované žiadne z povodí, keďže všetky vykazovali vysokú úroveň podobnosti. Výsledkom tejto kapitoly je výber troch navzájom najpodobnejších povodí, za ktoré boli metódou penalizácie označené Horné Orešany, Píla a Modra-Piesok.

3.2 Výpočty nemeráných prietokov s dvoma analógmi

3.2.1 Zrážkovo-odtokové modelovanie

Hydrologické modelovanie bolo vykonané s použitím zrážkovo-odtokového modelu TUW, ktorý je spracovaný v balíku TUW model (Viglione, Parajka, 2014). Keďže v tomto type výpočtov vystupovali 2 analogické povodia, bolo možné otestovať viacero spôsobov určenia výsledných prietokov a to:

1. Priemerovanie parametrov.

- Hľadanie kompromisných parametrov jedného výsledného hydrologického modelu pre obe analógové povodia. Toto sa realizovalo maximalizáciou **priemeru** ich Nash-Sutcliffe (NSE) koeficientov v objektívnej funkcii. Týmto modelom sa následne vypočítali výsledné neznáme prietoky na cieľovom povodí. Neznáme prietoky sa počítajú (ako aj v ostatných alternatívach) pri použití klimatických údajov z cieľového povodia.
- To isté ako bod 1a, len vylepšené multikriteriálnou optimalizáciou – okrem maximalizácie priemeru NSE koeficientov sa pri kalibrácii minimalizoval aj ich rozdiel, čo malo za cieľ eliminovať riešenia, kde sa optimalizácia príliš sústreďuje iba na jedno z povodí (napríklad

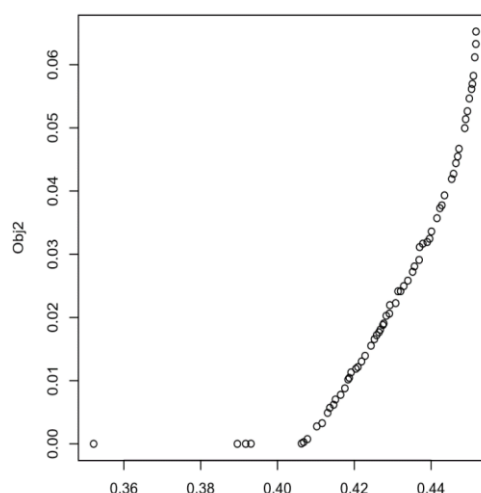
keď **priemerný NSE** je síce najvyšší, ale jeden z NSE je veľmi vysoký a druhý je relatívne nízky). Neznáme prietoky sa opäť určili na základe jedného kompromisného modelu s kompromisnými parametrami a s klimatickými údajmi cieľového povodia.

2. **Priemerovanie prietokov.** Výpočet dvoch (prípadne viacerých) verzií neznámych prietokov kalibrovaním hydrologického modelu postupne na každom z analógových povodí. Výsledné prietoky sa získajú ako priemerné prietoky z aplikovania týchto modelov na cieľové povodie. Takto je nakalibrovaných viacero modelov zväčša s rôznymi parametrami (na rozdiel od alternatívy 1, kde je výsledkom iba jeden model).

Pri prvom spôsobe je prítomnosť dvoch kalibračných povodí zakomponovaná v hlavnom kritériu objektívnej funkcie, ktorým bola aj pri predošlých výpočtoch prietokov podľa jedného analógu hodnota NSE. Cieľom objektívnej funkcie bolo nájsť najväčšiu dosiahnuteľnú hodnotu NSE, počítanú ako priemer NSE, získaných samostatnými výpočtami prietokov na dvoch kalibračných povodiach v rámci jednej objektívnej funkcie.

Pozmenenou verziou predošlého spôsobu je multikritériálna optimalizácia. Pri tomto spôsobe je v objektívnej funkcii okrem priemerného NSE zavedené ďalšie optimalizačné kritérium a to rozdiel jednotlivých NSE. Zatiaľ čo priemerný NSE je maximalizované kritérium, čo znamená, že objektívna funkcia hľadá jeho najvyššiu hodnotu, rozdiel NSE je minimalizovaný. Tento krok má zabezpečiť, aby model nebol nadmerne prispôbený jednému z analógových povodí na úkor toho druhého, čo by znamenalo, že nie je dostatočne všeobecný pre celú riešenú oblasť. Objektívna funkcia v tomto prípade teda hľadá kombináciu parametrov TUV modelu, ktoré dosahujú čo najvyšší priemerný NSE, pri zachovaní čo najmenšieho rozdielu hodnôt dielčích NSE. Na implementáciu multikritériálnej optimalizácie v R bola použitá funkcia *caRamel* z rovnomenného balíka (Le Moine, Monteil, Zaoui, 2019).

Pre viackritériálnu optimalizáciu neexistuje jediné riešenie, ktoré súčasne optimalizuje každé kritérium. Preto hľadá tzv. pareto optimálne riešenia, čo znamená, že žiadna z objektívnych funkcií nemôže byť už vylepšená bez toho, aby došlo k degradácii niektorých ďalších kritérií. Je teda akýmsi najlepším možným kompromisom. Bez ďalších subjektívnych preferenčných informácií sa všetky pareto riešenia považujú za rovnako dobré. Cieľom je získať reprezentatívny súbor pareto optimálnych riešení, ktoré môžu byť graficky vyjadrené ako tzv. pareto front znázornený na obrázku 3.10 (výpočet prietokov Modry). Na osi x a y sú hodnoty účelovej funkcie pre obidve zadané kritériá, vykreslené body sú pareto riešenia, ktoré formujú pareto front. Spomedzi nich sa volí (subjektívne) optimum – v našom prípade sme volili body v pravom hornom rohu grafu, kde je zaručená minimálna hodnota rozdielu oboch NSE a maximálna hodnota ich priemeru.



Obrázok 3.10: Pareto graf znázorňujúci vzťah medzi optimalizačnými kritériami – priemerným NSE (os X) a absolútnou hodnotou rozdielu NSE (os Y) získaných na dvoch kalibračných povodiach

Posledným testovaným spôsobom optimalizácie TUV modelu bolo vytvorenie dvoch nezávislých modelov pre každé zo zdrojových povodí zvlášť. Hľadanie optimálnych parametrov pre TUV model jednotlivých analogických povodí prebiehalo samostatne. Za použitia výsledných dvoch súborov najlepších dosiahnutých parametrov boli namodelované dve série prietokov, ktoré boli jednoducho aritmeticky spriemerované, následne prevedené na mesačné prietoky, pre ktoré bola zisťovaná zhoda s nameranými prietokmi cieľového povodia.

Súhrnné výsledky hodnôt NSE dosiahnutých prostredníctvom týchto troch popísaných spôsobov kalibrácie dvoch analógových povodí sú uvedené v tabuľke 3.6.

Tabuľka 3.6: Hodnoty koeficientu Nash-Sutcliffe dosiahnuté pri rôznych spôsoboch kalibrácie modelu

Povodie		Použitý spôsob kalibrácie		
Cieľové	Analogické	1	2	1b
Horné Orešany	Modra-Piesok	0.641	0.62	0.648
	Píla			
Píla	Horné Orešany	0.618	0.616	0.629
	Modra-Piesok			
Modra-Piesok	Horné Orešany	0.428	0.415	0.430
	Píla			

1 - priemerovanie koeficientov Nash-Sutcliffe (NSE) v objektívnej funkcii, 2 - použitie dvoch objektívnych funkcií, 1b - multikriteriálna optimalizácia.

Podľa výsledných hodnôt koeficientov NSE, možno za najmenej úspešnú metódu označiť metódu č.2. O niečo lepšie výsledky boli dosiahnuté metódou č.1a a najlepšie výsledky boli dosiahnuté multikriteriálnou optimalizáciou, teda metódou 1b.

Pri porovnávaní výsledkov hydrologického modelovania na základe jedného (podrobný popis a dosiahnuté výsledky sú uvedené v dizertačnej práci) a dvoch analogických povodií, bolo najvýraznejšie zlepšenie dosiahnuté pri výpočte prietokov povodia Píla, ktoré malo pri predošlom výpočte takmer vyrovnané hodnoty NSE 0,606 pri kalibrovaní na Orešanoch a 0,604 na Modre. Pri zahrnutí oboch analógov presnosť simulácie prietokov stúpla na 0,629.

