

SZEGEDI TUDOMÁNYEGYETEM

Természettudományi és Informatikai Kar

Bolyai Intézet

Gráfrepresentáció árhullámok
modellezésére idősoros adatokban

Országos Tudományos Diákköri Konferencia

Szerző: Dabis Marcell

Témavezető: Dr. Vizi Zsolt

2023

Tartalomjegyzék

1. Bevezetés	2
2. Adatok leírása	4
2.1. Állomások	4
2.2. Idősor	5
3. Gráfrepresentáció	8
3.1. Gráf konstrukció	8
3.2. Karakterisztikus gráfok és utak	15
4. Árhullámok és statisztikai elemzés	17
4.1. Implementáció	17
4.2. Mélységi Bejárás	19
4.3. Metrikák a statisztikai elemzéshez	20
4.4. Eredmények	24
5. Diskusszió és további tervek	29
Hivatkozások	32
A. Hőtérképek	33

1. Bevezetés

Az alkalmazott hidrológiában az árvízvédelem és -előjelzés kulcsszerepet játszik, és igen nagy érdeklődés övezi. Az Alsó-Tisza-vidéki Vízügyi Igazgatóság (ATI-VIZIG) munkatársai széleskörű kutatásokat végeztek az árvizekkel kapcsolatban, például vizsgálták, hogy milyen okok állhatnak az utóbbi időben a Tiszán levonuló árvizek megnövekedett száma [4] és megnövekedett nagysága [5] mögött. Általánosságban elmondható, hogy a Tiszán az árvizek gyakoribbá és nagyobb méretűvé váltak, ami a 19. században végrehajtott (legnagyobb) folyószabályozási intézkedésekkel magyarázható. A szabályozások eredménye egyrészt a folyó 485 kilométerrel való megrövidülése, másrészt az árvízi sebesség és a vízfelszín lejtésének növekedése volt. A folyót érő emberi beavatkozások hatásairól szóló elemzés a [7] cikkben található. Nemrégiben fedezték fel az árvizek megjelenésének időbeli változása és az éghajlatváltozás közötti összefüggést Európában, lásd a [1] tanulmány. Modelünk egyik célja, hogy pontosabban mérhetővé tegye az árhullámok, áresemények számát, aminek köszönhetően az ilyen időszakok könnyebben megkülönböztethetővé válnak. Ezen felül értékes metrikákat is ki lehet nyerni, például az árhullámok levonulási idejét. Ezek az adatok hasznosak lehetnek a folyó viselkedésének mélyebb megértéséhez.

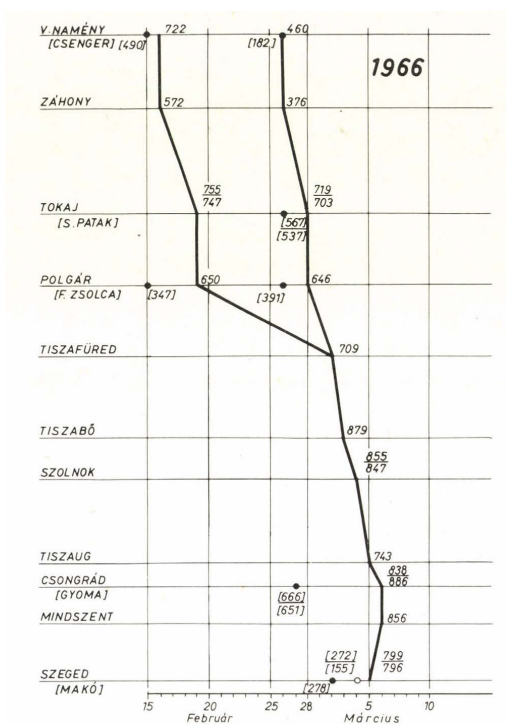
Napjainkban, a digitális mérőrendszerek terjedésének és fejlődésének köszönhetően, az adatalapú megközelítések egyre népszerűbbé válnak és segítik az árvizek megértését, bármilyen folyóról is legyen szó. Egy nemzetközi összefogás keretein belül sor került egy árvízi adatbázis létrehozására, amelynek célja az adatok, modellek és szakértelem megosztása, lásd [3]. Lokálisan a Tisza alsó folyásán bekövetkezett árvízi események miatt Kozák az árvízvédelem fejlesztését javasolta [6]. Ő kezdeményezte az adatvezérelt módszerek bevezetését, melyek hozzájárulhatnak a hatékonyabb intézkedések bevezetéséhez, ezért vette fel velünk a kapcsolatot.

Ebben a dolgozatban a Tisza mentén kialakuló árhullámokat vizsgáljuk, és egy keretrendszert mutatunk be a beérkező adatokból történő kinyerésre és elemzésre. Az árhullám a folyó mentén a mérési pontokon egymás után bekövetkező tetőzések sorozatát jelenti egy adott időszakban, amely a folyó áramlását követve terjed végig a mérőállomásokon. Mivel egy árhullámmal kapcsolatban számos nem várt jelenség is felmerülhet, a modellünkben használt formalizmus néhány feltételezéssel indul, amelyek automatikusan kizárják ezen jelenségek egy részét. Ugyanakkor a gyakran előforduló árhullámokat nem zárjuk ki, ami lehetővé teszi statisztikai vizsgálatok elvégzését (ezekre az árhullámokra), és betekintést nyújt a múltbeli folyószabályozási tevékenységek hatásaira is.

A vízszintbeli tetőzések terjedési ideje egy folyó viselkedésének leírására szolgáló mérőszám. Vágás a hidrológia alapjai tárgyaló könyvében [8] az árhullám terjedési

idejének kiszámítására javasol módszereket, de a módszertan bizonyos lépéseinél számos kivétel felmerülhet, ami nincs szigorú értelemben lekezelve. A jelen dolgozatban bemutatott fogalmak a mi modellezési szemléletünkhöz kötődnek, és teljesen eltér a korábban említett módszerektől, mind a számítás módjában, mind a formalizmus rigorózusságában.

Az első kísérletet az árhullámok formálisabb (de nem teljesen formális) elemzésére Vágás tette a Tisza folyón levonuló árvizekről szóló könyvében [10]. Ebben az időben még minden adatgyűjtést és számítást kézzel végzett el, így az elemzést a hidrológiai szempontból érdekes esetekre és időintervallumokra végezte. A tetőzési sorozat kialakítására vonatkozó megközelítése néhány további feltételezést is tartalmazott, és bizonyos szinten formális volt, de sok külső, szubjektív megfontolásokat is figyelembe vett, ezzel jelentősen megnehezítve ezen ábrák általánosságban vett használatát.



6—21 b ábra: Tetőzési vonal, 1966.

1. ábra. A [10] könyv 112. oldala, amely egy 1966 késő telén levonuló árhullámot ábrázol. A fekete vonal köti össze a vízszintbeli tetőzéseket.

Az 1. ábrán az x tengely az időt, az y tengely pedig az általunk használt mérési pontokat ábrázolja a folyón való elhelyezkedésük szerint rendezve. A mérési pontok közül Szeged a legközelebbi és Vásárosnamény a legtávolabbi a folyó torkolatától. A fekete pontok az adott állomás közelében a Tiszába torkolló mellékfolyó vízszintjeinek tetőzéseit jelölik. Az ilyen jellegű ábrákat nevezte el Vágás tetőzési vonalaknak. Ez a megközelítés számos, a Tisza áradásáról készített esettanulmányában is szerepelt, lásd [9].

Az jelenlegi dolgozat felépítése a következő: a 2. fejezetben tárgyaljuk a dolgozat alapjául szolgáló adathalmaz struktúráját, és bevezetjük a későbbi fejezetekben megjelenő formális ábrázolásokhoz szükséges jelöléseket. A 3. fejezetben bevezetünk egy gráfrepresentációt az árhullámok vizsgálatához: mivel a konstruált gráfok kizárólag az adatokon alapulnak, csak néhány általános tulajdonságot tudunk megállapítani és bizonyítani, például hogy a gráf irányított körmentes gráf, valamint néhány egyenlőtlenséget a csúcsok bemenő és kimenő fokszámaira. A 4. fejezetben megadjuk az árhullám definícióját, amelyet a projekthez implementált kóddal elemezzünk. Ez a fejezet a munka eddigi eredményeit mutatja be, amely az árhullámok számát és a hozzájuk tartozó átlagos terjedési időket vizualizálja különböző paraméterezések esetén. Az utolsó fejezetben összefoglaljuk az eddigi eredményeket, és megadjuk a további kutatás és fejlesztés irányait.

A munka az ATIVIZIG-gel folytatott együttműködés része, ők biztosították az adatokat az elemzéshez. Mivel az ipari együttműködések az esetek többségében termékorientáltak, az alábbiakban bemutatott módszertant tartalmazó szoftverprototípust fejlesztettünk. Az egyetemi oldalról Budai László, dr. Vizi Zsolt és Dabis Marcell alkotja a csapatot, partnereink az igazgatóságról pedig dr. Kozák Péter (igazgató), Fiala Károly, Lázár Miklós és Fehérváry István.

2. Adatok leírása

A dolgozatban az 1951.01.01. és 2020.12.31. közötti időszak adatait használjuk, amelyek a folyó mentén 23 mérési pontból származnak. A következőkben csak napi 1 adatot használunk fel, azonban az utóbbi időben az adatbekérés a kézi adatrögzítésről digitálisra váltott, ami sokkal sűrűbb adatsort eredményez. Mivel ezek a sűrűbb adatok csak az elmúlt 25-30 évben állnak rendelkezésre, ezért minden napból az első elérhető értéket használjuk. Az adatokban előfordulnak hiányosságok, például amikor az adott napon nem rögzítettek vízszint értéket, ennek kezelését később tárgyaljuk. Először a mérési pontokat vizsgáljuk, mivel azok elhelyezkedése értékes információkat hordoz számunkra.

2.1. Állomások

Az általunk vizsgált mérési pontok — más néven *állomások* — a Tisza mentén állnak rendelkezésre, és minden ilyen állomáshoz idősoros adat tartozik. Az állomások közötti természetes sorrendet szeretnénk megőrizni, amelyet a folyóparton való elhelyezkedésük ad. Ehhez a folyótorkolattól folyamkilométerben mért távolságukat használjuk. Egy másik értékes információ egy adott állomásról a nullpontjának magassága, amely az a hely, ahol a vízszintet nullának tekintjük, feltételezve, hogy az

alacsony vízállású folyó vízszintje negatív. A nullpont magassága lehetővé teszi, hogy a relatív vízszintméréseket abszolút értékre alakítsuk át. A következőkben megadjuk az állomások formális definícióját.

2.1. Definíció. Az *állomások* halmaza a következő típusú hármasokból áll:

$$\mathbf{S} := \{(\sigma, \rho, \nu) : \sigma \text{ egy karaktersorozat, } \rho \in \mathbb{R}^+, \nu \in \mathbb{R}\},$$

ahol σ a mérési pont azonosítója (legtöbb esetben a neve), ρ a hozzátartozó folyótorkolattól vett távolság folyamkilóméterben és ν jelöli a nullpont tengerszint feletti magasságát (méterben). Jelölje az állomások számát K , tehát $|\mathbf{S}| = K$.

A fent említett hármas elemei tartalmazzák az elemzésünkhöz szükséges összes információt az állomásokkal kapcsolatban, ezért a $\mathbf{S}$ elemeire állomásként fogunk hivatkozni. Fontos megjegyezni, hogy a nullpont magassága idővel változhat, ezért a ν értéke a nullpont magasságának legfrissebb értékét jelöli. Az általunk alkalmazott megközelítésben a folyó mentén haladunk lefelé, a folyó torkolatától mért legtávolabbi állomástól a legközelebbi állomásig, ezért az állomások elhelyezkedésük alapján történő összehasonlításához az $\mathbf{S}$ halmazon egy rendezést vezetünk be.

2.2. Definíció. Tekintsük az $s, \hat{s} \in \mathbf{S}$ állomásokat úgy, hogy $s = (\sigma, \rho, \nu)$, $\hat{s} = (\hat{\sigma}, \hat{\rho}, \hat{\nu})$. Azt mondjuk, hogy s megelőzi $\hat{s}$ -t akkor és csak akkor, ha $\hat{s}$ közelebb van a folyótorkolathoz mint s , azaz

$$s \leq \hat{s} \iff \hat{\rho} \leq \rho.$$

Megjegyezzük, hogy $(\mathbf{S}, \leq)$ egy jól rendezett halmaz, mivel a rendezési reláció reflexív és tranzitív, és mivel $\mathbf{S}$ bármely adott nem üres részhalmazának van legkisebb eleme, amely éppen a folyó torkolatától legtávolabbi állomás az adott részhalmazon belül.

2.2. Idősor

A továbbiakban a $\{1, 2, \dots, N\}$ halmazt, azaz az első N természetes szám halmazát $\mathbb{N}_N$ -nel, a k és l közötti természetes számok halmazát $\{t \in \mathbb{N} : k \leq t \leq l\}$, ahol $k, l \in \mathbb{N}$ és $k < l$, pedig $\mathbb{N}_{k,l}$ -l jelöljük.

2.3. Definíció. A *dátumok* halmazát a következőképpen definiáljuk:

$$\mathbf{D} := \{d_1, d_2, \dots, d_N\},$$

ahol d_t dátum $t - 1$ nappal van d_1 után és $t \in \mathbb{N}_N$.

Az árhullámok formális ábrázolásához bevezetünk egy leképezést, amely egy kiválasztott állomáshoz a megfelelő teljes vagy megszorított idősorral tér vissza.

2.4. Definíció. A s állomáshoz tartozó *idősor-adatokat* az X leképezés adja meg, a következőképpen:

$$X : \mathbf{S} \rightarrow (\mathbb{R} \cup \{\emptyset\})^N, \quad X(s) = \{x_t(s)\}_{t \in \mathbb{N}_N},$$

ahol $x_t(s)$ jelöli a vízállás értéket a d_t időpontban az s állomásnál, és $\emptyset$ a hiányzó adatot jelöli. Ezen kívül az X egy szűkebb időintervallumra való korlátozását a következőképpen jelöljük

$$X_{k,l} : \mathbf{S} \rightarrow (\mathbb{R} \cup \{\emptyset\})^{|\mathbb{N}_{k,l}|}, \quad X_{k,l}(s) = \{x_t(s)\}_{t \in \mathbb{N}_{k,l}},$$

amely a d_k és d_l közötti időpontok közötti intervallumra megszorított idősor.

