

## Halak fajkompozíciója és hidromorfológiai változók közötti összefüggések a Duna magyarországi szakaszán

Jandó Benedek<sup>1</sup>, Füstös Vivien<sup>2,3</sup>, Ermilov Alexander Anatol<sup>4</sup>, Szalóky Zoltán<sup>5,6</sup>, Erős Tibor<sup>7,8</sup>, Baranya Sándor<sup>3,4</sup>

<sup>1</sup> Állatorvostudományi Egyetem, Zoológiai tanszék, Budapest (e-mail: jandoben@gmail.com)

<sup>2</sup> HUN-REN-BME Vízgazdálkodási Kutatócsoport, Budapest

<sup>3</sup> Víztudományi és Vízbiztonsági Nemzeti Laboratórium, Budapesti Műszaki és Gazdaságtudományi Egyetem, Építőmérnöki Kar, Vízépítési és Vízgazdálkodási Tanszék, Budapest

<sup>4</sup> Budapesti Műszaki és Gazdaságtudományi Egyetem, Építőmérnöki Kar, Vízépítési és Vízgazdálkodási Tanszék, Budapest

<sup>5</sup> HUN-REN Ökológiai Kutatóközpont, Vízi Ökológiai Intézet, Budapest

<sup>6</sup> Víztudományi és Vízbiztonsági Nemzeti Laboratórium, HUN-REN Ökológiai Kutatóközpont, Vízi Ökológiai Intézet, Budapest

<sup>7</sup> HUN-REN Balatoni Limnológiai Kutatóintézet, Tihany

<sup>8</sup> Víztudományi és Vízbiztonsági Nemzeti Laboratórium, HUN-REN Balatoni Limnológiai Kutatóintézet, Tihany

DOI: 10.59258/hk.15656



### Kivonat

Hiányos ismeretekkel rendelkezünk a halfajösszegek szerkezetét meghatározó abiotikus tényezők jelentőségéről nagy folyókban. Kutatásunkban a Duna magyarországi szakaszán 2004. és 2022. között végzett halfaunisztikai felmérésekben szereplő 20 leggyakoribb halfaj előfordulásaihoz rendeltünk vízügyi adatsorokból, illetve hidrodinamikai szimulációkból adatokat és ezek között gépi tanulás (Machine Learning-ML) segítségével kerestünk összefüggéseket. Az elemzésben független változóként feltüntetett 9 abiotikus tényező közül a sodrás sebessége, a vízmélység és a mederanyag összetétele bizonyultak a legmeghatározóbb változóknak, ami összecseng a korábbi kutatások eredményeivel. Emellett, az általunk készített Random Forest modell alkalmasnak bizonyult adott körülmények mellett az elemzett halfajok számosságának prediktálására a Duna teljes hazai szakaszán. Ezek a becslések az egyes halfajok számára optimális élőhelyeket számszerűsítik. A ML elemzés mellett lehetőségeket mutattunk a nagy teret és időt lefedő halfaunisztikai adatbázis populációs kapcsolatok vizsgálatában (mint például invazív és őshonos fajok viszonya) való használatára klasszikus statisztikai módszerekkel. Az itt kapott eredmények sok esetben összeesnek a random forest modellel, azonban okot adnak a modell további független változókkal való bővítésére a dunai halfajok ökológiájának jobb megértése érdekében.

### Kulcsszavak

Hal-élőhely kapcsolatok, gépi tanulás, ökohidraulika, regressziós Random Forest, folyami élőhely értékelés.

## Relations between composition of fishes and hydromorphological variables in a very large river, in the Hungarian section of the Danube River

### Abstract

Understanding of the contribution of abiotic drivers to fish community structure in very large rivers is poor. In this study, we assigned the occurrence data of the 20 most common fish species in the Hungarian section of the Danube River from 2004 to 2022 to hydrodynamic variables from hydrological datasets and hydrodynamical simulations and looked for patterns among them using Machine Learning (ML). Among the nine abiotic factors used as independent variables in the analysis, the depth-averaged flow velocity, water depth and bed material composition were the most decisive variables, which aligns with the results of previous research. In addition, with our Random Forest model, we were able to predict the number of individuals of the 20 most common fish species in the given conditions in the entire Hungarian section of the Danube. These estimates refer to optimal habitat for fish species according to abiotic variables. In addition to the ML analysis, we showed the possibility of using the Danube fish faunistic database, which covers a large area and time, to investigate the relationships of the population (for example, the relationship between invasive and native species) using classical statistical methods. The results found here are in many cases consistent with the Random Forest model but give reason to extend the model with additional independent variables to better understand the ecology of the Danube fish species.

### Keywords

Ecohydrology, Machine Learning, fish-habitat relations, regression Random Forest, river habitat evaluation.

