

Porovnanie databáz zrážkových údajov pre zvolené územie

Ing. Barbora POVAŽANOVÁ

Katedra vodného hospodárstva krajiny
Stavebná fakulta

Slovenská technická univerzita v Bratislave
Radlinského 11, 810 05, Bratislava

barbora.povazanova@stuba.sk

Abstract

Nowadays, the topics of drought associated with rising temperatures and the resulting need for water retention and optimal water management are becoming more and more topical. Many projects are being implemented concerning measures to prevent the negative impact of rising temperatures and the more frequent occurrence of extreme climatic events (flood control measures, drought measures). These projects include detailed analyzes of the area, for which it is necessary to have a lot of data available. One of the most important types of data is the hydrological data, especially precipitation totals. Precipitation data can be obtained from direct measurements at precipitation measuring stations, the advantage of which is accuracy at the measuring point. The disadvantages are the irregular network of precipitation stations, measurement failures, insufficiently long measurements and the fact that the data are charged. The solution in these cases is the usage of climate grid databases, whose data are freely available, free of charge. Grid database networks are also regular and dense. This work aimed to determine which of the four examined databases provides data closest to the measured data. We examined the usability of the data for the area of interest of the Nitrianska Streda river basin as well as for the whole territory of Slovakia. Subsequently, we extrapolated data from a selected climate database using a regularized linear regression model and examined the effect of extrapolation on data accuracy.

Key words: precipitation, database, Rstudio, LASSO, ECA&D, CarpatClim, CHIRPS, MSWEP

Annotation

The aim of this paper is to present the possibilities of using climate grid databases as a source of data on precipitation totals for the territory of Slovakia. In my work I compared the measured data from precipitation stations with the data available in the climate grid databases ECA&D, CarpatClim, MSWEP and CHIRPS.

Anotácia

Cieľom tohto príspevku je priblížiť možnosti použitia klimatických gridových databáz ako zdroja údajov o zrážkových úhrnoch pre územie Slovenska. V práci som sa venovala porovnávaniu meraných údajov zo zrážkomerných staníc s údajmi dostupnými v klimatických gridových databázach ECA&D, CarpatClim, MSWEP a CHIRPS.

1 Úvod

V súčasnosti sú stále aktuálnejšie témy sucha spojeného so zvyšujúcimi sa teplotami. Je realizovaných mnoho projektov týkajúcich sa opatrení na predchádzanie negatívnych dopadov zvyšovania teplôt a častejšieho výskytu extrémnych klimatických javov. Súčasťou týchto projektov sú podrobné analýzy územia. Pre všetky takéto analýzy je potrebné mať k dispozícii mnoho podkladov, pričom najdôležitejšími sú hydrologické podklady, najmä zrážkové úhrny. Zrážky je možné získavať z priamych meraní, ktoré na našom území zabezpečuje Slovenský hydrometeorologický ústav. Zrážkové úhrny sú merané zrážkomermi v zrážkomerných staniciach. Výhodou týchto údajov je to, že sú pre bod merania presné, nakoľko pochádzajú z priamych meraní. Používanie týchto údajov so sebou zároveň prináša niekoľko nevýhod. Pri analýzach menších území, ktoré vyžadujú zrážkomernú stanicu vo vnútri územia sa môže stať, že sa na vhodnom mieste nijaká nenachádza. Analýzy veľkých území na druhej strane vyžadujú údaje z veľkého množstva staníc, čo je finančne náročné, nakoľko poskytovanie meraných údajov je spoplatnené. Ďalšou nevýhodou pri práci s meranými údajmi je to, že sú závislé od histórie fungovania zrážkomernej stanice, a v meraniach sa často nachádzajú výpadky, čo značne ovplyvní použiteľnosť údajov. Riešením uvedených problémov môžu byť klimatické gridové databázy, ktoré poskytujú údaje o zrážkach v pravidelnej sieti bodov pochádzajúce z interpolácie meraných údajov. Údaje nie sú až tak závislé na funkčnosti staníc, interpolácia umožňuje zväčša získať ich bez výpadkov. Okrem toho sú v mnohých databázach údaje zadarmo dostupné pre hustú sieť bodov, ktoré rovnomerne pokrývajú celé územia. Klimatické gridové databázy sa líšia spôsobom získavania údajov a metódami, ktorými sa merania dopĺňajú (napr. diaľkový prieskum Zeme, zohľadňovanie nadmorskej výšky). V tejto práci sme porovnávali údaje z databáz CarpatClim, ECA&D, MSWEP a CHIRPS s meranými údajmi zo SHMÚ. Na základe toho sme vyhodnotili databázu, ktorá sa svojimi údajmi najviac približuje údajom z SHMÚ zrážkomerných staníc a tú identifikovali ako najvhodnejšiu pre použitie na našom území. Následne sme údaje extrapolovali s použitím regularizovanej lineárnej regresie a skúmali vplyv extrapolácie na presnosť údajov.

2 Zdroje údajov o zrážkach

2.1 Priame merania

Zrážkové úhrny sa na našom území merajú klasickými zrážkomernými nádobami, váhovými zrážkomermi a totalizátormi. Intenzita zrážok sa zaznamenáva ombrografom [1].