3.2.2 Lineárne a nelineárne štatistické modely

Pri tomto type výpočtu je dôležité zdôrazniť, že napríklad výpočet prietokov Horných Orešian na základe prietokov z Modry a Píly, nie je to isté ako výpočet Orešian podľa Píly a Modry, pretože záleží na poradí analógových povodií. Pri dvoch analógových povodiach prvé povodie slúži ako zdroj pomocného prietoku medzi nezávislými premennými regresného modelu a druhé „supluje“ cieľové povodie, t.j. jeho prietok slúži ako závislá premenná. Na výpočet prietokov tohto supľujúceho povodia je teda trénovaný model za pomoci prietokov z prvého analógu a klimatických údajov. Keď je model natrénovaný, miesto druhého analógu zaujme nemerané povodie a počítajú sa cieľové prietoky. Keďže v reálnej situácii nie je možné vyhodnotiť, ktoré povodie je vhodnejšie ako prvý analóg a ktoré ako supľujúce (druhý analóg), bola vypočítaná aj alternatíva, pri ktorej sa prvé a druhé analogické povodia vymenia a inak zvyšok výpočtu prebieha rovnako. Výsledkom sú 2 odlišne namodelované rady prietokov toho istého cieľového povodia s prípadnou možnosťou kombinovania výsledku ich priemerovaním, prípadne váženým priemerovaním, čo je popísané neskôr.

Pri tomto type výpočtu boli vykonávané experimenty s rôznymi typmi vstupných údajov, interakciami a normalizovaním cieľovej premennej pomocou logaritmovania. V tabuľke 3.7 je uvedený prehľad vývoja koeficientu Nash-Sutcliffe (NSE) pri použití rôznych typov vstupov a ich modifikácie. Na uvedené varianty vstupov (základné, zrážkové, snehové údaje) reagovali všetky použité modely veľmi podobne, pre ukážku bolo zvolené podrobné vyhodnotenie metódy „Best subset regression“.

Tabuľka 3.7: Vyhodnotenie vplyvu rôznych vstupných údajov do modelu Best subset prostredníctvom hodnôt Nash-Sutcliffe koeficientu (NSE)

Cieľové povodie		Horné Orešany					
Analogické povodie		Píla+Modra			Modra+Píla		
Použité údaje:		B	P	S	B	P	S
Bez interakcií		0.691	0.708	0.714	0.501	0.561	0.568
S interakciami		0.722	0.731	0.734	0.547	0.642	0.661
Logaritmované		0.767	0.797	0.803	0.461	0.603	0.615

Cieľové povodie		Píla					
Analogické povodie		Modra+Horné Orešany			Horné Orešany+Modra		
Použité údaje:		B	P	S	B	P	S
Bez interakcií		0.496	0.552	0.559	0.694	0.686	0.686
S interakciami		0.538	0.643	0.656	0.694	0.695	0.692
Logaritmované		0.506	0.649	0.658	0.709	0.742	0.746

Cieľové povodie		Modra-Piesok					
Analogické povodie		Píla+Horné Orešany			Horné Orešany+Píla		
Použité údaje:		B	P	S	B	P	S
Bez interakcií		0.49	0.515	0.521	0.494	0.501	0.503
S interakciami		0.505	0.551	0.552	0.495	0.511	0.514
Logaritmované		0.461	0.517	0.52	0.386	0.45	0.461

Použité údaje: B – základné (base), P-zrážkové (precipitation), S – snehové (snow)

Z uvedenej tabuľky vyplýva, že pri lineárnej metóde „best subset regression“ pri údajoch bez interakcií a logaritmov, dosiahli v každom výpočte najvyššie presnosti modely založené na snehových vstupných údajoch (až na 1 výnimku, kde sa NSE vypočítané podľa snehových a zrážkových údajov rovnali). Rovnako tomu bolo aj pri využití interakcie premenných a pri logaritmovaných prietokoch.

Zakomponovanie interakcií zvýšilo presnosť výpočtu pri základných, zrážkových aj snehových údajoch, pričom najvyšší nárast NSE nastal pri výpočte prietokov Horných Orešian z Modry a Píly podľa snehových údajov a to z NSE 0,568 na 0,661.

Logaritmovanie prietokov malo taktiež kladný vplyv na presnosť prietokov, avšak nie vo všetkých prípadoch. Presnosť oproti interakciám naopak mierne klesla v prípade, že sa v úlohe nemeraneho povodia vyskytlo povodie Modra, alebo ak z neho boli brané analogické prietoky, teda vystupovalo vo výpočte ako prvý analóg. V prípade, že bolo povodie Modra v úlohe druhého analógu, na ktorého miesto bolo v ďalšom kroku dosádzané nemerane povodie, presnosti modelovania rástli. Úspech logaritmovania prietokov je pravdepodobne podmienený podobnosťou povodií, čo poukazuje na dôležitosť výberu vhodného analógu. Najväčší nárast NSE v tomto prípade nastal pri výpočte Horných Orešian z Píly a Modry, pričom pri logaritmovaní prietokov v súbore snehových údajov bola dosiahnutá hodnota NSE = 0,803, ktorá vysoko prevyšuje presnosť získanú akoukoľvek inou metódou, vrátane hydrologického modelovania.

Vplyv interakcie premenných a logaritmovania prietokov bol skúmaný podrobnejšie aj pri modeloch strojového učenia – Random Forest, CatBoost a AutoML, pretože autorka predpokladala, že môžu tieto zložité, nelineárne závislosti vystihnúť lepšie ako lineárne modely. Vo výpočtoch boli používané výlučne snehové údaje, keďže ich najväčšiu vhodnosť preverili predchádzajúce výpočty. Vyhodnotenie modelov strojového učenia prostredníctvom NSE je uvedené v tabuľke 3.8.

Tabuľka 3.8: Vyhodnotenie vplyvu modifikácií vstupov pri nelineárnych modeloch prostredníctvom Nash-Sutcliffe koeficientu (NSE)

Metóda	Vstupy/Povodia	Orešany		Píla		Modra	
		Píla	Modra	Modra	Orešany	Píla	Orešany
Metóda	Vstupy/Povodia	Modra	Píla	Orešany	Modra	Orešany	Píla
TUW	Multikriteriálna optimalizácia		0.648		0.629		0.430
Random Forest	Bez interakcií	0.74	0.728	0.722	0.653	0.552	0.515
	Interakcie	0.749	0.743	0.734	0.68	0.559	0.527
	Logaritmy	0.731	0.738	0.746	0.662	0.526	0.507
CatBoost	Bez interakcií	0.724	0.701	0.709	0.67	0.555	0.509
	Interakcie	0.727	0.715	0.711	0.677	0.56	0.51
	Logaritmy	0.754	0.717	0.719	0.75	0.537	0.5
AutoML	Bez interakcií	0.732	0.74	0.751	0.684	0.568	0.539
	Interakcie	0.742	0.755	0.746	0.7	0.559	0.52
	Logaritmy	0.808	0.765	0.753	0.741	0.557	0.52

Na základe tabuľkových výsledkov možno potvrdiť, že modely strojového učenia poskytujú pre tento typ výpočtu najlepšie výsledky pri výpočte neznámych prietokov. Výrazne prevyšujú presnosť hydrologického modelovania, ktoré je vyhodnotené v tabuľke 3.6. Až na pár výnimiek (AutoML, výpočet prietokov Modry) možno pri každom modeli pozorovať, že využitie interakcie premenných poskytlo lepšie presnosti vypočítaných prietokov ako vstupné údaje bez interakcií. Miera zlepšenia však bola menšia ako v predošlom prípade. Veľkosť zvýšenia presnosti sa väčšinou pohybovala len okolo jednej stotiny NSE, čo znamená, že uvedené modely vedú v tomto prípade vytvoriť dostatočne dobrý model aj bez interakcií, čo môže prispieť k ich lepšej interpretácii. V skutočnosti sa jedná o to, že svojim spôsobom je už interakcia premenných zahrnutá priamo v algoritmoch niektorých týchto modelov a to najmä tých, ktoré sú založené na regresných stromoch.

Logaritmovanie prietokov malo na výslednú presnosť väčší vplyv, čo vidno hlavne pri výpočte prietokov Orešian a Píly, modelmi CatBoost a AutoML. Napríklad pri výpočte prietokov Píly podľa Orešian a Modry bola základná hodnota NSE zlepšená z 0,67 na 0,75. Model Random Forest tu naopak vykazoval nižšie presnosti modelovaných prietokov.

Čo sa týka celkového porovnania modelov, za najúspešnejší možno označiť AutoML, ktorý vo všetkých výpočtoch, okrem výpočtu prietokov Píly, podľa Orešian a Modry, dosiahol najvyššie hodnoty NSE a to 4-krát s logaritmickou modifikáciou prietokov a 2-krát s údajmi bez interakcií (pri Modre), pričom bol výrazne úspešnejší ako hydrologický alebo lineárny model.

Otvorenou otázkou zostáva, či sa nedá dopredu zistiť, ktoré povodie je vhodné ako prvý analóg a ktoré ako druhý analóg. Autorka tento problém v tejto práci neriešila, keďže by bolo vhodnejšie riešiť ho v širšej štúdii, kde sa otestuje viac povodí ako v tejto pilotnej štúdii, ktorá má za cieľ iba zmapovať metodologické možnosti výpočtu neznámych prietokov.

