

A Duna mederanyag összetételének predikciója gépi tanulási módszerrel

Schrott Márton IV. évf.

Konzulensek: Dr. Baranya Sándor, Ermilov Anatol Alexander

Vízépítési és Vízgazdálkodási Tanszék

2022/23 őszi félév



M Ű E G Y E T E M 1 7 8 2



Tartalomjegyzék

1.	Tartalmi kivonat.....	2
2.	Bevezetés	3
2.1.	A téma relevanciája	3
2.2.	Mederanyag kutatási módszerek	4
2.3.	A kutatás célja	6
2.4.	Adatok származási területének bemutatása	6
3.	Módszerek	7
3.1.	Fizikai mederanyag mintázás.....	7
3.2.	Számítógépes hidrodinamikai modell.....	9
3.3.	A Random Forest modell.....	10
3.3.1.	Döntési fák.....	10
3.3.2.	Az RF modell tanítása	10
3.4.	Adatfeldolgozás és előkészítés az RF modellhez	11
3.4.1.	Hidrodinamikai adatok előkészítése	11
3.4.2.	Mederanyag adatok előkészítése	12
3.5.	Az RF modellek paraméterezése.....	14
4.	Eredmények	16
4.1.	Ren et al. (2020) modelljének adaptálása.....	16
4.2.	A kutatás keretében felállított klasszifikációs RF modell.....	18
4.3.	A kutatás keretében felállított regressziós RF modell	18
5.	Eredmények értékelése	21
6.	Következtetések	26
7.	Köszönetnyilvánítás.....	26
8.	Ábrajegyzék	27
9.	Irodalomjegyzék	28

1. Tartalmi kivonat

A mederanyag összetétele fontos jellemzője a vízfolyásoknak. Többek között hatással van a folyó áramlási viszonyaira, a hordalékelragadó erő területi eloszlására, a part menti vízbázisok működésére, valamint befolyásolja a folyóbeli és folyó menti élőhelyek minőségét is. Kiemelt szerepe ellenére ismereteink korlátosak, hiszen a közvetlen, hagyományos fizikai mintavételi eljárások – az időszakos vízjárású folyóktól eltekintve – nehezen, sok munkaigénnyel végezhetők el, eredményük pedig kevés információval bír ahhoz, hogy az áramlás és a folyómeder komplex kölcsönhatását megbízhatóan leírassuk. Éppen ezért, folyamatos a kutatás az új módszerek iránt.

Mivel a mesterséges intelligencia (artificial intelligence) és a gépi tanulás (machine learning) a tudomány számos területén áttörő sikereket segített elő, ezen a téren is sok felhasználási kísérlet volt már világ szerte. Dolgozatomban a faszerkezetű, döntési erdőkön alapuló Véletlen Erdő (Random Forest, továbbiakban RF) eljárással, modellel foglalkozom. Korábbi, a Dunán végzett hidromorfológiai felmérések adatainak összegyűjtése és csoportosítása után, a modellel kapcsolatot állítok fel a különböző hidromorfológiai (pl. vízmélység, áramlási sebesség, fenéksúsztató feszültség) adatok és a mederanyag összetétele között. Mivel a hidromorfológiai paraméterek felmérése ma már gyorsan és megbízhatóan történik (ellentétben a mederanyag mintázással), így a modell alkalmas lehet mederanyag információk nyújtására (i) olyan helyeken, ahol mintavételezés nem történt, valamint (ii) jövőbeli, numerikus modellek által számított, vízjárási állapotok esetében. A dolgozatban egy, az Egyesült Államokban fejlesztett RF modellt adaptálok a hazai vizsgálatokra, hogy az általam adatbázisba rendezett adatokkal milyen eredményeket szolgáltat. Az eredményeket továbbá összevetem más módszerekkel is, a modellen pedig érzékenységvizsgálatot hajtok végre.

2. Bevezetés

2.1. A téma relevanciája

A mederanyag kiemelten fontos fizikai jellemzője a vízfolyásoknak hiszen a különböző hidromorfológiai jellemzőkkel kölcsönösen hatnak egymásra. A mederanyag szemcséi ellenállást keltenek, így lassítva a rajtuk elmozduló vízrészecskéket. A kölcsönhatás másik fele, hogy a víz, sebessége függvényében különböző méretű hordalékokat tud magával ragadni, majd egy lassabb szakaszon lerakni. Ezen folyamatok által azt is nagyban befolyásolja a mederanyag, hogy az ember és az élővilág számára hogyan hasznosíthatóak a vízfolyások.

A partiszűrészű kutak kiemelt szerepet töltenek be az ivóvízellátásban Magyarországon és világszerte is. A hazai vízellátás közel 40% partiszűrészű kutakból származik (Szűcs et al., 2020), így minden szempontból fontos ezen vízbázisok fennmaradása. Amellett, hogy a tisztítási folyamatban fontos szerepet játszik a mederanyag (Ray et al., 2011), ha annak pórusai kolmatáció során eltömődnek nagyban sérül a kitermelhető víz mennyisége, és (a talajvíz részarányának növekedése miatt) a minősége is (Ulrich et al., 2015). További veszélyforrást jelent a medermélyülés, mely a talajvízszint csökkenése által a kutak hozamát is csökkentheti.

A hajózás esetében viszont fontos, hogy az év nagyrésztében járható vízi útvonalak legyenek, melyet a lassabb szakaszokon lerakott hordalékból kialakuló gázlók veszélyeztetnek. Mivel a Duna egy kiemelt fontosságú hajózási útvonal, nemcsak hazai gazdasági szempontból, de európai szinten is, ennek megóvása és fejlesztése is egy fontos építőmérnöki feladat (UTIBER et al., 2020). A hajózási viszonyok fejlesztését célzó beavatkozások hatékonysága nagyban függ attól, hogy a beavatkozással érintett területen milyen a mederanyag összetétele, ezért annak ismerete ilyen jellegű vizsgálatoknál kiemelten fontos.

Környezet- és természetvédelmi szempontból is jelentős kérdés a mederfelszín fizikai összetétele, ezt bizonyítják a halak ívási szokásai (Gabiele és Guti, 2010, Füstös et al., 2019). A különböző fajok különböző élőhelyeket igényelnek, és ez egyedfejlődésük elején kifejezetten fontos, hiszen olyankor kisebb a tűrőképességük. A halivadékok kutatása pl. bemutatta, hogy a vízmélység, sebesség, nyírófeszültség mellett a mederanyag is szerepet játszik a szaporodás helyében. Ezen területek, így közvetve az őshonos fajok védelme érdekében fontos alaposan megismernünk az emberi beavatkozások hatásait a mederanyag összetételére.

A felsorolt környezeti erőforrások, és élővilág megóvása érdekében fontos ismernünk, az egyes beavatkozások hatásait a mederanyagra és medermorfológiára. A hordaléktranszportba való

beavatkozás nagy kiterjedésben hat a folyóra. Ha például egy vízlépcső duzzasztott tere csapdázza a hordalékot, az alvízi szakaszok egyensúlyi állapotából hiányozni fog, ezzel ott is változást idéz elő a műtárgy. Így a pontos tervezéshez a mederanyag összetételét egy a beavatkozási helyszínt méretben meghaladó szakaszon szükséges ismerni. Erre a tradicionális módszerek nem alkalmasak, idő és munkaigényük miatt.

2.2. Mederanyag kutatási módszerek

Kiemelt jelentősége miatt ez egy élénken kutatott téma, számos módszer létezik mederanyag vizsgálására. A hagyományos eljárások jelentős többsége fizikai mintavételen alapul, amikor egy mintavételi eszközzel a mederfelszín egy részéről kiemelnek egy darabot (Bunte és Abt., 2001). A vett minták vizsgálatához a szemcse számláló módszerek kategóriájába tartozik a rács és vonalmenti eljárás, mely során állandó, a feltételezhető legnagyobb szemátmérőt meghaladó lépésközként véletlenszerűen egy mintát veszünk. Ez a 2 mm-t meghaladó kavicsos esetleg görgeteges medreknél alkalmazható. A terepi mintagyűjtés után az egyes kavicsok megmérése következik, majd a méret és darabszám alapján készítenek szemeloszlási görbét. A következő nagy kategória a felületi módszereké, mely egy gyűjtőnév a körülbelül 0,1-1 m² területről történő mintavételekre. Ennek legegyszerűbb változata, amikor egy erre a célra készített keret segítségével lehatárolt területről kézzel begyűjtik a mintát. Ennek során figyelni kell, hogy csak a felső réteget távolítsuk el. A kiértékelés az előzőekhez hasonlóan történik. Ide tartozik az adhéziós módszer is, mely során egy lapot epoxigyantával, viasszal vagy agyaggal vonunk be majd elhelyezzük a mintavételi pontban, és nyomást helyezünk rá. A kötőanyagba maradt szemcséket a lemezzel együtt emelhetjük ki, majd oldóanyag használatával, vagy mechanikai úton eltávolítjuk, majd elemezzük. Ezek az eljárások szárazra került, vagy sekély vizű szakaszokon végezhetőek, mely a Dunán, és ahhoz hasonló nagy folyókon nem nyújt lehetőséget a teljes szelvényben történő felmérésre.

Nagy előrelépés a fagyasztásos mintavételi eljárás, mely során folyékony nitrogén vagy széndioxiddal egy körülbelül 2 cm mély és 15 cm sugarú korongot fagyasztunk össze a mederanyagból, melyet így kiemelve zavartalan mintaként is elemezhetünk, vagy kiolvasztás után szabványos szitasorral szemeloszlási görbét készíthetünk (pl. Baranya és Józsa, 2006). Mély vízben viszont kitűzőhajó és nagyteljesítményű gépezetet igényel, mely megnöveli az energiaigényét és költségét. Ezzel ellentétben az ún. serleges mintavevővel a folyó bármely pontjában végezhető mintagyűjtésre ad lehetőséget relatív egyszerű eszközökkel. A harang alakú mintagyűjtőt hajóról a vízbe engedve, a mederfenéken végig húzzuk. Ha kellőképpen

megtelt kiemeljük és laborba szállítjuk, ahol szabványos szitasor segítségével szemeloszlási görbét készíthetünk. Ezt a módszert a dolgozat 3.1. fejezetében részletesebben is kifejtem, mivel a kutatásom során ezt használtam.