2.2 Klimatické gridové databázy

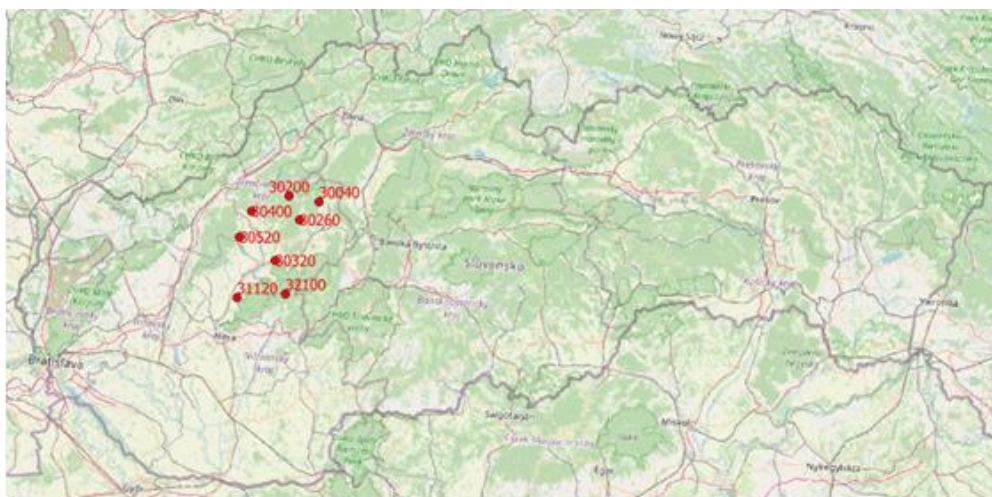
Ak z nejakého dôvodu nemáme prístup k vyhovujúcim meraným zrážkovým dátam, je možné využiť klimatické gridové databázy, ktoré spracovávajú údaje zo zrážkomerov, meteorologických radarov, predpovedacích modelov alebo diaľkového prieskumu Zeme [2]. Následne sa z týchto údajov interpoláciou vytvárajú dáta pre siete bodov. V každom bode takejto siete je časový rad údajov; v

našom prípade sa zaujímate o historické časové rady údajov. V tejto práci sme používali údaje z databáz CarpatClim, ECA&D, MSWEP a CHIRPS.

3 Príprava údajov

3.1 Výber územia, zrážkomerných staníc a bodov klimatických databáz

Prvým krokom bol výber záujmového územia. Pre účely tejto práce bola zvolená oblasť Hornej Nitry, nakoľko boli pre toto územie dostupné údaje vo vyhovujúcej dĺžke (min. 30 rokov), využitie územia je prevažne homogénne (poľnohospodárska pôda), s čím sa spájajú aj podobné nadmorské výšky staníc a bodov. Po výbere záujmového územia boli identifikované zrážkomerné stanice zo siete SHMÚ, ktoré sa v záujmovom území nachádzajú (Obr. 1). Ku každej dostupnej zrážkomernej stanici SHMÚ sme následne vyberali blízke body z každej vyhodnocovanej klimatickej databázy, ktoré boli neskôr použité pri porovnávaní údajov databáz s údajmi z priamych meraní. Body klimatických databáz prislúchajúce jednotlivým zrážkomerným staniciam sú vypísané v Tab. 1, kde SHMÚ = „s“, CarpatClim = „CC“, ECA&D = „EC“, MSWEP = „MS“ a CHIRPS = „CH“.



Obr. 1 Poloha vybraných zrážkomerných staníc SHMÚ na záujmovom území

Tab. 1 Body klimatických databáz priradené k zrážkomerným staniciam SHMÚ

SHMU	s30040	s30200	s30260	s30320	s30400	s30520	s31120	s32100
CC1	c1127	c1126	c1228	c1428	c1225	c1326	c1527	c1531
CC2	c1128	c1125	NA	c1328	NA	c1325	NA	c1530
EC1	e24	e23	e24	e48	e21	e32	e73	e63
EC2	e25	e24	e25	e49	e22	e33	e74	e64
EC3	e26	e13	e36	e63	e33	e47	e75	e78
MS1	m24	m23	m24	m48	m21	m32	m73	m63
MS2	m25	m24	m25	m49	m22	m33	m74	m64
MS3	m26	m13	m36	m63	m33	m47	m75	m78
CH1	ch90	ch67	ch111	ch213	ch106	ch153	ch290	ch268
CH2	ch91	ch68	ch112	ch214	ch105	ch154	ch291	ch269
CH3	ch113	ch89	ch135	ch185	ch107	ch155	ch292	ch270
CH4	ch114	ch90	ch136	ch186	ch130	ch181	ch318	ch297

3.2 Získavanie údajov

3.2.1 Klimatické gridové databázy

Všetky údaje o polohe bodov z gridových databáz a ich nadmorskej výške (okrem CarpatClim, ktorá má svoj originálny formát) sú voľne dostupné na stránkach jednotlivých databáz vo formáte NetCDF, ktorý má príponu „.nc“, rovnako ako údaje o zrážkach.

3.2.2 Zrážkomerné stanice SHMÚ

Údaje o polohe SHMÚ staníc a ich nadmorských výškach sú dostupné na stránke <http://www.SHMÚ.sk/File/metaklin/zrasta.pdf>. Použité údaje o denných zrážkových úhrnoch pochádzajú zo Slovenského hydrometeorologického ústavu. V údajoch každej zrážkomernej stanice sa nachádzali výpadky meraní, ktoré ovplyvňujú použiteľnosť údajov. Bolo preto potrebné tieto údaje analyzovať a vyhodnotiť ich vhodnosť na ďalšie výpočty, čo popisuje nasledujúca kapitola.

3.3 Analýza údajov SHMÚ

Všetky analýzy bolo potrebné robiť pre obdobie, v ktorom sú dostupné údaje zo všetkých zdrojov, pretože to boli údaje, s ktorými sme ďalej pracovali pri regresných výpočtoch. V tomto prípade to znamená, že bolo potrebné vybrať také obdobie, v ktorom boli dostupné aj merané údaje zo SHMÚ, aj údaje zo všetkých klimatických databáz. Tejto podmienke vyhovovalo obdobie 1.1.1981-31.12.2010.

3.3.1 Popisná štatistika

Popisná štatistika pre merané údaje zo zrážkomerných staníc bola počítaná v rozšírení Excelu XLSTAT, pomocou funkcie Descriptive statistics. Výsledky popisnej štatistiky sú uvedené v Tab. 2.