### BEVEZETÉS

Hazánk vízrajza Duna-központú, az ország folyóhálózatának főtengegyét ez a folyó alkotja, illetve a felszín alatti vízhalózathoz is szervesen kapcsolódik (*A Víz keretirányelv hazai megvalósítása 2009*). Ebből adódóan hatalmas szerepet játszott hazánk jelenlegi tájszerkezetének kialakításában. A hazai tájtípológiai rendszer jelentős részét teszik ki folyókkal kapcsolatba hozható tájtípusok, amelyek

nagy része a Duna mentén is megtalálható (*Tózsza 1998*). Ezenkívül központi folyónkként a kereskedelmi, a vízgazdálkodási, a lakossági és a mezőgazdasági kérdésekben is évszázadok óta jelentős szerepet kap, aminek eredménye lett a folyószabályozás és árvízmentesítés a teljes hazai szakaszon (*A Víz keretirányelv hazai megvalósítása 2009*), amely jelentősen átformálta a Dunához kapcsolódó vizes élőhelyek működését (*Farkas-Iványi és Trájer 2015*).

A mesterséges intelligencia valósággal berobbant a tudományba. Mivel rengeteg új lehetőséget hozott szinte minden területen, így nem meglepő, hogy hamar elkezdtek használni az ökológiában is. Napjainkra gyakorlatilag a mesterséges intelligencia minden megközelítését (neurális hálózatok, genetikai algoritmusok, random forestrendszerek, deep learning algoritmusok) használták már ökológiai kutatásokban (Sylvain és társai 2019).

A Machine Learning (ML) algoritmusok alkalmazásának fő célja, hogy mintázatot fedezzenek fel adatsorok között. Az ökológusok számára ez azért rendkívül előnyös, mert így lehetővé válik a bonyolult, nem lineáris adatsorok elemzése, ami gyakran előfordul komplex ökológiai kérdéseknél (Olden és társai 2008). A Random Forest (RF) módszer a ML megközelítések egy típusa, amely remekül alkalmazható ökológiai felhasználásban is változók előre jelzésére (Bergström és társai 2011, Liu és társai 2018, Elmahdy és társai 2020), ugyanis a klasszikus statisztikai módszerekkel szemben, könnyen vizsgálhatóak vele komplex, nem lineáris kapcsolatok is.

A módszer általunk alkalmazott verziója, a regressziós RF az ún. „regressziós fákon” (regression tree – RT) alapul, amelyek a „döntés fák” (decision tree – DT) numerikus változókból álló típusai. Ezek a fák egy adott változó becsülését, vagy klaszterezését teszik lehetővé számos változó előre kiszámított küszöbértéke mentén történő logikai döntések segítségével (Olden és társai 2008). Az egyes adatsorokra felépített RT-k a racionális emberi döntések analógiáját követik és a komplexebb problémákat lineáris szegmensekre bontva segítenek az előrejelzésben. Hátrányuk azonban, hogy egyáltalán nem rugalmasak. Jól használhatók azon az adatsoron, amin létrehozták, azonban nem lehet velük interpolálni más adatbázisokat, mert nem képesek alkalmazkodni az ott tapasztalható más küszöbértékekhez, vagy a kategóriák eltérő kapcsolataihoz. Itt jön képbe a Random Forest (RF), ami ötvözi az RT-k hatékonyságát és a ML flexibilitását, bár extrapolálni a betanításra használt adatsor tartományán kívülre ez sem képes. Az RF lényege, hogy egy adott tanító adatbázis alapján random választott változók és küszöbértékek segítségével létrehoz számos RT-t. Majd az adott változó(k) új adatsorok alapján történő becsüléséhez az új adatbázis minden egyes magyarázó változójának értékét lefuttatja az összes, korábban elkészített RT-n. Regressziós RF esetében a fák által megszavazott

értékek átlaga lesz az előrejelzett érték. Az RF tehát több száz, vagy akár ezer RT összessége, amelyek mindegyikét felhasználja az új becslésekhez, ezáltal képes alkalmazkodni az új adatsorokban található ismeretlen esetekhez is (Olden és társai 2008).

Kutatásunk célja egy olyan RF modellt felállítása volt, amellyel képesek vagyunk különböző körülmények esetén megbecsléni az egyes halfajok számára optimális élőhelyeket a Duna magyarországi szakaszán, valamint számszerűsíteni tudjuk az egyes abiotikus faktorok halak előfordulásában betöltött szerepét. Célunk volt emellett alapot nyújtani további élőhelyalkalmassági modellek finomításához.