	Szolnok	Tiszaug	Csongrád	Mindszent	Algyő	Szeged
2007-06-20	77.19	75.72	75.37	75.02	74.86	74.81
2007-06-21	77.09	75.56	75.18	74.87	74.77	74.71
2007-06-22	77.21	75.58	75.15	74.87	74.80	74.76
2007-06-23	77.24	75.56	75.15	74.87	74.77	74.73
2007-06-24	77.21	75.59	75.15	74.85	74.77	74.71
2007-06-25	77.14	75.59	75.18	74.88	74.80	74.72
2007-06-26	77.33	75.65	75.19	74.89	74.82	74.74
2007-06-27	77.26	75.63	75.18	74.88	74.79	74.73
2007-06-28	77.58	75.62	75.12	74.84	74.76	74.70
2007-06-29	77.55	75.81	75.26	74.90	74.81	74.73
2007-06-30	77.33	75.72	75.27	74.92	74.79	74.74
2007-07-01	76.81	75.49	75.14	74.83	74.74	74.70
2007-07-02	76.52	75.22	74.95	74.73	74.65	74.64
2007-07-03	76.33	75.02	74.78	74.62	74.55	74.57
2007-07-04	76.28	74.92	74.66	74.53	74.50	74.53
2007-07-05	76.46	74.86	74.63	74.50	74.47	74.50
2007-07-06	76.99	74.99	74.67	74.53	74.54	74.53
2007-07-07	77.32	75.30	74.93	74.68	74.68	74.61

2. ábra. 6 állomáson mért vízállás értékek idősorai abszolút magasságban (méterben megadva), táblázatba rendezve. Az állomások sorrendje balról jobbra követi az állomások folyás irány szerinti sorrendjét, az érték rögzítésének időpontja az első (index) oszlopban szerepel.

Mint említettük, az adatokban gyakran vannak hiányosságok, ami azt jelenti, hogy valamilyen okból kifolyólag az adott napon nem történt adatrögzítés. A X leképezés definiálásánál a programozási környezetben a NaN (vagy n.a.) értékeket reprezentáló $\emptyset$ jelölést ennek okán vezettük be. Mivel módszertanunk feltételezi, hogy a vizsgált állomásokra vonatkozóan rendelkezésre állnak az adatok, bevezetünk néhány jelölést a $\emptyset$ értékeket nem tartalmazó intervallumokra.

2.5. Definíció. Legyen adott $s \in \mathbf{S}$ és $k, l \in \mathbb{N}$ úgy hogy $k < l$. Azt mondjuk, hogy (k, l) egy *érvényes időablak*, ha

$$\forall t \in \mathbb{N}_{k,l} : x_t(s) \neq \emptyset.$$

Az egyes állomások nullpontjainak magassága idővel változhat a folyómeder alakját befolyásoló természeti hatások miatt, amelyek a mérőeszköz felemelkedéséhez vagy más hatásokhoz vezetnek. Ezért ebben a munkában *nullpont-korrigált* idősorokat használunk, ami azt jelenti, hogy minden vízszint érték a legutóbbi nullpont magasságához van korrigálva. Ez úgy történik, hogy a magasságkülönbséget hozzáadjuk azokhoz a vízállás értékekhez, amelyeket egy olyan múltbeli időintervallumban rögzítettek, amikor a nullpont nem volt azonos magasságban a jelenlegi nullponttal. Amikor egy vízállást vizsgálunk, mindig feltételezzük, hogy az nullpont-korrigált. A vízállás értékeket centiméterben fogjuk használni, ez a pontosság elegendő a céljainkhoz.

A *vízállás értékek abszolút magasságban* vett idősorát létrehozhatjuk a nullpont-korrigált értékek felhasználásával, úgy, hogy eltoljuk őket a jelenlegi nullpont magasságával, így megkapva a tengerszint feletti abszolút vízszint értékét. Ezt az idősort méterben értelmezzük két tizedesjegy pontossággal, ugyanolyan pontossággal, mint a nullpont-korrigáltat. Formálisan, ha vesszük az $s = (\sigma, \rho, \nu)$ állomást és ennek a (nullpont-korrigált) megszorított idősorát $X_{k,l}(s) = \{x_t(s)\}_{t \in \mathbb{N}_{k,l}}$, ahol (k, l) egy érvényes időablak, akkor a vízállás értékeket abszolút magasságban úgy kapjuk meg, hogy minden értékhez hozzáadjuk a ν jelenlegi nullpont magasságot, azaz

$$\left\{ \frac{x_t(s)}{100} + \nu \right\}_{t \in \mathbb{N}_{k,l}},$$

mivel ν méterben van megadva. Ez az idősor lehetővé teszi a hidraulikus gradiens, azaz a vízfelszín meredekségének kiszámítását, amely a folyószabályozás szempontjából hasznos statisztika.

Az idősorokat egy $N \times K$ mátrixba rendezhetjük, ahogy ezt a 2. táblázatban is láthatjuk. Itt az első sor és oszlop, vagyis az állomások neve, illetve a dátumok illusztrációs céllal szerepelnek. Ez a táblázat a vízállás értékeket abszolút magasságban tartalmazza. Mivel a rendszert általában egy kijelölt időintervallumból származó idősoros adatokra alkalmazzuk, az elemzésben egy közös érvényes időablakot feltételezünk, lásd a példát a 3. táblázatban.

Célunk, hogy az 1. ábrán látható ábrázolást formálisan elő tudjuk állítani az árhullámokra. Emiatt bevezetünk egy jelölést az állomások közötti "rákövetkezőség" kapcsolatra, mivel folyótorkolattól nézve a legtávolabbitól a legközelebbiig haladunk.

	Szolnok	Tiszaug	Csongrád	Mindszent	Algyő	Szeged
2007-06-20	77.19	75.72	75.37	75.02	NaN	74.81
2007-06-21	77.09	75.56	75.18	74.87	NaN	74.71
2007-06-22	77.21	75.58	75.15	74.87	NaN	74.76
2007-06-23	77.24	75.56	75.15	74.87	74.77	74.73
2007-06-24	77.21	75.59	75.15	74.85	74.77	74.71
2007-06-25	77.14	75.59	75.18	74.88	74.80	74.72
2007-06-26	77.33	75.65	75.19	74.89	74.82	74.74
2007-06-27	77.26	75.63	75.18	74.88	74.79	74.73
2007-06-28	77.58	75.62	75.12	74.84	74.76	74.70
2007-06-29	77.55	75.81	75.26	74.90	74.81	74.73
2007-06-30	77.33	75.72	75.27	74.92	74.79	74.74
2007-07-01	76.81	75.49	75.14	74.83	74.74	74.70
2007-07-02	76.52	75.22	74.95	74.73	74.65	74.64
2007-07-03	76.33	75.02	74.78	74.62	74.55	74.57
2007-07-04	76.28	74.92	74.66	74.53	74.50	NaN
2007-07-05	76.46	74.86	74.63	NaN	74.47	NaN
2007-07-06	76.99	74.99	74.67	NaN	74.54	74.53
2007-07-07	77.32	75.30	NaN	74.68	74.68	74.61

3. ábra. A vízállás értéket abszolút magasságban a 2007-06-20 és 2007-07-07 közötti időintervallumra vonatkozóan, és (mesterségesen, szemléltetés céljából) bevezettük a hiányzó adatokat (NaN jelöléssel). A zöld terület az érvényes ablakok közös részét jelöli ezen a példán, ahol nincs $\emptyset$ érték.

2.6. Definíció. Az $s, \hat{s} \in \mathbf{S}$ és $s \neq \hat{s}$ állomások közötti $\rightsquigarrow$ reláció a következőképpen definiálhatjuk

$$s \rightsquigarrow \hat{s} \iff \nexists r \in \mathbf{S} \setminus \{s, \hat{s}\} : s \leq r \leq \hat{s}.$$

3. Gráfrepresentáció

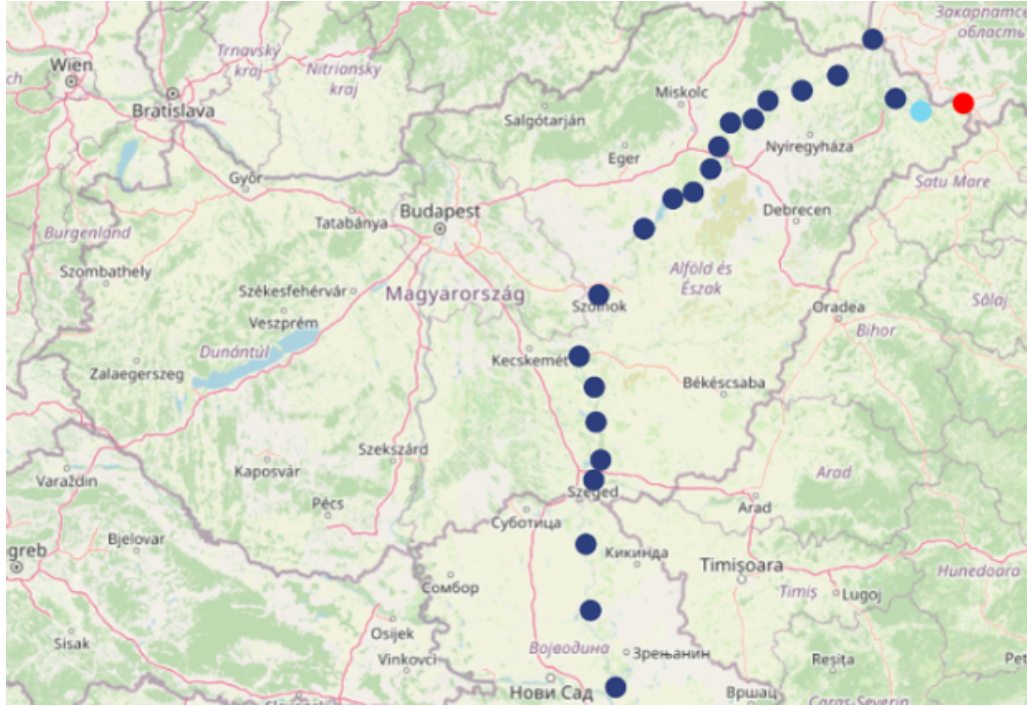
Ahogy a bevezetőben említettük, célunk a rendelkezésre álló idősoros adatok alternatív, több információval szolgáló reprezentációjának felírása. Az erre adott megoldásunk egy irányított gráf, amely tartalmazza az árhullámok leszűréséhez szükséges valamennyi információt.

3.1. Gráf konstrukció

Először is olyan mintázatokat keresünk az idősorokban, amelyek csúcspontjai egy meghatározott időintervallum közepén helyezkednek el. Mivel az adatok gyakran nagyon zajosak, nem akarunk minden ingadozást csúcspontnak tekinteni.

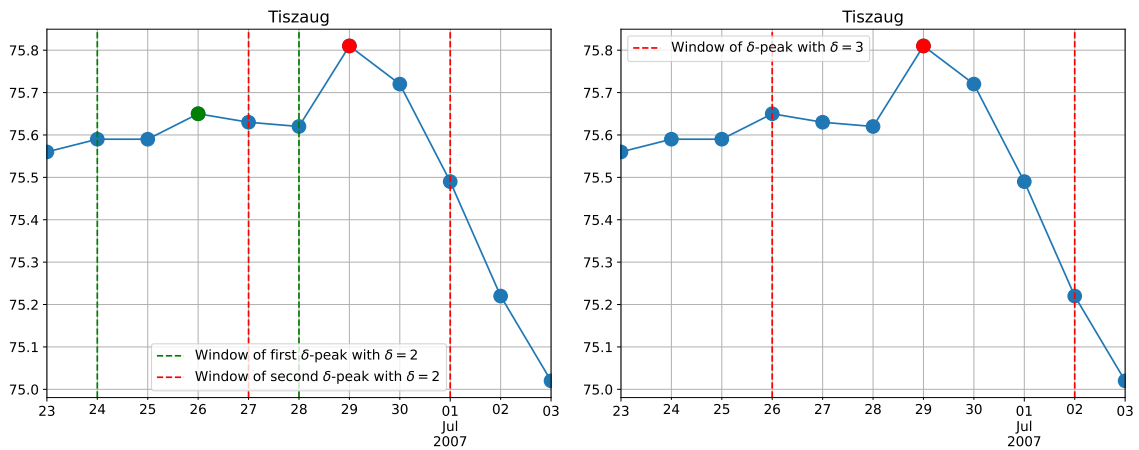
3.1. Definíció. Adott $X_{k,l}(s)$ úgy, hogy (k, l) egy érvényes időablak. Azt mondjuk, hogy $x_t(s)$ vízállás érték egy δ -csúcs egy rögzített $\delta \in \mathbb{Z}^+$ (napokban értendő) sugar esetén, ha

1. $k \leq t - \delta < t + \delta \leq l$,
2. $\forall i \in \mathbb{N}_{t-\delta, t-1} : x_i(s) < x_t(s)$,
3. $\forall j \in \mathbb{N}_{t+1, t+\delta} : x_j(s) \leq x_t(s)$.



4. ábra. Ebben a munkában felhasznált 23 állomás helye, ebből 3 Szerbia, 20 pedig Magyarország területén található. A piros pont az első állomás Tiszabecsnél, a világoskékkel jelölt rákövetkező állomás pedig Tivadarnál található.

Következésképpen, ha x_t értékét δ -csúcsnak tekintjük, akkor ez a legmagasabb érték (nem szigorú értelemben véve) $2\delta + 1$ napos időablakban, amelynek az x_t a középső eleme. Ez kiszűri az ingadozások nagy részét, és minél nagyobb δ -t használunk, annál kevesebb δ -csúcsot kapunk a csúcserőterek keresése során.



5. ábra. Idősoros adatok a 2007-06-23 és 2007-07-03 közötti időintervallumból a Tiszaug nevű állomásról. A bal oldalon látható, hogy $\delta = 2$ esetén ezen az intervallumon ezt a két vízállás értéket tekintjük δ -csúcsnak, mivel nincsenek benne egymás időablakában. A jobb oldalon a δ értéket 3-ra növeltük, aminek eredményeként csak az utóbbit tekintjük δ -csúcsnak

A következőkben magyarázatot adunk arra, miért törjük meg szimmetriát δ -csúcs definíciójában a bal és a jobb intervallumban lévő vízállásokra vonatkozóan. Mivel az általunk használt idősorokban gyakori, hogy az egymást követő értékek egyenlőek, ezt a mintát *platónak* nevezzük, és a következőkben definiáljuk.

3.2. Definíció. Adott $X_{k,l}(s)$, ahol (k, l) egy érvényes időablak. Legyenek $t \in \mathbb{N}_{k,l}$ és $p \in \mathbb{Z}^+$ olyan indexek, melyekre $k \leq t-1 < t+p \leq l$. A d_t időpontban egy p -napos platót kapunk, ha p darab egymást követő egyenlő érték van d_t -től kezdve, azaz:

$$x_t(s) = x_{t+1}(s) = \dots = x_{t+p-1}(s)$$

és az értékek a plató körül különbözőek, formálisan: $x_{t-1}(s) \neq x_t(s)$ és $x_{t+p}(s) \neq x_t(s)$.

Mielőtt továbblépnénk, a δ -csúcs definíciójából kimondható egy hasznos következmény.