3.2.2.1 Ensemblový model pri použití dvoch zdrojových povodí

Keďže v reálnej aplikácii tohto riešenia nie je možné s istotou povedať, ktorá kombinácia analógov je vhodnejšia, možno zabezpečiť väčšiu stabilitu výpočtu použitím ensemblového modelu. Ensemble bol realizovaný jednoduchým aritmetickým priemerom dvoch radov rôzne vypočítaných prietokov toho istého cieľového povodia. Výsledky takto vytvorených ensemblov jednotlivých modelov sú uvedené v tabuľke 3.9. Keďže pri hydrologickom modelovaní neexistuje ensemblový model, ktorý by principiálne úplne zodpovedal ensemblovému modelu ostatných použitých metód, ako alternatíva pre účely porovnania boli použité presnosti prietokov dosiahnutých viackriteriálnou optimalizáciou pri kalibrovaní hydrologického modelu.

Tabuľka 3.9: Vyhodnotenie ensemblových modelov prostredníctvom Nash-Sutcliffe koeficientu

Cieľové povodie	TUW	Best subset	Random forest	CatBoost	AutoML
Orešany	0.65	0.750	0.764	0.772	0.804
Píla	0.63	0.734	0.732	0.750	0.763
Modra-Piesok	0.43	0.503	0.521	0.533	0.539

Pri porovnaní najúspešnejšieho modelu z tabuľky 3.8, ktorým je AutoML, s jeho výsledkami v tabuľke 3.9, vidieť, že ensemblový model dosiahol pri výpočte Modry presnosť rovnajúcu sa priemernej presnosti prietokov dosiahnutých v predošlej tabuľke. Pri Orešanoch bola presnosť ensemblových modelovaných prietokov takmer rovnako dobrá ako pri vhodnejšom z povodí a pri Píle dosiahla dokonca lepšiu presnosť ako pri separátnych výpočtoch, na základe čoho možno tento ensemblový spôsob odporučiť ako najvhodnejší.

Podrobnejšie vyhodnotenie celkovo najlepších modelov tejto kapitoly je podľa rozšíreného súboru štatistických ukazovateľov v nasledujúcich tabuľkách 3.10 – 3.12.

Tabuľka 3.10: Rozšírené štatistické vyhodnotenie ensemblových modelov - Orešany

Model	ME	RMSE	PBIAS%	NSE	cp	R2	VE
TUW	-0.03	0.41	-3.9	0.65	0.55	0.65	0.64
Best Subset	0.04	0.34	5.1	0.75	0.68	0.75	0.69
Random Forest	-0.06	0.33	-8.2	0.76	0.7	0.77	0.71
AutoML	-0.05	0.3	-7.2	0.8	0.75	0.81	0.73
CatBoost	-0.06	0.33	-7.4	0.77	0.71	0.78	0.71

Tabuľka 3.11: Rozšírené štatistické vyhodnotenie ensemblových modelov – Píla

Model	ME	RMSE	PBIAS%	NSE	cp	R2	VE
TUW	-0.02	0.4	-2.8	0.63	0.58	0.63	0.65
Best Subset	-0.05	0.34	-6.5	0.73	0.7	0.74	0.72
Random Forest	0.02	0.34	3.3	0.73	0.71	0.74	0.72
AutoML	-0.03	0.32	-3.6	0.76	0.73	0.77	0.74
CatBoost	-0.02	0.33	-2.8	0.75	0.72	0.75	0.73

Tabuľka 3.12: Rozšírené štatistické vyhodnotenie ensemblových modelov – Modra

Model	ME	RMSE	PBIAS%	NSE	cp	R2	VE
TUW	-0.14	0.68	-16	0.43	0.41	0.47	0.58
Best Subset	-0.18	0.64	-20.9	0.5	0.48	0.57	0.63
Random Forest	-0.16	0.62	-18.6	0.52	0.5	0.57	0.63
AutoML	-0.14	0.61	-16.2	0.54	0.52	0.52	0.64
CatBoost	0.31	0.62	-14.9	0.54	0.51	0.57	0.64

Z uvedených výsledkov vyplýva, že hlavným predpokladom pre úspech výpočtu nemeráných prietokov metódami strojového učenia podľa dvoch analógov je tvorba odvodených premenných a zohľadnenie interakcií medzi nimi. Logaritmované prietoky presnosť navýšili pri výpočte prietokov Horných Orešian a Píly. Toto zlepšenie sa ale neopakuje pri výpočte prietokov najmenej podobného povodia v regióne, čo poukazuje na dôležitosť výberu vhodného analógu.

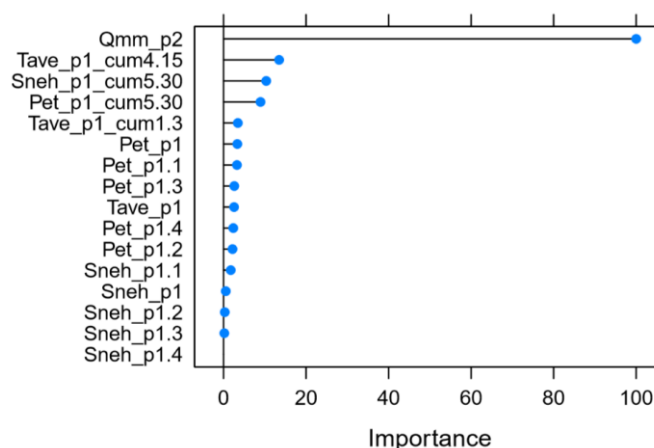
Presnosť modelovania možno zvýšiť využitím údajov o súčte zrážok a topenia snehu namiesto základných údajov o zrážkach. Tiež je nápomocná voľba vhodného modelu, od ktorého závisia ďalšie modifikácie vstupov a výpočtu. Pri hydrologickom modelovaní modelom TUW sa ako najlepší kalibračný spôsob ukázala multikriteriálna optimalizácia. Najlepšie výsledky lineárnej metódy Best subset boli dosiahnuté pri použití interakcie premenných a logaritmovania prietokov. Výnimkou bol prípad, že cieľovým povodím bola Modra, alebo Modra poskytla pomocný prietok ako nezávislú premennú. Algoritmus Random Forest dosahoval najlepšie výsledky s interakciou premenných. Pri metódach AutoML a Catboost zvýšili presnosť modelovania logaritmované, čiže normalizované prietoky. Použitím ensemblového modelu možno zaručiť väčšiu stabilitu alebo istotu výpočtu, aj keď sa môže stať, že konečná presnosť je o niečo nižšia ako pri vhodnejšom z analógov. Na základe uvedených

štatistických ukazovateľov možno za najúspešnejšie modely označiť AutoML a CatBoost; tejto práci je však dôležitejšie overenie použiteľnosti prietokov z nich získaných, v bilančnom výpočte závlahovej nádrže, ktorému sa venuje kapitola 3.4.

3.2.2.2 Posúdenie fyzikálnej opodstatnenosti modelov

Jedným zo základných nástrojov pre posúdenie fyzikálnej opodstatnenosti modelu je vyhodnotenie dôležitosti jednotlivých nezávislých premenných (variable importance).

Ako vidno na obrázku 3.11, ktorý znázorňuje najlepší model algoritmu Random Forest počítajúci prietoky Orešian podľa povodí Píla a Modra, na prvom mieste v dôležitosti premenných je analogický prietok, ktorého dôležitosť výrazne prevyšuje ostatné premenné. Toto je v celku logický výsledok, ktorý prispieva k dôveryhodnosti „čiernej skrinky“ akou model Random Forest je. Číselné vyjadrenie dôležitosti premenných je uvedené v tabuľke 3.13. Na druhom až piatom mieste sa nachádzajú kumulované premenné, ktoré majú za úlohu indikovať stav nasýtenosti povodia vodou, ktorý tak isto môžeme označiť za významný faktor odtoku. Jedná sa o kumulované hodnoty priemerných teplôt za 4. až 15. deň pred dňom, ktorého prietoky sú počítané. Ďalej sú to kumulované hodnoty kombinovanej premennej zrážky/topenie snehu a evapotranspirácia späťne za 5. až 30. deň, ktorá indikuje vysušanie povodia a teda tiež jeho vlhkostný stav v čase výpočtu prietoku. Nasleduje potenciálna evapotranspirácia z predchádzajúceho dňa a viaceré premenné o základných veličinách z predchádzajúcich dní („posunuté premenné“). Na prvý pohľad sa môže zdať prekvapujúce, že zrážky sú voči premennej zahŕňajúcej topenie snehu pri výpočte prietokov o niečo menej významné. Autorka predpokladá, že je to tým, že zrážky sú už nepriamo zahrnuté v najdôležitejšej premennej – analógom prietoku, keďže ten rastie a klesá v závislosti od zrážkovej činnosti (koreluje so zrážkami).