A technológia fejlődésével megjelentek a fizikai mintavételek mellett a különböző számítógépes elemzési módszerek. A digitális kamerák, és a számítógépek gyors fejlődésének köszönhetően, költséghatékonyan nagy mennyiségű felvétel készíthető, melyeket gyorsan lehet elemezni. Az ezredforduló elején, a már más területeken bevált módszerként kezelt fotogrammetriát is igyekeztek felhasználni medermorfológiai felmérésekhez (Chandler et al., 2000). A hivatkozott tanulmányban sikeres laborkísérletek után a kanadai Sunwapta folyón végeztek terepi mérést. Kisvizes időszakban ismert koordinátájú referencia pontokat helyeztek el, majd egy közeli magaslatról képet készítettek. Ez a módszer sikeresen lecsökkentette a terepi munkaigényt, de a vízborította mederszakaszokat továbbra is hagyományos módszerekkel vizsgálták.

A Zürichi Szövetségi Műszaki Főiskolán fejlesztett MATLAB alapú BASEGRAIN program képelemzéssel határozza meg a szemösszetételt (Detert et al., 2013). A program szürkeárnyalatosított fényképeken detektálja a szemcsék éleit, így körülhatárolva őket a felhasználó által bevitt mm/pixel arány alapján átmérőt számít, melyből a Fehr-képlet segítségével tömegszerinti eloszlást számít és elkészíti a szemeloszlási görbét. Az eljárást a sikeres laboratóriumi tesztek után, terepen is bevetették (Rüther, et al.2013). A Surna nevű szabályozott vízfolyás tiszta vize miatt megfelelő körülményeket biztosított a kísérletekhez. Biztató eredményei ellenére, mivel sok paramétert és jó képminőséget igényel, nagyobb, lebegtetett hordalékban gazdag folyókban való használatra ebben a formájában nem alkalmas.

A hasonló elven működő Wavelet-módszer (Buscombe, 2013) kevesebb paramétert igényel, így nagymennyiségű adat feldolgozására is alkalmas lehet. A módszer vízalatti alkalmazását a BME Vízépítési és Vízgazdálkodási Tanszék (továbbiakban VVT) munkatársai is vizsgálták (Ermilov et al., 2020),

A tudomány számos területén áttörő sikereket hoz a mesterséges intelligencia (továbbiakban MI), így ennek használata a mederanyag elemzésében is kutatott téma (pl. Benkő et al., 2021). Benkő et al. (2021) tanulmányában bemutatott módszer Mély Tanulás eljárással a fényképeken felismeri az adott részeken jellemző mederanyagot. Az így előállítható területarányos szemösszetételt összehasonlítva a Wavelet és a serleges mintavétel eredményével jó egyezést

mutatott. Ez bizonyítja az MI alapú eljárások potenciálját vízalatti mérésekhez történő felhasználását (Tikász és Schrott, 2020).

A hagyományos módszerekkel ellentétesen az MI alapú eljárások gyorsan, kevés helyszíni munkával és laboratóriumi feldolgozás nélkül is kielégítő adatokat biztosítására is alkalmasak lehetnek. Így ezek további alkalmazási módjait vizsgálom ebben a tanulmányban egy másik MI eljárás, a faszerkezetű, döntési erdőkön alapuló Random Forest (továbbiakban RF) módszer segítségével.

2.3. A kutatás célja

A kutatás célja, hogy egy, az Egyesült Államokbeli Columbia folyóra felépített RF modell (Ren et al. 2020) hazai Dunára való alkalmazását megvizsgálja, majd az alapján kifejlesszen egy új, saját adatokból tanított modellt. Végso célként, egy a teljes magyarországi Duna szakaszt lefedő mederanyag térkép előállítását tűztük ki, amelyhez a kutatást befogadó tanszék által az elmúlt években gyűjtött mederanyag minták mellett egy 2D hidrodinamikai modell eredményezői (vízmélység, áramlási sebesség és fenék-csúsztatófeszültség) szolgáltatnak bemeneti adatokat. Sikeres modellfejlesztés esetén egy rendkívül gyors és hatékony mederanyag térképezési eljárást kapunk, amely alkalmas lehet különböző térléptékű (1D-2D-3D) áramlási- és morfordinamikai modellek paraméterezésére, tervezési feladatok megalapozására, vagy éppen adathiányos területekre szolgáltat információt.

2.4. Adatok származási területének bemutatása

A kiválasztott RF modell alkalmazhatósági korlátainak tanulmányozásához fontos tisztában lenni a mintaterületek geohidromorfológiai jellemzőivel. A kiinduláshoz használt amerikai modell alapjául szolgáló Columbia folyó forrása Kanadában található, majd az Egyesült Államokban folytatja útját Washington és Oregon államokon át, végül a Csendes-óceánba torkollik. A Kaszkádok magmás kőzetein, valamint glaciális, alluviális és tengeri üledékein folyik leeresztül. Vízjárás tekintetében 6800 m³/s középvízi hozam és 2970 m³/s kisvízi hozam jellemzi, de a klímaváltozás, és vízkivételek miatt rendkívüli időszakokban elérte már a 280 m³/s értéket is. A vízmélység egyes helyeken a 9 m-t is meghaladja (Sherwood és Creager, 1990).

A magyarországi Duna szakasz jellemzően üledékes kőzeteken át halad, kivétel ez alól a Duna-kanyar, ahol a Visegrádi-hegység magmás kőzet is található. A VVT által végzett mérések gyakori helyszíne Gönyű (1792-1796), itt a jellemző vízhozamok körülbelül a következők legkisebb hajózási vízhozam $1400 \text{ m}^3/\text{s}$, a középvíz $2000 \text{ m}^3/\text{s}$, és a legnagyobb víz $10000 \text{ m}^3/\text{s}$ (Baranya és Józsa, 2006).

3. Módszerek

3.1. Fizikai mederanyag mintázás

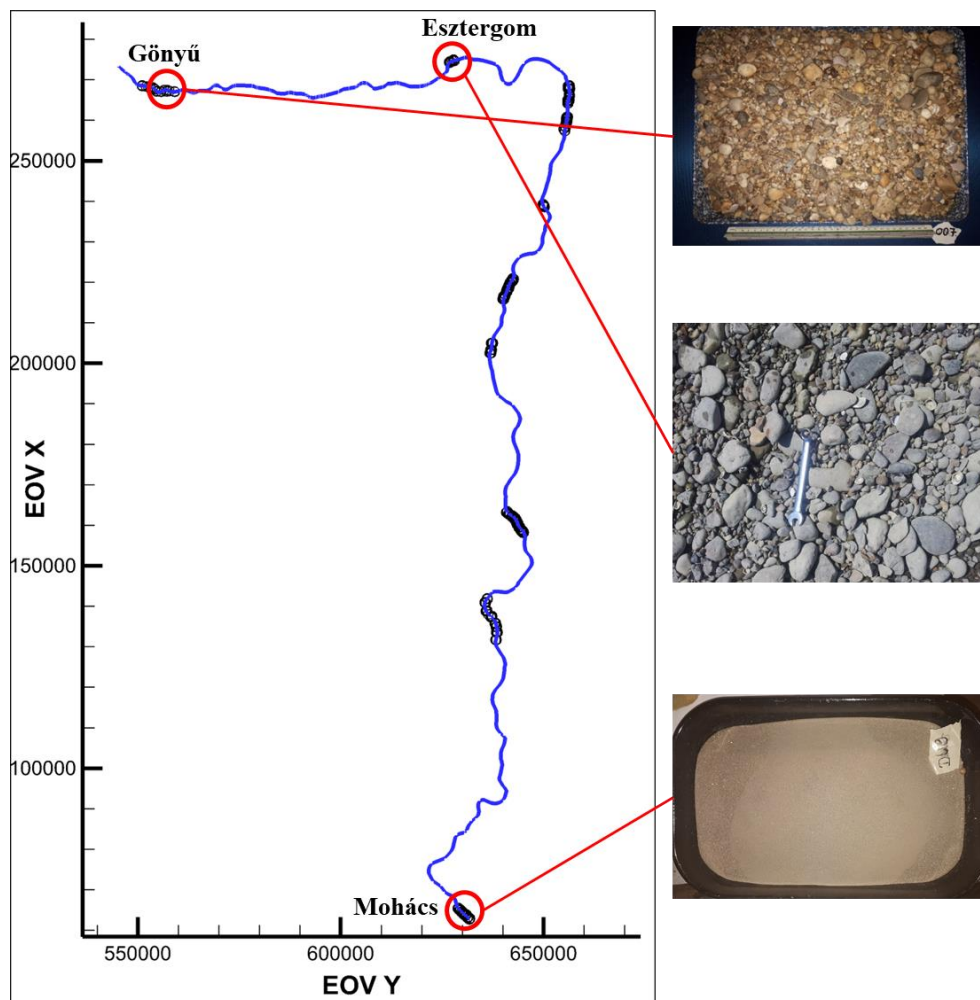
Kutatásunkban az RF modell tanításának egyik pillérét a Duna mederanyag összetételének jellemzői képzik. Az adatokat az elmúlt években, részben általam végrehajtott mederanyag mintavételekből állítottam elő. Vizsgálatainkban serleges mintavevő (1. ábra) segítségével történt az adatgyűjtés. Ez egy régóta alkalmazott és széleskörben elterjedt eljárás, mellyel a mederfenékről lehet mintát gyűjteni. A mintavételhez egy csipkézett szélű, félig nyitott hengerre van szükség. Ezt csónakból leeresztve végig húzzuk a folyó fenekének felszínén, a csipkézett széle bele-belekap a mederanyagba, mely így a serleg belsejébe kerül. A kiemelés után, laborba szállítva elemezzük a mintát. Szárítókemencében tömegállandóságig szárítjuk $105 \text{ }^\circ\text{C}$ -on. Ezután eltávolítjuk a kagylókat és amennyiben megkötött, mechanikai úton a finom szemcséket szétválasztjuk, vigyázva a nagyobb kavicsok épségére. Az így előkészített mederanyag mintát szabványos szitasoron rázzuk át, és az egyes szitákon fennmaradt szemcsék tömegét lemérjük. A kapott részeredményekből szemeloszlási görbét szerkesztettünk.