3.1. Lemma. Tekintsük az s állomást, egy (k, l) érvényes időablakot, és a hozzátartozó $X_{k,l}(s)$ idősort. Ha $x_t(s)$ és $x_\tau(s)$ egyaránt δ -csúcsok az s állomásnál és $t \neq \tau$, akkor $|t - \tau| > \delta$.

Bizonyítás. Először is, $t > \tau$ esetén, ha $x_t(s)$ egy δ -csúcs, akkor az őt megelőző δ vízállás érték szigorúan kisebb, mint $x_t(s)$, tehát minden $i \in \mathbb{N}_{t-\delta, t-1}$ esetén $x_i(s) < x_t(s)$. Tegyük fel, hogy $x_\tau(s)$ az egyik ilyen korábbi vízállásérték a δ közül, tehát: $t - \tau \leq \delta$, illetve $\tau \in \mathbb{N}_{t-\delta, t-1}$. Mivel $x_\tau(s)$ is δ -csúcs, ezért az $x_\tau(s)$ -t követő δ vízállás érték nem lehet nagyobb, mint $x_\tau(s)$, tehát minden $j \in \mathbb{N}_{\tau+1, \tau+\delta}$ esetén $x_j(s) \leq x_\tau(s)$. Vegyük észre, hogy $t \in \mathbb{N}_{\tau+1, \tau+\delta}$, ami azt jelenti, hogy $x_\tau(s) < x_t(s)$ és $x_t(s) \leq x_\tau(s)$ ami ellentmondáshoz vezet, így $t - \tau > \delta$ áll fenn.

A másik irány analóg módon bizonyítható. ■

Következésképpen a csúcsokat a δ sugár választja el egymástól. Ez látható az 5. ábrán, ahol a δ paraméter eggyel való változtatásának eredményeképp a bal oldali ábrán található első csúcs a jobb oldali ábrán már nem tekinthető csúcsnak.

A δ -csúcsokat választjuk a gráfunk csúcsainak, majd a gráfot az 1. ábrához hasonlóan egy koordináta-rendszerben akarjuk elhelyezni. A felépítendő gráf csúcshalmazának bemutatásához már minden szükséges előkészületet megtettünk.

3.3. Definíció. Rögzítsünk egy $\delta \in \mathbb{Z}^+$ sugarat és egy (k, l) érvényes időablakot. A *csúcsok halmazát* a következőképpen definiáljuk:

$$\mathbf{V}_{k,l}(\delta) := \{(d_t, s) : x_t(s) \in X_{k,l}(s) \text{ egy } \delta\text{-csúcs az } s \in \mathbf{S} \text{ állomás esetén}\}.$$

Továbbá, ha egy $v = (d_t, s)$ csúcsot tekintünk, akkor a következő jelöléseket fogjuk használni

- $dt(v)$ -t a t -re (a d_t dátum indexére) való hivatkozáshoz,
- $s(v)$ -t a v -hez tartozó s állomásra való hivatkozáshoz.

Két csúcs közé húzott él a tetőzés vonulását jelenti számunkra, a következőben az ehhez szükséges feltételeket adjuk meg.

3.4. Definíció. Rögzítsünk egy $\delta \in \mathbb{N}$ sugarat és $\alpha, \beta \in \mathbb{N}_0$ egészeket.

Az $\mathbf{E}_{k,l}(\alpha, \beta, \delta) \subseteq \mathbf{V}_{k,l}(\delta) \times \mathbf{V}_{k,l}(\delta)$ *élek halmazát* a következőképpen definiáljuk:

$$\mathbf{E}_{k,l}(\alpha, \beta, \delta) := \{(u, v) : s(u) \rightsquigarrow s(v) \text{ és } -\alpha \leq dt(v) - dt(u) \leq \beta\}.$$

Vegyük észre, hogy ha $\alpha \neq 0$, akkor a u -ból v -be húzott élek akkor is létezhetnek, ha az u csúcs időpontja később van, mint a v -nek megfelelő időpont. Ahogy az 1. ábrán látjuk, egy ilyen helyzet a hidrológiai kontextusban is értelmezhető, ezért ezt lehetővé tettük a formalizmusban. Fontos kiemelni, hogy Vágás a [10]-es könyvében a vízszintbeli csúcsérték vonulását (amelyet árhullám folytatásaként is említ) lazábban és szubjektívebben értelmezte, az ábráin semmilyen formálisan rögzített szabályt nem feltételezett két talált csúcs összekötésénél. Az ő munkájához képest a mi megközelítésünk formális szabályokat állít fel, parametrikus és rugalmas, de a szigorú leírás megtöri a szubjektív döntést, hogy mikor kell összekötni a csúcsokat a gráfban.

	Szolnok	Tiszaug
2007-06-20	77.19	75.72
2007-06-21	77.09	75.56
2007-06-22	77.21	75.58
2007-06-23	77.24	75.56
2007-06-24	77.21	75.59
2007-06-25	77.14	75.59
2007-06-26	77.33	75.65
2007-06-27	77.26	75.63
2007-06-28	77.58	75.62
2007-06-29	77.55	75.81
2007-06-30	77.33	75.72
2007-07-01	76.81	75.49
2007-07-02	76.52	75.22
2007-07-03	76.33	75.02
2007-07-04	76.28	74.92
2007-07-05	76.46	74.86
2007-07-06	76.99	74.99
2007-07-07	77.32	75.30

6. ábra. A vízállás értékek abszolút magasságban a 2007-06-20 és 2007-07-07 időintervallumra nézve Szolnok és Tiszaug szomszédos állomásokról. Ha $\alpha = 0, \beta = 3, \delta = 2$ értéket használunk, a zöld intervallum azt jelzi, hogy a szolnoki csúcsoktól milyen intervallumban lévő tiszaugi csúcsokhoz húznánk be élt.

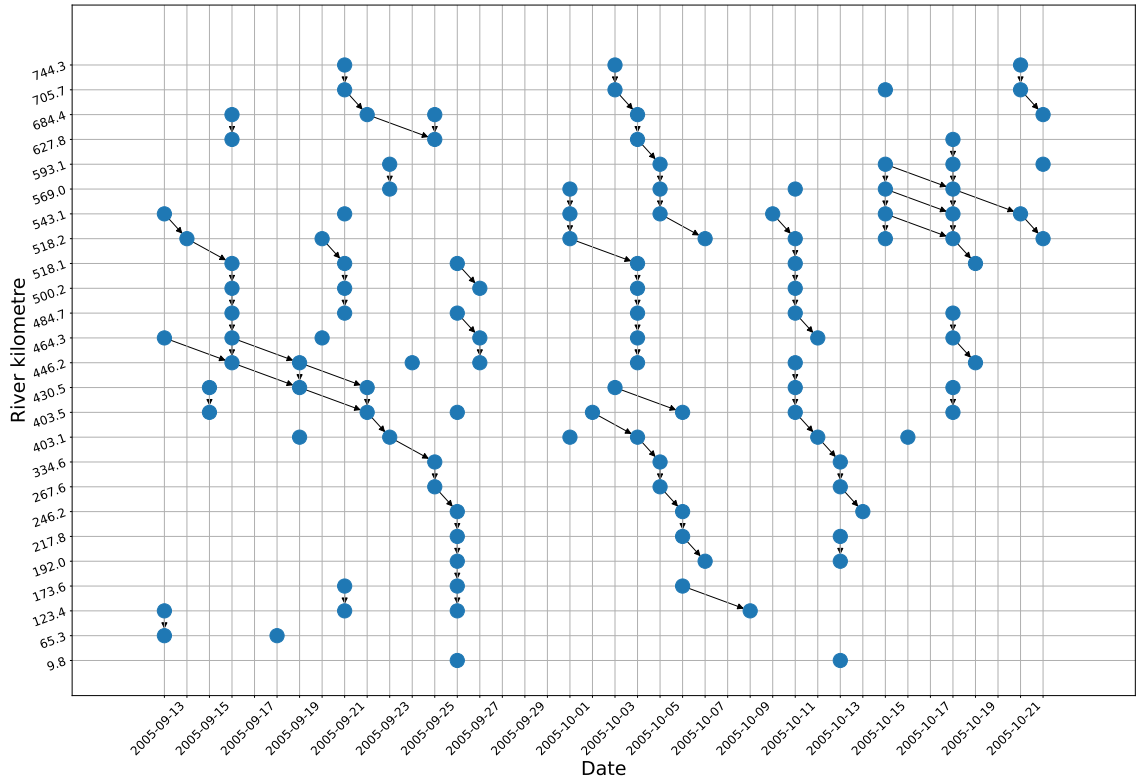
A csúcsok és élek korábban bevezetett halmazaiából egy irányított gráfot tudunk konstruálni, amely (a csúcsok definíciójának köszönhetően) koordináta-rendszerben is ábrázolható. Az x koordinátát a dátum, az y koordinátát pedig az állomások fent meghatározott sorrendje alapján kap minden csúcs (ennek következményeként az y tengelyt úgy is elképzelhetjük, mintha "kiegyenesítenénk" a Tisza folyót).

3.5. Definíció. Rögzítsük a $\delta \in \mathbb{N}$ sugarat, $\alpha, \beta \in \mathbb{N}_0$ egészeket és egy (k, l) érvényes időablakot. Definíáljuk a következő irányított gráfot:

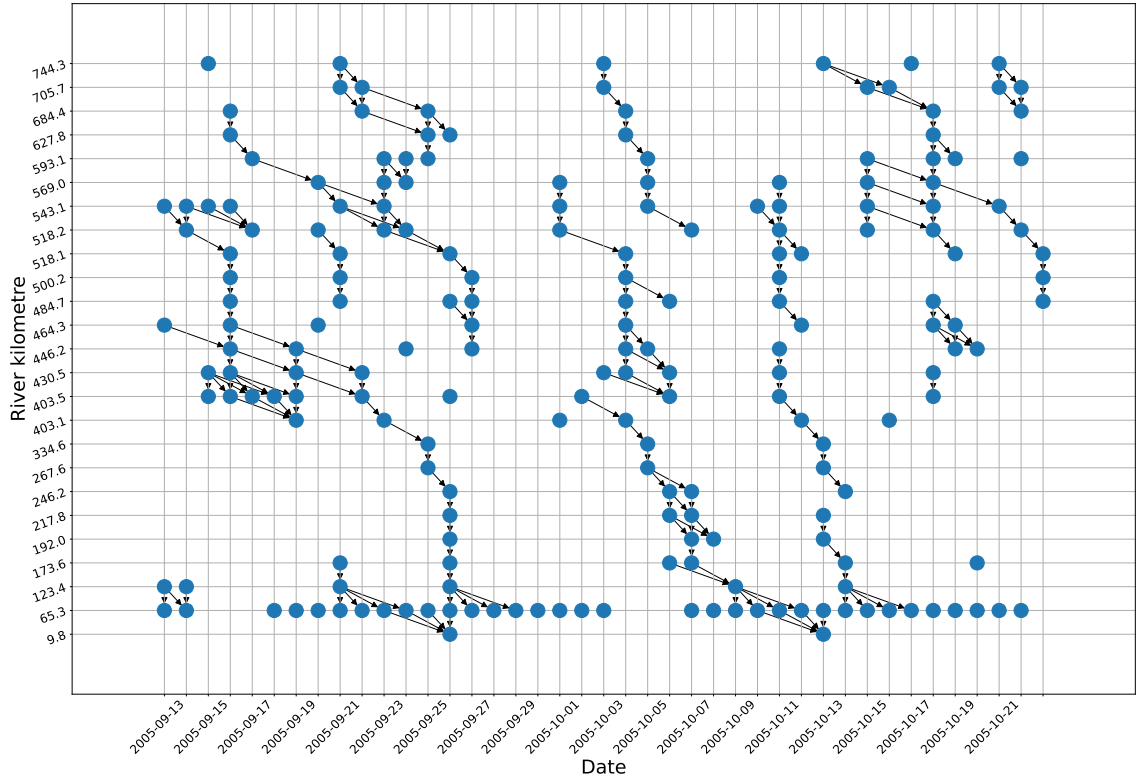
$$\mathbf{G}_{k,l}(\alpha, \beta, \delta) := (\mathbf{V}_{k,l}(\delta), \mathbf{E}_{k,l}(\alpha, \beta, \delta)).$$

Amint a jelölésből is láthatjuk, ez a gráf egy adott időintervallumra vonatkozik, amely akár évtizedeket is lefedhet az adatokból. Egy ilyen, koordináta-rendszerben elhelyezett gráfot láthatunk a 7. ábrán.

Vegyük észre, hogy a δ -csúcs aszimmetrikus definíciójából következik, hogy a platók esetén csak az első elem tekinthető δ -csúcsnak, mivel a platón az $x_t(s)$ utáni vízállás értékek nem felelnek meg a δ -csúcs definíció utolsó részének. A 7. és 8. ábrák azt a különbséget mutatják, hogy egy p -nap hosszú plató minden egyes értékét δ -csúcsnak tekintjük-e vagy sem.



7. ábra. $\mathbf{G}_{k,l}(0, 3, 2)$ gráfja, ahol k a 2005-09-12 dátum indexének, l pedig a 2005-10-21 dátum indexének felel meg. Amint azt az 1. ábrán láttuk, a vízszintes tengelyen az időpontokat, a függőleges tengelyen az állomásokat ábrázoljuk. A csúcsokat a rácspontokon jelöljük, az élek olyan állomások között vannak, ahol $\rightsquigarrow$ reláció fennáll.



8. ábra. $\mathbf{G}_{k,l}(0, 3, 2)$ gráfja, ahol k a 2005-09-12 dátum indexének, l pedig a 2005-10-21 dátum indexének felel meg. Az ábrán az látható, mi történne ha lazítanánk a δ -csúcs definíció megelőző értékekre vonatkozó szigorúan kisebb követelményén, így a platók minden pontját figyelembe kellene veyük. Úgy döntöttünk, hogy nem ezt a megközelítést használjuk, mivel a gráf számos jó tulajdonsága, például a 3.1-es lemma és annak minden további következménye sérül, és ez nehezebben áttekinthető ábrákhoz, valamint hibás statisztikai eredményekhez vezethet.

3.6. Definíció. Tekintsük a $G = (V, E)$ irányított gráfot. A $v \in V$ csúcs bemenő foksámát, azaz a v -be bemenő élek számát $\deg^-(v)$ -vel jelöljük. Hasonlóan, a $v \in V$ csúcs kimenő fokszáma (amely a v -ből kimenő élek száma) legyen $\deg^+(v)$.

A következő két tétel a $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ irányított gráf egy jellemzését adják. Be-látjuk, hogy a gráf paraméterei szoros kapcsolatban állnak egymással és $\mathbf{V}_{k,l}(\delta)$ bármelyik csúcsának be- és kimenő foksámával.

3.7. Tétel. Tekintsük a $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ irányított gráfot. Ekkor bármely $v \in \mathbf{V}_{k,l}(\delta)$ csúcsra:

$$\deg^+(v) \leq \left\lfloor \frac{\alpha + \beta}{\delta + 1} \right\rfloor + 1$$

áll fenn.