Obrázok 3.11: Dôležitosť premenných modelu Random Forest, počítajúceho prietoky Horných Orešian podľa povodí Píla a Modra

V tabuľke 3.13 je okrem dôležitosti premenných pri výpočte prietokov Horných Orešian podľa Píly a Modry, uvedená aj dôležitosť z výpočtu Orešian s vymenenými analógmi, teda podľa Modry a Píly. V druhom prípade sa na 1. mieste rovnako umiestnil analogický prietok s dôležitosťou 100, na 2. až 5. mieste sa nachádzajú totožné kumulované premenné ako v prvom stĺpci, avšak v mierne pozmenenom poradí. Možno teda povedať, že analogické prietoky majú na zlepšenie presnosti výpočtu najväčší vplyv a nasledované sú kumulovanými premennými a klimatickými premennými z predchádzajúcich dní.

Tabuľka 3.13: Porovnanie dôležitosti premenných pri výpočte prietokov Horných Orešian

Orešany z Píly a Modry		Orešany z Modry a Píly	
Dôležitosť	Premenná	Dôležitosť	Premenná
100.00	Qmm_p1	100.00	Qmm_p2
20.87	Tave_p2_cum4.15	13.47	Tave_p1_cum4.15
13.56	Pet_p2_cum5.30	10.35	Sneh_p1_cum5.30
8.09	Sneh_p2_cum5.30	8.97	Pet_p1_cum5.30
4.42	Tave_p2_cum1.3	3.47	Tave_p1_cum1.3
3.19	Tave_p2	3.34	Pet_p1
2.29	Pet_p2.1	3.26	Pet_p1.1
2.27	Pet_p2	2.60	Pet_p1.3
2.19	Pet_p2.4	2.56	Tave_p1
1.94	Pet_p2.2	2.36	Pet_p1.4
1.81	Pet_p2.3	2.17	Pet_p1.2
0.86	Sneh_p2.1	1.75	Sneh_p1.1
0.26	Sneh_p2.2	0.55	Sneh_p1
0.16	Sneh_p2	0.30	Sneh_p1.2
0.02	Sneh_p2.3	0.19	Sneh_p1.3
0.00	Sneh_p2.4	0.00	Sneh_p1.4

3.3 Vodná bilancia nádrže

Táto časť sa venuje overeniu použitia získaných prietokov pri vodohospodárskej aplikácii - bilancovaní závlahovej nádrže. Vodohospodárska bilancia je riešená pre nádrž Horné Orešany, ktorá bola pôvodne vybudovaná pre účely závlahy.

3.3.1 Softvérové riešenie vodohospodárskej funkcie závlahovej nádrže – R balík

HydBal

V balíku *HydBal* (*Hydrological Balance*) bol použitý model bilančného výpočtu nádrže vytvorený na základe existujúceho modelu tohto výpočtu v programe Microsoft Excel (ktorého autorom je vedúci dizertačnej práce). Verzia v R má mnohé vylepšenia a výhody, nakoľko takúto funkciu je možné spúšťať v kontexte ďalších možných výpočtov, ktoré R poskytuje a vytvoriť tak veľmi efektívne pracovné postupy so širším záberom.

3.3.1.1 Výpočet bilancie nádrže

Vodohospodársku bilanciu v balíku vykonáva funkcia s názvom *balance()*, ktorá simuluje priebeh akumulácie a spotreby vody, pri známych vstupoch a výstupoch vody z nádrže. Tieto údaje sú pre výpočet známe, pretože sa realizuje systémom retrospektívnej bilancie, teda v určitom historickom období, pre ktoré sú k dispozícii údaje o potrebných veličinách.

Tento model umožňuje posúdiť schopnosť nádrže plniť všetky požiadavky na odbory. Bilančný výpočet prebieha v mesačnom kroku, ktorý je na základe skúseností pre tento typ výpočtu dostačujúci. Takýmto spôsobom možno vypočítať zásobu vody v nádrži pre každý mesiac pozorovaného radu a definovať celkové percento zabezpečených závlahových nárokov.

Výpočtový algoritmus funkcie *balance()* je založený na rovnici vodnej bilancie, ktorá udáva objem vody V v mesačnom časovom kroku i :

$$V_i = V_{i-1} + (TV_i - RAW_i - IRN_i - E_i) \cdot \Delta_t \quad (3.1)$$

kde, V_i je objem vody v nádrži v mesiaci i [m^3]; TV_i je celkový objem vody, ktorá prítiekla do nádrže v mesiaci i [m^3]; RAW_i je potrebný objem vody v mesiaci i , čo zahŕňa biologický prietok a ostatné prietoky okrem vody, ktorá je potrebná na závlahu [m^3]; IRN_i je objem vody potrebnej na závlahu v mesiaci i [m^3]; E_i je evaporácia z hladiny nádrže a iné straty vody z nádrže v mesiaci i [m^3] a Δ_t je časový krok (1. mesiac).

Počas simulácie musí byť voda v nádrži udržiavaná na vyššej úrovni, ako je definovaná minimálna úroveň hladiny a zároveň na nižšej, ako je maximálna úroveň hladiny. Toto sa dá docieľiť

znížením odberov z nádrže, alebo uvoľnením prebytočnej vody z nádrže. Pri uvoľnení nadbytočnej vody z nádrže je potrebné posúdenie kapacity bezpečnostného priepadu, ktorý sa vykonáva inou funkciou – spomínanou *floodWave()*. Zníženie (závlahových) odberov z nádrže za situácie, že objem vody v nádrži sa znižuje a blíži k minimálnemu objemu je v zmysle posudzovania bilancie nádrže nepriaznivým stavom, ktorý sa vo výpočte eviduje a následne vyhodnocuje. Percento zníženia požadovaných odberov možno v modeli vyhodnotiť pomocou funkcie *summBalance()*, ktorá okrem iného poskytuje aj informácie o každom poklese hladiny vody pod minimálnu (maximálnu) úroveň.

3.3.1.2 Postup výpočtu bilancie nádrže v R

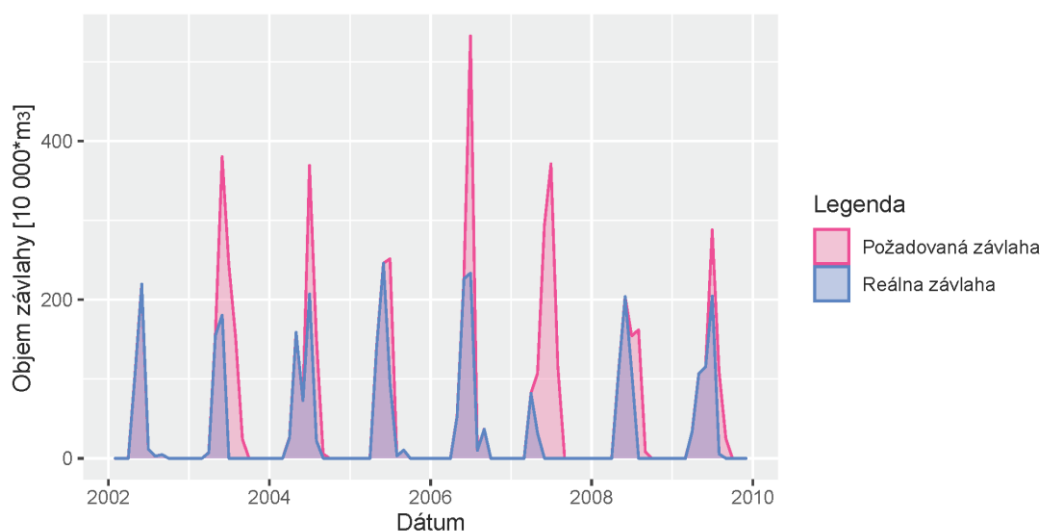
V úvode bilančného výpočtu v prostredí R je potrebné načítať alebo vytvoriť dátový rámec so základnými údajmi definujúcimi parametre nádrže. Pre uľahčenie tohto kroku bola navrhnutá pomocná funkcia s názvom *createBasicInputs()*. Užívateľ zadá parametre, ktoré funkcia požaduje v jednotkách, ktoré sú definované v pomocníkovi tejto funkcie. Ide o charakteristiky nádrže popisujúce jej základné vlastnosti a počiatočné podmienky: maximálnu a minimálnu nadmorskú výšku hladiny v nádrži (m n. m.), nadmorskú výšku počiatočného stavu vodnej hladiny a objemy vody zodpovedajúce týmto stavom (m^3), ďalej je potrebné zadať celkovú výmeru zavlažovanej plochy (ha), hodnotu ekologického prietoku a koeficientu spoľahlivosti prietokov.