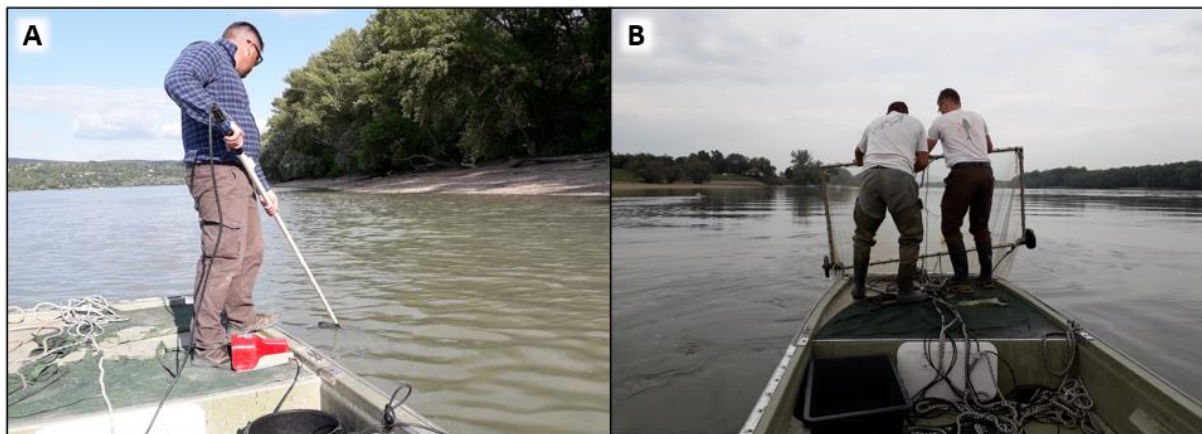
## ANYAG ÉS MÓDSZER

### Duna halfaunisztikai adatbázisa

#### Általános jellemzés és mintavételi módszerek

Az elemzésekhez használt halfaunisztikai adatbázis a Balatoni Limnológiai Kutatóintézet, valamint a HUNREN Ökológiai Kutatóközpont munkatársai által, különböző projektekhez kapcsolódóan 2004 és 2022 között végzett felmérésekből került összeállításra és az 1668 mintavétel nagyjából egyenletesen lefedi a Duna teljes hazai szakaszát. A mintavételezés bár standard módszerekkel, de nem standard időközönként történt és a minták térbeli eloszlása sem egyenletes.

A mintavételezés két módszerrel történt április és november között, nappal és éjjel is. Az egyik a partközeli, vagy parti elektromos módszer (PE) (1. ábra), mindkét napszakban. Ebben az esetben a part mellett, az áramlással együtt sodródó csónakból közvetlenül a parti zónából történik a halak gyűjtése, leggyakrabban 500 méteres, míg általánosságban 45-800 méter hosszú szakaszon. Az elektromos bentikus keretes húzóhálóval (EBKH) pedig a parttól távoli területek vizsgálata történt nappal (1. ábra). Ebben az esetben egy halászgép (leggyakrabban a Hans-Grassl EL65 IIGI eszköz) alkalmazásával, a legtöbb esetben 500 méteres, míg általánosságban 300-870 méteres szakaszon történt a halászat (Szalóky és társai 2014). A két módszer együtt megfelelően lefedi a mederszelvényt, így kombináltan alkalmasak a folyók halfaunájának értékelésére (Zajicek és Wolter 2018).



1. ábra. Parti elektromos halászat (PE) (A) és az elektromos bentikus keretes húzóháló (EBKH) (B) a gyakorlatban (Fotó: Füstös V.)

Figure 1. Shoreline electrofishing (SE) (A) and an electric benthic frame trawl (EBFT) (B) in practice (Photo by V. Füstös)

### Adatbázis tisztításának módszerei

Az adatbázisból eltávolítottuk a pontatlan, hiányos, vagy az általunk vizsgált területen, tehát a Duna magyarországi szakaszán túlmutató adatokat. A redundáns, vagy a további elemzések szempontjából irreleváns változókat (például a vízparti növényzet becslései) szintén töröltük. Végül az adatbázis szöveges megjegyzéseit feldolgozva szűrtünk ki a további rekordokat.

Az elemzésekhez használt koordináta-párokat az alábbiak szerint alkalmaztuk: abban az esetben, amikor mindkét koordináta-pár rendelkezésre állt (minták 70%-a), ott a kettő számtani közepeként előálló szakaszfelezőpontot számítottuk ki. Abban az esetben, ahol csak a kezdő pont állt rendelkezésünkre (minták 30%-a), ezt a koordináta-párt vettük alapul. Az elemzésekhez használt koordinátákat az Egységes Országos Vetületi rendszerben (EOV) értelmeztük.

### Fajpárok közötti kapcsolatok elemzése

A halfaunisztikai adatbázis elemzéséhez az RF mellett használtunk klasszikus statisztikai módszereket is, amelyekkel lehetőségeket mutatunk a populációs kapcsolatok nagyléptékű faunisztikai adatbázisokon keresztül történő vizsgálatára és összehasonlításra az itt kapott értékeket az RF eredményeivel. Ehhez az elemzéshez két fajpárnál határoztuk meg, hogy amikor mindkettő jelen van egy adott helyen, akkor az egyik faj milyen hatással van a másik egyedszámára.