Bizonyítás. A bizonyítás lelke a 3.1-es lemma megfelelő alkalmazása. A következők-ben tekintsünk egy $v \in \mathbf{V}_{k,l}(\delta)$ csúcsot.

Először is feltesszük, hogy $\alpha + \beta < \delta + 1$. A 3.1-es lemma alapján tudjuk, hogy két δ -csúcs legalább $\delta + 1$ nap távolságra van egymástól. Ha $\alpha + \beta < \delta + 1$, akkor legfeljebb egy δ -csúcs lehet egy $\alpha + \beta$ méretű időablakon belül (mivel az ablak mérete kisebb, mint két δ -csúcs közötti minimális távolság). Ez azt jelenti, hogy v -ből legfeljebb egy kimenő élünk lehet, tehát

$$\deg^+(v) \leq 1 = \left\lfloor \frac{\alpha + \beta}{\delta + 1} \right\rfloor + 1$$

érvényes.

Most tekintsük az $\alpha + \beta \geq \delta + 1$ esetet. Tegyük fel, hogy $\alpha + \beta = (Q - 1)(\delta + 1) + R$, ahol $Q, R \in \mathbb{N}_0$, $Q \geq 2$ és $0 \leq R < \delta + 1$. Ebből adódik, hogy legfeljebb Q darab δ -csúcs fér bele egy $\alpha + \beta$ méretű időablakba, azaz

$$\begin{aligned} Q - 1 &\leq \frac{\alpha + \beta}{\delta + 1} \\ Q &\leq \frac{\alpha + \beta}{\delta + 1} + 1 \\ Q &\leq \left\lfloor \frac{\alpha + \beta}{\delta + 1} \right\rfloor + 1 \end{aligned}$$

következik, mivel Q egy egész szám. Ha a v csúcs esetén az $\alpha + \beta$ méretű időablakban talált δ -csúcsok száma Q , kimenő foksám megegyezik ezzel az értékkel, így

$$\deg^+(v) = Q \leq \left\lfloor \frac{\alpha + \beta}{\delta + 1} \right\rfloor + 1$$

teljesül. ■

A gráfnak ezen a tulajdonsága szimmetrikus abban az értelemben, hogy ez a felső korlát a $\mathbf{V}_{k,l}(\delta)$ tetszőleges csúcsának kimenő foksámának is felső korlátja.

3.8. Tétel. *Tekintsük $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ irányított gráfot. Ekkor bármely $v \in \mathbf{V}_{k,l}(\delta)$ csúcsra:*

$$\deg^-(v) \leq \left\lfloor \frac{\alpha + \beta}{\delta + 1} \right\rfloor + 1$$

fennáll.

Bizonyítás. Itt ugyanazt a logikát használjuk, mint az előző tétel bizonyításánál. A fontos megfigyelés, hogy ha az $u_1, u_2, \dots, u_n$ csúcsokból megy él v -be, akkor

$$-\beta \leq \text{dt}(u_i) - \text{dt}(v) \leq \alpha, \quad \forall i \in \mathbb{N}_n, \quad (1)$$

mivel $(u_i, v) \in \mathbf{E}_{k,l}(\alpha, \beta, \delta)$ minden $i \in \mathbb{N}_n$ esetén. Nyilvánvaló, hogy a $\{u_i\}_{i \in \mathbb{N}_n}$

csúcsok ugyanazon állomáshoz tartoznak, ahol az egymást követő δ -csúcsok legalább $\delta + 1$ nap távolságra vannak egymástól, ahogy azt a 3.1-es lemmában is láttuk. Következésképpen, ha u_1^* a legkorábbi dátummal rendelkező csúcs, és hasonlóan u_n^* a legkésőbbi dátummal rendelkező csúcs, akkor $\Delta = dt(u_n^*) - dt(u_1^*) \geq (n-1)(\delta+1)$. Továbbá Δ nem lehet nagyobb, mint $\alpha + \beta$, mivel az 1. egyenletből következik, hogy

$$\begin{aligned} dt(u_n^*) - dt(v) &\leq \alpha \\ -dt(u_1^*) + dt(v) &\leq \beta \end{aligned}$$

így $\Delta = dt(u_n^*) - dt(u_1^*) \leq \alpha + \beta$. Összefoglalva,

$$(\deg^-(v) - 1)(\delta + 1) = (n - 1)(\delta + 1) \leq \Delta \leq \alpha + \beta,$$

Innen a tétel állítása átrendezéssel kapható. ■

3.2. Karakterisztikus gráfok és utak

A gráf megkonstruálása után a gráf különböző komponenseit fogjuk vizsgálni. Ehhez vezessük be a következő jelöléseket.

3.9. Definíció. Tekintsünk két állomást, s -t és s' -t. A $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ gráfban a *karakterisztikus út* egy olyan p irányított út, azaz $p = (u_1, u_2, \dots, u_n)$ csúcsok sorozata s és s' között, amelyre $s(u_1) = s$ és $s(u_n) = s'$ teljesül, továbbá minden szomszédos csúcs rákövetkező állomásnál van, formálisan $s(u_j) \rightsquigarrow s(u_{j+1})$ minden $j \in \mathbb{N}_{n-1}$ esetén.

3.2. Lemma. Tekintsük az $s, s' \in \mathbf{S}$, $s \neq s'$ állomásokat. A $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ irányított gráfban tetszőleges irányított út az s és s' állomások között egy karakterisztikus út.

Bizonyítás. Vegyük észre, hogy u -ból v -be csak akkor húztunk be élt, ha $s(u) \rightsquigarrow s(v)$ fennáll, tehát a v az u -hoz képest a rákövetkező állomáshoz tartozik. Ez azt jelenti, hogy nincsenek "visszafelé" mutató élek, vagyis az élek a folyó irányát követik. Ezzel ekvivalens módon, ha veszünk egy $(u_1, u_2, \dots, u_n)$ irányított utat, akkor

$$s(u_1) \rightsquigarrow s(u_2), s(u_2) \rightsquigarrow s(u_3), \dots, s(u_{n-1}) \rightsquigarrow s(u_n)$$

áll fenn. ■

3.9.1. Következmény. $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ -ben tetszőleges irányított út hossza legfeljebb $K - 1$ lehet, ahol K a felhasznált állomások száma.

Az irányított gráfok területén fontos szerepet játszanak az irányított körmentes gráfok, és az előző tulajdonságok alapján megállapítható, hogy a fenti gráf a gráfok ezen osztályába tartozik.

3.10. Definíció. A G irányított gráf egy *irányított körmentes gráf* — röviden DAG (directed acyclic graph) — ha G nem tartalmaz irányított kört.

3.11. Tétel. $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ egy irányított körmentes gráf.

Bizonyítás. Meg kell mutatnunk, hogy a $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ irányított gráf nem tartalmaz irányított kört. Tekintsünk egy tetszőleges $(u_1, u_2, \dots, u_n)$ irányított utat a $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ -ban. Az előző állításokból kapjuk, hogy

$$s(u_1) \rightsquigarrow s(u_2), s(u_2) \rightsquigarrow s(u_3), \dots, s(u_{n-1}) \rightsquigarrow s(u_n),$$

továbbá tudjuk, hogy

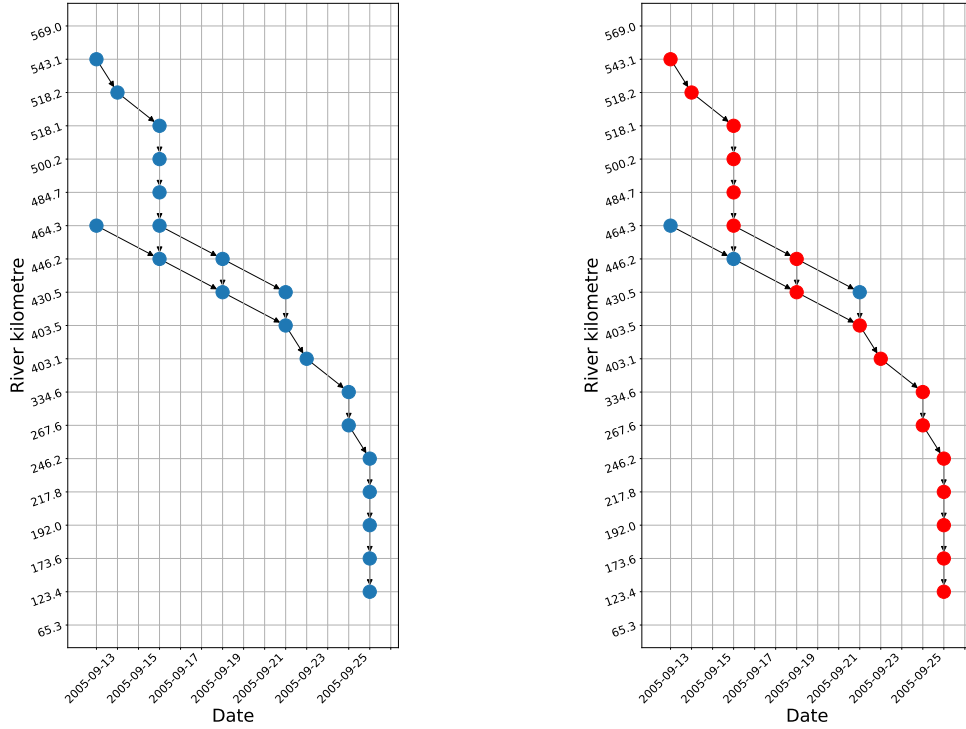
$$s(u_1) \neq s(u_2) \neq s(u_3) \neq \dots \neq s(u_n),$$

mivel $(u_1, u_2, \dots, u_n)$ egy karakterisztikus út. Ez azt jelenti, hogy nincs olyan irányított út, amely kétszer ugyanazt a csúcsot tartalmazza, tehát nem lehet irányított körünk. Ebből adódik, hogy $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ egy irányított körmentes gráf. ■

3.12. Definíció. A $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ egy tetszőleges gyengén összefüggő komponensét *karakterisztikus gráfnak* nevezzük.

A következő fejezetben a karakterisztikus gráfra úgy fogunk hivatkozni, mint $C = (V, E)$ és egy $p \subseteq C$ karakterisztikus út csúcshalmazát $V(p)$ fogja jelölni.

Világos, hogy a 9. ábrán látható p nem az egyetlen karakterisztikus út, amely a legfelső és a legalsó csúcsokat összeköti, ezzel a jelenséggel a következő fejezetben foglalkozunk. Ez a gráf a folyóra gyakorolt globális hatások elemzésére alkalmas rövidebb időintervallumot figyelembe véve, mivel sok különböző kiinduló-csúcsból származó, egymást metsző karakterisztikus utat tartalmazhat.



9. ábra. Ez a $\mathbf{G}_{k,l}$ $(0, 3, 2)$ részgráfja, ahol k a 2005-09-12 dátum indexének, l a 2005-09-26 dátum indexének felel meg. Balra egy karakterisztikus gráf, jobbra pedig egy pirossal jelölt p karakterisztikus út látható

4. Árhullámok és statisztikai elemzés

Az árhullámokkal kapcsolatos kérdések megválaszolásához először a gyakorlatban is fell kell építenünk a gráfot. A következő bekezdésekben röviden összefoglaljuk az implementáció részleteit, majd a gráfon számolandó metrikákra/statisztikai mutatókra javasolt koncepciókkal foglalkozunk. Látni fogjuk, hogy a gyakorlati megközelítés általában a formális konstrukciótól eltérő megoldások kíván.

4.1. Implementáció

Az előző fejezetekben vizsgált teljes folyamatot Pythonban, az elejétől indulva valósítottuk meg. A gráfot felépítő keretrendszer a következő lépésekből áll:

0. Az α, β, δ értékek, illetve az érvényes időablak végpontjainak rögzítése.
1. Napi idősor-adatok betöltése az előbb meghatározott időintervallumra.
2. δ -csúcsok keresése minden egyes állomáshoz (a rögzített δ értéke alapján), majd a hozzájuk tartozó dátumok tárolása. Ennek eredményeként kapjuk a csúcsok halmazát.

3. Egymást követő állomásokhoz tartozó csúcsok közötti kapcsolat keresése (a megadott α és β értékek alapján), majd az állomáspáronként csoportosított csúcspárok tárolása. Ettől a ponttól kezdve elhagyjuk az összes olyan csúcsot, amelynek be- és kimenő fokszáma nulla.

Az adatokat az ATIVIZIG általában CSV formátumú fájlokban bocsátotta rendelkezésünkre, amelyekből egy SQL adatbázist hoztunk létre. Mivel az adatok több napi rögzített értéket tartalmaznak 1990-től, szűrünk kell az adatokat, hogy olyan táblázatot kapjunk, ahol minden dátumhoz és állomáshoz legfeljebb egy érték áll rendelkezésre. A csúcsok és élek keresését klasszikus Python csomagokkal (`numpy`, `pandas`) valósítottuk meg, majd a gráf ábrázolásához és vizualizálásához a `networkx` könyvtárat használtuk.

A kódot objektumorientált módon írtuk meg a könnyebb karbantartás és bővíthetőség érdekében. Az algoritmusok futtatására és a grafikonok létrehozására `Jupyter` notebook környezetet használunk, ezért létrehoztunk egy kódbázist, amely tartalmazza a különböző osztályok implementációit és a `Jupyter` notebook-okat. A projekt felépítése a 10. ábrán látható.

```
flood-wave-detector
├── generated
│   ├── build_graph
│   ├── find_edges
│   └── find_vertices
├── notebooks
│   └── flood_wave_detector.ipynb
└── src
    ├── analysis.py
    ├── database.ini
    ├── flood_wave_data.py
    ├── flood_wave_detector.py
    ├── flood_wave_handler.py
    ├── gauge_data.py
    ├── graph_builder.py
    ├── json_helper.py
    └── plotter.py
```

10. ábra. A kódbázis struktúrája. A `generated` mappa és annak almappái a szkriptek által generált dictionary-eket és fájlokat tartalmazzák, hogy a későbbi elemzés során futási időt takarítsunk meg a betöltésekkel. A `notebooks` mappa tartalmazza a futtatáshoz és elemzéshez létrehozott `Jupyter` notebook-okat, végül a `src` mappa tartalmazza az összes osztály implementációját és egy `database.ini` fájlt, amely az adatbázis eléréséhez szükséges hitelesítő adatokat tartalmazza.

Mivel a legtöbb statisztikai elemzésben a karakterisztikus utakat és gráfokat használjuk, fel kell építenünk az utakat, és ezek unióját kell képeznünk, hogy megkapjuk

azokat részgráfokat, amelyek a számítások alapjául szolgálnak. A karakterisztikus utak felépítését a nulláról programoztuk le, `networkx` kizárólag arra szolgál, hogy az eredményt gráfként tároljuk a memóriában. Előfordulhat bizonyos paraméterezés esetén, hogy egy csúcshoz egynél több kifelé vezető élt is lehet találni, ilyenkor egy hálószerű struktúrát vesz fel a gráf. Ebből az okból kifolyólag szükségünk van egy olyan algoritmusra, amely nem hagy ki egyetlen utat sem. Erre a célra mélységi keresést alkalmaztunk.