V prípade, že sa užívateľ rozhodne vytvoriť si vstupy pre bilančnú funkciu sám, či už v externom programe alebo v R, a nie je si istý ich správnou štruktúrou, má ku každej výpočtovej úlohe k dispozícii po jednom vzorovom datasete, ktorý možno načítať jednoducho volaním funkcie *data(názovDatasetu)*. Okrem kontroly štruktúry dátovej tabuľky je možné ich použitie pre testovacie spúšťanie funkcií balíka.

Výsledkom funkcie *balance()* je tabuľka s informáciou pre každý mesiac vo forme časových radov, obsahujúca vstupné údaje a z nich dopyčítane údaje bilancie, ako napríklad splnené a nesplnené požiadavky na závlahu, reálnu závlahu, reálny odtok z nádrže a priame bilancovanie objemu vody v nádrži. Údaje z výstupnej tabuľky možno graficky zobrazit' s použitím funkcií *plotHistBal()* a *plotTimeSeries()*.

Funkcia *plotTimeSeries()* umožňuje vykresliť porovnanie priebehu akýchkoľvek kontinuálnych premenných v ľubovoľnom časovom rozsahu. Na obrázku 3.12 je znázornené porovnanie časových radov požadovanej a reálne dodanej závlahy vo vybranom časovom úseku 2002 - 2010.

Popísaná funkcia *balance()* je využiteľná pre výpočet kvantitatívnej vodohospodárskej bilancie zásobnej funkcie nádrže a jej štatistického a grafického vyhodnotenia. Môže byť užitočná pri navrhovaní základných parametrov závlahovej nádrže alebo jej vodohospodárskom prehodnotení.



Obrázok 3.12: Ukážka výstupu funkcie *plotTimeSeries()*

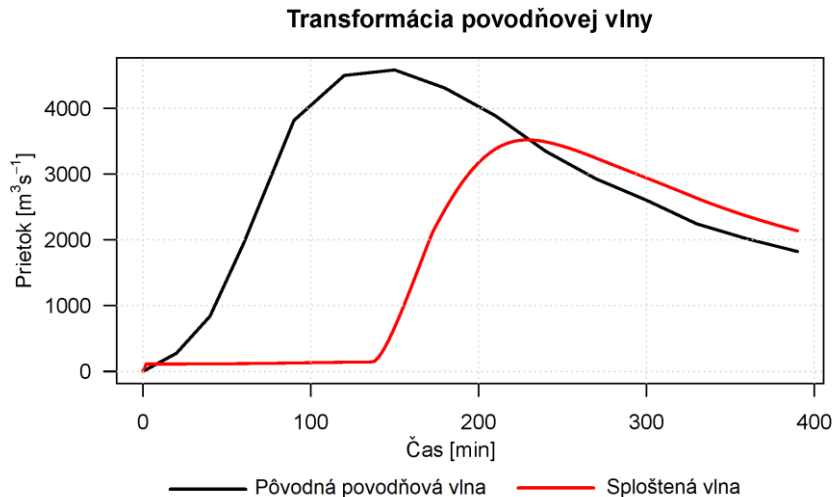
3.3.1.3 Prechod povodňovej vlny nádržou

Ako už bolo spomínané, balík tiež obsahuje funkciu s názvom *floodWave()*, ktorá slúži na simulovanie prechodu povodňovej vlny nádržou alebo poldrom; počíta ich transformačný účinok. Metóda výpočtu vychádza z posúdenia pomeru zachytného priestoru nádrže a objemu návrhovej povodňovej vlny so zahrnutím hydraulického riešenia odtoku vody výpustným potrubím, prípadne bezpečnostným prepacom. Spracovanie tohto riešenia formou funkcie R balíka uľahčuje vykonanie väčšieho počtu variantných riešení, čo umožňuje nájsť optimálny pomer medzi kulmináciou prítoku a odtoku vody z nádrže, respektíve zabezpečiť, aby kulminálny odtok z nádrže neprevýšil hodnotu neškodného prítoku z hľadiska chránenej lokality pod nádržou. To je možné dosiahnuť buď voľbou veľkosti zachytného priestoru alebo priemeru odpadového potrubia, respektíve ich vzájomnou kombináciou. Výsledkom riešenia je možnosť výberu najvhodnejšieho variantu technických parametrov nádrže (výška hrádze, priemer a výškové umiestnenie odpadového potrubia, rozmery a výškové umiestnenie bezpečnostného prepadu) a získanie informácií o maximálnom odtoku vody pod hrádzou pre návrhovú povodňovú vlnu príslušnej doby opakovania.

3.3.1.4 Postup výpočtu povodňovej vlny v R

Použitie funkcie pre výpočet transformačného účinku nádrže v R opäť pozostáva z úvodnej tvorby vstupov požadovaných hlavnou funkciou. Pri inicializácii dátového rámca so základnými údajmi popisujúcimi charakteristiky nádrže, možno zvoliť možnosť jeho tvorby pomocou pomocnej funkcie alebo načítaním hotového formátovaného vstupu pripraveného v jednom z textových editorov.

Ďalšími požadovanými vstupmi sú batygrafické charakteristiky nádrže, ktoré možno dopočítať vytvorenou funkciou *areaVolume()* a súbor popisujúci priebeh návrhovej vlny (dátový rámec s prítokom povodňovej vody do nádrže v zodpovedajúcom časovom kroku). Nasleduje samotné spustenie výpočtu. Výstupom funkcie *floodWave()* je dátový rámec s transformovanou povodňovou vlnou, ktorej priebeh je tiež možné znázorniť graficky (obrázok 3.13).



Obrázok 3.13: Grafické znázornenie transformovanej povodňovej vlny, vypočítanej funkciou *floodWave* z balíka *HydBal*

3.3.1.5 Ďalšie funkcie balíka *HydBal*

Ďalšími funkciami nachádzajúcimi sa v balíku *HydBal*, ktoré tiež možno používať nezávisle od riešenia nádrže, sú prevodové funkcie. Prvá je zameraná na prevod v rámci bežne zaužívaných jednotiek prietokov, druhá funkcia umožňuje premenu časových radov z údajov v dennom kroku na mesačný a ročný krok, pričom je možné zvoliť, či sa údaje budú kumulovať súčtom alebo priemerovaním. Všetky spomenuté funkcie sú spracované ako súčasť R balíka, ktorý je možné po jeho lokálnej inštalácii používať rovnako ako balíky stiahnuté z úložiska CRAN. Ďalším cieľom vo vývoji balíka je dokončenie jeho podrobnej dokumentácie v angličtine a požiadanie o jeho umiestnenie a distribúciu prostredníctvom siete CRAN.

3.3.2 Výpočet bilancie nádrže Horné Orešany s použitím vypočítaných a meraných prietokov

Vodohospodárska bilancia vodnej nádrže bola vykonaná pre obdobie 1980 – 2017 na základe nasledovných vstupných údajov:

- priemerných mesačných prietokov na toku Parná, v týchto variáciách:
 1. reálne merané prietoky,
 2. prietoky vypočítané hydrologickým modelom,
 3. prietoky vypočítané najlepšimi modelmi strojového učenia,
- údaje o manipulácii s vodou v nádrži (najmä úžitkový priestor v nádrži, minimálna a maximálna hladina a ďalšie údaje prevzaté z manipulačného poriadku),
- zmluvne dohodnuté odbery z nádrže a údaj o povinnom biologickom odtoku z nádrže (manipulačný poriadok),
- krivka plôch a objemov nádrže (manipulačný poriadok),
- požadované závlahové odbery z nádrže na 1 hektár, určené výpočtom pomocou simulačného programu CROPWAT prevzaté z (Szolgay, 2015).

Overenie využiteľnosti vypočítaných prietokov bolo hodnotené podľa nasledovných ukazovateľov:

1. percenta zabezpečenia závlhovej vody v bilančnom období 1980 – 2017, ktoré predstavuje pomer dodanej a požadovanej závlahy v percentách:

$$\text{Zabezpečenosť} = (\text{ZDodaná} / \text{ZPožadovaná}) * 100 \quad (3.2)$$

2. koreláciou medzi dodaným ročným množstvom závlhovej vody na základe nameraných a vypočítaných prietokov.

Samotná bilancia nádrže bola počítaná prostredníctvom funkcií balíka HydBal, ktorého tvorba je súčasťou práce.