Tab. 2 Výsledky popisnej štatistiky pre merané údaje

Statistic	sh_30040	sh_30200	sh_30260	sh_30320	sh_30400	sh_30520	sh_31120	sh_32100
Nbr. of observations	10957	10957	10957	10957	10957	10957	10957	10957
Nbr. of missing values	31	121	120	31	335	92	31	61
Sum of weights	10926	10836	10837	10926	10622	10865	10926	10896
Minimum	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Maximum	63,000	70,500	72,200	74,200	57,600	90,500	62,500	91,500
1st Quartile	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Median	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3rd Quartile	2,200	2,100	1,600	1,000	1,800	1,500	1,300	1,600
Sum	27453	26772	23121	18034	21828	21194	19689	22693
Mean	2,513	2,471	2,134	1,651	2,055	1,951	1,802	2,083
Variance (n-1)	32,944	32,193	27,470	18,952	23,617	23,866	21,664	25,890
Standard deviation (n-1)	5,740	5,674	5,241	4,353	4,860	4,885	4,655	5,088
Skewness (Pearson)	3,954	3,985	4,456	5,025	4,133	4,879	4,801	4,561
Kurtosis (Pearson)	21,013	21,335	28,163	38,053	23,197	37,179	33,115	31,765

3.3.2 Trendy

Na zistenie toho, či v radoch údajov existujú trendy, t.j. či údaje postupom času narastajú alebo klesajú, sme použili Man-Kendallov test (Tab. 3). Ak je hodnota p-value väčšia, ako hladina významnosti $\alpha=0,05$, platí základná hypotéza, t.j. že v rade údajov neexistuje trend. Ak je hodnota menšia, platí alternatívna hypotéza, v rade sa trend nachádza.

Tab. 3 Výsledky Man-Kendallovho testu pre merané údaje

ID	sh_30040	sh_30200	sh_30260	sh_30320	sh_30400	sh_30520	sh_31120	sh_32100
alpha	0,050	0,050	0,050	0,050	0,050	0,050	0,050	0,050
p-value	0,440	0,004	0,234	<0,0001	<0,0001	0,961	0,194	0,000

V staniciach 30200, 30320, 30400 a 32100 sa nachádza trend vývoja zrážok. V ostatných staniciach sa trend vývoja nenachádza. Okrem týchto výsledkov je potrebné zohľadniť aj to, akú veľkosť má trend v radoch údajov. Toto je možné určiť na základe hodnoty Sen's slope (Tab. 4), ktorej výpočet je súčasťou Man-Kendallovho testu. Ak je hodnota veľmi blízka nule, môžeme trend v danom rade údajov zanedbať, ako tomu bolo aj v tomto prípade.

Tab. 4 Výsledky Sen's slope pre merané údaje

ID	Sen's slope
sh_30040	-0,000009174
sh_30200	-0,000009171
sh_30260	-0,000009152
sh_30320	-0,000009158
sh_30400	-0,000009130
sh_30520	-0,000009132
sh_31120	-0,000009161
sh_32100	-0,000009168

3.3.3 Homogenita

Cieľom testu homogenity je zistiť, či sa v rade údajov o denných zrážkach nachádza bod, v ktorom sa výrazne mení niektorá charakteristika údajov. V našom prípade bolo najvhodnejšie použiť SNHT test (Standard Normal Homogeneity Test).

Tab. 5 Výsledky SNHT testu homogenity

ID	sh_30040	sh_30200	sh_30260	sh_30320	sh_30400	sh_30520	sh_31120	sh_32100
SNHT test	0,053	0,079	0,132	0,032	0,054	0,105	0,036	0,018

V staniciach 30040, 30200, 30260, 30400, 30520 je hodnota p väčšia ako hladina významnosti, preto nie je možné vylúčiť základnú (nulovú) hypotézu (údaje sú homogénne). V staniciach 30320, 31120 a 32100 je hodnota p menšia ako 0,05, mali by sme teda základnú hypotézu vylúčiť (údaje považujeme za nehomogénne). V tu neprezentovaných výpočtoch sme preverovali rozsah zmeny priemernej hodnoty zrážok („change point“), a dospeli k zisteniu, že sa vždy jedná o veľmi malé, zanedbateľné zmeny.

Údaje sú na základe vykonaných analýz vhodné na ďalšie výpočty a je teda možné ich použiť pri určovaní najvhodnejšej gridovej klimatickej databázy.

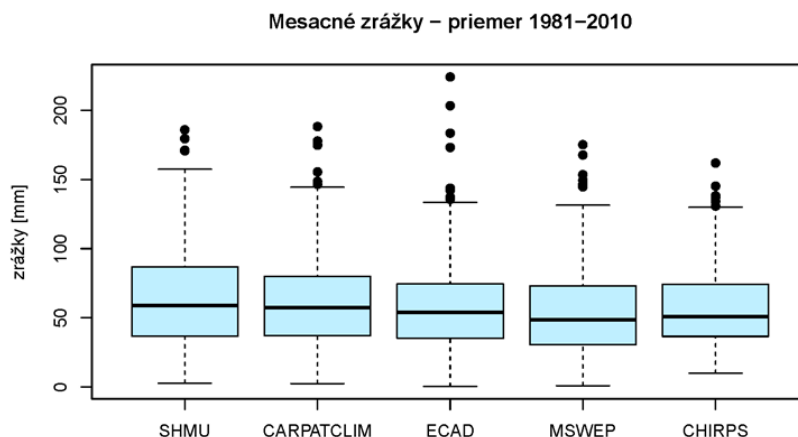
4 Výber najvhodnejšej klimatickej databázy

4.1 Porovnanie údajov SHMÚ a údajov klimatických databáz

Cieľom tejto kapitoly bolo na základe porovnávania údajov získaných priamymi meraniami v zrážkomerných staniciach s údajmi z klimatických gridových databáz vyhodnotiť, ktorá databáza poskytuje údaje najviac sa približujúce meraným údajom. Porovnávanie bolo robené pre všetky databázy v jednotnom období od 1.1.1981 do 31.12.2010. Každá zrážkomerná stanica bola porovnávaná s bodmi klimatických databáz, ktorých výber je popísaný v kap.3.1 a zobrazený na Tab. 1.