4.2. Mélységi Bejárás

A mélységi keresés, angolul Depth-First Search algoritmus — röviden DFS — alapvetően a leghosszabb utat keresi egy gráfban, és mivel az egész gráfot átvizsgálja, az összes lehetséges utat megtalálhatjuk vele. Az eredeti DFS algoritmus egy gráfon fut, ahol minden egyes csúcs jelölésére egy színezési sémát alkalmazunk, amely elkülöníti egymástól a még fel nem fedezett, az aktuális iterációban felfedezett vagy korábban felfedezett állapotokat. Ebből a célból minden egyes csúcsot kibővítünk egy további attribútummal a szín jelölésére. Egy csúcs esetében $u = (d_t, s, c)$, ahol c az adott csúcs színe, és amikor $\text{color}(u)$ -t írunk, akkor c -re hivatkozunk. Minden csúcs fehér színű az algoritmus legelején, majd amikor azt írjuk, hogy $\text{color}(u) = \text{BLACK}$, akkor azt értjük, hogy az u csúcsot feketére színezzük.

Ezt a bejáró logikát alkalmazzuk a karakterisztikus utak felépítésére, majd unióképzéssel kapjuk a gyengén összefüggő komponenseket, azaz a karakterisztikus gráfokat. Megjegyezzük, hogy ez nem az eredeti DFS algoritmus, mivel eredetileg egy rögzített élhalmazzal rendelkező gráfon van futtatva, mi csak kihasználjuk, hogy bejárja a gráfot. A következő pszeudokód foglalja össze ezeket az ötleteket.

Algorithm 1 Graph building with DFS logic

Require: $\mathbf{V}_{k,l}(\delta)$ and α, β fixed

```

1: for each vertex  $u \in \mathbf{V}_{k,l}(\delta)$  do
2:    $\text{color}(u) = \text{WHITE}$ 
3: end for
4:  $\mathbf{E} = \emptyset$ 
5: for each vertex  $u \in \mathbf{V}_{k,l}(\delta)$  do
6:   if  $\text{color}(u) == \text{WHITE}$  then
7:     DFS-Visit( $\mathbf{G}(\mathbf{V}_{k,l}(\delta), \mathbf{E}), u$ )
8:   end if
9: end for

```

Algorithm 2 DFS-Visit

Require: $G(V, E)$, and $u \in V$

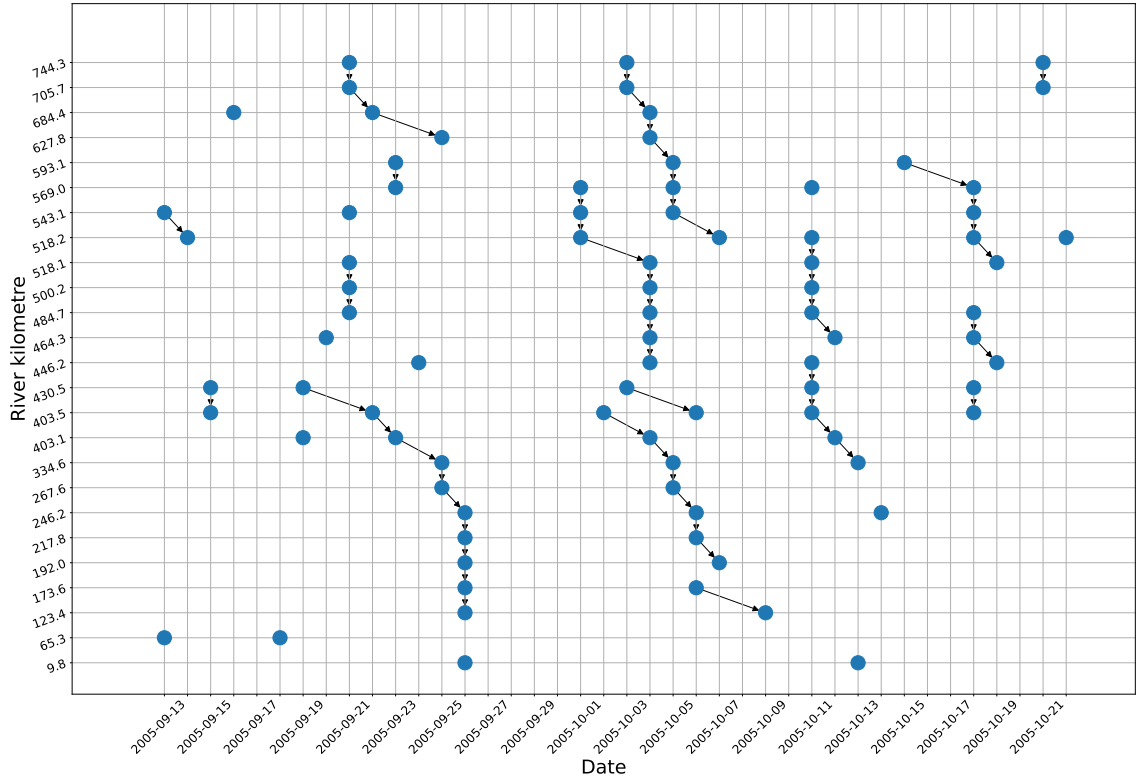
```
1: color( $u$ ) = GRAY
2: Adj[ $u$ ] =  $\{w \in \mathbf{V}_{k,l}(\delta) : s(u) \rightsquigarrow s(w), -\alpha \leq \text{dt}(w) - \text{dt}(u) \leq \beta\}$ 
3: for each vertex  $v \in \text{Adj}[u]$  do
4:    $\mathbf{E}.\text{append}((u, v))$ 
5:   if color( $v$ ) == WHITE then
6:     DFS-Visit( $\mathbf{G}(\mathbf{V}_{k,l}(\delta), \mathbf{E}), v$ )
7:   end if
8: end for
9: color( $u$ ) = BLACK
```

A 2. algoritmusban, egy adott $\mathbf{E}$ élhalmaz esetén, az $\mathbf{E}.\text{append}((u, v))$ művelet az (u, v) él hozzáadását jelenti a jelenlegi $\mathbf{E}$ élhalmazhoz.

A színezési séma helyett a mi implementációnkban egy verem (másképp Last-In First-Out, azaz LIFO) adatszerkezetet használunk az elért elágazási pontok tárolására. Az eredeti DFS algoritmus és a verem részletesebb leírása a [2] könyvben található.

4.3. Metrikák a statisztikai elemzéshez

A Tisza folyó árhullámokkal kapcsolatos viselkedésével vonatkozó kérdésekre szeretnénk választ adni a korábban tárgyalt grafikus ábrázolás segítségével. Konkrétan az árhullámok számát és egy adott időintervallumra számított átlagos levonulási időt elemezzük különböző paraméterbeállítások mellett. Mivel a helyi igazgatósággal való együttműködés ebben a témában 2022-ben kezdődött, az alábbi elképzelések kiindulópontot adnak a további munkához, és fejlesztésre szorulnak.



11. ábra. $G_{k,l}(0,3,3)$ gráfja, ahol k a 2005-09-12 dátum indexének, l a 2005-10-21 dátum indexének felel meg. Mivel $\alpha + \beta < \delta + 1$ fennáll, tetszőleges csúcsnak legfeljebb egy bemenő és egy kimenő éle lehet. Ez jelentősen megszorítja a 7. ábrán látható gráfot. A hasonló paraméterezések ezt az egyszerű struktúrát eredményezik, amely könnyen kezelhető és vizsgálható, de sok különleges esetet kizár, így nem korlátozhatjuk a vizsgálatot kizárólag ilyen paraméterbeállítások használatára.

Először is formalizáljuk, mit értünk árhullám alatt. A mi megfontolásunkban az s és $\hat{s}$ állomások közötti árhullám alatt egy adott karakterisztikus gráf olyan részgráfjával lesz kapcsolatos, amelynek mind az s , mind a $\hat{s}$ állomásoknál van csúcsa. A 9. ábrán láthatjuk, hogy vannak olyan csúcsok, amelyekhez karakterisztikus út ugyanabból a kezdeti csúcsból vezet. Mivel a 3.7 és 3.8 tételekből levezetett $\alpha + \beta \geq \delta + 1$ teljesül, ez az eset fennállhat. Ezt a jelenséget, amikor egy karakterisztikus gráf egyes csúcsainál több be- vagy kimenő élt kapunk, a következőkben *keresztkapcsolódásnak* fogjuk nevezni. Nyilvánvaló, hogy ha $\alpha + \beta \geq \delta + 1$ teljesül akkor előfordulhat keresztkapcsolódás (lásd a 7. ábrát), ha viszont $\alpha + \beta < \delta + 1$ fennáll, akkor nem lehet keresztkapcsolódás, ahogy ezt a 11. ábra is mutatja. Egy kijelölt időintervallumban egy s kiinduló állomáshoz tartozó csúcsokat fogjuk tekinteni és minden egyes csúcsból megszámloljuk a $\hat{s}$ beérkezési állomáshoz talált csúcsokhoz vezető utakat, de kezelni akarjuk azokat az eseteket is, amikor több útvonal köti össze ugyanazt a kiinduló és beérkezési állomást. A következőkben a formális vizsgálatokhoz szükséges jelöléseket vezetjük be.

4.1. Definíció. Tekintsük az $s, \hat{s} \in \mathbf{S}$ állomásokat és a $C = (V, E) \subseteq \mathbf{G}_{k,l}(\alpha, \beta, \delta)$ karakterisztikus gráfot. A $(s, \hat{s})$ -megszorított karakterisztikus utak halmaza C -ben a következő:

$$P(C|s, \hat{s}) := \left\{ p \subseteq C \left| \begin{array}{l} p \text{ egy karakterisztikus út} \\ v, \hat{v} \in V(p) \text{ olyanok, hogy } s(v) = s \text{ és } s(\hat{v}) = \hat{s} \end{array} \right. \right\}$$

és az összes, $C \subseteq \mathbf{G}_{k,l}(\alpha, \beta, \delta)$ -re vett $(s, \hat{s})$ -megszorított karakterisztikus út halmazának halmazát $\mathbf{P}(s, \hat{s})$ -vel jelöljük, formálisan

$$\mathbf{P}(s, \hat{s}) := \{P(C|s, \hat{s}) : C \text{ egy karakterisztikus út } \mathbf{G}_{k,l}(\alpha, \beta, \delta) \text{-ben}\}$$

Megjegyezzük, hogy $P(C|s, \hat{s})$ azon karakterisztikus utak halmazát jelöli a C karakterisztikus gráfban, amely tartalmaz egy-egy csúcsot az s illetve az $\hat{s}$ állomásnál, és $\mathbf{P}(s, \hat{s})$ pedig az összes ilyen halmazt tartalmazza $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ -re nézve.

Az árhullám alatt nem csak egy karakterisztikus utat értünk (mivel több útvonal is összekötheti ugyanazt a csúcspárt), ezért az alábbi fogalmakra van szükségünk az árhullám pontos meghatározásához.

4.2. Definíció. Tekintsük a $p, \hat{p} \in P(C|s, \hat{s})$ karakterisztikus utakat, ahol $p = (v_1, v_2, \dots, v_n)$ és $\hat{p} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n)$. Azt mondjuk, hogy

$$p \sim \hat{p} \iff v_1 = \hat{v}_1 \text{ és } v_n = \hat{v}_n$$

Megjegyezzük, hogy ez egy ekvivalenciareláció, mivel szimmetrikus, reflexív és tranzitív. Továbbá, a következő jelöléseket vezetjük be

- $P(C|s, \hat{s}) / \sim$ jelöli $P(C|s, \hat{s})$ faktorhalmazát (ez a halmaz tartalmazza a $\sim$ reláció által meghatározott ekvivalenciaosztályokat)
- $F_{u,v}$ jelöli egy adott $p = (u, \dots, v)$ karakterisztikus út osztályát $P(C|s, \hat{s}) / \sim$ halmazban.

4.3. Definíció. Tegyük fel, hogy $F_{u,v}$ egy eleme $P(C|s, \hat{s}) / \sim$ faktorhalmaznak. Ez az elem egy *árhullám* az s és $\hat{s}$ állomások között a kijelölt időszakban.

4.4. Definíció. Tekintsük a $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ irányított gráfot és $s, \hat{s} \in \mathbf{S}$ állomásokat, úgy hogy $s \neq \hat{s}$. Azt mondjuk, hogy (egy d_k és d_l által meghatározott időintervallumra) az *árhullámok száma* az s és $\hat{s}$ állomások között a következő:

$$\sum_{P(C|s, \hat{s}) \in \mathbf{P}(s, \hat{s})} |P(C|s, \hat{s}) / \sim|$$

Világos, hogy ezzel a definícióval helyesen tudjuk megszámolni a kiindulási-beérkezési csúcspárokat, mivel az ugyanazon pár közötti többszörös útvonalakat nem számoljuk többször az összegzésben. Mint már említettük, a keresztkapcsolódás felépése esetén nagy mennyiségű árhullámot kapunk, ha a δ -csúcsok valóban sűrűn helyezkednek el (ami például $\delta = 1$ esetén elérhető), vagy ha az él behúzásához használt ablak $\alpha + \beta$ mérete túl nagy (lásd 9. ábra). Vegyük észre, hogy minél kisebb a δ -csúcs sugara, annál nagyobb az árhullámok száma, ami összhangban van az elvártakkal.

Amint azt a bevezetőben is tárgyaltuk, az árhullámok levonulási ideje egy másik hasznos mérőszám lehet a folyószabályozás szempontjából. Akárcsak az árhullámok számának esetében, a következőkben javasolt megközelítés is csak egy kezdőlépés, további egyeztetésekkel finomítható. Az átlagos (napokban mért) időt kell kiszámolni, amely alatt az árhullám eljutott egy adott s állomásról az $\hat{s}$ állomásig. Most ki kell zárunk azokat az eseteket, ahol negatív levonulási idő előfordulhat — tehát minden olyan paraméterbeállítást, ahol $\alpha > 0$ —, mivel ez nem értelmezhető ebben a megközelítésben.

A korábban bevezetett fogalmakat alkalmazva a $\mathbf{P}(s, \hat{s})$ halmaz minden egyes P halmazára az $\hat{s}$ és s állomásokhoz tartozó csúcsok közötti időbeli eltéréseket átlagoljuk a $P/\sim$ egyes ekvivalenciaosztályaira, amit a következő definícióban foglalunk össze.