A módszer részletes eredményekkel szolgál, azonban számos hiányossága és korlátja van. Bár a mérés során a hajó koordinátája számos eszközzel elég pontosan rögzíthető, addig a mintavevő a vízmélységtől függően, 5-15 m távolságra is lehet. A kellő mennyiségű minta gyűjtéséhez, hosszú távon kell húzni a serleget, ahonnan nem egyenletesen kerül bele mederanyag, így bár pontbéli mintavételnek nevezzük, 10 méternél pontosabban nem lehet megállapítani a mintavétel helyét. A csipkézés a mederfenéknél mélyebbre is bele tud marni, így alsóbb rétegekből is kerülhet a mintákba. Sok terepi munkával és hosszas laboratóriumi elemzéssel jár a módszer és ezek a folyamatok nem automatizálhatóak, így nagyobb területek lefedésére nem alkalmas, jellemzően pontbeli információt szolgáltat. A VVT munkatársai ezzel a módszerrel sok éven át vettek mintákat a Duna különböző szakaszairól. Saját méréseim mellett ezeket a korábbi mintavételi eredményeket is felhasználtam ebben a kutatásban. A

rendelkezésre álló, összesen 302 db mederanyag minta a Duna teljes hazai szakaszát lefedik, és megfelelően reprezentálják a változatos mederanyag összetételi viszonyokat (2. ábra). Az ábrán illusztrációként bemutatott fényképeken jól látszik pl. a Gönyű környéki szakaszon található finom kavicsos mederanyag, az Esztergomnál lokálisan feldurvuló kavics, és a Mohácsnál jellemző tiszta homok.



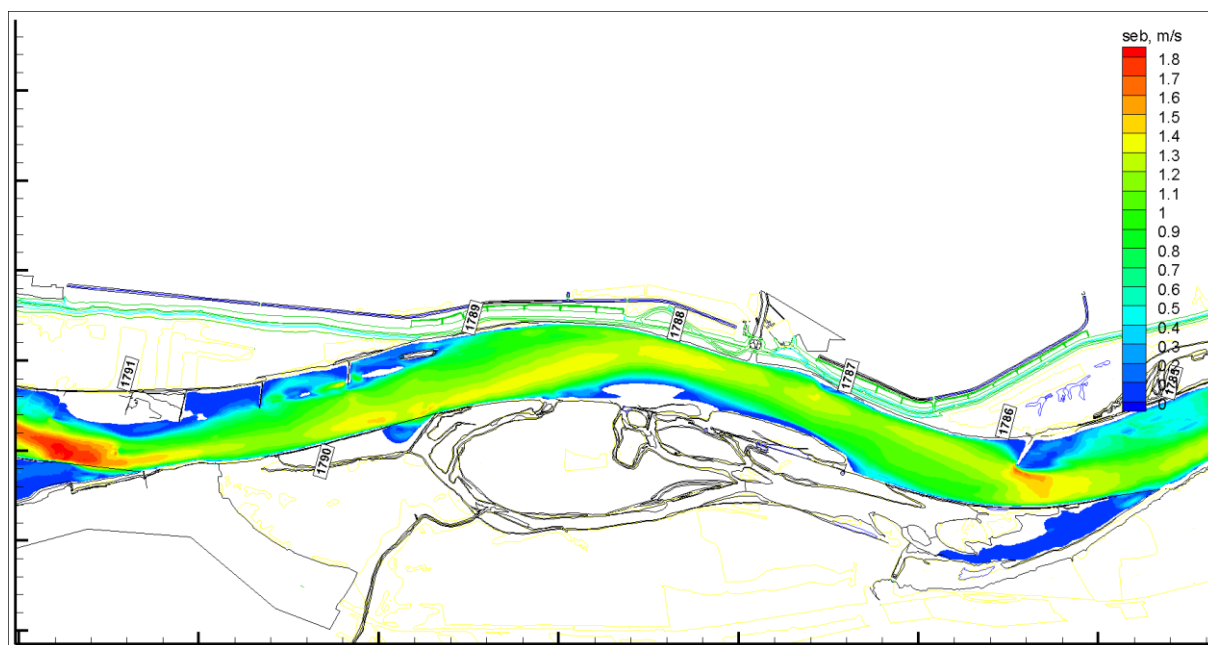
1. Ábra: Serleges mintavevő eszköz



2. Ábra: A duna különböző szakaszain vett jellemző mederanyag

3.2. Számítógépes hidrodinamikai modell

Az RF alapú mederanyagterképezés másik fontos bemeneti adata a fizikai mederanyag minták mellett egy olyan hidrodinamikai modell volt, amely a teljes magyarországi Dunára szolgáltat hidromorfológiai információkat. Vizsgálataimhoz egy, a VVT által készített, validált 2D áramlási modell eredményeit használtam fel. A modellt a közelmúltban befejezett hajózásfejlesztési célú tervhez (Utiber et al., 2020) kapcsolódóan építették fel és igazolták kis-, közép- és nagyvízi viszonyokra. A modelleredmények is erre a három állapotra vonatkoznak, nevezetesen a legkisebb hajózási vízhozamra (LKHQ), a középvízi vízhozamra (KÖQ) és az eddig mért legnagyobb vízhozamokra (LNQ). Ezek az értékek a Duna hossza mentén természetesen változnak, az eredménymezők a Duna egy-egy szakaszára vonatkoznak. A modell által szolgáltatott eredménymezők a vízmélység és az áramlási sebesség (3. ábra) területi eloszlásait adják meg.

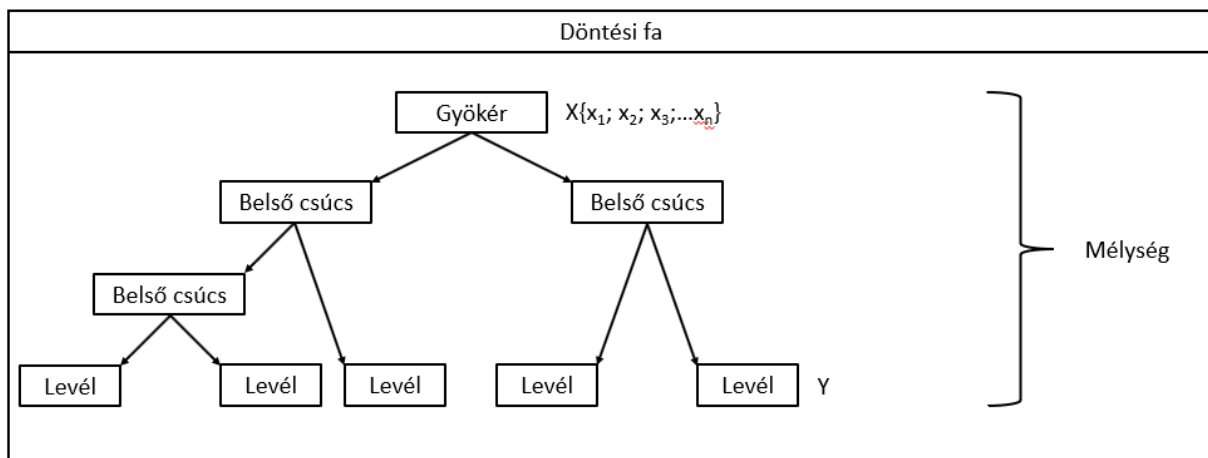


3. Ábra: A 2D modellel szimulált áramlási sebesség megoldásmezője a Duna kisvízi állapotánál a folyó 1785-1791 fkm közötti szakaszán.

3.3. A Random Forest modell

3.3.1. Döntési fák

Az RF modell döntési fákon alapul (4. ábra) (Biau és Scornet, 2016), mely az alakját tekintve egy irányított gyökeres fagraf, melynek minden csúcsából pontosan 2 további gyerek ágazik ki. Azokat a csúcsokat, melynek pontosan egy éle van levélnek nevezzük, ezek helyezkednek el a fa végén, ezek adják az eredményt. A bemeneti adata egy attribútumhalmaz ($X=\{x_1; x_2; x_3; \dots x_n\}$), és az eredmény lehet osztály vagy numerikus érték (Y).



4. Ábra: Döntésifa működési mechanizmusának sematikus ábrája

3.3.2. Az RF modell tanítása

A modellépítéshez a kódot a Google Colab felületén írtam, python nyelven. Valamint az `sk.learn` könyvtár RF moduljait használtam fel.

Az RF több, véletlenszerűen készített döntési fából áll. A tanítás során több, ismert eredményű, azonos attribútumhalmazzal rendelkező adat kerül a gyökérbe. Innen egy véletlenszerű attribútum alapján kettéosztja az adathalmazt, ezzel javítva a gyerekekben az adatok tisztaságát, azaz homogenitást. Ezt addig ismétli, amíg a fagraf mérete eléri a megadott maximális mélységét, vagy teljesen homogén leveleket kap. Ez a folyamat megismétlődik annyiszor, amennyi fából szeretnénk, hogy álljon a Random Forest.

Az RF modellek, eredményük alapján két nagy csoportba sorolhatóak az általuk adott eredmény alapján. Klasszifikációs (Random Forest Classification – RFC) során minden gráf „szavaz” egy osztályra, és a legtöbb szavazatot kapó osztályba sorolja be az adatot. Regressziós (Random Forest Regression – RFR) esetén minden fa ad egy numerikus eredményt, és ezek átlaga lesz

az attribútumhalmazhoz rendelt érték. RFR esetén a fagráf minden ága a maximális mélységig megy.