4.1.1 Rozloženie údajov pomocou krabicových grafov

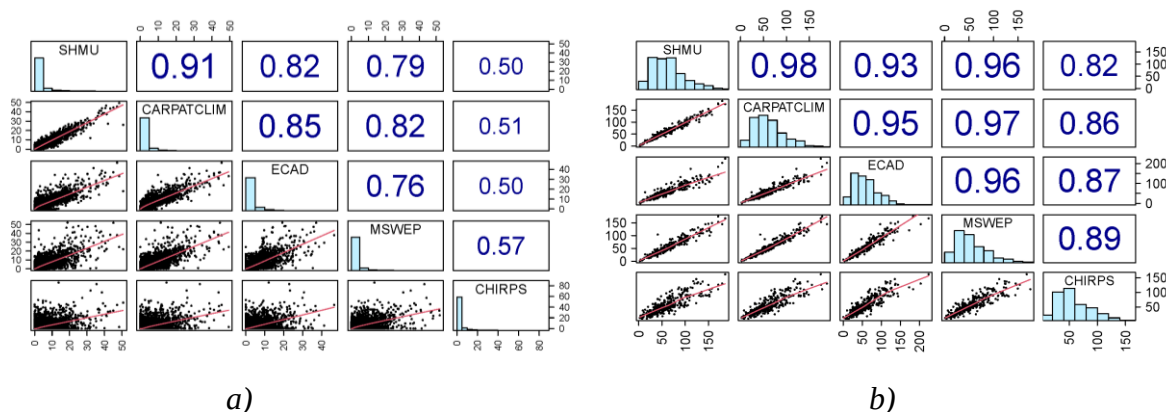
Vhodným nástrojom na porovnávanie dlhších radov údajov sú boxploty (krabicové grafy). Pre každú databázu sme vytvorili jeden priemerných denných údajov, z ktorého bol vytvorený boxplot. Najjednoduchšie je odčítať výsledok z boxplotu pre mesačné úhrny (viď. Obr. 2), pretože pri mesačných úhrnoch je eliminovaná väčšina nulových a odľahlých hodnôt. Na základe Obr. 2 je možné povedať, že databáza SHMÚ sa najviac približuje databáza CarpatClim.



Obr. 2 Boxploty pre mesačné údaje

4.1.2 Korelácie denných údajov

Ďalšia korelačná analýza porovnáva merané údaje zo zrážkomerných staníc s údajmi z klimatických databáz takým spôsobom, že pre každý deň boli opäť najskôr vytvorené priemerné rady pre každú klimatickú stanicu a následne porovnávané s meranými dennými úhrnmi zo staníc SHMÚ na základe korelačného koreficientu. Rovnaký výpočet bol realizovaný aj pre mesačné úhrny. Výsledky sú zobrazené na Obr. 3.

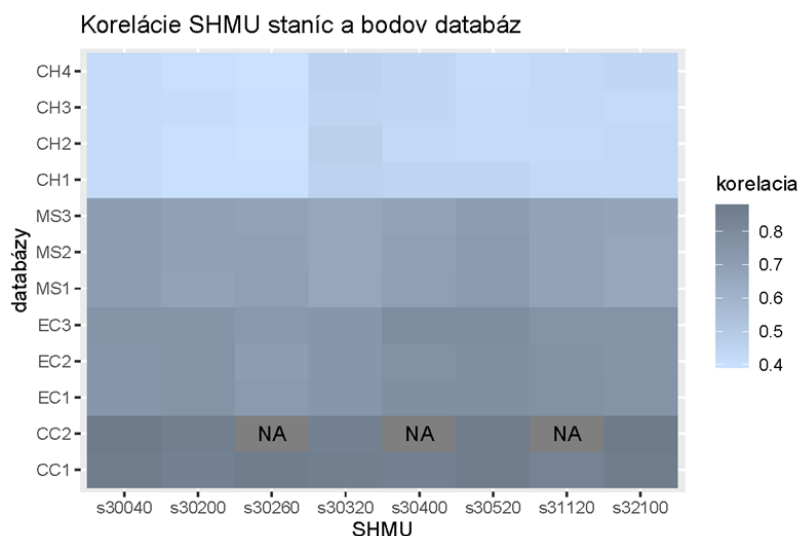


Obr. 3 Korelačné matice pre a) denné a b) mesačné úhrny

V prvom riadku sú vypočítané korelačné koeficienty pre dvojicu SHMÚ – klimatická gridová databáza. Z týchto hodnôt sa hodnote +1 najviac približuje hodnota 0,98 prislúchajúca dvojici SHMÚ – CarpatClim pre mesačné údaje. Je preto možné povedať, že údaje z databázy CarpatClim sa najviac približujú meraným údajom zo zrážkomerných staníc SHMÚ.

4.1.3 Určenie dní s výskytom zrážok

Hodnotiť spoľahlivosť klimatickej databázy je možné aj na základe toho, s akou presnosťou je schopná určiť dni, v ktorých sa vyskytovali zrážky (za správne boli opäť považované údaje z SHMÚ). Výsledky sú spracované a graficky zobrazené v podobe heatmap (teplotnej mapy) zobrazenej na Obr. 4. Tmavšia farba znamená vyššiu hodnotu korelačného koeficientu. Na základe vykonanej korelácie je možné konštatovať, že databáza CarpatClim najpresnejšie určuje výskyt zrážok.



Obr. 4 Heatmap pre koreláciu dní s vyskytujúcimi sa zrážkami

Na základe vykonaných analýz je možné povedať, že databáza CarpatClim poskytuje pre naše záujmové územie najpresnejšie údaje. Nevýhodou tejto databázy je to, že merania sú vedené len po rok 2010. V ďalšej kapitole bude vykonaná regularizovaná lineárna regresia, na základe ktorej bude následne možné zrážky extrapolovať a zlepšiť tak ich použiteľnosť.

5 Regularizovaná lineárna regresia

Cieľom tejto kapitoly je z kombinácie údajov CarpatClim a inej, na regresiu najvhodnejšej, klimatickej databázy vytvoriť model regularizovanej lineárnej regresie LASSO a vypočítať rovnicu regresnej priamky, na základe ktorej bude potom možné údaje extrapolovať a vytvoriť tak dlhší rad zachytávajúci aj zrážkové úhrny v súčasnosti. Výhodou LASSO je regularizácia, ktorá spočíva v regulovaní počtu nezávislých premenných vstupujúcich do výpočtu. V našom prípade sú nezávislými premennými body klimatickej databázy, na základe ktorej extrapolácia prebieha.