V prvom kroku bola bilancia nádrže počítaná s použitím skutočných nameraných prietokov a udávala hodnoty dodaného závlhového množstva, s ktorými by malo byť dodané množstvo závlahy vypočítané na základe modelovaných prietokov, čo najzhodnejšie. Následne prebehli rovnaké výpočty s prietokmi z najúspešnejších regresných modelov, modelov strojového učenia a hydrologického modelu. Všetky výpočty prebiehali v mesačnom kroku, pre súhrnné vyhodnotenie boli použité ich ročné priemery. Vyhodnotenie týchto výpočtov je uvedené s tabuľke 3.14.

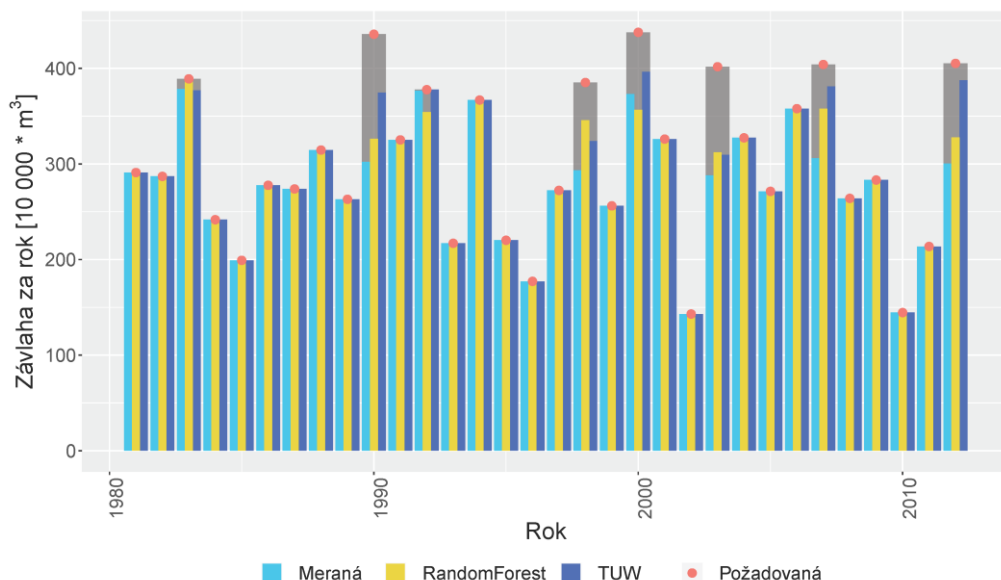
Tabuľka 3.14: Súhrnné vyhodnotenie plnenia požiadaviek na závlahu, vypočítaného podľa prietokov získaných rôznymi modelmi

Zdroj použitých prietokov	Reálna závlaha	Počet nesplnených požiadaviek	Zabezpečenosť závlahy v %	Korelácia priemerných ročných závlah
Merané	107171730	14	93.52	1
TUW	110893760	14	96.77	0.944
CatBoost	109425073	12	95.49	0.967
Best subset	110005752	10	96.00	0.959
Random forest	108995889	10	95.11	0.970
autoML	109619487	12	95.66	0.966

Z tabuľky vidieť, že reálna závlaha, rovnako ako aj zabezpečenosť závlahy v percentách je pri každom použití modelovaných prietokov nadhodnotená. Pri reálne poskytnutej závlahy sa referenčnej hodnote najviac priblížil model Random Forest. Pri počte nesplnených požiadaviek bol s hodnotou 14 výskytov s referenčnou hodnotou zhodný TUW model. Percento zabezpečenia najviac nadhodnotil TUW model (o 3,25%) a najmenej Random Forest (1,59%), ktorý tiež vykazuje najväčšiu mieru korelácie medzi dodaným ročným množstvom závlhovej vody na základe nameraných a vypočítaných prietokov.

Zabezpečenosť závlahy podľa jednotlivých rokov, porovnávajúca hydrologický a najúspešnejší machine learning model je zobrazená na obrázku 3.14, na ktorom možno pozorovať výrazné nadhodnocovanie zabezpečenia zo strany zrážkovo-odtokového modelu TUW.

Podľa percenta zabezpečenia závlahy a korelácie medzi dodaným ročným množstvom závlahovej vody na základe nameraných a vypočítaných prietokov možno vyhodnotiť, že všetky z použitých modelov poskytli relatívne dobrú presnosť. Referenčným hodnotám sa v porovnaní s ostatnými modelmi o niečo viac priblížil model Random Forest, pri porovnaní objemu neposkytnutej závlahy bol najpresnejším hydrologický model.



Obrázok 3.14: Grafické porovnanie zabezpečenia závlahy v rozmedzí rokov 1980 – 2017 vypočítanej podľa rôznych metód výpočtu prietokov

4 Záver

Na základe porovnania použitých metód sa popri hydrologickom zrážkovo-odtokovom modeli ako vhodné javili viaceré regresné metódy a algoritmy strojového učenia. V práci bolo zistené, že pri ich použití sa môže pri vhodne zvolených vstupných údajoch a manipulácii s nimi zásadne zvýšiť presnosť modelovania. Boli navrhnuté a testované viaceré pokročilé regresné postupy, jednak z hľadiska použitých metód strojového učenia a tiež z pohľadu spomenutých úprav, či preprocesingu údajov. Najvýznamnejší prínos z pohľadu presnosti modelovania priniesla tvorba nových premenných (vytvorených na základe znalosti hydrológie), vytvorenie premenných zohľadňujúcich vzájomnú interakciu základných premenných a logaritmovanie prietokov. Výrazné zlepšenie presnosti bolo dosiahnuté nahradením údaju o zrážkach, údajom zohľadňujúcim zrážky plus topenie snehu. Tieto postupy možno považovať za originálny prínos práce, keďže podľa vedomostí autorky neboli doteraz použité ani v svetovej literatúre na tému určenia prietokov v nameraných povodiach a pomohli významnou mierou zvýšiť presnosť modelovania. Medzi najdôležitejšie premenné, ktoré boli identifikované pomocou pokročilej črty moderného strojového učenia (tzv. „variable importance“), patrili špecifickým spôsobom použité prietoky z vedľajších povodí (analogov), konštruované kumulované premenné (ide o tzv. „feature engineering“) charakterizujúce hydrologickú históriu povodia a klimatické premenné z predchádzajúcich dní, ktoré majú vplyv na bezprostredný odtok po zrážke.

Na základe uskutočnených výpočtov možno zhrnúť, že v prípade, ak sú na toku k dispozícii uskutočnené aspoň krátke merania (postačujúce sú väčšinou už 2 ročné) a zároveň sú k dispozícii prietoky z blízkych (analogových) povodí, najlepšou metódou na výpočet nameraných prietokov sú regresné metódy, z ktorých sa najviac osvedčil ensembleový model strojového učenia AutoML a regularizovaná regresia typu Lasso.

V prípade, že na tokoch neboli vykonávané vôbec žiadne merania, na základe ktorých by bolo možné upresniť výpočet, je dôležité, či je k dispozícii jedno, alebo aspoň dve blízke povodia, vhodné na využitie v analógii. Vhodnosť týchto povodí je podmienená ich podobnosťou s nemeraným povodím a v práci bola určovaná na základe ich topografických, pedologických, hydrogeologických a klimatických charakteristík, spôsobu využitia územia a schopnosti kalibrovania povodia samotného blízkeho meraného povodia. Na základe študovaných kategórií podobnosti možno z kandidátnych povodí vybrať najvhodnejšie; v práci boli vybrané tri povodia z piatich. Pre lepšie overenie navrhutej metodiky, bolo za „nemerané“ povodie postupne použité každé z nich.

V prípade, že je k dispozícii jedno analógové povodie, je v práci porovnané využitie analógu (podobného toku) pre hydrologické a regresné modelovanie. Bolo zistené, že pri regresných postupoch je možné dosiahnuť presnosť podobnú s hydrologickým modelovaním. Presnosť regresného modelovania (myslia sa aj metódy strojového učenia) možno zvýšiť zahrnutím prietokov z ďalšieho analogického povodia, prípadne z viacerých analogických povodí. Túto výhodu prakticky hydrologické modelovanie nemá.

Zahrnutie dvoch analogických povodí spôsobilo zvýšenie presnosti vypočítaných prietokov. Najpresnejšie výsledky priniesli metódy AutoML (NSE = 0.804), CatBoost (NSE=0.77) a Random Forest (NSE=0.76). Aj keď to už v práci nebolo testované, dá sa očakávať, že väčší počet pomocných prietokov pochádzajúcich z viac než dvoch analogických povodí by mohol ešte viac spresniť modelované prietoky v nemeranom povodí. Hlavným dôvodom je fakt, že ide o povodia priestorovo blízke, ktorých prietoky navzájom korelujú, avšak v rôznej miere a v rôznych hydrologických situáciách inak a to v závislosti od ich topografických, pôdných a iných vlastností.