A két fajpár esetén listáztuk azokat a mintákat, amelyekben mindkét faj előfordul, ugyanis abban az esetben, amikor egy faj nincs jelen egy mintában, a probléma ökológiai komplexitása miatt nehéz meghatározni, hogy ez a másik faj miatt van, vagy valami teljesen más, általunk nem is mért változó miatt. Ezt követően közös szórásdiagramon ábrázoltuk az adatpárokat. Minden szórásdiagramra illesztettünk egy  $f(x) = x$  referenciaegyenest is, amely azt a hipotetikus esetet szemlélteti, amikor a két faj egyedszáma egyenesen arányos egymással. Az alapján, hogy a pontok milyen mintázatot követtek, illetve az egyenes melyik oldalán sűrűsödtek, felállítottuk az ellenhipotézist, majd kiszámítottuk a Pearson-féle korrelációs együtthatót a két változó szimmetrikus kapcsolatának értékeléséhez. A korrelációs együttható után fajonként kiszámítottuk az egyedszám-arányokat (a két faj együttes egyedszámának mekkora részét teszi ki az egyik, és mekkora részét a másik). Az így kapott párosított minták átlagát kétmintás, páros t-próbával hasonlítottuk össze egyoldali ellenhipotézissel, vagyis R-ben t.test(paired = T). Mivel az egyedszám-párok minden esetben a kis értékeknél sűrűsödtek és a nagyobbak felé haladva minden irányban erősen ritkultak, így a különbségeik sem követtek normális eloszlást, azonban a különbségek tapasztalati eloszlása leginkább egyenletes volt és több csúcsú, nem pedig ferde. Viszont a nagy mintaelemszám (30 <= n) köszönhetően közelítőleg tudtuk alkalmazni a kétmintás, páros t-próbát (Reiczigel és társai 2010).

### Random Forest modellezés

#### A 20 leggyakoribb halfaj vizsgálata

Az RF modellben a 20 leggyakoribb (legnagyobb egyedszámban jelenlévő) halfaj egyedszám-adatait, a minták fajszámát, illetve egyedszámát használtuk fel függő változóként, valamint négy mintavételi körülményt leíró és öt abiotikus változót független változóként. A négy körülményt leíró változó a dunai halfaunisztikai adatbázisból származott. A 20 leggyakoribb halfajt az alábbiak alkották: szélhajtó kűsz (*Alburnus alburnus*), bagolykeszeg (*Blicca bjoerkna*), paduc (*Chondrostoma nasus*), selymes durbincs (*Gymnocephalus schraester*), balin (*Leuciscus aspius*), jászkeszeg (*Leuciscus idus*), menyhal (*Lota lota*), folyami géb (*Neogobius fluviatilis*), kerekfejű géb (*Neogobius melanostomus*), csapósügér (*Perca fluviatilis*), kessler-géb (*Ponticola kessleri*), halványfoltú küllő (*Romanogobio vladkyovi*), bodorka (*Rutilus rutilus*), leánykoncér (*Rutilus virgo*), fogassüllő (*Sander lucioperca*), kőszüllő (*Sander volgensis*), fejes domolykó (*Squalius cephalus*), szilvaorrú keszeg (*Vimba vimba*), német bucó (*Zingel streber*), magyar bucó (*Zingel zingel*).

#### Abiotikus változók a modellben

Az öt abiotikus változóból egy, a vízmélység, a dunai halfaunisztikai adatbázisból származott, a többi hidrodinamikai modellek segítségével, illetve vízügyi adatbázisokból rendeltük az adatokhoz. A vízügyi adatsorból nyert változó a vízállás és a vízjárás, amelyek a kis-közép-nagy, illetve az árad-apad-stagnál értékek valamelyikét vették fel. A sodrási sebességet (m/s) az egyes mintavételi pontokra a vízállás és a vízjárás, továbbá az adott szakaszra jellemző esés segítségével becsültük a Duna mentén elhelyezett vízmércék visszamenőleges mérései alapján. Mivel a mintákat pontszerűen értelmeztük, így a sebesség becslése is pontszerűen, a transekt kezdő-, vagy középpontjában történt. Ez az egyszerűsítés azonban a nagy adatsokaság miatt nem okoz számottevő eltérést az elemzésekben.

A mederanyag-összetétel egy kétdimenziós (2D) hidrodinamikai modellből lett a mintavételi pontokhoz hozzárendelve Baranya és társai (2018), Füstös és társai (2019), valamint Füstös és társai (2021) közleményeiben használt módszerek és modellek segítségével:

A lokális mederanyagbecslés lényege, hogy a modell egy lefektetett számítási rácsháló minden egyes rácspontjára a vízhozam alapján megbecsüli a mederfenék és az áramló víztömeg közötti nyírófeszültséget, ez alapján meghatározható a mederanyag is. Az adott területen jellemző vízhozam gyakorlatilag maximum likelihood (legnagyobb valószínűség) becslési módszerrel adódik, amely alapján a kettő közötti összefüggés miatt a nyírófeszültségre is becsülhető egy érték valószínűségi úton. A korábban vett mederanyag minták és mérések alapján pedig a szemcseméret szerint csoportosított mederanyag típusok a becsült nyírófeszültséghez hozzárendelhetőek minden egyes rácspontban, ami a mederanyag nagyon aprólékos leképezését adja.