A modell tanítása során több változót lehet állítani. Meghatározható, mennyi döntési fát tartalmazzon a modell. Beállítható a döntésifák maximális mélysége is. Az ún. „bootstrap” segítségével eldönthető, hogy egy fa készítése során minden attribútumot fel kelljen-e használni, vagy sem. Meg lehet határozni azt is, hogy milyen feltétel rendszer alapján minősítse az osztás tisztaságát, és milyen minimális javulást várunk minden csúcs után. Ezen kívül még számos funkció van, melyet ebben a tanulmányban nem használtam.

A tisztaságot többek között az ún. GINI indexszel lehet meghatározni, a tanulmányban ezt használtam. Annak a valószínűségét mutatja, hogy egy adat rosszul van kategorizálva az adott csúcsban. 0 és 1 között veheti fel az értékét, ahol a 0 a legjobb, ekkor teljesen homogén, így nincs esély arra, hogy a modell tévesen kategorizálja be az adatot. Az RF modellben a változók fontosságát is nagyon jól mutatja, hogy az összes fában átlagosan mennyivel csökkent a GINI index a változóval történő osztás esetén (feature importance).

Az általam kialakított RF modell az előző pontokban bemutatott adatokat használja fel tanításra és igazolásra: i) mederanyag összetétel a fizikai mintavételezések alapján, amihez a kutatás jelen fázisában egyrészt a szemcseméret osztályt, és a medián szemátmérőt (D_{50}) vesszük alapul; másfelől pedig ii) a hidrodinamikai jellemzőket: vízmélység, áramlási sebesség, fenék-csúsztatófeszültség. Korábbi vizsgálatok alapján (pl. Baranya et al., 2018) ugyanis azt várjuk, hogy elsősorban az áramlástani viszonyok determinálják a mederfelszín szemcseösszetételét, és egy folyóra, amely geohidromorfológiai szempontból homogénnek tekinthető, általánosíthatóvá válik egy ilyen modell.

3.4. Adatfeldolgozás és előkészítés az RF modellhez

3.4.1. Hidrodinamikai adatok előkészítése

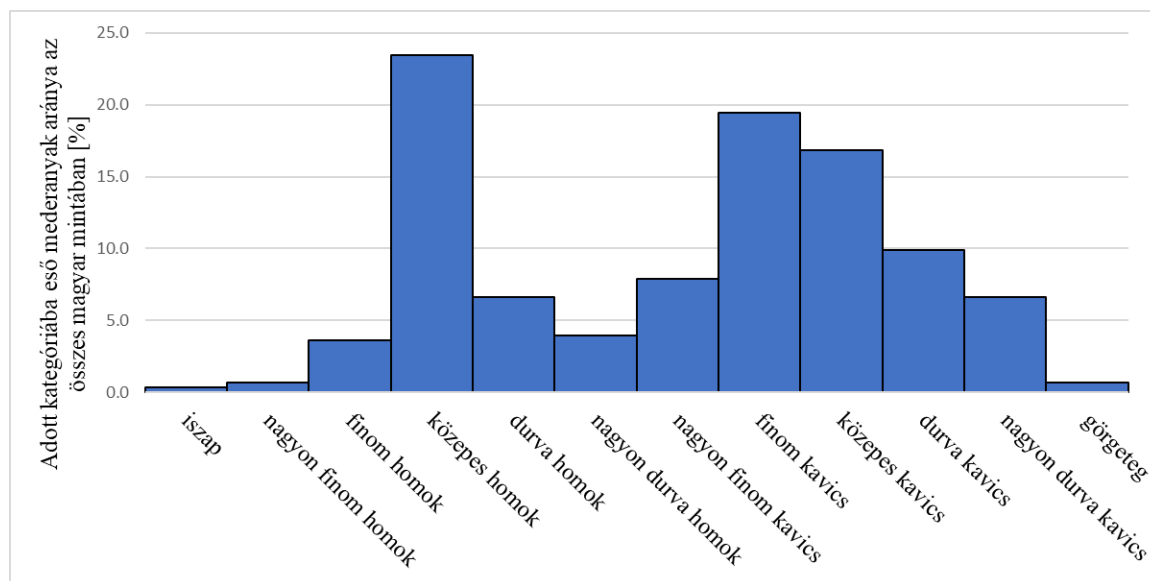
A VVT-n készített 2D áramlási modell a számomra szükséges adatokból a vízmélységet (h) és az áramlási sebességet (v) tartalmazta, valamint a szakaszokra jellemző Manning-féle érdességeket (n), ezek felhasználásával pedig kiszámítottam a fenék-csúsztatófeszültség értékeit is (τ) a Manning-féle képlettel közelítve:

$$\tau = \frac{\rho_{\text{víz}} * g * n^2}{h^{\frac{7}{3}}} * v^2$$

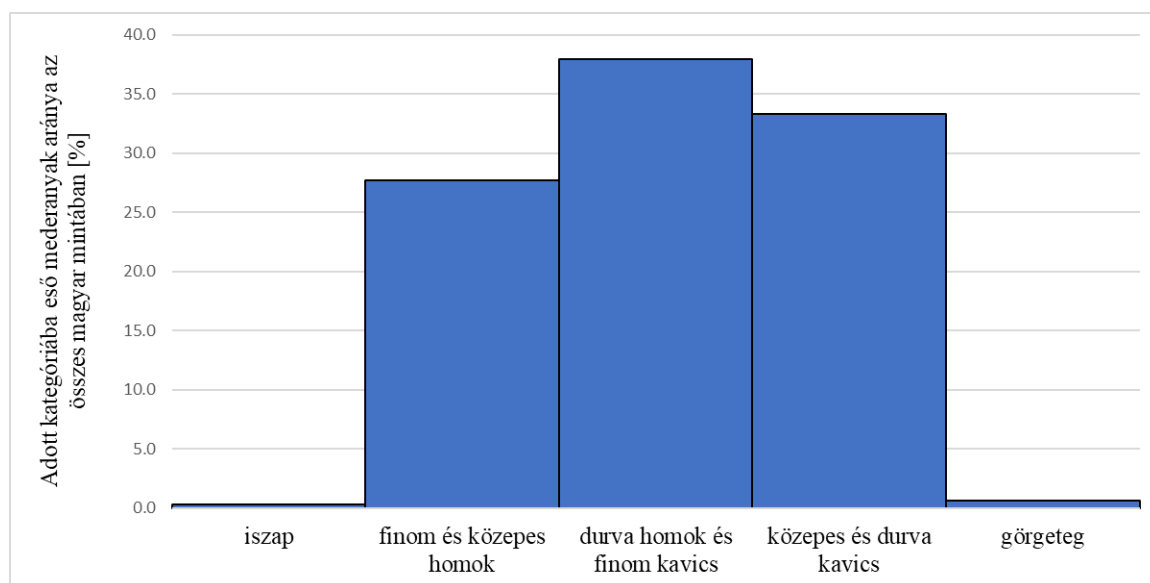
Mind a három szimulált vízjárási állapothoz (lásd korábban: LKHQ, KÖQ, LNQ) további statisztikai adatokat számítottam: átlagot, szórásnégyzetet, szórást és ferdeséget. Ennek az előkészítő lépésnek az eredményeképpen tehát a teljes hazai Dunára, térképszerű formában rendelkezésre álltak a jellemző vízjárási állapotokhoz tartozó vízmélység, áramlási sebesség és fenék-csúsztatófeszültség értékek, továbbá azok statisztikai jellemzői. Az RF modell tanításához és igazolásához minden mederanyag mintavételi pontban lekérdeztük az előbbi adatokat, míg a teljes 2D modell mezőit a későbbi mederanyag térkép készítéséhez használtuk fel.