4.5. Definíció. Tekintsük a (k, l) érvényes időablakot, a $\beta, \delta \in \mathbb{N}$ egészeket és a hozzájuk tartozó $\mathbf{G}_{k,l}(0, \beta, \delta)$ irányított gráfot. Az s és $\hat{s}$ állomások közötti árhullámok átlagos levonulási ideje $\mathbf{G}_{k,l}(0, \beta, \delta)$ -ben a következő:

$$\frac{1}{|\mathcal{P}(s, \hat{s})|} \sum_{P \in \mathbf{P}(s, \hat{s})} \left(\sum_{F_{u,v} \in P/\sim} (\text{dt}(v) - \text{dt}(u)) \right)$$

ahol $s(u) = s$ és $s(v) = \hat{s}$, továbbá

$$\mathcal{P}(s, \hat{s}) = \bigcup_{P \in \mathbf{P}(s, \hat{s})} P/\sim.$$

Látható, hogy az előző definíció minden árhullámot egyszer számol az átlagolásban, de azt is érdemes lehet figyelembe venni, hogy hány útvonal van az adott árhullámnak megfelelő ekvivalenciaosztályban. Ez az ötlet a következőképpen formalizálható.

4.6. Definíció. Tekintsük $\mathbf{G}_{k,l}(0, \beta, \delta)$ irányított gráfot. Az s és $\hat{s}$ állomások közötti

árhullámok *súlyozott átlagos levonulási ideje* $\mathbf{G}_{k,l}(0, \beta, \delta)$ -ben a következő:

$$\frac{1}{|\mathcal{F}(s, \hat{s})|} \sum_{P \in \mathbf{P}(s, \hat{s})} \left(\sum_{F_{u,v} \in P/\sim} \left(|F_{u,v}| \cdot (\text{dt}(v) - \text{dt}(u)) \right) \right)$$

ahol $s(u) = s$ és $s(v) = \hat{s}$, továbbá

$$\mathcal{F}(s, \hat{s}) = \bigcup_{P \in \mathbf{P}(s, \hat{s})} \left(\bigcup_{F \in P/\sim} F \right).$$

Megjegyzendő, hogy a 4.5-ös és 4.6-os definíciók ugyanazt az eredményt adják, ha minden ekvivalenciaosztályban csak egy árhullám van, például ha nincs kereszt-kapcsolódás. A következőkben mindkét definícióra vonatkozó számolást elvégezzük és összevetjük egymással.

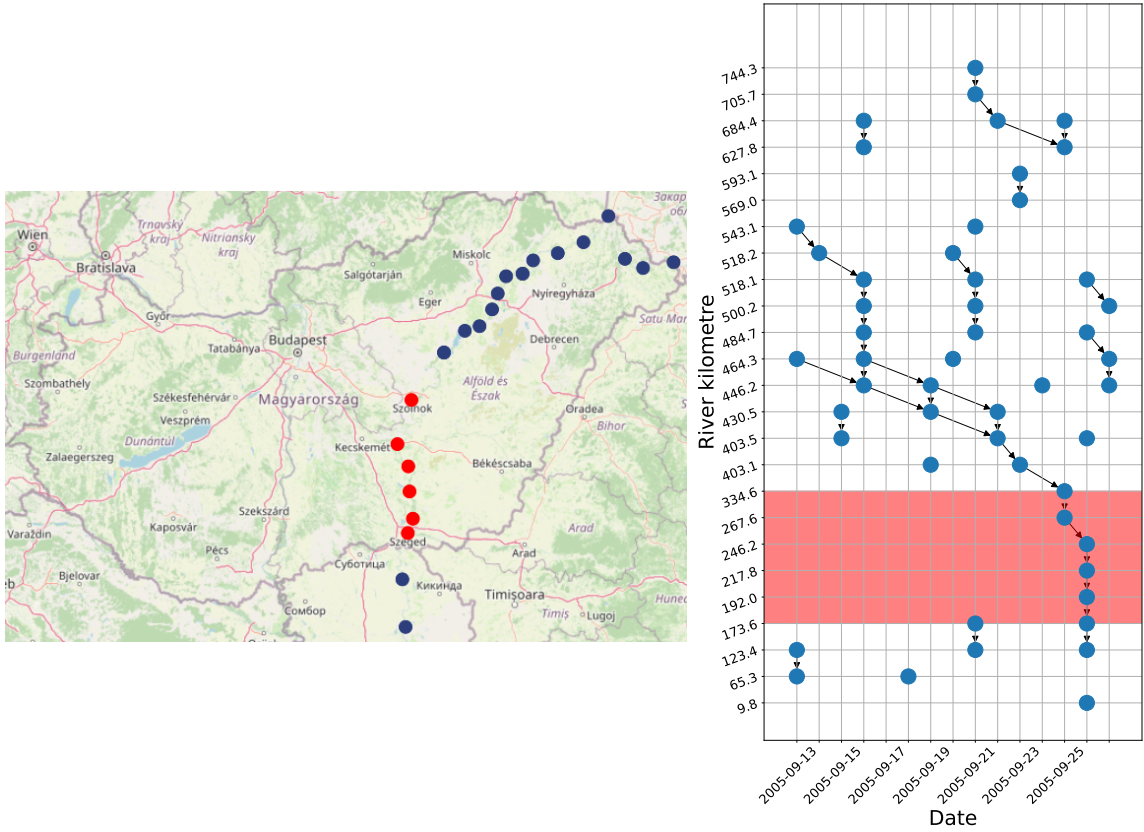
4.4. Eredmények

A továbbiakban a Szolnok és Szeged közötti folyószakaszt vizsgáljuk (amely két szomszédos vízügyi igazgatóság központját köti össze), az említett folyószakasz a következő oldalon látható a 12. ábrán

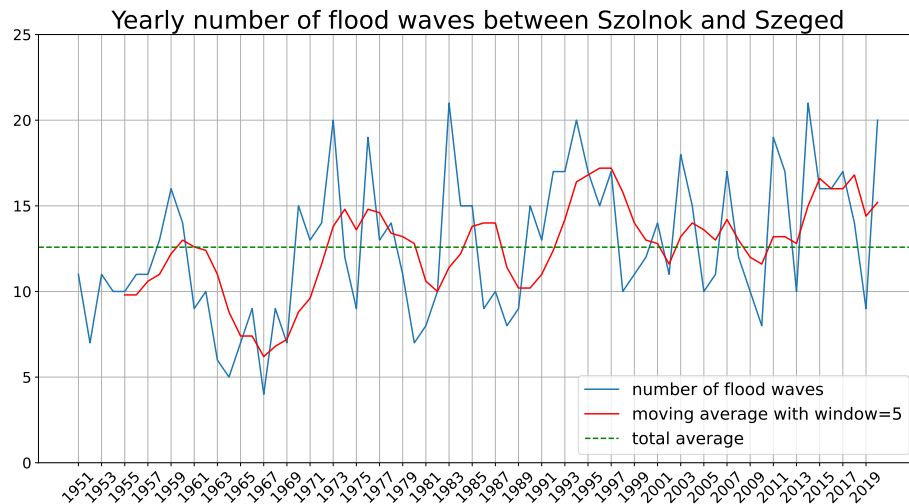
Megjegyezzük, hogy ha a fent bevezetett mennyiségeket a teljes folyószakaszra számolnánk, ha bármelyik állomás adatai hiányosak lennének, akkor a számítások nem lennének relevánsak. Például az árhullámok éves száma esetén egy adott állomás adatai között egy évnnyi hiány van — ami a gyakorlatban előfordulhat —, akkor az adott év árhullámainak száma nulla lenne, még akkor is, ha az adott állomás alatt és felett is fel tudjuk építeni a gráfot. Tapasztalatból tudjuk, hogy a Szolnok és Szeged közötti folyószakaszon viszonylag kevés adat hiányzik, így számításaink pontosabbak és megbízhatóbbak.

Fontos megemlíteni, hogy az alábbi elemzés a képek értelmezésére koncentrál, és most nem kíséreljük meg hidrológiai szempontból megmagyarázni az érdekes jelenségek vagy globális tendenciák kiváltó okait (ezt a közeljövőben a vízügyi igazgatóság szakembereivel közösen fogjuk elvégezni).

Az árhullámok számának vagy a terjedési időnek a szemléltetésének egyik módja egy éves összesített idősor létrehozása. Ekkor rögzített α, β, δ értékekre vesszük a $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ gráfot évente és kiszámítjuk az árhullámok számát minden egyes gráfra. Az idősoros ábrázolással betekintést nyerhetünk abba, hogy az árhullámok száma vajon nőtt-e az idő elteltével. A jobb olvashatóság kedvéért az x tengelyen csak minden második beosztást rajzoltunk be, de az adatokat évente ábrázoltuk.

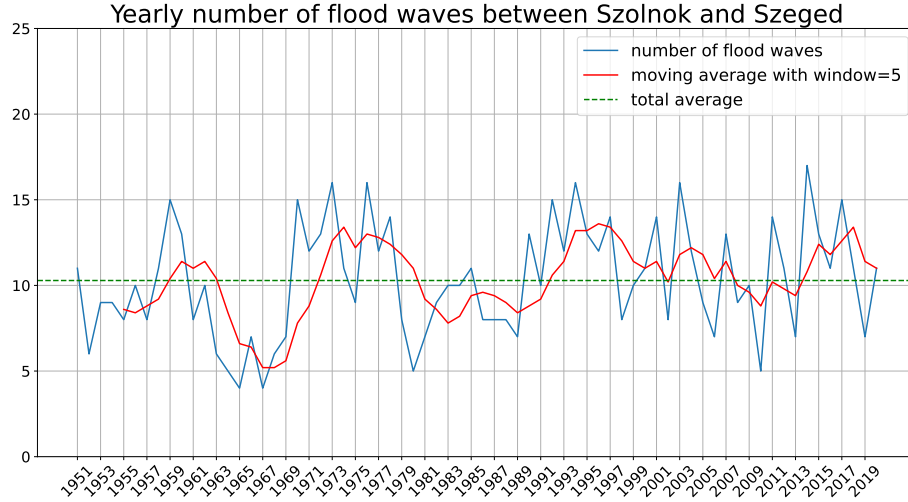


12. ábra. A Szolnok és Szeged közötti folyószakasz pirossal kiemelve. A következő elemzéshez csak ennek a 6 állomásnak az adatait használjuk fel. Ez jelentősen csökkenti a futási időt, mivel számításainkat csak $\mathbf{G}_{k,l}(\alpha, \beta, \delta)$ piros sávba eső részén kell elvégeznünk.



13. ábra. A Szolnok és Szeged állomások közötti árhullámok éves száma a $\mathbf{G}_{k,l}(0, 3, 2)$ -ben számítva, ahol k az adott év első dátumának, l pedig az adott év utolsó dátumának indexe. A zöld szaggatott vonal az árhullámok éves számának egész időszakra vett átlagát mutatja.

Láthatjuk, hogy az elmúlt 30 évben az árhullámok száma nem csökkent jelentősen a mindenkori átlag alá. Ha megnézzük a piros vonalat, amely a mozgóátlagot mutatja 5 év méretű időablakkal, akkor az elmúlt 12 évben növekvő tendenciát és az egész időszakra pedig emelkedő trendet olvashatunk le.

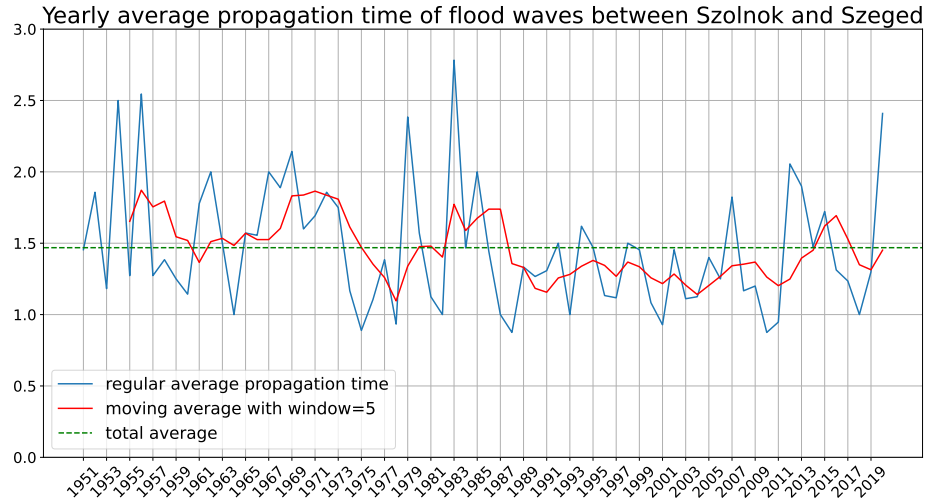


14. ábra. A Szolnok és Szeged állomások közötti árhullámok éves száma $\mathbf{G}_{k,l}(0, 3, 3)$ -ban számítva, ahol k az adott év első dátumának, l pedig az adott év utolsó dátumának indexe. A zöld szaggatott vonal az árhullámok éves számának egész időszakra vett átlagát mutatja.

Ha összehasonlítjuk a 13. és 14. ábrákat, láthatjuk hogy a megfelelő görbék alakja nagyon hasonló, de összességében nagyobb δ esetén kevesebb árhullámot kapunk, ami összhangban van az elvárattal. Az alaki hasonlóságok miatt itt arra következtethetünk, hogy ennél a paraméterbeállításnál még mindig növekvő tendenciákat látunk, ami úgy tűnik, hogy összhangban van a [4] tanulmány megállapításaival, de ezt a hipotézist a későbbiekben további elemzéssel fogjuk alátámasztani.

A 15. ábrán megfigyelhető, hogy az elmúlt 30 évben az éves átlagos levonulási idők csak egyszer haladták meg az egész időszakra vett átlagot. Bár közel van az átlaghoz, mégis azt mondhatjuk, hogy az elmúlt 30 évben csökkent az egész időszakra vett átlag. Ha $\delta = 3$, akkor kevesebb δ -csúcsot vizsgálunk, de ezekre szigorúbb feltételek vonatkoznak. A 16. ábrán látható, hogy a 15. ábrához képest sok szélső érték eltűnt vagy lecsökkent, ami alacsonyabb összesített átlagot és kevesebb nagy ingadozást eredményezett. Ezen az ábrán a megfelelő mozgóátlagon jól látható, hogy ezen paraméterbeállítások mellett a levonulási idők átlaga nagyon stabil volt 2015-ig, majd az árhullámok néhány évig lassabbá váltak.