#### Modell paraméterei

A kutatás során egy regressziós RF modellt alkalmaztunk, melynek betanítása a rendelkezésre álló adatok 80%-

ával történt meg, míg a maradék 20%-on validáltuk a betanítás sikerességét. Az RF modell 150 fát tartalmazott és maximálisan öt szintig hatolhatott le. A modellezéshez a sklearn könyvtárat használtuk, amely az egyik leghasznosabb és egyben ingyenes gépi tanulási eszköztár Python programozási nyelvben (Pedregosa és társai 2011). A sklearn a NumPy, Matplotlib, illetve SciPy csomagokra épül. A kódolás, futtatás, ábrázolás és értékelés felülete a Google Colab volt (<https://colab.google/>), ahova maga a Random Forest Regresszor, a tanító és tesztelő halmazok létrehozására, valamint a statisztikai értékelések elvégzésére szolgáló eszközöket importáltuk a fent említett könyvtárból. Az adatok rendezésére és az Excel táblából való ki

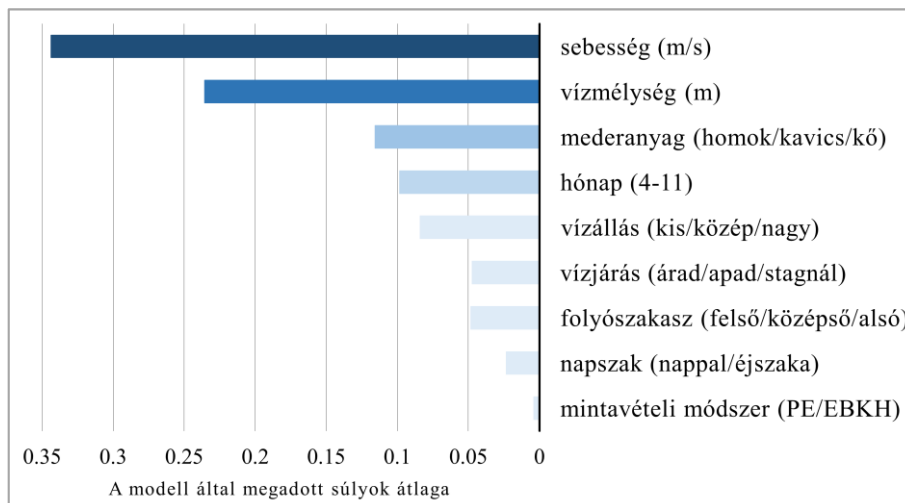
nyerésére a pandas Python könyvtár került használatra (<https://pandas.pydata.org/>).

### Modell empirikus tesztelése

A modell pontosságának és hatékonyságának tesztelését egy halfaj, validáló adatsorból származó adatain végeztük el. A tesztelés során válogatott mintavételezéseket tettünk össze egy ritkított 2D grid segítségével végzett lekérdezéssel. Az egyes mintavételezésektől maximum 150 méteres távolságon belüli becsült értékek mediánját hasonlítottuk a mért értékekhez.

## EREDMÉNYEK ÉS KIÉRTÉKELÉS

### RF elemzés eredményei



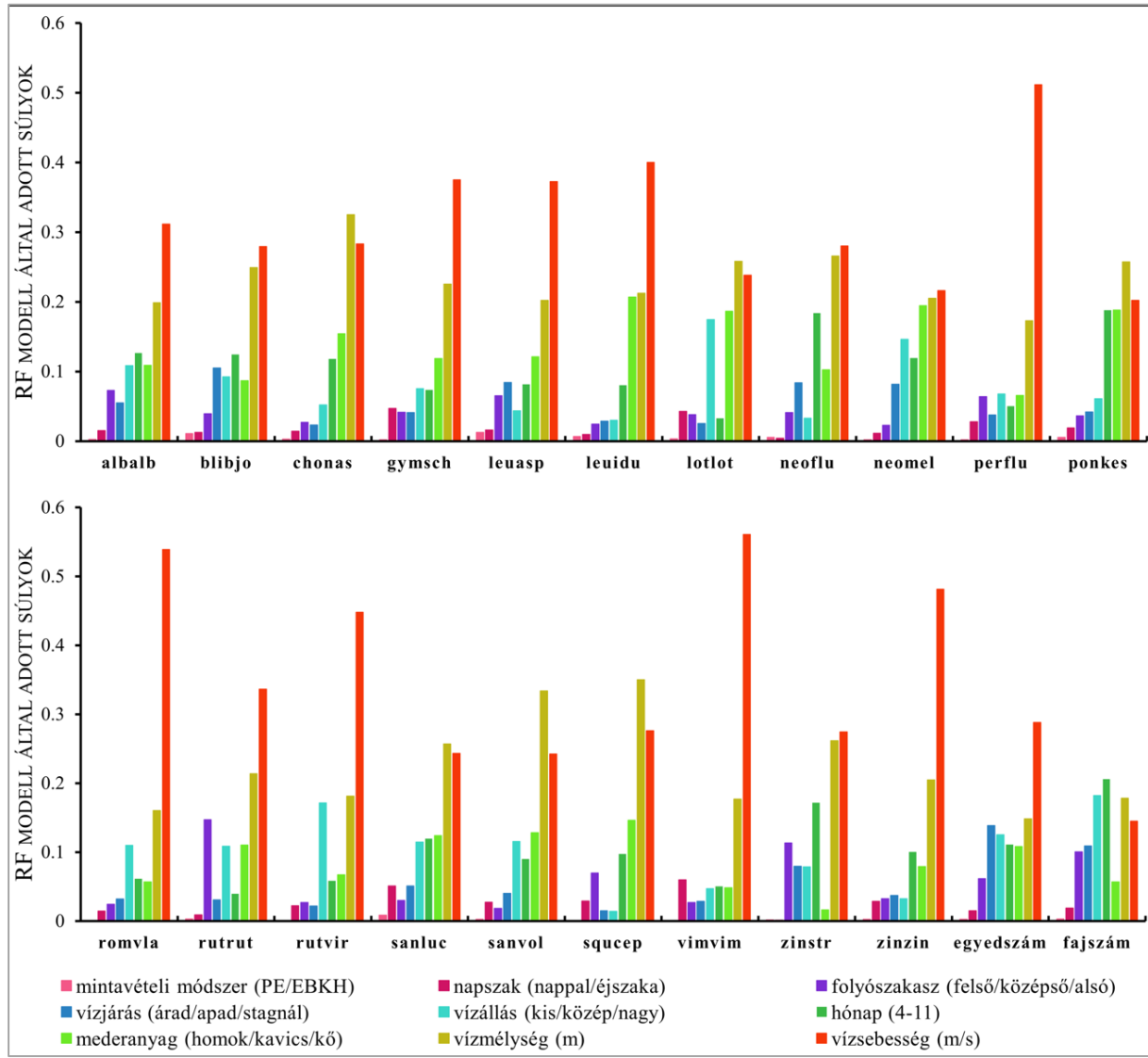
2. ábra. A független változók átlagos szerepe a 20 leggyakoribb halfaj előfordulásában az RF modell eredményei szerint  
Figure 2. Average weight (feature importance) in the presence of the most common fish species in the Danube of the nine abiotic variables used in the model