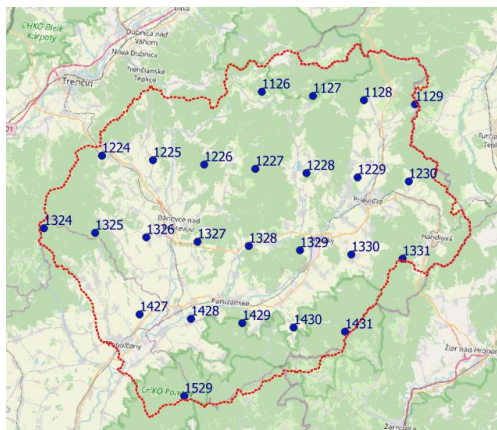
5.1 Výber klimatickej databázy a bodov pre regresný výpočet

Prvým krokom je výber klimatickej databázy, na základe ktorej je najvhodnejšie vykonať extrapoláciu. Pri výbere môžeme vychádzať z už vykonaných analýz. Na základe korelačnej matice (

a)

b)

Obr. 3) je zrejmé, že údaje databázy CarpatClim najlepšie korelujú s údajmi databázy ECA&D. Táto databáza bude preto použitá pri výpočte regularizovanej lineárnej regresie. Body databázy CarpatClim sme volili na základe ich polohy na záujmovom území (Obr. 5). Rozhodli sme sa do výpočtov pre každý bod databázy CarpatClim vybrať 8 najbližších bodov databázy ECA&D.



Obr. 5 Vybrané body CarpatClim na záujmovom území

5.2 Príprava údajov pre výpočet LASSO

Dáta o zrážkových úhrnoch klimatických databáz boli rozdelené na testovacie a trénovacie (kalibračné). Trénovacie údaje slúžia na kalibrovanie modelu regularizovanej lineárnej regresie (na určenie regresnej rovnice). Na tento účel boli vyhradené údaje od 1.1.1961 – 31.12.2000. Po kalibrácii je potrebné vzniknuté dáta otestovať, čo v tomto prípade znamená porovnanie s originálnymi dátami databázy CarpatClim. Za testovacie obdobie bolo určené obdobie od 1.1.2001 – 31.12.2010.

5.3 Výpočet LASSO pre oblasť Dolnej Nitry

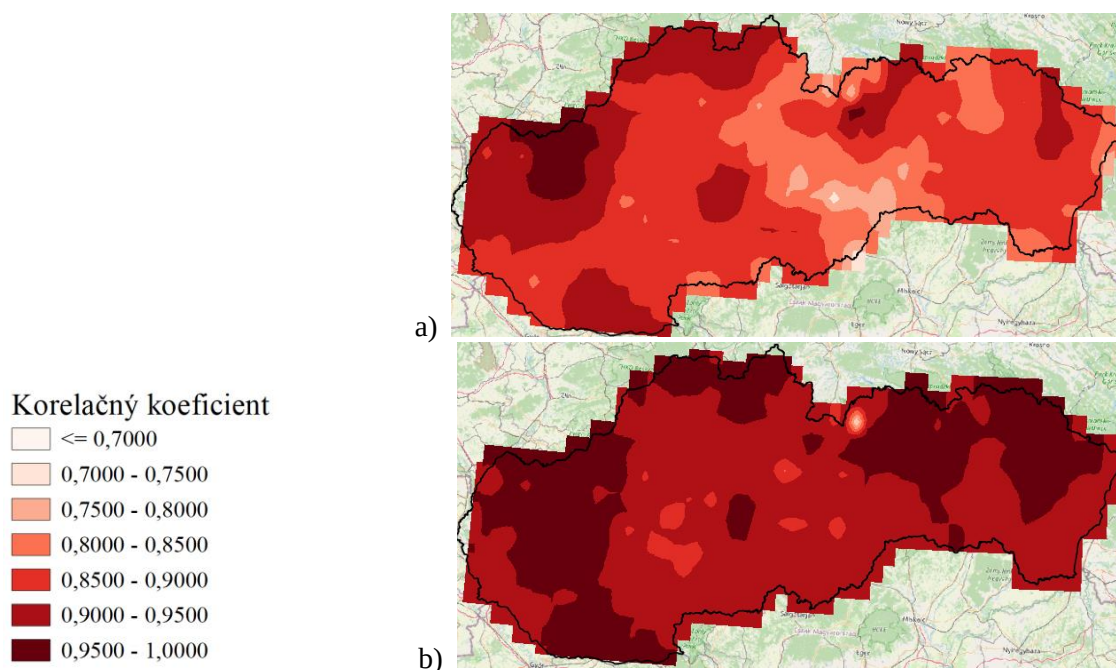
Výpočet LASSO sme realizovali v prostredí RStudio. Tento počas svojho behu zistí mená bodov databázy ECA&D, ktoré budú využité vo výpočte LASSO. Pre výpočet (zostavenie rovnice) sa použili trénovacie (kalibračné) údaje. Výpočtom sme dostali hodnoty denných úhrnov od roku 2001 po 2010. Vypočítané údaje sme previedli na mesačné úhrny, keďže sme chceli porovnať presnosť extrapolácie mesačných aj denných hodnôt. Po výpočte regularizovanej lineárnej regresie je potrebné údaje overiť porovnaním s testovacími údajmi pochádzajúcimi z databázy CarpatClim. Toto overenie sme robili aj pre denné aj pre mesačné úhrny a je možné na základe neho posúdiť, či je databáza ECA&D vhodná pre regresné výpočty databázy CarpatClim. Výsledky sú uvedené v Tab. 6. Na základe výsledkov je možné konštatovať, že databáza ECA&D je pomerne vhodná pre regresné výpočty extrapolácie databázy CarpatClim.