Ahogy láttunk (és vártunk is) különbségeket az eltérő paraméterezések esetén, meg akartuk vizsgálni a paraméterek változásának hatását, ezért szimulációkat fut-



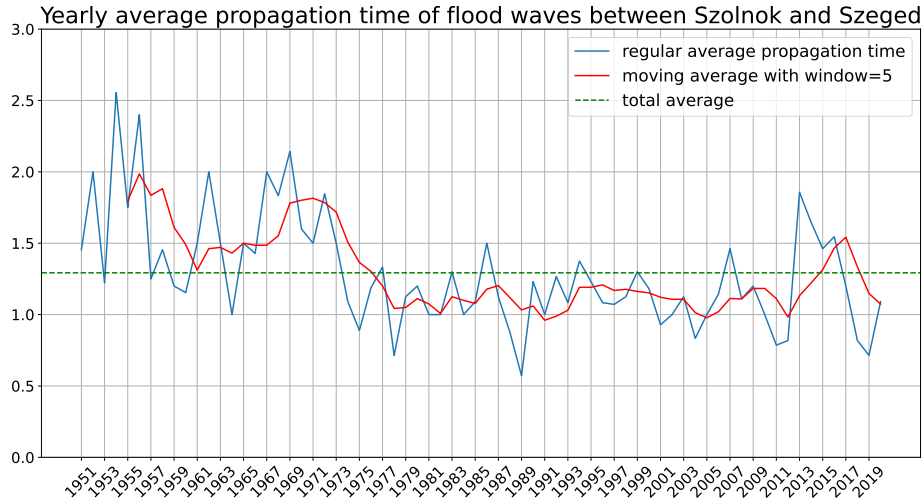
15. ábra. Az árhullámok éves átlagos levonulási ideje Szolnok és Szeged állomások között a $G_{k,l}(0,3,2)$ -ben számítva, ahol k az adott év első dátumának, l pedig az adott év utolsó dátumának indexe. A számítás a 4.5-ös definícióban ismertetett koncepciót használja. A zöld szaggatott vonal az árhullámok éves levonulási idejének egész időszakra vett átlagát mutatja.

tattunk a paraméterek szisztematikus változtatására. Természetesen vannak haladóbb technikák is egy-egy paraméter rendszerre vett érzékenységeinek vizsgálatára, most kiindulópontként egy egyszerű megközelítést használtunk. Az ábrákhoz az adatokat a következő két intervallumra osztottuk.

1. 1951.01.01.-től 1989.12.31.-ig,
2. 1990.01.01.-től 2020.12.31.-ig

A felosztás egyik oka, hogy bár a keletkezett intervallumok hossza nem azonos, azonban 1990 körül kezdték meg működésüket az első digitális adatrögzítő eszközök, amelyeket a folyó mentén telepítettek. Ez azt jelenti, hogy az mérések gyakoribbak, megbízhatóbbak és kevésbé valószínű, hogy $\emptyset$ értékeket tartalmazó hiányosságok fordulnak elő, ezért ez technológiai szempontból egy új korszakot jelent. A másik ok, hogy szeretnénk összehasonlítani a múltbeli és a közelmúltbeli mérőszámokat, és megnézni, hogy levonható-e valamilyen következtetés a változásukra vagy értékükre vonatkozóan.

Elemzésünkhöz hőterképeket készítünk, ahol az x tengely a különböző β értékeket, az y tengely a különböző α értékeket jelöli és külön hőterképeket készítünk a különböző δ értékekre, amelyeket az ábra címében jelölünk. Ez utóbbi szétválasztás lehetővé teszi, hogy a paramétertérben egy tégltest egyes rácspontjaiból vizsgálhassuk a célértéket (most az árhullámok számát) minden paraméterkombinációra.



16. ábra. Az árhullámok éves átlagos levonulási ideje Szolnok és Szeged állomások között, a $\mathbf{G}_{k,l}(0, 3, 3)$ -ben számítva, ahol k az adott év első dátumának, l pedig az adott év utolsó dátumának indexe. A számítás a 4.5-ös definícióban ismertetett koncepciót használja. A zöld szaggatott vonal az árhullámok éves levonulási idejének egész időszakra vett átlagát mutatja.

Példaként az A.1. ábrán látható hőtérképet tekintve a második sor, harmadik oszlop cellájának értéke a Szolnok és Szeged állomások közötti árhullámok számát jelenti 1951.01.01. és 1989.12.31. között a $\mathbf{G}_{k,l}(1, 2, 1)$ -ben. Balról jobbra haladva minden oszlop (az első kivételével) tartalmazza az előző oszlopban lévő összes hullámot az élek behúzásának megfelelő feltételei miatt. Ebben a vizsgálatban megengedjük, hogy $\alpha > 0$ legyen, mert ez része volt a [10]-ban látott megközelítésnek, ami látható az 1. ábra alsó részén is. Ez a paraméter érték indokolt lehet, hiszen a folyón előfordulhat visszaduzzasztás egy mellékfolyó vagy egy duzzasztógát miatt, így ha bármelyik hatás miatt árhullám keletkezik, akkor szeretnénk, ha ezek is szerepelnének az ábrázolásban.

Világos, hogy érvényes az a korábban tett megfigyelés, hogy minél kisebb a δ -csúcs sugara, annál nagyobb számú árhullámot kapunk. A legnagyobb ugrás a négy hőtérkép bármelyikén az első és a második oszlop között van, ami az ablak sugarának időtartamban növelését jelenti. Ez akkor is érvényes, ha a δ -csúcs sugarát 4-ig növeljük, ami azt jelenti, hogy egy 9 napos időablakban csak egy (feltételeknek megfelelő) δ -csúcsot veszünk figyelembe. A $\delta = 4$ -hez tartozó hőtérkép a A.1. ábrán azt mutatja, hogy még a jelentős árhullámokat figyelembe véve is csak egy töredékük éri el Szegedet ugyanazon a napon (vagy előtte, ha $\alpha > 0$), mint amikor elindult. Az is megfigyelhető, hogy az ablak sugarát időben előre felé növelve 3-ról 4-ra a hullámok száma nem sokat növekszik. Ugyanezeket a megfigyeléseket az A.2. ábra esetében is

meg lehet tenni.

Az A.3. és A.4. ábrákon a két időszakra vonatkozóan ábrázoltuk az árhullámok átlagos éves számát. Jól látható, hogy a számok minden felsorolt paraméterkombinációban emelkedtek az utóbbi időszakban, ami összhangban van [4] megállapításaival.

Végül az árhullámok átlagos levonulási idejét vizsgáljuk a korábban bemutatott két fogalom segítségével. A 4.5-ös definíció által leírt koncepcióra leírt megközelítésre *reguláris* módszerként, míg a 4.6-os definícióban ismertetettre *súlyozott* módszerként fogunk hivatkozni. A A.5. ábrán az x tengely a különböző δ értékeket, az y tengely pedig a különböző β értékeket jelöli. Mint korábban említettük, mivel időmérésről van szó, kizárunk minden olyan esetet, amikor $\alpha > 0$, így csak a $\alpha = 0$ esetet vizsgáljuk. Az első hőtérkép második sorának és harmadik oszlopának cellájában lévő érték a Szolnok és Szeged állomások közötti árhullámok átlagos levonulási idejét tartalmazza az 1951.01.01. és 1989.12.31. között a $\mathbf{G}_{k,l}(0, 1, 3)$ gráfban, amelyet a reguláris módszerrel számítottunk ki és a fent említett hőtérkép címében is megjegyeztünk.

Nyilvánvaló, hogy ha $\beta = 0$, akkor mindenhol 0 értéket kapunk, mivel az árhullámok csak ugyanazon a napon haladhatnak előre, amikor elindultak. Megjegyezzük, hogy azonos időszakból származó hőtérképek esetén a 3.7-es és 3.8-as tételek miatt a hőtérképek felső trianguláris része illetve a tridiagonális alsó része is megegyezik, hiszen azokban az esetekben $\alpha + \beta < \delta + 1$ fennáll, így nem fordulhat elő kereszt-kapcsolódás, ami mindkét átlagolási fogalom esetében azonos értékeket eredményez. Az időszakokat összehasonlítva nincs olyan jellegű különbség, amely minden elemre igaz lenne, de általánosságban elmondható, hogy az újabb időszakok az esetek többségében mindkét fogalom esetén kisebb átlagértékeket mutatnak, ami összhangban van a [5] megállapításaival. A fogalmakat összehasonlítva ott, ahol különbség jelenik meg (azaz $\beta > \delta + 1$), az eltérések nagyobbak a δ alacsonyabb értékei esetén, amikor a kereszt-kapcsolódás sűrű hálószerű struktúrát okoz, így a súlyozás nagyobb szerepet kap.

5. Diszkusszió és további tervek

A dolgozatban tárgyalt szoftvert az ATIVIZIG számára fejlesztettük, és jelenleg is fejlesztés alatt áll. A szellemi tulajdonjogok miatt a forráskód nem nyilvános, de igényelhető a cikk szerzőjétől. Az hidrológiai szakértőkkel betekintést nyerhetünk a továbbfejlesztés lehetséges irányába, így számos olyan terület van, amire a jelenlegi rendszert kiterjeszthetnénk. Az egyik már megvalósított, de a fenti modellben le nem írt funkció az egyes állomásokra vonatkozó információk felhasználása. Ezek között vannak olyanok, amelyek duzzasztógátak között vannak, vagy amelyekbe az állomás közelében egy mellékfolyó torkollik: ebben az esetben, ha egy árhullám érkezne a mel-

lékfolyón, az visszaduzzasztaná a folyót néhány állomással hátrébb. Világos, hogy ez az árhullám nem a Tisza felől érkezett, de tetőzést okoz az idősorban, ezért nem szeretnénk kizárni az ilyen eseményeket, és itt kerülnek előtérbe az $\alpha > 0$ értékek. Ha általánosságban megengednénk egy egységes visszafelé irányuló ablak sugarát, az elég sok nem kívánt él gráfba kerülését eredményezné. Ennek paramétere ssé tétele minden egyes állomás esetében lehetővé teszi az egyes állomások profiljainak létrehozását. Példaként egy új α paraméterről beszélünk, ahol $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)$ és hasonlóan, az új β paraméter pedig $\beta = (\beta_1, \beta_2, \dots, \beta_K)$ lenne, ahol K a felhasznált állomások száma.

Hasonlóan felmerül (bár vizsgálatok még nem történtek a kérdés kapcsán) a különböző δ paraméterek használata. Lehetnek olyan állomások, amelyek hajlamosabbak vízállás nagyobb ingadozására, mint a többi. Ebben az esetben, hogy ne maradjanak ki jelentős, egymáshoz túl közel eső csúcsértékek, minden egyes állomáshoz különböző δ értékeket kellene beállítani, így az új paraméter analóg módon a $\delta = (\delta_1, \delta_2, \dots, \delta_K)$ lenne.

Egy másik érdekes irány lehet a mellékfolyók információinak beépítése, ami azt jelenti, hogy az árhullámok keresését ezekre a folyókra is kiterjesztenénk, és integrálnánk őket a jelenlegi modellünkbe. Ezt Vágás valamilyen módon meg is tette, ahogyan azt az 1. ábrán láthatjuk, ezt jelölik az elhelyezett fekete színű pontok, amelyek a Tiszába csatlakozó mellékfolyó vízállás tetőzésére utalnak az adott mérési pontnál.

Az egyik legfontosabb terv a bevezetett koncepciók és mérőszámok számításának pontosítása, vagy akár új mérőszámok felállítása a hidrológiai szakértők bevonásával. Ez nemcsak megbízhatóbbá tenné a metrikáinkat, hanem megerősítene minket arról, hogy módszertanunk és tervezésünk abba az irányba mutat, amelyet a jövőben az ágazat is hasznosítani tud. A korábban leírt metrikák segítségével ez a szoftver arra is használható, hogy egy előrejelző modell számára - amelyet szintén az együttműködés részeként fejlesztünk - érdekes időintervallumokat találjunk, ahol tanítani vagy tesztelni lehet a modellt, ami kulcsfontosságú ahhoz, hogy a modell jól teljesítsen, amikor éles árvízi esemény következik be.

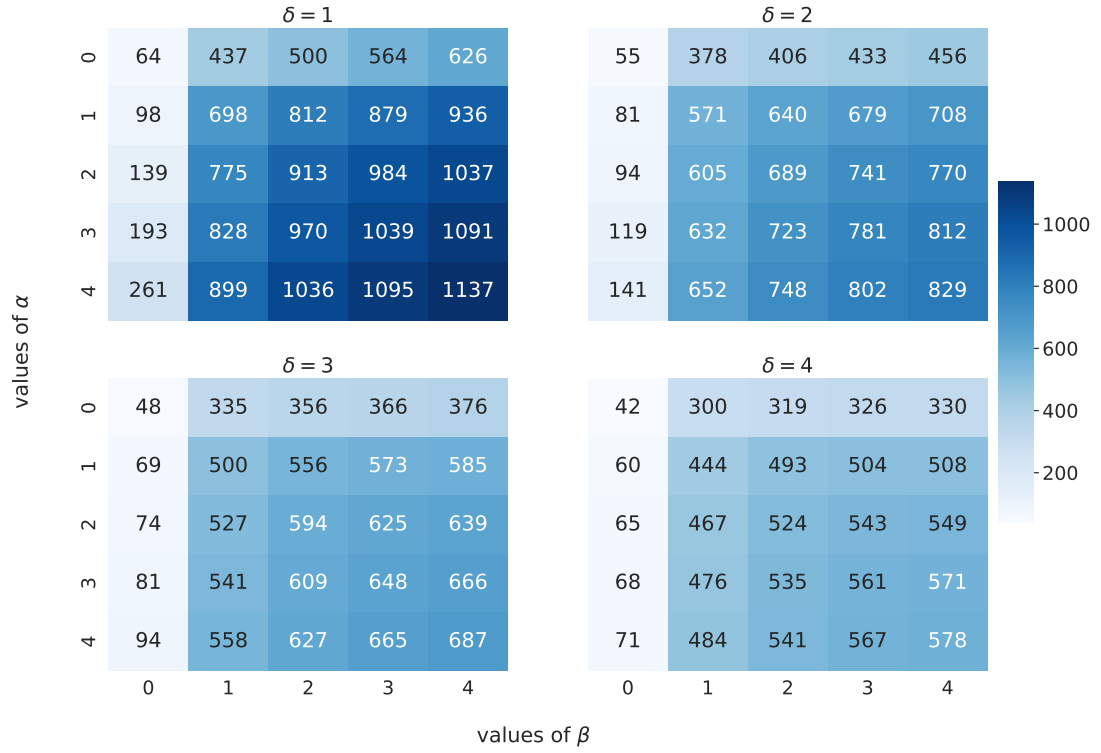
Ötletként felmerült a jövőre nézve egy olyan előrejelző modell megalkotása, amely az első néhány csúcspont térbeli és időbeli elhelyezkedése és a csúcsértékek alapján megjósolja, hogyan fog kinézni egy árhullám. Természetesen a valós idejű vagy napi szintű előrejelzés csak akkor lenne lehetséges, ha a szoftver az adatfolyamhoz kapcsolódna, ami jelenleg nem megoldott. Vegyük észre, hogy a leolvasási értékeket kizárólag arra használtuk, hogy δ -csúcsokat találjunk, egyik metrikánk vagy ábrázolásunk sem használja jelenleg a csúcsban mért értékeket. A folyó hidraulikus gradiensét a vízállás értékek abszolút magasságban vett idősorának segítségével le-

het kiszámítani, és ez információt adhat a folyó viselkedéséről az állomások között, ami releváns bemeneti adat egy előrejelző rendszer számára. Jelenleg csak a napi adatot használunk, de az utóbbi időben minden állomásról naponta több leolvasást is kaptunk így jelenleg az adatoknak csak töredékét használjuk fel. Ha akár 12 vagy 6 órás leolvasásokat használnánk, az átlagos levonulási idő számítása sokkal pontosabb lenne, mivel ha egy árhullám ugyanazon a napon kezdődik és ér véget, akkor azt (jelenleg) 0 napnak számítjuk.