Az RF elemzés szerint a vizsgálatra kiválasztott abiotikus változók közül kiemelt fontosságú a vízsebesség és a vízmélység, harmadik helyre pedig a mederanyag típusa került (2. ábra). Az első két paraméter ennyire kiemelt szerepe egybevág a ténnyel, hogy az élőhely-megfelelőség kutatásának kezdeti időszakában amerikai kutatók ugyanezen két paraméteren kezdték az élőhelyek minősítését (U.S. Fish and Wildlife Service 1985), a mederanyag minőségét pedig már később vették be a halmazba (Baranya és társai 2018). Szintén kiemelt jelentősége adódott a mederanyag összetételének a használt adatbázisra épülő korábbi, klasszikus statisztikai módszerekkel tett elemzésben (Szalóky és társai 2021). A további paramétereket illetően van olyan, amit fenntartással kell kezelnünk. Saját tapasztalatainkból ismeretes, hogy a napszak és a mintavételi mód is befolyásolja a mintázott halállomány fajösszetételét, ami nem jelenik meg e paraméterek súlyánál. Fontos azonban kiemelni, hogy vizsgálatunk függő változóiban egyedszámok és fajszámok szerepeltek, nem fajok klasszifikációja. Ezzel a megközelítéssel tehát a megváltozott összetételt nem, csak a fajszámban, és az egyedszámban beállt változást mutathatjuk ki (amely a modelleredmények szerint nem volt jelentős a két napszak és a két mintázási módszer között). Ezenkívül lényeges az is, hogy a modellből számos

ritkább halfaj kimaradt, amelyeknél bizonyára nagyobb súllyal jelent volna meg a mintavételi módszer. Erre jó példa a kecsge (*Acipenser ruthenus*), amely szinte kizárólag az EBKH módszerrel halászható sikeresen, ugyanis a nyílt vízi élőhelyeket kedveli (Szalóky és társai 2014).

Az egyes fajokra külön meghatározva a kilenc paraméter súlyát, szembevetünk, hogy bizonyos fajoknál az áramlási sebesség többszörösen meghatározó a többi paraméterhez képest (3. ábra).

Ismeretes, hogy a csapósüger jellemzően az állóvizeket, míg a szilvaorrú keszeg éppen ellenkezőleg, a nagyobb sodrású területeket kedveli (Harka és Sallai 2004). Mindkét faj előfordulásában egyaránt fontos limitáló tényező tehát a sebesség, amelyet kimutatott a modellünk is. Egy másik megfigyelhető jelenség a kerekfejű gébhez kapcsolódik, melynek előfordulásában homogén módon alacsony szerepet határozott meg az RF a három legfontosabb paraméternek (3. ábra). Ez alapján adódik a következtetés, hogy e fajnak nincsenek különösebb igényei a hidrológiai, morfológiai változókat illetően. Ezt alátámasztja, hogy egy generalista, idegenhonos fajról beszélünk, amely rövid idő alatt el tudott terjedni a Duna hazai szakaszán (Szalóky és társai 2015); az inváziós siker valóban származhat ebből a fajta környezeti paraméterekkel szembeni széles toleranciából is.