Tab. 6 Výsledné hodnoty porovnania vypočítaných hodnôt s testovacími

CC	r (-)	pbias (%)	NSE (-)	mes_r (-)	mes_pbias (%)	mes_nash (-)	param	nVar
1126	0.874	1,50	0.763	0.915	1,50	0.811	0	30
1127	0.863	2,40	0.744	0.903	2,40	0.781	0.002	18
1128	0.872	2,40	0.757	0.918	2,40	0.81	0.001	21
1129	0.871	1,10	0.757	0.922	1,10	0.829	0.002	16
1224	0.942	1,00	0.887	0.973	1,00	0.943	0.001	16
1225	0.92	2,30	0.845	0.965	2,30	0.927	0.001	21
1226	0.894	1,40	0.799	0.938	1,40	0.872	0	5
1227	0.865	-0,70	0.749	0.889	-0,70	0.775	0	32
1228	0.873	1,70	0.762	0.918	1,70	0.826	0.002	19
1229	0.881	0,70	0.777	0.928	0,70	0.858	0.003	15
1230	0.883	-1,00	0.78	0.927	-1,00	0.858	0.002	15
1324	0.926	-1,70	0.858	0.934	-1,70	0.863	0	21
1325	0.934	-0,30	0.873	0.948	-0,30	0.893	0.001	24
1326	0.899	2,60	0.808	0.95	2,60	0.896	0.001	6
1327	0.872	3,50	0.76	0.941	3,50	0.875	0.006	5
1328	0.879	2,30	0.772	0.938	2,30	0.87	0.005	4
1329	0.874	1,10	0.764	0.932	1,10	0.866	0.001	11
1330	0.884	-0,90	0.781	0.926	-0,90	0.857	0	32
1331	0.872	-1,90	0.761	0.909	-1,90	0.822	0.005	11
1427	0.904	0,70	0.817	0.955	0,70	0.912	0.021	4
1428	0.889	1,60	0.79	0.949	1,60	0.899	0.018	5
1429	0.859	-1,80	0.737	0.92	-1,80	0.846	0.022	5
1430	0.870	-0,90	0.757	0.931	-0,90	0.866	0.002	7
1431	0.865	-2,20	0.748	0.879	-2,20	0.759	0	34
1529	0.832	-4,90	0.691	0.887	-4,90	0.779	0.025	6

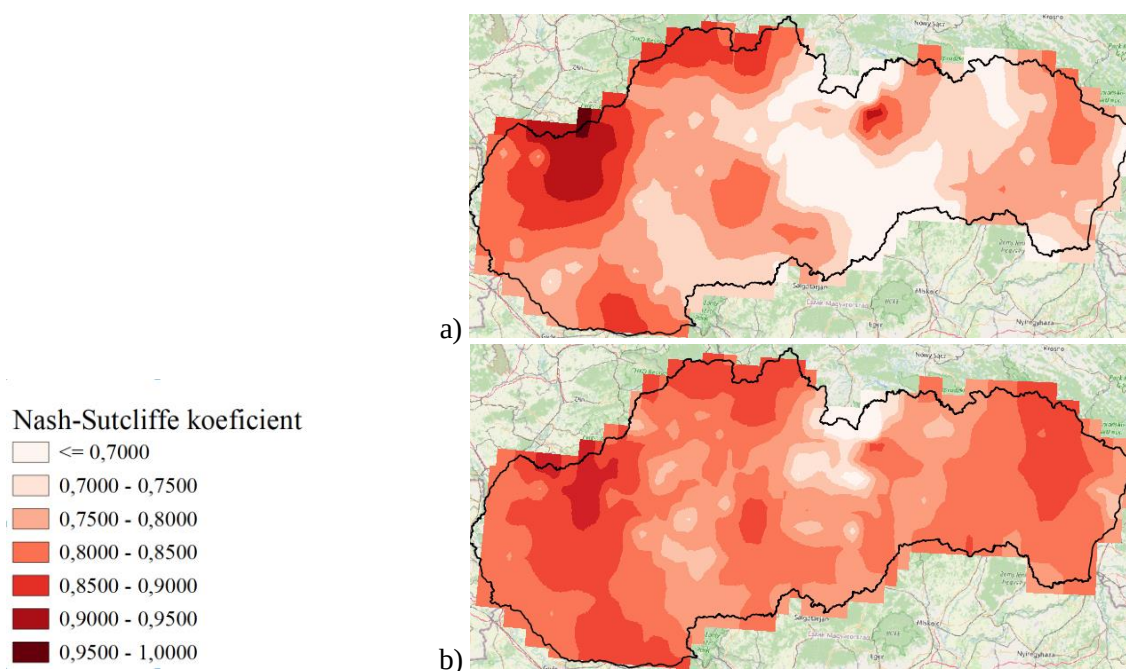
5.4 Výpočet LASSO pre územie celého Slovenska

Cieľom tejto kapitoly bolo rovnakým spôsobom ako v predchádzajúcej kapitole nakalibrovať a otestovať model LASSO pre územie celého Slovenska. Týmto spôsobom sme overili, nakoľko je možné túto metódu použiť pre rôzne časti Slovenska. Princíp výpočtu bol rovnaký a takisto obdobie, s ktorým sme pracovali. V tejto kapitole sme pracovali so všetkými bodmi databázy CarpatClim, ktoré sa nachádzajú na území Slovenska (599 bodov). Ku každému bodu CarpatClim sme vybrali 8 najbližších bodov databázy ECA&D. Na údajoch z obdobia 1960 – 2000 bol model natrénovaný a následne boli na základe zistenej regresnej rovnice vypočítané dáta pre obdobie 2001 – 2010 pre každý bod CarpatClim na Slovensku. Dáta vypočítané na základe zistenej regresnej rovnice boli porovnané s testovacími údajmi databázy CarpatClim. Výsledky sú spracované graficky. Na Obr. 6 je možné vidieť, že na väčšine Slovenska korelačné koeficienty dosahovali minimálne hodnotu 0,8, čo signalizuje silnú väzbu medzi údajmi vypočítanými modelom LASSO a testovacími údajmi pochádzajúcimi priamo z databázy CarpatClim. Ešte lepšie hodnoty (0,9-0,95) korelačných koeficientov boli dosiahnuté pri porovnávaní mesačných úhrnov (Obr. 6 b).