3.4.2. Mederanyag adatok előkészítése

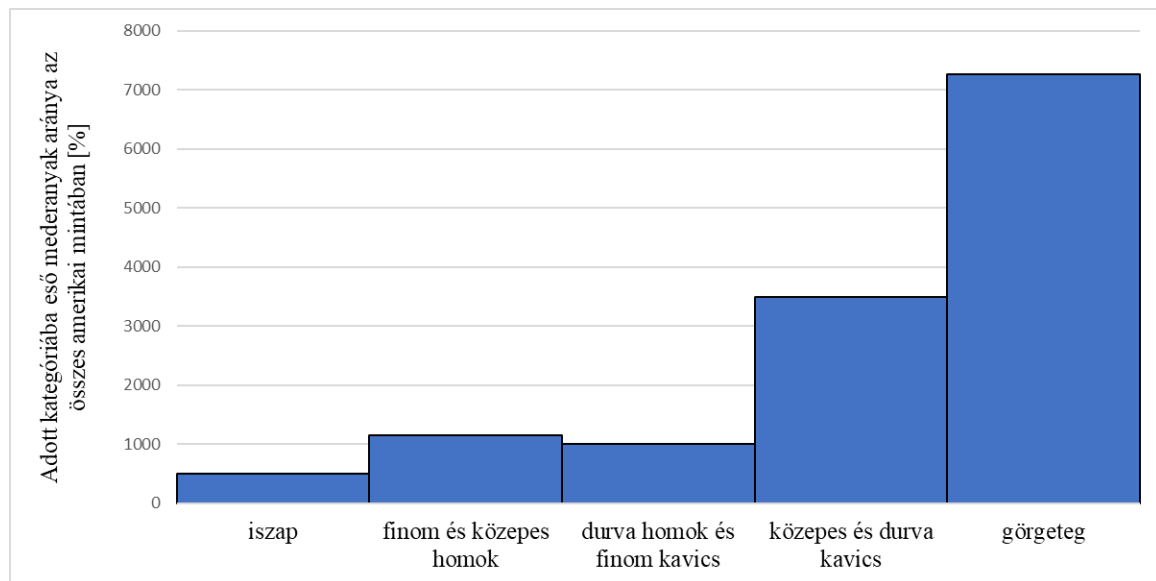
Az RF modellek tanításhoz használt különböző osztályokhoz tartozó adatok aránya (pl. különböző szemcseméret tartományban lévő adatok mennyisége) befolyásolja a modell pontosságát. A tanítás során ritkábban előforduló elemek az eredményben is ritkán fognak előfordulni, mivel a döntési fák kevesebb levelén jelenik meg eredményként (az RF modell elméleti hátterét a következő pont mutatja be). Emiatt az adatok lehetőség szerinti arányos eloszlására van szükség, esetünkben tehát fontos, hogy a tanításhoz használt adatok jellege a Duna hazai szakaszára valóban reprezentatív legyen. Az RF alapú mederanyag térképezési módszertől éppen azt várjuk, hogy a Duna változatos mederanyag összetételét leképezze, hiszen ott a homoktól a durva kavicsig széles tartomány megtalálható. Ahogy az 5. ábra illusztrálja a hazai mintaterületre vonatkozó adatokban a közepes homok és kavics jelenik meg leggyakrabban. Az osztályok összevonásával egyenletesebb eloszlást az alábbi kategóriákkal értem el (6. ábra): Iszap: $D_{50} < 0,0625$ mm, finom és közepes homok: $0,0625 \text{ mm} < D_{50} < 0.5$ mm, durva homok és finom kavics: $0,5 \text{ mm} < D_{50} < 2$ mm, közepes és durva kavics: $2 \text{ mm} < D_{50} < 64$ mm, görgeteg: $64 \text{ mm} < D_{50}$. Látszik, hogy a Duna adatok között iszap és görgetegből nincs megfelelő mennyiségű, ezt csak további mintavétellel lehet javítani, bár szerepük valószínűleg nem számottevő. A mederanyag szemcseméret itt bemutatott osztályozása az amerikai adatokban egyértelműen a görgeteg osztály dominanciáját jelzi (7. ábra), és mivel a többi osztály sokkal kisebb arányban van jelen, már azt sugallja, hogy a modell nem lesz adaptálható a Dunára.



5. Ábra: Adott kategóriába eső mederanyag aránya az összes hazai Dunára rendelkezésre álló mintában [%]



6. Ábra: Összevont kategóriákba eső mederanyag aránya az összes hazai Dunára rendelkezésre álló mintában [%]



7. *Ábra: Összevont kategóriába eső mederanyag aránya az összes Columbia folyóra rendelkezésre álló mintában [%]*

Az RFR modell átlagos hibája a finomabb szemcseméretben jelentős eltéréseket okozhat, ennek kiküszöbölésének érdekében a szemcseméret (D , [mm]) kettes alapú logaritmusát véve kiszámítottam a ψ értékeket. Az így származtatott adatokból logaritmikus szemeloszlási görbét készítettem. Ennek az előnye, hogy míg a szemcseméreték tömegszerinti eloszlásában az átlag és a medián nagy mértékben eltérhet, ez a kevésbé jelentkezik a ψ értékeknél (Bunte és Abt, 2001).

3.5. Az RF modellek paraméterezése

Az eredeti, Ren és társai (2020) által készített a Columbia folyóra felállított klasszifikációs RF modell nem került publikálásra, de az adatbázis igen. Az elérhető adatokból a vízmélység, sebesség és fenék-csúsztatófeszültségek átlagait, szórásnégyzeteit, szórásait és ferdeségeit (centrális momentumok) használtam fel, hogy reprodukálhassam a modellt. 750 fát állítottam be, 10-es maximum szintmélységgel, GINI index alapján tanítottam, úgy, hogy a fáknak nem kellett az összes tulajdonságnak szerepelnie. Az amerikai adatbázis 70%-át a modell tanítására, míg 30%-át az igazolásra fordítottam. A hazai Duna szakaszra rendelkezésre álló adatokkal a az előbb leírtakkal azonos beállításokkal futtattam új, saját modelleket is. Ez tehát azt jelenti, hogy összeállítottam egy adatbázist a dunai adatokból, majd azokat szintén tanító- és igazoló halmazokra bontottam.

A korábban leírtak szerint, két különböző típusú RF modellt állítottam fel. Egyfelől, egy klasszifikációs (RFC) modellt, amelyhez a mederanyag kategóriákat a 3.4.2. pontban leírtak szerint határoztam meg. Felállítottam továbbá két regressziós RF (RFR) modellt is, az egyikben a hidrodinamikai adatokat a mederanyag D_{50} értékeivel (mederanyag szemeloszlás görbe 50%-os értékéhez tartozó szemátmérő) kapcsoltam össze, a másikban a D_{50} kettes alapú logaritmusával (ψ), az előző pontban leírtak szerint. A két utóbbi paraméter fizikai tartalma nem tér el, de mivel az adathalmazban lévő értékek más tartományt fednek le, eltérő eredményt vártam. Az RFC és az RFR modellek között az alapvető különbség az, hogy míg az előbbi osztályokat ad eredményül (esetünkben a legvalószínűbb mederanyag osztályokat), tehát egyfajta kvalitatív jellemzést ad, a regressziós modell számértéket ad eredményül, tehát kvantitatív becslést tesz, esetünkben a lokális D_{50} értékekre.

4. Eredmények

4.1. Ren et al. (2020) modelljének adaptálása

Mivel a hivatkozott tanulmányhoz kapcsolódó modell közvetlenül nem állt rendelkezésre, de az adatai igen, magam reprodukáltam az RFC modellt az elérhető adatbázisuk alapján, melynek igazoló eredményeit az 1. táblázat mutatja be. Az egyes sorokban látható, hogy az adott mederanyag kategóriából hány mintát becsült az RFC modell a különböző mederanyag kategóriákba (oszlopok). A zöld háttérrel kiemelt átló mutatja a jól prediktált adatok számát. Az utolsó sor (Helyes predikciók aránya [%]) az egyes oszlopokba helyesen sorolt (zöld) mintaszámot és az összes abba a kategóriába prediktált mintaszám (oszlop szummája) arányát mutatja százalékos értékben. Az utolsó oszlop (Helyes beazonosítás aránya [%]) pedig a helyesen besorolt és az összes valóságban is oda tartozó adat (sor szummája) arányát tartalmazza. A kapott eredmény egybevágott a Ren és társai (2020) által publikált eredményekkel, így igazolást nyertünk a reprodukálás sikerességéről. A helyes predikciók aránya jól mutatja, hogy a modell által történt besorolás elfogadható, a véletlenszerű besorolásnál (ami 5 kategória esetében $\frac{1}{5} = 20\%$) jóval magasabb arányban helyes. Ez a lépés tehát azt jelentette, hogy sikeresen reprodukáltam a külföldi tanulmányban felállított modellt azzal a céllal, hogy megvizsgálhassam annak adaptálhatóságát a hazai Dunára.

A helyes beazonosítás aránya kevésbé biztató. A kisebb szemcséknél látható, hogy a mérés szerint odatartozó mintákat gyakran téves kategóriába sorolta be. Emiatt a valósághoz képest alulreprezentált lesz az eredmény. Ez a 7. ábra látható, mért mederanyagok eloszlásával egybevág. Ebből egy bizonyos mértékű hibára lehet számítani a modelleredményekben.

US	Predikció						Helyes beazonosítás aránya [%]
Validációs adat	Kategória	iszap	finom és közepes homok	durva homok és finom kavics	közepes és durva kavics	görgeteg	
	iszap	14	6	23	46	55	9.7
	finom és közepes homok	2	65	43	59	195	17.9
	durva homok és finom kavics	0	16	132	90	65	43.6
	közepes és durva kavics	1	12	57	506	455	49.1
	görgeteg	2	32	35	258	1856	85.0
Helyes predikciók aránya [%]		73.7	49.6	45.5	52.8	70.7	

1. táblázat: Az amerikai adatok alapján tanított RFC modell validációs eredményei

A következő lépés tehát az RFC modell Dunára való alkalmazhatóságát vizsgálja. A 2. táblázat az elsőhöz hasonlóan épül fel. Ez bemutatja, hogy a külföldi kutatásban felállított modell a hazai Dunán vett mederanyag mintavételi helyeken milyen sikerrel prediktálja a domináns mederanyag kategóriát. Ebben a vizsgálati lépésben a mintavételi helyeken jellemző hidrodinamikai viszonyokat már a 2D modellből vettük fel. Látható, hogy az amerikai modell egy kivételével minden hazai pontra görgeteg típusú mederanyag osztályt jósolt, holott a tényleges (teszt) adatok ennél jóval változatosabbak, ráadásul görgeteg a hazai Dunán nem jellemző. A modell adaptálhatósági korlátai tehát egyértelműen látszódnak. Az eltérő hidrológiai, hidrodinamikai, mederanyag összetételi és medermorfológiai viszonyok, összefoglalóan tehát az eltérő geohidromorfológiai jellemzők nem teszik lehetővé a modell átültetését.

HU	Predikció						Helyes beazonosítás aránya [%]
Teszt adat	Kategória	iszap	Finom és közepes homok	Durva homok és finom kavics	Közepes és durva kavics	Görgeteg	
	iszap	0	0	0	0	1	0.0
	finom és közepes homok	0	0	0	0	81	0.0
	durva homok és finom kavics	0	0	0	0	116	0.0
	közepes és durva kavics	0	0	0	1	100	1.0
	görgeteg	0	0	0	0	2	100.0
Helyes predikciók aránya [%]		-	-	-	100.0	0.7	

2. táblázat: Az amerikai adatok alapján tanított RFC modell magyar adatokra adott eredményei

4.2. A kutatás keretében felállított klasszifikációs RF modell

A hazai Dunára rendelkezésre álló adatok és a 2D modellből lekérdezett hidrodinamikai jellemzők alapján felállítottam egy új klasszifikációs RF modellt. Az új modell eredményeit a 3. táblázat tartalmazza, hasonló felépítésben, mint a korábbiakban bemutatott esetben. Látható, hogy mind a helyes predikciók, mind a helyes beazonosítás aránya messze jobb eredményt mutat a külföldi modellhez képest, minden esetben 85%-ot meghaladó az egyezés. Mivel görgeteg és iszap típusú mederanyag minták alig álltak rendelkezésre, az RF modell valójában csak három mederanyag osztályt fed le effektíven.