3. ábra. A 20 leggyakoribb faj előfordulásában szerepet játszó kilenc abiotikus paraméter súlya az egyes fajokra, valamint a mintaegetedszámra, mintafajszámra külön tekintve (RF)

Megjegyzés: A fajnév kódok magyar megfelelői: albalb = szélhajtó kiusz, blibjo = karikakeszeg, chonas = paduc, gymsch = selymes durbincs, leuasp = balin, leuidu = jászkeszeg, lotlot = menyhal, neoflu = folyami géb, neomel = kerekfejű géb, perflu = csapósüger, ponkes = Kessler-géb, romvla = halványfoltú küllő, rutrut = bodorka, rutvir = leánykoncér, sanluc = süllő, sanvol = kősüllő, squcep = domolykó, vimvim = szilvaorrú keszeg, zinstr = német bucó, zinzin = magyar bucó

Figure 3. Weights of the nine abiotic parameters (feature importance) involved in the occurrence of the 20 most abundant fish species for each species and for the number of individuals and number of species (RF)



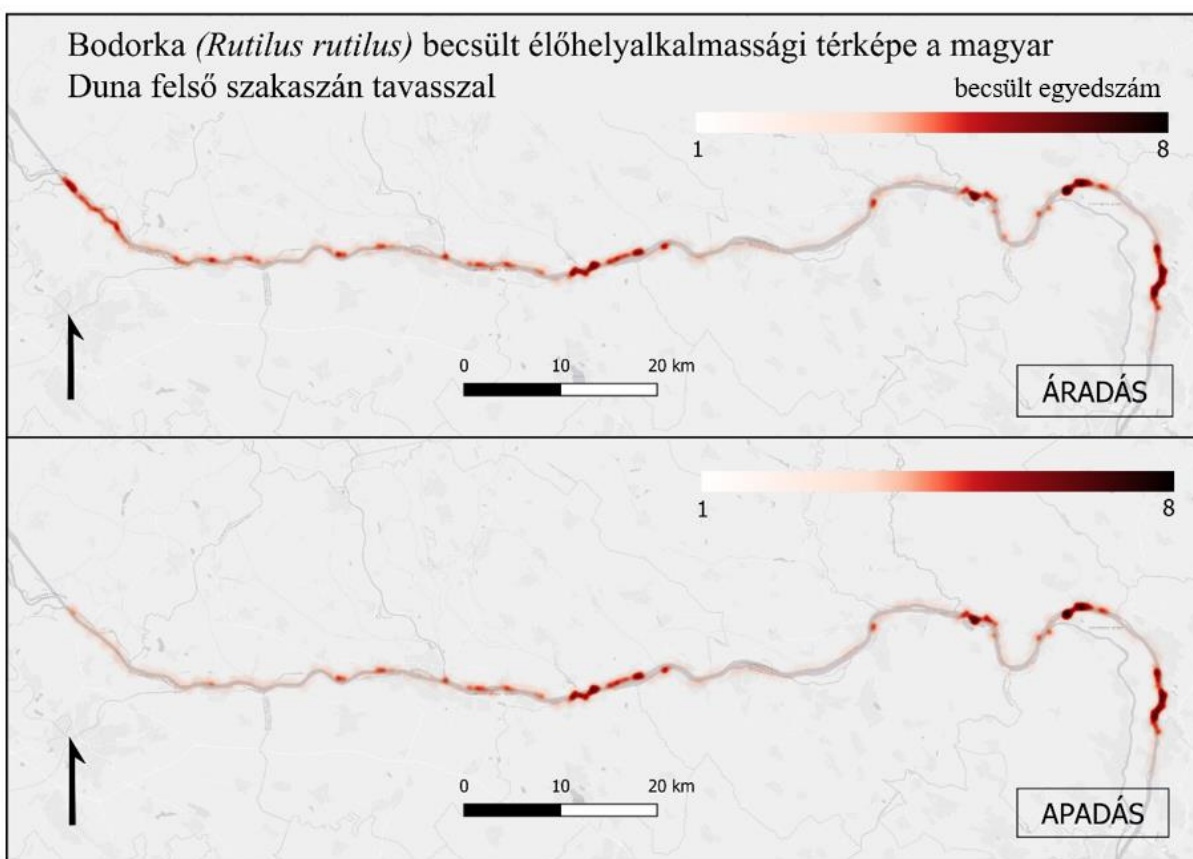
### Modell használata

Az RF modell segítségével létrehozhatóak olyan, fiktív helyzetek, lekérdezések, amelyekben mi állítjuk be a körülményeket (pl. évszak, vízállás) és a tapasztalt összefüggések alapján becslést kapunk az általunk vizsgált halfajok egyedszámaira a Duna egy megadott pontjában. Ennek segítségével képesek vagyunk térben is modellezni a halak abiotikus változók szerinti elterjedését, azaz gyakorlatilag az egyes fajoknak megfelelő élőhelyeket.