Je dôležité zdôrazniť, že táto práca je len pilotný projekt zameriavajúci sa na zmapovanie metodologických možností určovania prietokov v nemeraných povodiach pre potrebu bilancovania nádrže. Robia sa v nej výpočtové experimenty iba pre skupinu malých susedných povodí nachádzajúcich sa na západnom Slovensku. Pre potvrdenie použiteľnosti navrhnutých metód, odporúčaných algoritmov a tiež vytvorených softvérových riešení je nevyhnutné vykonať v ďalších prácach aplikáciu týchto postupov a nadväzujúce výpočty v širšom zábere na ďalších skupinách povodí. Autorka však verí, že poskytnuté zmapovanie možností, bude pri ďalšom prieskume v tejto oblasti nápomocné.

Použiteľnosť vypočítaných prietokov vo vodohospodárskej praxi bola verifikovaná v bilančnom výpočte závlahovej nádrže Horné Orešany, ktorý bol uskutočnený prostredníctvom softvérového balíka, vytvoreného autorkou v rámci riešenia dizertačnej práce za účelom prehodnocovania, prípadne návrhu nových závlahových nádrží. Tieto softvérové riešenia predstavujú praktický prínos práce, keďže sú vytvorené v programovacom jazyku R, ktorý je populárnym nástrojom v hydrologickej komunite a vytvorené programy sú voľne k dispozícii.

Bilančný výpočet závlahovej nádrže zisťuje stupeň plnenia požiadaviek na závlahu. Výsledkom bilancie je percento zabezpečenosti pri použití prietokov vypočítaných a meraných. Ďalším ukazovateľom použitým v tejto práci bola korelácia medzi dodaným ročným množstvom závlahovej vody vypočítanej na základe nameraných a vypočítaných prietokov. Uvedené ukazovatele potvrdili vhodnosť modelovaných prietokov, pričom s výsledkom bilancie podľa meraných údajov bol najvhodnejší výpočet metódou strojového učenia Random Forest a dobré výsledky priniesol aj podobný model CatBoost. V niektorých ukazovateľoch preukázal dobré výsledky koncepčný hydrologický model TUV a v prípade potreby ho tiež možno využiť pre danú úlohu – v práci je prezentovaná originálna metodika jeho kalibrácie na dvoch povodiach pomocou viackriteriálnej optimalizácie.

5 Literatúra

AMUTHA, R. - PORCHELVAN, P. 2009. Estimation of surface runoff in Malattar sub-watershed using SCS-CN method. In Journal of the Indian Society of Remote Sensing. 2009, 37 (2), s. 291.

BERGSTRÖM, S. 1995. The HBV model. Computer models of watershed hydrology.

- BLÖSCHL, G. - SIVAPALAN, M. 1995. Scale issues in hydrological modelling: a review. In Hydrological processes, 1995, roč. 9 (3-4), s. 251-290.
- BREIMAN, L. 2001. Random forests, In Machine Learning, roč. 45, č. 1, s. 5-32.
- DOROGUSH, A. – ERSHOV, V. – GULIN, A. 2018. CatBoost: gradient boosting with categorical features support. <https://arxiv.org/abs/1810.11363>.
- FUKA, D. R. et al. 2018. EcoHydRology: A Community Modeling Foundation for Eco-Hydrology. R package version 0.4.12.1. <https://CRAN.R-project.org/package=EcoHydRology>.
- LEDELL, E. et al. 2019. h2o: R Interface for 'H2O'. R package version 3.26.0.2. <https://CRAN.R-project.org/package=h2o>.
- LUMLEY, T. 2017. based on Fortran code by MILLER, A. leaps: Regression Subset Selection. R package version 3.0.
- ROTH, V. - FISCHER, B. 2008, July. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In Proceedings of the 25th international conference on Machine learning, s. 848-855.
- SOLOMATINE, D. P. 2006. Data-driven modeling and computational intelligence methods in hydrology. In Encyclopedia of hydrological sciences.
- SZOLGAY et al. 2015. Analýza a zhodnotenie hydrologickej bilancie na toku Parná a vodohospodárskej sústavy Horné Orešany, SvF STU.
- VIGLIONE, A. - PARAJKA, J. 2014. TUWmodel: Lumped hydrological model for educational purposes. Version 0.1-4. R package version 1.0-1. <https://CRAN.R-project.org/package=TUWmodel>.
- WALTER, M.T. et al. 2005. Process-based snowmelt modeling: Does it require more input data than temperature-index modeling? In Journal of Hydrology 300(1-4), s. 65-75.
- YASEEN, Z. M. et al. 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. In Journal of Hydrology. roč. 530, s. 829–844.
- ZAMBRANO-BIGIARINI, M. 2020. hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling. R package version 0.6-0. <https://github.com/hzambran/hydroTSM>. DOI:10.5281/zenodo.839864.

Zoznam publikácií

Vedecké práce v ostatných zahraničných časopisoch

HRUDKA, J. - ČERVEŇANSKÁ, M. [NAHÁLKOVÁ, M.] - RUSNÁK, D. - CZÖLDEROVÁ, M. - SOLDÁNOVÁ, V. Regulation facilities at the sewer network as a tool in reducing the effects of heat islands in urban areas. In Czech Journal of Civil Engineering. Roč. 3, č. 2 (2017), s. 66-72. ISSN 2336-7148.

Vedecké práce v zahraničných recenzovaných vedeckých zborníkoch, monografiách

ČISTÝ, M. - SOLDÁNOVÁ, V. Ensemble prediction of stream flows enhanced by harmony search optimization. In Time Series Analysis and Applications. 1. vyd. Rijeka : InTech, 2018, s. 153-169. ISBN 978-953-51-3742-9. V databáze: DOI: 10.5772/intechopen.71192.

Publikácia v currentovanom/recenzovanom/odbornom časopise – zahraničnom

ČISTÝ, M. - SOLDÁNOVÁ, V. 2020. Suspended Sediment Modelling with Hydrological and Climate Input Data. Journal of Hydroinformatics [v recenznom procese].

Publikované príspevky na zahraničných vedeckých konferenciách

ČISTÝ, M. - SOLDÁNOVÁ, V. Comparison of the hydrologic and data driven river flow generators. In SGEM 2017. 17th International Multidisciplinary Scientific GeoConference. Volume 17. Water Resources. Forest, Marine and Ocean Ecosystems : conference proceedings. Vienna, Austria, 27 - 29 November 2017. 1. vyd. Sofia : STEF 92 Technology, 2017, s. 111-118. ISSN 1314-2704. ISBN 978-619-7408-27-0. V databáze: SCOPUS: 2-s2.0-85063106649 ; DOI: 10.5593/sgem2017H/33/S12.014.

ČISTÝ, M. - SOLDÁNOVÁ, V. Flow prediction versus flow simulation using machine learning algorithms. In Machine Learning and Data Mining in Pattern Recognition: Part II, proceedings of the 14th International Conference MLDM 2018. New York, NY, USA, July 15-19, 2018. 1. vyd. Cham : Springer, 2018, s. 369-382. ISBN 978-3-319-96133-0. V databáze: SCOPUS: 2-s2.0-85050502234 ; DOI: 10.1007/978-3-319-96133-0_28.

ČISTÝ, M. - CYPRICH, F. - SOLDÁNOVÁ, V. Prediction of soil moisture data by various regression techniques. In SGEM 2018. 18th International Multidisciplinary Scientific GeoConference. Volume 18. Science and Technologies in Geology, Oil and Gas Exploration, Water Resources, Forest Ecosystems : conference proceedings. Vienna, Austria, 3 - 6 December 2018. 1. vyd. Sofia : STEF 92 Technology, 2018, S. 383-389. ISSN 1314-2704. ISBN 978-619-7408-72-0. V databáze: SCOPUS: 2-s2.0-85063075166.

ČISTÝ, M. - CYPRICH, F. - SOLDÁNOVÁ, V. A comparison of linear and machine learning models for the simulation of soil moisture. In Machine Learning and Data Mining in Pattern Recognition: Volume I, proceedings of the 15th International Conference MLDM 2019. New York, NY, USA, July 20-25, 2019. 1. vyd. Leipzig : ibai-publishing, 2019, s. 104-116. ISSN 1864-9734. ISBN 978-3-942952-62-0.

ČISTÝ, M. - SOLDÁNOVÁ, V. - POVAŽANOVÁ, B. Irrigation Reservoir Modeling in Catchments without Measurements of Stream Flow. In ADVCOMP 2019: the Thirteenth International Conference on Advanced Engineering Computing and Applications in Sciences. September 22-26, 2019, Porto, Portugal. 1. vyd. Wilmington, DE : International Academy, Research and Industry Association (IARIA), 2019, s. 12-18. ISBN 978-1-61208-737-5.