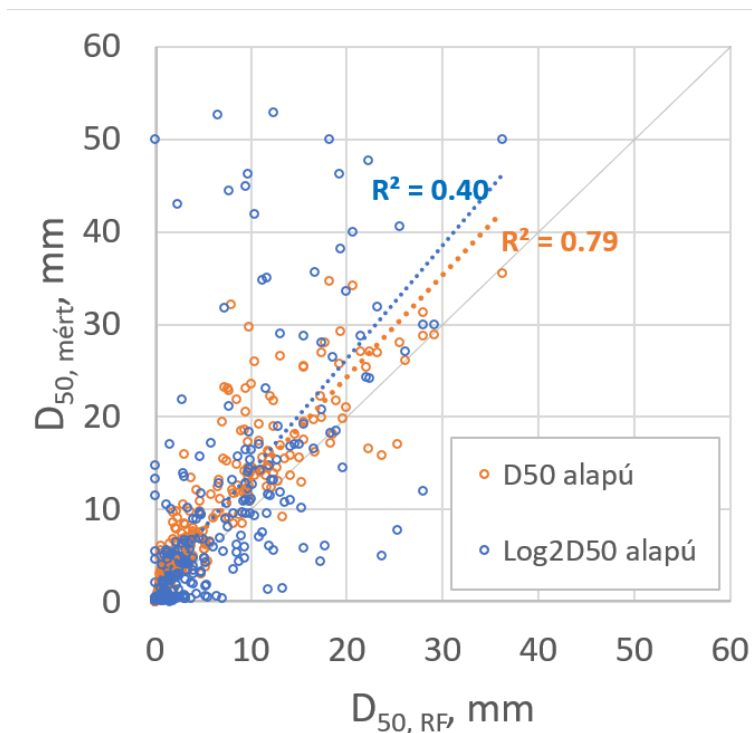
HU	Predikció						Helyes beazonosítás aránya [%]
Teszt adat	Kategória	iszap	finom és közepes homok	durva homok és finom kavics	közepes és durva kavics	görgeteg	
	iszap	0	0	0	0	0	-
	finom és közepes homok	0	29	2	0	0	93.5
	durva homok és finom kavics	0	3	35	2	0	87.5
	közepes és durva kavics	0	0	1	19	0	95.0
	görgeteg	0	0	0	0	0	-
Helyes predikciók aránya [%]		-	90.6	92.1	90.5	-	

3. táblázat: A hazai Duna mederanyag osztály adataival tanított RFC modell validációs eredményei.

4.3. A kutatás keretében felállított regressziós RF modell

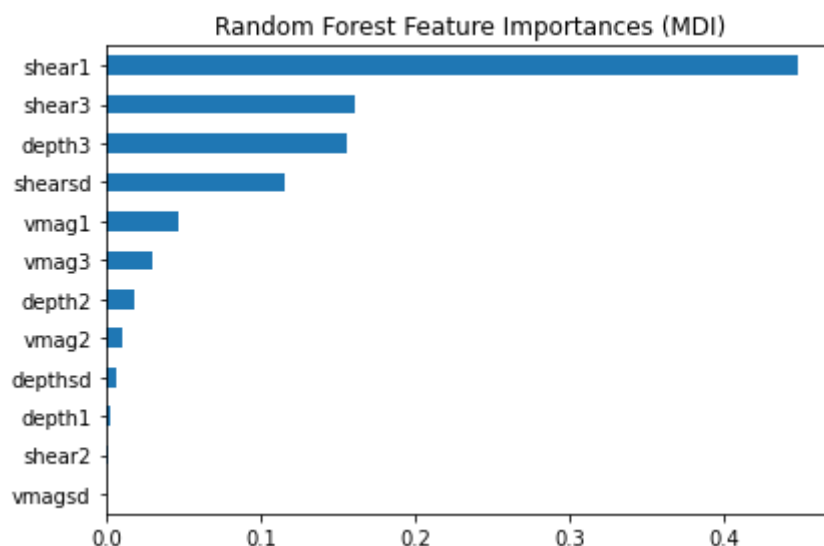
A két regressziós modell esetében az eredményeket másképp érdemes szemléltetni, mint a klasszifikációs esetben, mivel itt a konkrét D_{50} értékeket tudjuk összevetni a mért adatokkal. Az alábbi grafikon az igazoló adatokra kapott egyezéseket mutatja be a két tanítási módszer esetén. Azt látjuk, hogy az RF modell mindkét esetben kismértékben alulbecsli a medián szemátmérőket. A pontok szórása nagyobb, a regressziós együttható pedig alacsonyabb abban az esetben, ha a tanításnál nem közvetlenül a medián értékeket, hanem annak logaritmusát

alkalmazzuk (8. ábra). Az egyezés jóval erősebb abban a változatban, ahol közvetlenül a D_{50} értékekkel tanítjuk a modellt.

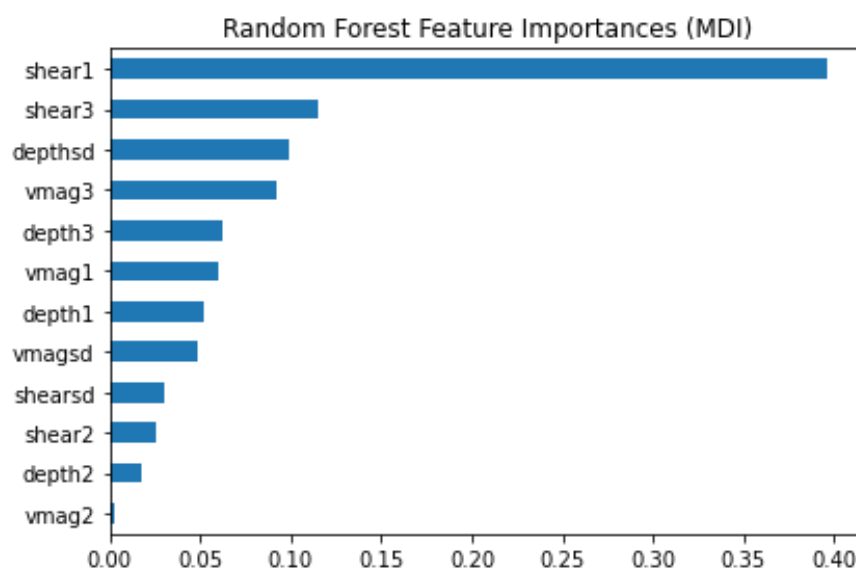


8. Ábra: A regressziós RF modell ellenőrzése a két tanítási módszer esetén.

Az RF modell lehetőséget ad arra, hogy megvizsgáljuk, a bemeneti adatok közül melyik, milyen súllyal vesz részt a döntések meghozatalában, vagyis jelen vizsgálatnál mely paraméterek determinálják a mederanyag összetételét. Ennek bemutatására az RF modell ún. Feature Importance elemzését használjuk, ami minden bemeneti paraméternek ad egy súlyszámot, melyek összege egy. Az alábbi grafikonok (9. és 10. ábra) alapján igazolást nyer, hogy a mederanyag összetételét leginkább a folyó adott pontjában jellemző fenék-csúsztatófeszültség értéke határozza meg. Ez az eredmény összhangban van más kutatások, pl. Miedema (2010) kutatási eredményeivel, de a hazai Duna vizsgálatára Baranya és társai (2018) is erre a következtetésre jutottak, bár más módszerrel. Mind a D_{50} , mind a $\log_2 D_{50}$ (vagyis ψ) alapú megközelítésnél ugyanerre a következtetésre jutottunk.



9. Ábra: A hazai Duna D50 adataival tanított RFR modell alapján kapott döntési súlysúlyszámok (depth=vízmélység, vmag=sebesség, shear=fenék-csúsztatófeszültség, 1=átlag, 2=szórásnégyzet, sd= szórás, 3= ferdeség)



10. Ábra: A hazai Duna ψ adataival tanított RFR modell alapján kapott döntési súlysúlyszámok (depth=vízmélység, vmag=sebesség, shear=fenék-csúsztatófeszültség, 1=átlag, 2=szórásnégyzet, sd= szórás, 3= ferdeség)