Obr. 6 Korelačné koeficienty a) denných údajov b) mesačných údajov

Ďalším koeficientom, na základe ktorého sme porovnávali vhodnosť použitia výpočtu LASSO na Slovensku bol Nash-Sutcliffe koeficient. Opäť boli porovnávané aj denné aj mesačné úhrny (Obr. 7)

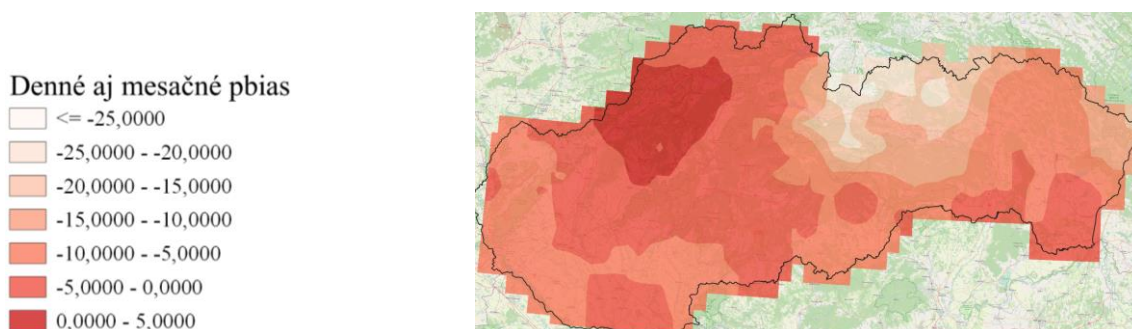


Obr. 7 Nash Sutcliffe koeficienty pre a) denné úhrny b) mesačné úhrny

Pre denné úhrny boli na väčšine územia dosiahnuté koeficienty minimálne 0,7, čo znamená silnú väzbu medzi vypočítanými a testovacími údajmi. Dosiahnuté výsledky boli lepšie pre mesačné úhrny, kde údaje väčšiny územia dosahovali hodnoty 0,8 - 0,95.

Ďalšou skúmanou veličinou bola hodnota pbias, čiže percentuálna odchýlka. V našom prípade predstavovala hodnota pbias odchýlku vypočítaných údajov od testovacích údajov. Táto hodnota bola

rovnaká pre denné a mesačné úhrny, nakoľko do vzorca pre jej výpočet vstupujú sumy údajov za celé obdobie, ktoré sú rovnaké pri práci s dennými aj mesačnými údajmi. Výsledné hodnoty percentuálnych odchýlok pbias sú opäť graficky znázornené na Obr. 8.



Obr. 8 Hodnoty pbias pre denné aj mesačné úhrny

Hodnoty pbias sa považujú za dobré, ak sa pohybujú v intervale $(0; \pm 5)$. Príčinou vyšších odchýlok sú pravdepodobne väčšie rozdiely v nadmorskej výške, s ktorou súvisia aj extrémne zrážky, ktorých zachytenie pomocou zrážkových databáz je náročnejšie.

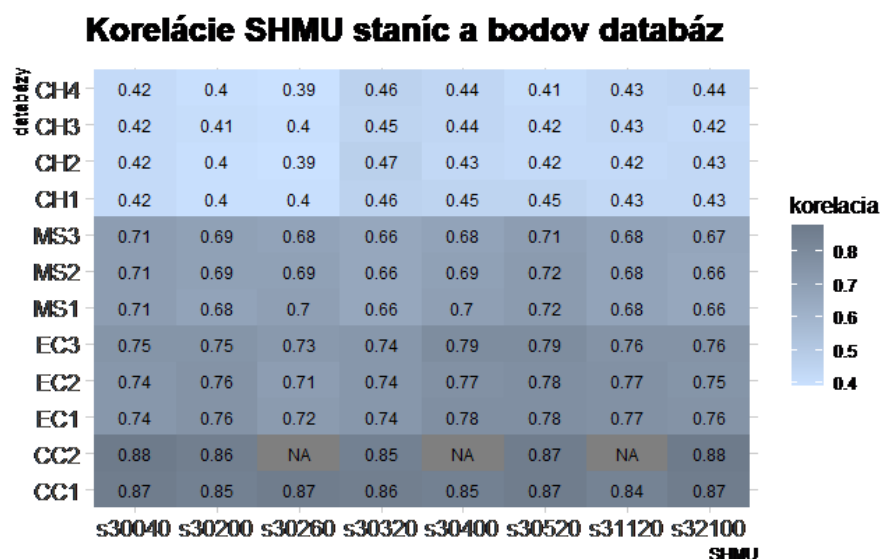
Na základe výsledkov v tejto kapitole je možné povedať, že na väčšine Slovenska je vhodné použiť model regularizovanej lineárnej regresie na výpočet zrážok databázy CarpatClim pomocou databázy ECA&D, nakoľko je možné týmto spôsobom vytvárať rady zrážok, ktoré sú v dobrej zhode s údajmi pochádzajúcimi priamo z databázy CarpatClim (testovacími údajmi).

5.5 Extrapolácia údajov databázy CarpatClim

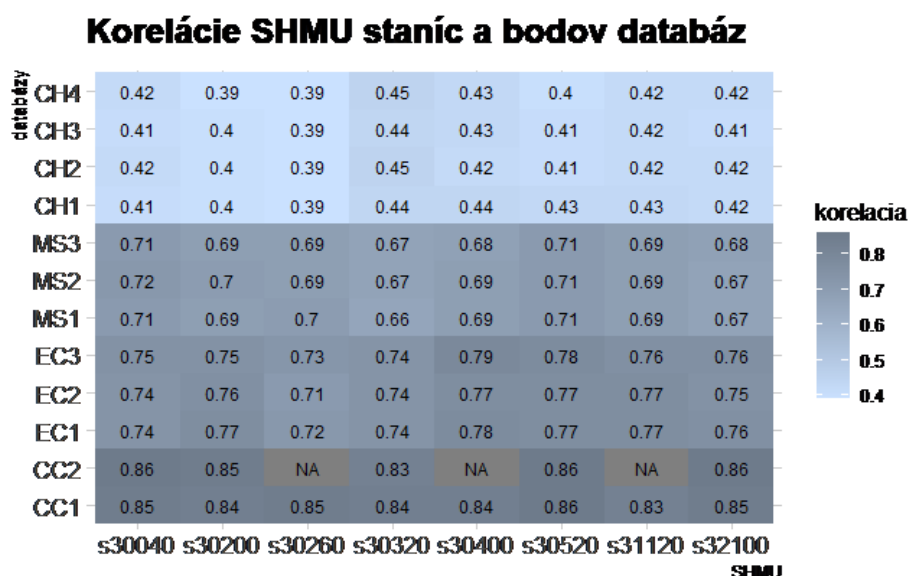
V tejto kapitole sme opäť pracovali s meranými údajmi SHMÚ a s údajmi z klimatických gridových databáz CarpatClim, ECA&D, MSWEP a CHIRPS. V tejto kapitole sme pomocou regresnej rovnice modelu LASSO vytvorili údaje databázy CarpatClim od 1.1.2011 po 31.10.2017, pretože v tomto období sú dostupné dáta všetkých ostatných klimatických databáz a zároveň merané údaje SHMÚ. Cieľom tejto kapitoly bolo zistiť, či v prípade, že časť údajov databázy CarpatClim pochádzala z extrapolácie, bol významnejšie ovplyvnený korelačný vzťah medzi údajmi databázy CarpatClim a meranými údajmi zo zrážkomerných staníc SHMÚ. Kalibračným obdobím v tomto prípade bolo obdobie 1960 – 2010.