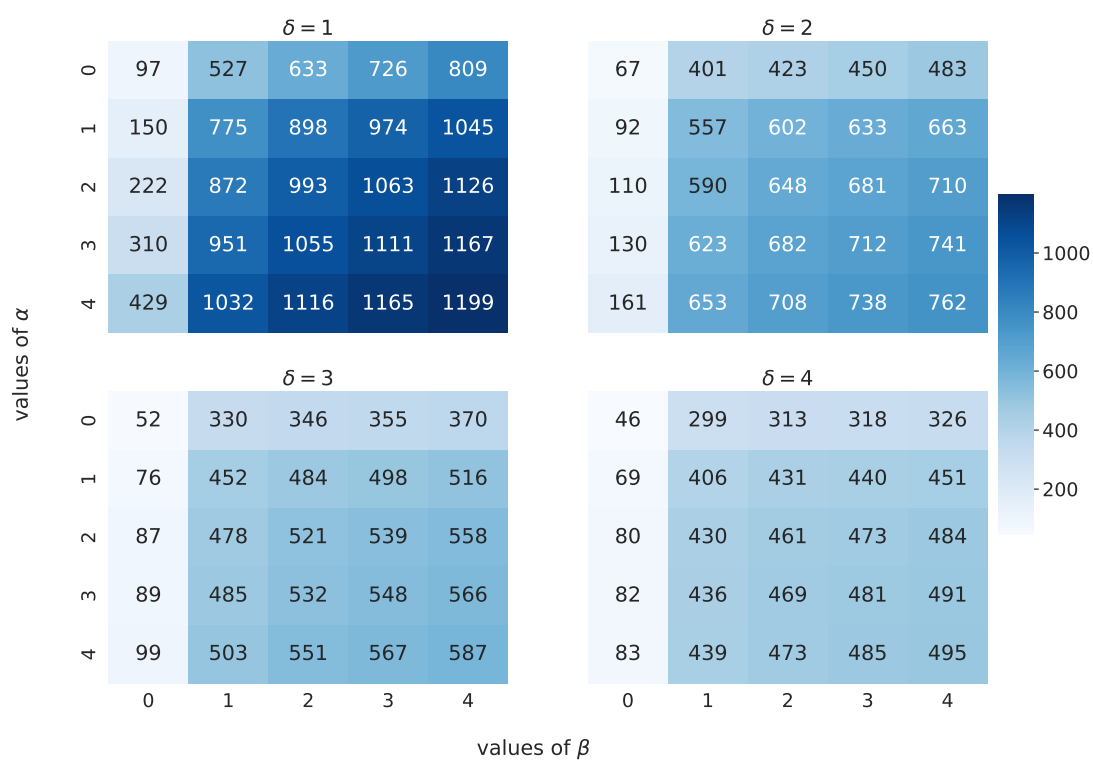
Hivatkozások

- [1] Günter Blöschl, Julia Hall, Juraj Parajka, Rui AP Perdigão, Bruno Merz, Berit Arheimer, Giuseppe T Aronica, Ardian Bilibashi, Ognjen Bonacci, Marco Borga, Károly Fiala, et al. Changing climate shifts timing of european floods. *Science*, 357(6351), 2017.
- [2] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.
- [3] Julia Hall, Berit Arheimer, Giuseppe T Aronica, Ardian Bilibashi, M Boháč, Ognjen Bonacci, Marco Borga, Paolo Burlando, Attilio Castellarin, Giovanni B Chirico, Károly Fiala, et al. A european flood database: facilitating comprehensive flood research beyond administrative boundaries. *Proceedings of the International Association of Hydrological Sciences*, 370, 2015.
- [4] Tímea Kiss, Károly Fiala, and György Sipos. Alterations of channel parameters in response to river regulation works since 1840 on the lower tisa river (hungary). *Geomorphology*, 98(1-2), 2008.
- [5] Tímea Kiss, Judit Nagy, István Fehérvári, Gabriel J Amissah, Károly Fiala, and György Sipos. Increased flood height driven by local factors on a regulated river with a confined floodplain, lower tisa, hungary. *Geomorphology*, 389:107858, 2021.
- [6] Péter Kozák. A szegedi belvárosi árvízvédelmi rendszer megújítása. *Mérnök Újság*, 6, 2015.
- [7] János Rakonczai and Péter Kozák. The consequences of human impacts on hungarian river basins. *Zeitschrift für Geomorphologie, Supplementary Issues*, pages 95–107, 2011.
- [8] István Vágás. *Hidrológiai statisztika*. Tankönyvkiadó, 1974.
- [9] István Vágás. Adatok az 1876–1975 időszak tiszavölgyi árvizeiről. a tisa fontosabb árhullámai 1876 és 1911 között. *Hidrológiai Közlöny*, 1978.
- [10] István Vágás. *A Tisza árvizei*. VIZDOK, 1982.

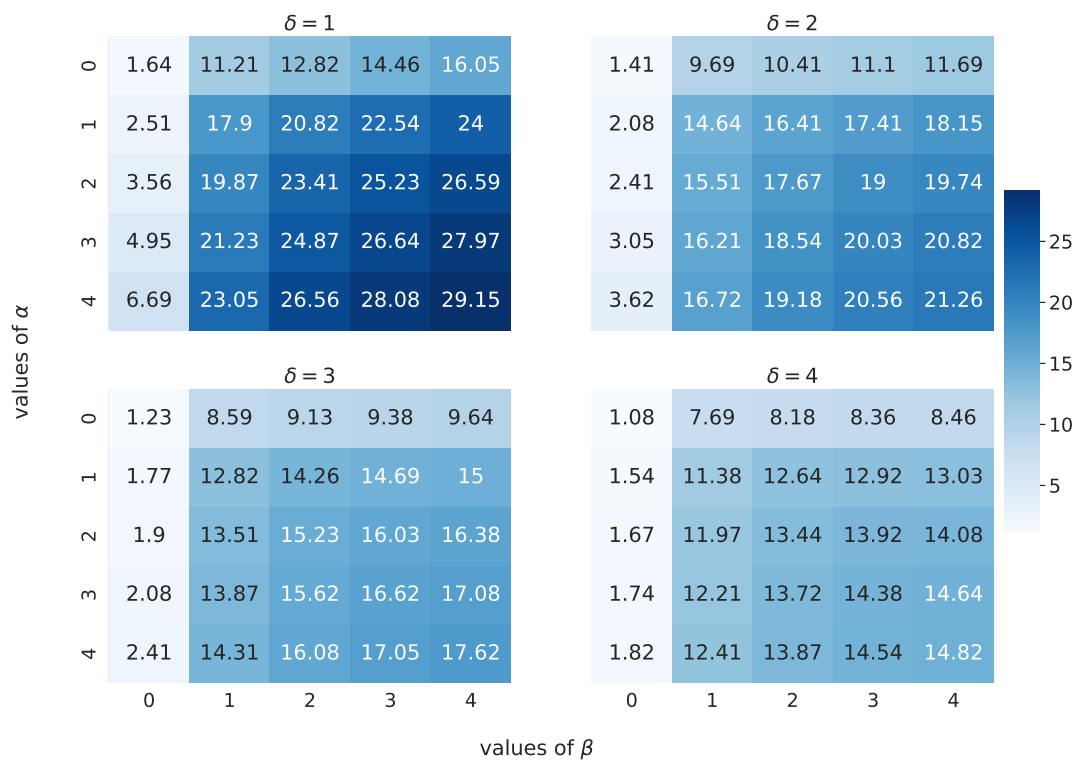
A. Hőterképek



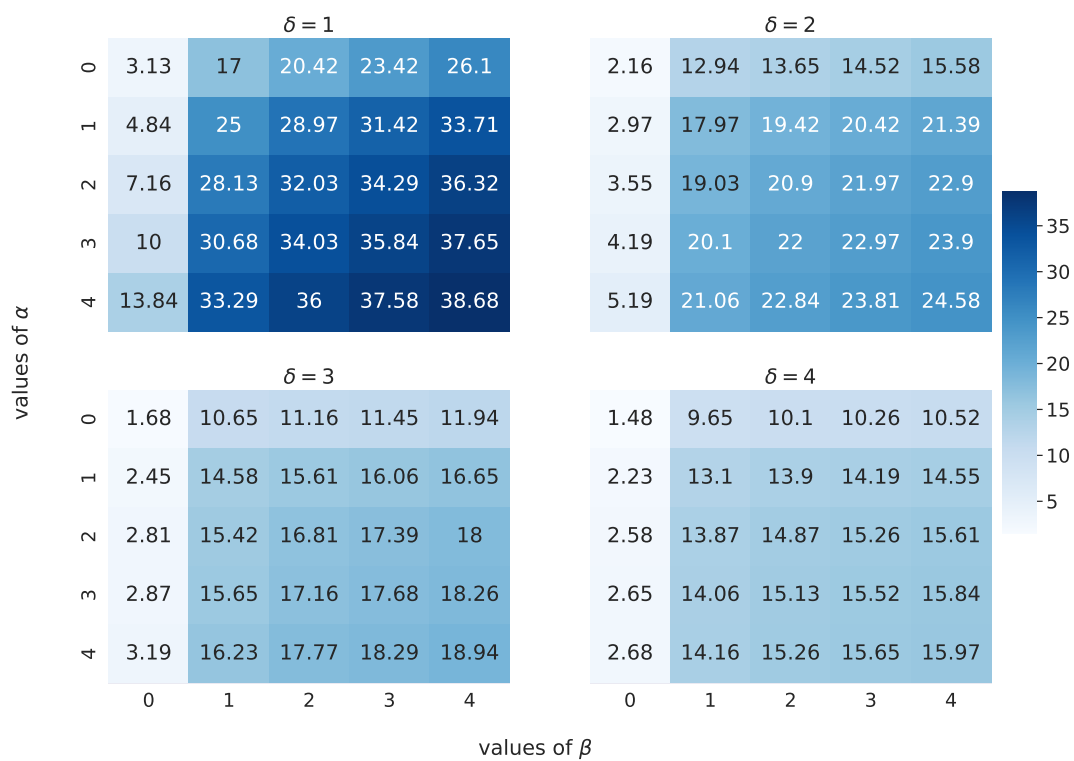
A.1. ábra. Az 1951.01.01. és 1989.12.31. közötti árhullámok száma különböző paraméterbeállításokra kiszámítva. Az eredmények hőterképekbe vannak rendezve, a különböző δ paraméterekhez külön hőterképek tartoznak.



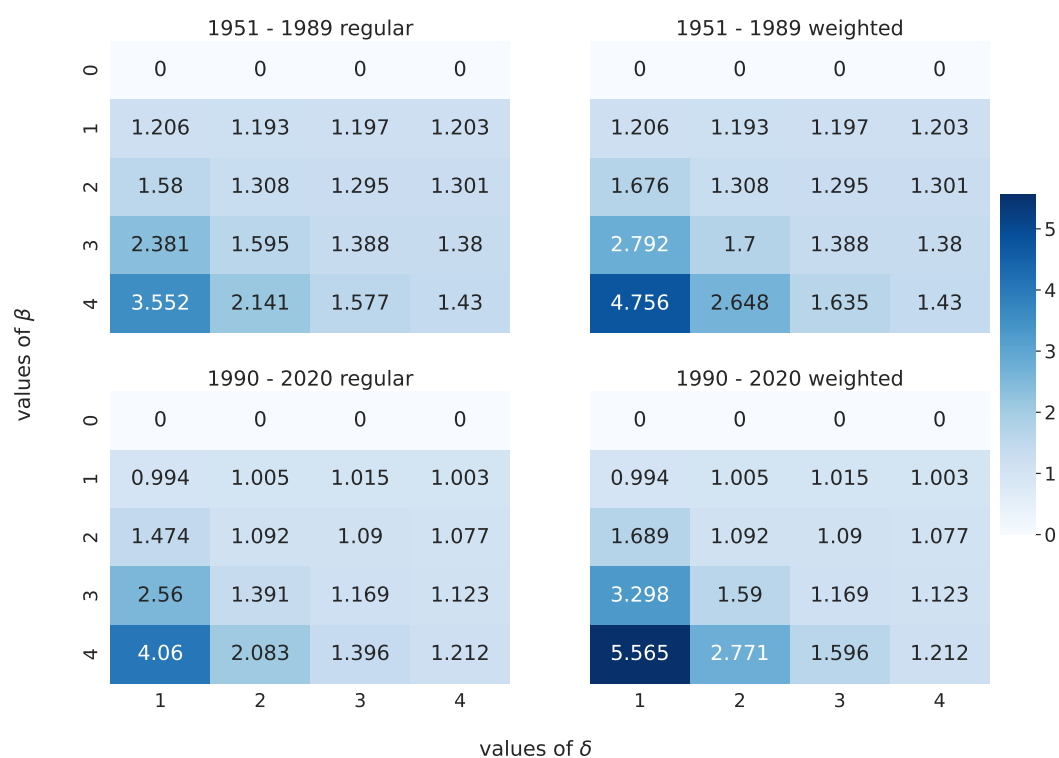
A.2. ábra. Az árhullámok száma 1990.01.01. és 2020.12.31. között, különböző paraméterbeállításokra kiszámítva. Az eredmények hő térképekbe vannak rendezve, a különböző δ paraméterekhez külön hő térképek tartoznak.



A.3. ábra. Az árhullámok éves átlagos száma 1951.01.01. és 1989.12.31. között, különböző paraméterbeállításokra kiszámítva. Az eredmények hőtérképekbe vannak rendezve, a különböző δ paraméterekhez külön hőtérképek tartoznak.



A.4. ábra. Az árhullámok éves átlagos száma 1990.01.01. és 2020.12.31. között, különböző paraméterbeállításokra kiszámítva. Az eredmények hő térképekbe vannak rendezve, a különböző δ paraméterekhez külön hő térképek tartoznak.



A.5. ábra. Az árhullámok átlagos levonulási ideje látható, a hőterképek felső sorában 1951.01.01. és 1989.12.31. között, az alsó sorban 1990.01.01. és 2020.12.31. között. A bal oldali hőterképeknél a számításokhoz a 4.5-ös definícióban ismertetett, A jobb oldali hőterképeknél a számításokhoz a 4.6-os definícióban ismertetett koncepciót használtuk.