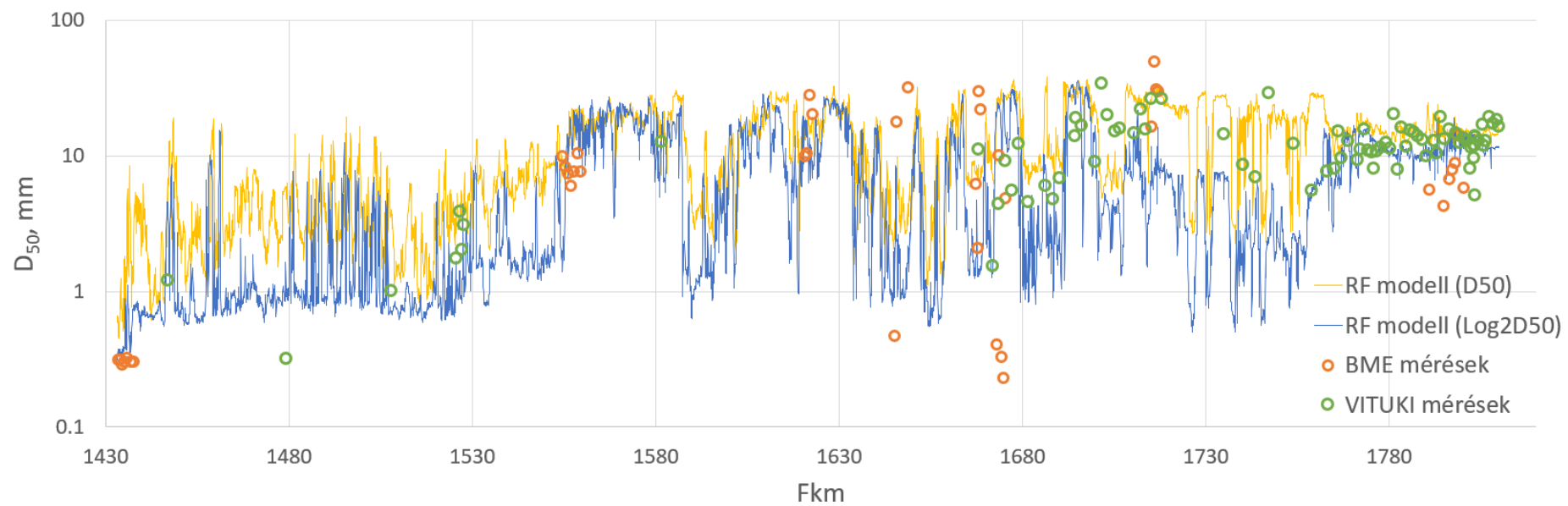
A fentiek alapján elmondható, hogy az új, általam felállított RF modellek megfelelően reprodukálják mind a mederanyag osztályokat, mind a medián szemátmérő lokális értékeit, továbbá igazoltam, hogy a hidrodinamikai jellemzők közül a hordalékelragadó erő játssza a legfontosabb szerepet a mederanyag összetételének alakulásában.

5. Eredmények értékelése

A vizsgálatok alapján belátható, hogy az amerikai Columbia folyóra felépített RF modell nem adaptálható megfelelően a hazai viszonyokra. Ennek oka többértű lehet. Bemutattam, hogy a jellemző mederanyag szemcseösszetétel jelentősen eltér a két mintaterületen, mert míg a hazai Duna szakaszon a homok-homokoskavics-kavics osztályok a meghatározók, a külföldi mintaterületen a görgeteg, tehát jóval durvább szemcseméret dominál. Ahogy leírtam, az RF modell pontossága nagyban függ a tanító adatok reprezentativitásától, és mivel a Dunára jellemző szemcseméret tartomány alulreprezentált volt a külföldi kutatásban, a pontosság ebben a tartományban lecsökkent. Fontos továbbá figyelembe venni, hogy jelentősen eltérő geohidromorfológiai jellemzőkkel bírt az amerikai folyószakasz, más közettípus és geometriai adottságokkal. Új paraméterek bevonásával mindazonáltal, mint pl. a közettípus, véleményem szerint elképzelhető, hogy felállítható egy olyan RF modell, ami univerzálisan alkalmazható. Ehhez célszerűen több folyó adataiból kell meríteni, melyet a sokféle tanítási adat bárhol bevethetővé tenne.

A hazai Duna mederanyag mintavételi adataira épített klasszifikációs RF modellel (RFC) öt mederanyag osztályba való kategorizálást hajtottam végre, tehát itt nem a szemcseméretet, hanem az osztályt modelleztem. Bemutattam, hogy az ötből három, a hazai Dunára kimondottan jellemző osztályt (i) finom és közepes homok, ii) durva homok és finom kavics, iii) közepes és durva kavics) a modell nagy megbízhatósággal (jellemzően > 90%-os egyezés) azonosította be. A részletesebb osztályozást az adatmennyiség korlátai miatt nem végeztem el, de a modell több tanító adat segítségével (tehát további helyszíni mederanyag mintavételek után) a további részletesebb osztályokba sorolásra is alkalmas lehet.

A kutatás legnagyobb előrelépését a regressziós RF modell (RFR) sikeres felépítése jelentette. Az RF modell segítségével kidolgoztam egy olyan módszert, amivel a 2D hidrodinamikai modell eredményezői alapján a medián (D_{50}) szemátmérő területi eloszlására tehetünk becslést. A betanított modell lehetőséget ad pl. a medián szemátmérő hazai Duna hossza mentén változó szemcseméretek becslésére. Az alábbi grafikon (11. ábra) a két tanítási módszerrel (D_{50} és $\log_2 D_{50}$ alapú) kapott eloszlásokat mutatja be. A grafikonon felhasználtuk a VITUKI 1970-90 között végzett mérésének eredményeit is, amit Kutai (2017) tanulmányából vettem át.



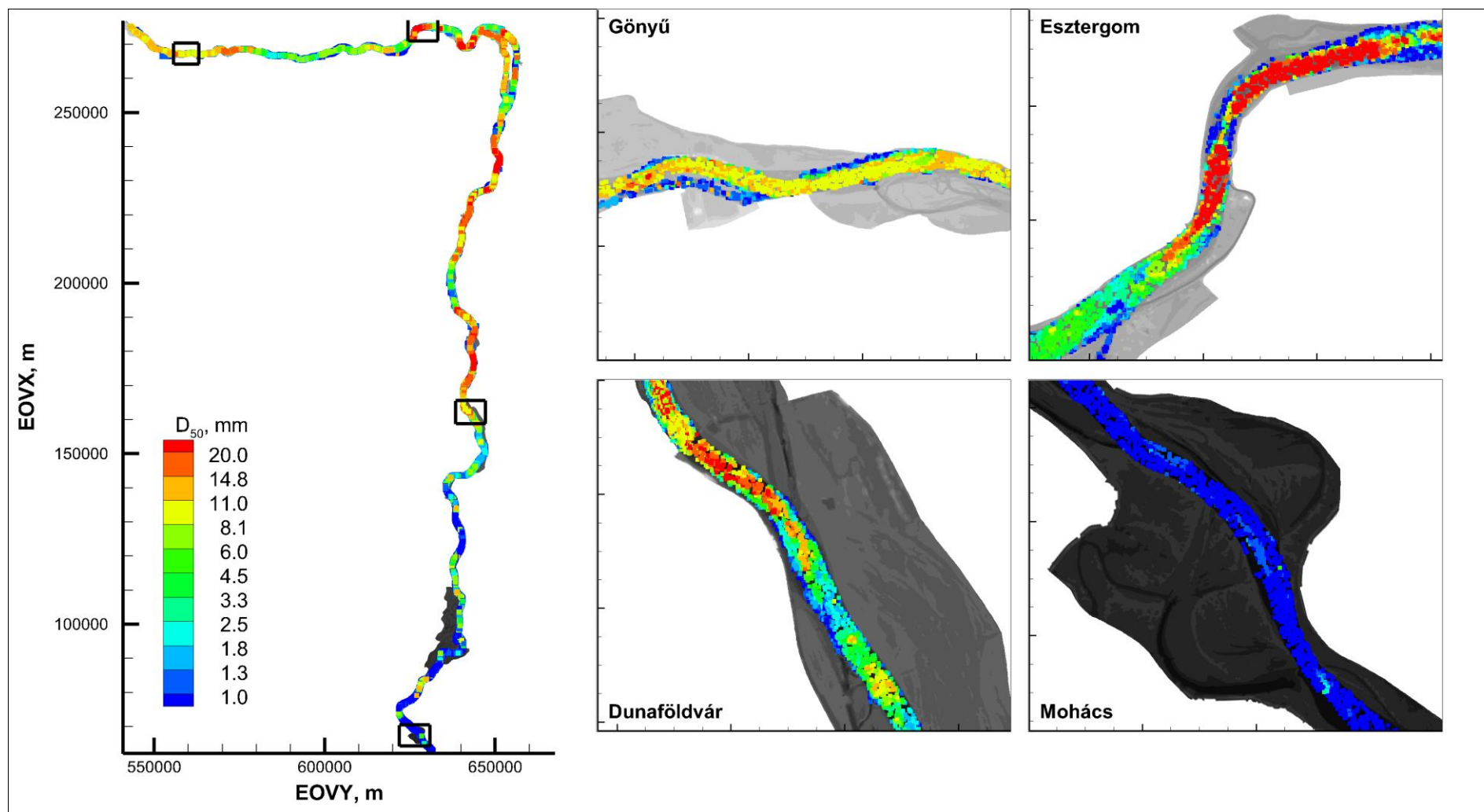
11. Ábra: Az RF modellek által adott és korábbi mérések eredményeiből felvett medián szemátérők hossz-menti alakulása a hazai Duna szakaszon.

Ezen a grafikonon jól érthető, hogy miért választottuk a $\log_2 D_{50}$ alapú megközelítést a modellben. Hiába mutattam be, hogy összességében a másik megközelítés pontosabb becslést tesz, a Duna alsó szakaszán található homokmedrű részen (kb. Pakstól indulva, 1530 fkm környezetétől) a modell rendre felülbecsli a szemátmérőket. A logaritmus alapú tanításnál viszont ez a változás a Duna mederanyagában jóval pontosabban tetten érhető.

A betanított modell eredményeit a Duna teljes magyarországi szakaszára kiterjesztve a 2D szimulációk alapján, elő tudtam állítani a D_{50} területi eloszlásának mezőjét is (12. ábra). Megfigyelhető, hogy az RF modell megfelelően tükrözi a mederanyag mintázásból kapott szemcseméreteket a folyó egyes szakaszain. Különösen jól megmutatkozik a mederanyag változékonny jellege, a térképről kiragadott néhány területen vett mederminták képeit is megvizsgáljuk (13. ábra). A modell jól visszaadja pl. az Esztergomnál lokálisan feldurvuló mederanyagot, a Dunaföldvári szakaszon lévő eséstörés mederanyagra kifejtett hatását, vagy a Mohács környékén található tisztán homok medret.

A jövőbeli terveim egyike ennek a modellnek a továbbfejlesztése, további adatok bevonásával, és további terepi mintagyűjtéssel. Az RFR modell ezen eredmények alapján akár képes lehet a D_{50} -hez hasonlóan D_{10} és D_{90} , vagy éppen a szemösszetételi görbe szórásának (σ) meghatározására, amelyből szemeloszlási görbe is szerkeszthető lenne.