Otestovanie vypočítaných údajov neprebiehalo ako súčasť LASSO, nakoľko údaje z databázy CarpatClim pre toto obdobie nie sú dostupné a nie je preto možné s nimi porovnať vypočítané údaje, ako v predchádzajúcich kapitolách. Namiesto toho boli otestované porovnaním s meranými údajmi pre obdobie 1981 – 2017. Okrem databázy CarpatClim boli s meranými údajmi porovnané aj ostatné použité klimatické databázy, aby bolo možné zistiť, či aj pri použití extrapolovaných dát bude mať databáza CarpatClim najlepšie výsledky.

V údajoch databázy CarpatClim boli údaje v období 2011 – 2017 získané extrapoláciou pomocou údajov databázy ECA&D a nepochádzali priamo z klimatickej databázy. Napriek tomu sa údaje databázy CarpatClim najviac približujú meraným údajom zo zrážkomerných staníc SHMÚ. Aby bolo možné zhodnotiť vplyv, aký malo použitie extrapolovaných údajov na koreláciu s meranými údajmi, vytvorili sme heatmap korelačných koeficientov len pre údaje z SHMÚ a klimatických databáz (bez extrapolovaných údajov), viď. Obr. 9 a heatmap korelácii po doplnení extrapolovaných údajov databázy CarpatClim (Obr. 10). Je možné konštatovať, že použitie extrapolovaných dát malo iba minimálny vplyv na silu väzby údajov databázy CarpatClim s meranými údajmi.



Obr. 9 Korelačné koeficienty denných údajov pre obdobie 1981-2010 bez extrapolovaných dát



Obr. 10 Korelačné koeficienty denných údajov pre obdobie 1981 - 2017 pri použití aj extrapolovaných dát databázy CarpatClim

Záver

Cieľom tejto práce bolo zistiť, ktorá z voľne dostupných klimatických gridových databáz poskytuje zrážkové údaje najviac sa približujúce údajom pochádzajúcim z priamych meraní v zrážkomerných staniciach. Zistili sme, že meraným údajom zo zrážkomerných staníc SHMÚ sa najviac približujú údaje databázy CarpatClim. Nevýhodou tejto databázy je to, že poskytuje údaje len po rok 2010. Z tohto dôvodu sme sa rozhodli, že ich budeme extrapolovať. Na extrapoláciu sme si vybrali metódu regularizovanej lineárnej regresie LASSO. Na nakalibrovanie a otestovanie modelu sú potrebné údaje z dvoch rôznych zdrojov. Z databázy, ktorej údaje chceme extrapolovať (CarpatClim) a databázy, ktorá má dostupné údaje v období pre ktoré chceme extrapoláciou získať údaje. Ako druhú databázu sme zvolili ECA&D, pretože jej údaje najlepšie korelovali s údajmi databázy CarpatClim. Aby sme overili, či je zvolená metóda LASSO vhodná pre extrapoláciu, vytvorili sme model pre záujmové

územie povodia Nitrianska Streda a model sme natrénovali a otestovali na dátach, ktoré sme mali dostupné. Vypočítané údaje sme s testovacími porovnávali na základe korelačného koeficientu, Nash Sutcliffe koeficientu a percentuálnej odchýlky p_{bias} . Keďže vypočítané údaje sa približovali údajom z klimateckej databázy, rozhodli sme sa model použiť aj na všetky body databázy CarpatClim nachádzajúce sa na Slovensku. Týmto testom sme chceli dokázať, že metódu LASSO je možné použiť na väčšine územia Slovenska s veľmi dobrými výsledkami. Najlepšie výsledky boli dosiahnuté na nížinatých územiach. Korelačné väzby sme overovali pre denné aj mesačné údaje, pričom lepšie výsledky boli dosiahnuté s mesačnými. Po overení modelu LASSO sme extrapolovali údaje databázy CarpatClim a znova overili korelačné väzby s meranými údajmi, čím sme dokázali, že údaje získané extrapoláciou sú veľmi dobre použiteľné v prípade, že klimatecká databáza neposkytuje údaje v potrebnom období. Prácou sme dokázali, že v prípade nedostatočných alebo nevyhovujúcich meraných údajov zo zrážkomerných staníc je možné použiť údaje z klimateckých gridových databáz. Tieto údaje majú široké uplatnenie pri analýzach malých aj veľkých povodí, nakoľko ich siete bodov sú husté, údaje sú voľne dostupné a nesplošné.

6 Zdroje

- [1] „Princípy a prístroje na meranie zrážok,“ [Online]. Available: <https://bioclio.com/principy-a-pristroje-na-meranie-zrazok/>.
- [2] G. e. a. Blöschl, Runoff prediction in ungauged basins, Cambridge: Cambridge University Press, 2013.
- [3] A. Keszeliová, „Porovnanie územného výparu vo vybraných povodiach Slovenska,“ 2019. [Online]. Available: https://www.shmu.sk/File/KMO/KeszeliovaA_Porovnanie_uzemneho_vyparu_vo_vybranych_povodiach_Slovenska.pdf.
- [4] „Home. carpatclim-eu.org,“ [Online]. Available: <http://www.carpatclim-eu.org/>, Home. carpatclim-eu.org. [Online].