ČISTÝ, M. - VÝLETA, R. - SOLDÁNOVÁ, V. A Comparison of Methods for Calculating Monthly Flows on Small Catchments. In World Multidisciplinary Earth Sciences Symposium (WMESS 2019) : proceedings. 9-13 September 2019, Prague, Czech Republic. 1. vyd. Bristol : IOP Publishing, 2019, s., art. no. 012102. ISSN 1755-1307. V databáze: SCOPUS: 2-s2.0-85076612729 ; DOI: 10.1088/1755-1315/362/1/012102.

HRUDKA, J. - CZÖLDEROVÁ, M. - SOLDÁNOVÁ, V. - ŠKULTÉTYOVÁ, I. - STANKO, Š. The influence of surface runoff on the quality of the environmental environment. In SGEM 2018. 18th International Multidisciplinary Scientific GeoConference. Volume 18. Water Resources. Forest, Marine and Ocean Ecosystems : conference proceedings. Albena, Bulgaria, 2 July - 8 July 2018. 1. vyd. Sofia : STEF 92 Technology, 2018, S. 739-744. ISSN 1314-2704. ISBN 978-619-7408-42-3. V databáze: SCOPUS: 2-s2.0-85058903845 ; DOI: 10.5593/sgem2018/3.1/S12.096.

SOLDÁNOVÁ, V. Špecifika simulácie a predpovedania prietokov pomocou učiacich sa algoritmov. In Juniorstav 2018: zborník príspevků. 20. odborná konferencie doktorského studia. Brno, ČR, 25. 1. 2018 = Juniorstav 2018, proceedings of the 20th International Conference of Ph.D. Students. 1. vyd. Brno : ECON publishing, 2018, s. 833-837. ISBN 978-80-86433-69-1.

SOLDÁNOVÁ, V. - ČISTÝ, M. Extrapolation of CarpatClim data for engineering purposes. In SGEM 2018. 18th International Multidisciplinary Scientific GeoConference. Volume 18. Water Resources. Forest, Marine and Ocean Ecosystems : conference proceedings. Albena, Bulgaria, 2 July - 8 July 2018. 1. vyd. Sofia : STEF 92 Technology, 2018, S. 305-312. ISSN 1314-2704. ISBN 978-619-7408-42-3. V databáze: SCOPUS: 2-s2.0-85058871946 ; DOI: 10.5593/sgem2018/3.1/S12.040.

TAKÁČ, J. - SOLDÁNOVÁ, V. - ČISTÝ, M. Comparison of an analysis of drought occurrence using climatic indices and agro-ecological modelling in Slovakia. In SGEM 2017. 17th International Multidisciplinary Scientific GeoConference. Volume 17. Water Resources. Forest, Marine and Ocean Ecosystems : conference proceedings. Albena, Bulgaria, 29 June - 5 July 2017. 1. vyd. Sofia : STEF 92 Technology, 2017, S. 145-156. ISSN 1314-2704. ISBN 978-619-7408-04-1. V databáze: SCOPUS ; DOI: 10.5593/sgem2017/31/S12.019.

Publikované príspevky na domácich vedeckých konferenciách

HRUDKA, J. - ČERVEŇANSKÁ, M. [NAHÁLKOVÁ, M.] - RUSNÁK, Dušan - CZÖLDEROVÁ, Marianna - SOLDÁNOVÁ, Veronika. Regulácia odtoku dažďovej vody z intravilánu mesta ako nástroj pri eliminácii UHI efektu. In Konferencia mladých výskumníkov - KOMVY 2017 : zborník prednášok. Chvojníca, SR, 20. - 22. 11. 2017. 1. vyd. Bratislava : Spektrum STU, 2017, s. 40-48. ISBN 978-80-227-4749-3.

HRUDKA, J. - RUSNÁK, D. - ČERVEŇANSKÁ, M. [NAHÁLKOVÁ, M.] - SOLDÁNOVÁ, V. - CZÖLDEROVÁ, M. Analýza povrchového odtoku na území mesta Bratislava. In Zborník prednášok z Konferencie mladých výskumníkov - KOMVY 2018 : Podhájska, SR, 17. - 19. september 2018. 1. vyd. Bratislava : Spektrum STU, 2018, s. 55-62. ISBN 978-80-227-4847-6.

SOLDÁNOVÁ, V. Tvorba profilov dna tajchu Klinger s využitím GIS. In 28. konferencia mladých hydroológov, 15. konferencia mladých vodohospodárov, 17. konferencia mladých meteorológov a klimatológov : zborník príspevkov. Bratislava, SR, 9. 11. 2016. 1. vyd. Bratislava : Slovenský hydrometeorologický ústav, 2016, ISBN 978-80-88907-94-7.

SOLDÁNOVÁ, V. - ČISTÝ, M. Výpočet prítokov do závlahovej nádrže. In Konferencia mladých výskumníkov - KOMVY 2017 : zborník prednášok. Chvojnice, SR, 20. - 22. 11. 2017. 1. vyd. Bratislava : Spektrum STU, 2017, s. 111-117. ISBN 978-80-227-4749-3.

SOLDÁNOVÁ, V. Tvorba generátora prietokov s využitím kombinácie hydrologického a data driven modelovania. In Advances in Architectural, Civil and Environmental Engineering: 27th Annual PhD Student Conference on Applied Mathematics, Applied Mechanics, Geodesy and Cartography, Landscaping, Building Technology, Theory and Structures of Buildings, Theory and Structures of Civil Engineering Works, Theory and Environmental Technology of Buildings, Water Resources Engineering. 25. October 2017, Bratislava, Slovakia. 1. vyd. Bratislava : Spektrum STU, 2017, s. 192-197. ISBN 978-80-227-4751-6.

SOLDÁNOVÁ, Veronika. Využitie priestorových databáz v krajinnom inžinierstve. In Advances in Architectural, Civil and Environmental Engineering : 28th Annual PhD Student Conference on Applied Mathematics, Applied Mechanics, Building Technology, Geodesy and Cartography, Landscaping, Theory and Environmental Technology of Buildings, Theory and Structures of Buildings, Theory and Structures of Civil Engineering Works, Water Resources Engineering. October 24th 2018, Bratislava. 1. vyd. Bratislava : Spektrum STU, 2018, s. 215-221. ISBN 978-80-227-4864-3.

SOLDÁNOVÁ, V. Vodohospodárska bilancia nádrže pri nameraných prietokoch. In Advances in Architectural, Civil and Environmental Engineering : 29th Annual PhD Student Conference on Applied Mathematics, Applied Mechanics, Building Technology, Geodesy and Cartography, Landscaping, Theory and Environmental Technology of Buildings, Theory and Structures of Buildings, Theory and Structures of Civil Engineering Works, Water Resources Engineering. October 16th 2019, Bratislava. 1. vyd. Bratislava : Spektrum STU, 2019, s. 182-188. ISBN 978-80-227-4972-5.

Abstrakty príspevkov zo zahraničných konferencií

ČISTÝ, M. - BAJTEK, Z. - ČELÁR, Ľ. - SOLDÁNOVÁ, V. New methods versus the smart application of existing tools in the design of water distribution network. In Geophysical Research Abstracts. Volume 19/2017 : the open-access abstracts of the EGU General Assemblies. Göttingen : Copernicus Publications, 2017, s. ISSN 1607-7962.

ČISTÝ, M. - SOLDÁNOVÁ, V. - CYPRICH, F. Advantages of the Simultaneous Use of the Hydrologic and Stochastic Flow Generators. In Geophysical Research Abstracts. Volume 20/2018: the open-access abstracts of the EGU General Assemblies. Göttingen : Copernicus Publications, 2018, s. ISSN 1607-7962.

ČISTÝ, M. - SOLDÁNOVÁ, V. - CYPRICH, F. Random forest based reclassification of soil texture for hydrological modelling. In Geophysical Research Abstracts. Volume 21/201 : the open-access abstracts of the EGU General Assemblies. Göttingen : Copernicus Publications, 2019, ISSN 1607-7962.

Postery zo zahraničných konferencií

SOLDÁNOVÁ, V. - ČISTÝ, M. - CYPRICH, F. On the use of grid-based databases for r-r modelling. In HydroCarpath 2018. Catchment Processes in Regional Hydrology: Field Experiments and Modelling in Carpathians Basins : posters and abstracts of the conference. Vienna/Bratislava/Sopron, 12. 11. 2018. 1. vyd. Sopron : University of Sopron Press, 2018, s. 36. ISBN 978-963-334-199-5.

Citácie

CHEN, X. et al. 2020. The importance of short lag-time in the runoff forecasting model based on long short-term memory. In Journal of Hydrology, s.125359.

MEZOUARI, A.E. - NAJIB, M. 2019. A Hadoop Based Framework for Soil Parameters Prediction. In 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), s. 681-687, IEEE.