Fontos, és korábbi vizsgálatok alapján (Baranya et al., 2018) várható eredmény volt, hogy a mederanyagban domináló szemcseméreteket a hordalékelragadó erő, vagy másnéven a fenék-csúsztatófeszültség határozza meg leginkább. Az RF modell segítségével számszerűen kiértékelhető volt a szemcseméretben játszott súlya ennek a paraméternek. Ezen eredmény alapján továbbfejlesztési irányként megfogalmazható, hogy az RF modell kiértékeléséhez felhasznált 2D hidrodinamikai modell helyett akár 1D, akár 3D szimulációs adatokat használni. Előbbi kifejezett előnye lenne, hogy teljes folyórendszerek vizsgálatára teheti alkalmassá a módszert, bár eredményként csak folyószelvényekre jellemző szemösszetételt adhat, tehát a területi változékonyság kiértékeléséről le kell mondanunk. 3D modellektől ezzel szemben a fenék-csúsztatófeszültség pontosabb becslését várjuk, amivel a mederanyag összetételének még pontosabb modelljét kaphatjuk.



12. Ábra: Az RF modellek által modellezett medián szemátmérő területi eloszlása a teljes hazai Dunán és néhány kiragadott helyszínen.

Gönyű



Esztergom



Dunaföldvár



Mohács



13. Ábra: Tipikus mederanyag összetételek a Duna egyes szakaszain

6. Következtetések

A TDK kutatás keretében felállítottam egy Random Forest alapú modellt folyók mederanyag szemösszetételének vizsgálatára. A modellt a hazai Duna szakaszon végzett mederanyag mintavételek alapján tanítottam be, és igazoltam annak alkalmazhatóságát mind az ún. klasszifikációs mind az ún. regressziós vizsgálatokra. A modell alapjául egy a teljes Dunára rendelkezésre álló 2D hidrodinamikai modell szolgált, aminek segítségével a betanított algoritmus a becsült szemcseméret osztályok és medián értékek területi eloszlásait állítottam elő. A modell megfelelően reprodukálta a Duna hazai szakaszára jellemző nagyon változatos mederanyag összetételt, képes volt a lokálisan feldurvuló szakaszok beazonosítására, és az alvízi irányban jellemző finomodás bemutatására. Az eredmények jelentős tudományos értéket képviselnek, mivel segítségükkel pontosabban megérthetjük folyók geomorfológiai sajátosságainak alakulását, alkalmas lehet adathiányos területeken adatpótlásra, térképszerű bemeneti adatokat képes szolgáltatni morfordinamikai szimulációk számára, de alapját képezheti folyórestaurációs tervezéseknek, vagy az ún. élőhely-hidraulikai vizsgálatoknak.

7. Köszönetnyilvánítás

Ezúton szeretném köszönetemet kinyilvánítani konzulenseimnek: Dr Baranya Sándornak és Ermilov Alexander Anatolnak, valamint VVT további munkatársai felé, akik lehetővé tették a helyszíni mintavételeket. Továbbá köszönet a Kék Bolygó alapítványnak, mely anyagi támogatást nyújtott a kutatás ideje alatt.

8. Ábrajegyzék

1. Ábra: Serleges mintavevő eszköz	8
2. Ábra: A duna különböző szakaszain vett jellemző mederanyag	8
3. Ábra: A 2D modellel szimulált áramlási sebesség megoldásmezője a Duna kisvízi állapotánál a folyó 1785-1791 fkm közötti szakaszán.	9
4. Ábra: Döntésifa működési mechanizmusának sematikus ábrája	10
5. Ábra: Adott kategóriába eső mederanyag aránya az összes hazai Dunára rendelkezésre álló mintában [%]	13
6. Ábra: Összevont kategóriákba eső mederanyag aránya az összes hazai Dunára rendelkezésre álló mintában [%]	13
7. Ábra: Összevont kategóriába eső mederanyag aránya az összes Columbia folyóra rendelkezésre álló mintában [%]	14
8. Ábra: A regressziós RF modell ellenőrzése a két tanítási módszer esetén.	19
9. Ábra: A hazai Duna D50 adataival tanított RFR modell alapján kapott döntési súlyszámok (depth=vízmélység, vmag=sebesség, shear=fenék-csúsztatófeszültség, 1=átlag, 2=szórásnégyzet, sd= szórás, 3= ferdeség)	20
10. Ábra: A hazai Duna ψ adataival tanított RFR modell alapján kapott döntési súlyszámok (depth=vízmélység, vmag=sebesség, shear=fenék-csúsztatófeszültség, 1=átlag, 2=szórásnégyzet, sd= szórás, 3= ferdeség)	20
11. Az RF modellek által adott és korábbi mérések eredményeiből felvett medián szemátérők hossz-menti alakulása a hazai Duna szakaszon.	22
12. Az RF modellek által modellezett medián szemátmérő területi eloszlása a teljes hazai Dunán és néhány kiragadott helyszínen.	24
13. Tipikus mederanyag összetételek a Duna egyes szakaszain	25

9. Irodalomjegyzék

1. Baranya, S, Fleit, G, Józsa, J, et al. (2018): Habitat mapping of riverine fish by means of hydromorphological tools. *Ecohydrology*.; <https://doi.org/10.1002/eco.2009>
2. Baranya, S., és Józsa, J. (2006). Flow analysis in river Danube by field measurement and 3D CFD turbulence modelling. *Periodica Polytechnica Civil Engineering*, 50(1), 57-68.
3. Benkő, G., Baranya, S., Török, T.G., Molnár, B. (2020): Folyami mederanyag szemössze-tételének vizsgálata Mély Tanulás eljárással drónfelvételek alapján (in English: Analysis of composition of river bed material with Deep Learning based on drone video footages). *Hidrol. Közlöny* 2020, 100, 61–69.
4. Biau, G., és Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197-227.
5. Bunte, K., Abt, R.S. (2001): Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analyses in sediment transport, hydraulics, and streambed monitoring. *Gen. Tech. Rep. RMRS-GTR-74*. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 428 p
6. Buscombe, D. (2013): Transferable wavelet method for grain-size distribution from images of sediment surfaces and thin sections, and other natural granular patterns. *Sedimentology* 2013, 60, 1709–1732.
7. Chandler, J., Lane, S. N. , Ashmore, P. (2000): Measuring river-bed and flume morphology and parameterising bed roughness with a Kodak DCS460 digital camera. January 2000. *Archives of Photogrammetry, Cartography and Remote Sensing XXXIII*.
8. Detert, M., Weitbrecht, V. (2013). User guide to gravelometric image analysis by BASEGRAIN. In: *Advances in Science and Research*, S. Fukuoka, H. Nakagawa, T. Sumi, H. Zhang (Eds.), Taylor és Francis Group, London, ISBN 978-1-138-00062-9, pp. 1789-1795.
9. Ermilov, A.A, Baranya, S., Török, G.T. (2020): Image-Based Bed Material Mapping of a Large River. *Water* 2020, 12(3), 916; <https://doi.org/10.3390/w12030916>
10. Füstös, V., Baranya, S., Fleit, G., Erős, T., Szalóky, Z., és Józsa, J. (2019). A felső-magyarországi Duna élőhelyszempontú hidrodinamikai vizsgálata *PISCES HUNGARICI*, (13), 81-90.

11. Gaebele, T., és Guti, G. (2010). Halivadék fajegyüttesek élőhelyhasználata a gödi Duna-szakaszon PISCES HUNGARICI, 4, 9-16.
12. Hong Han – Xiaoling Guo – Hua Yu: Variable selection using mean decrease accuracy and mean decrease gini based on random forest, 2016. 08
13. Kutai R. D. (2017), A Duna mederanyagának vizsgálata
14. Ray, C., Jasperse, J., és Grischek, T. (2011). Bank filtration as natural filtration. In Drinking Water Treatment (pp. 93-158). Springer, Dordrecht.
15. Ren, H., Hou, Z., Duan, Z., Song, X., Perkins, W. A., Richmond, M. C., ... és Scheibe, T. D. (2020). Spatial Mapping of Riverbed Grain-Size Distribution Using Machine Learning. *Frontiers in Water*, 2, 551627.
16. Sape A. Miedema : Constructing the Shields curve, a new theoretical approach and its applications. See also other papers 29nt he Shields curve. Conference: World Dredging Conference XIXAt: Beijing, China, 2010.
17. Sherwood, C. R., és Creager, J. S. (1990). Sedimentary geology of the Columbia River estuary. *Progress in Oceanography*, 25(1-4), 15-79.
18. Szűcs, P., és Mikita, V. (2020). Felszín alatti vízkészleteink és a hidrogeológiai kutatások helyzete hazánkban. *Hidrológiai közlöny*, 96 évf., 1.szám, pp 7-20.
19. Tikász G., Schrott M. (2021) Folyók mederanyagának vizsgálata kép alapú eljárásokkal <http://tdk.bme.hu/EMK/viz9/Folyok-mederanyaganak-vizsgalata-kep-alapu>
20. Ulrich, C., Hubbard, S.S., Florsheim, J., Rosenberry, D., Borglin, S., Trotta, M., Seymour, D., Riverbed Clogging Associated with a California Riverbank Filtration System: An Assessment of Mechanisms and Monitoring Approaches, *Journal of Hydrology*, Volume 529, Part 3, 2015, pp 1740-1753,
21. UTIBER-Viziterv-BME (2020): „A NIF Zrt. beruházásában megvalósuló a Transzeurópai Közlekedési Hálózat – TEN-T belvízi út fejlesztéséhez kapcsolódó tervezői feladatok ellátása tervezési szerződés keretében” 2014-HU-TMC-0606-S. HELYZETÉRTÉKELOŐ TANULMÁNY. Szakértői jelentés. Budapest. 2020.