Egy lekérdezés során a modell a Duna hazai szakaszára fektetett, kis léptékű 2D grid minden pontjában becslést ad a 20 leggyakoribb halfaj egyedszámára (Füstös és társai 2021). A kapott érték gyakran nem egész szám, ugyanis az egyes regressziós fák eredményeinek átlagát adja a modell, azonban jól alkalmazható arra, hogy az egyes változók közötti összefüggések alapján az adott halfaj szempontjából rendszerezzük a folyami élőhelyeket.

### Nagy léptékű lekérdezések: Bodorka (*Rutilus rutilus*) gyakorisági térképe tavasszal áradó és áradó vízjárásnál

A bodorka egész Európában elterjedt, gyakori halfaj. Stagnofil, de a lassú áramlású víztestek között nem válogat, ezáltal a csatornák és gátak építése kedvező lehet számára (Kottelat és Freyhof 2007). Magyarországon a sebes patakok és a gyorsabb folyású folyószakaszok kivételével mindenhol előfordul (Harka és Sallai 2004). Ivadéka és fiatal egyedei elsősorban a sekély (max. 1,0 m), hínárnövényzetben és növényi törmelékben gazdag, lenitikus víztesteket kedvelik, míg kifejlett (1+) egyedei a partmenti zóna nyíltabb vízfelületeit részesítik előnyben. Ennek a korfüggő habitatpreferencia-változásnak köszönhetően a bodorka számára a fokozatosan és lassan mélyülő, partközeli hínárnövényzetben és növényi törmelékben gazdag, valamint lassú sodrású litorális területek a legmegfelelőbbek (Copp 1990, Garner 1995, Rossier és társai 1996).



4. ábra. Bodorka (*Rutilus rutilus*) modell által becsült élőhelyalkalmassági térképe a Duna felső szakaszán tavasszal áradó és apadó vízjárásnál

Figure 4. Predicted habitats (RF) of the roach (*Rutilus rutilus*) in the upper section of the Danube on spring during flooding and receding flow regime

Az RF modell súlyai alapján ennél a fajnál a sodrás sebesség és a vízmélység mellett a mederanyag, a folyószakasz és a vízállás is jelentősen befolyásolja a jelenlétét. A 4. ábrán is jól látszik, hogy áradás esetén összességében megnő az abundanciája, ami kifejezetten jól látszik a Szigetközönél. Ez összefügg azzal, hogy az árhullámmal az arra alkalmas helyeken kiszélesedik a partmenti zóna,

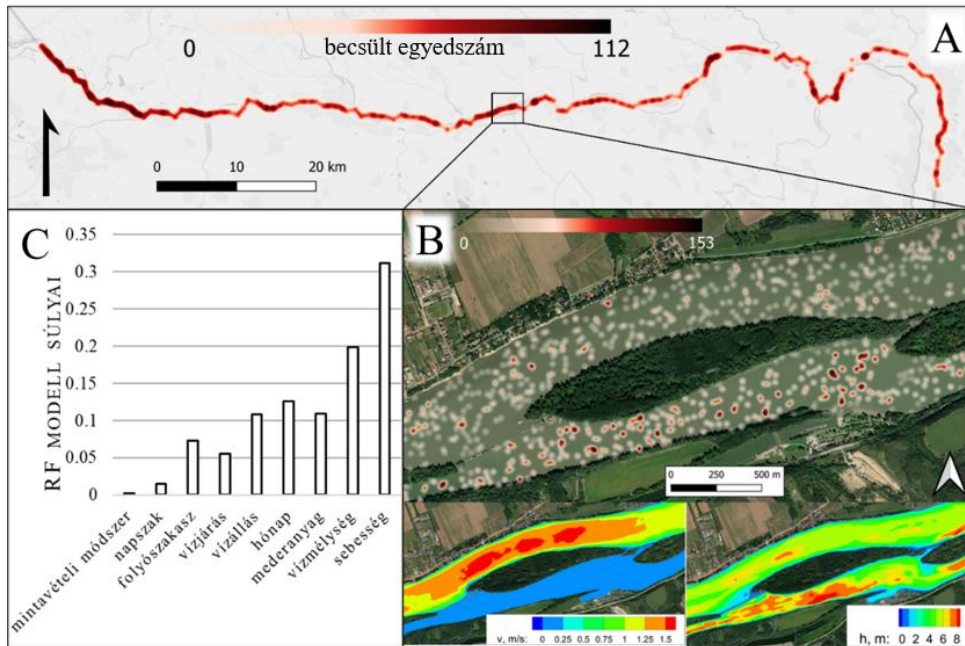
amely megfelelő élőhely a bodorkának, különösen az ivadékoknak és juvenilis egyedeknek.

Emellett az is látszik (4. ábra), hogy a vízjárástól függetlenül a legnagyobb sűrűsödési pontok állandóak. Ezek olyan területek (pl. Neszmély), ahol a szigetek, vagy öblök miatt megtalálhatóak a bodorkának legjobban megfelelő élőhelyek.

**Kisléptékű lekérdezések: szélhajtó kűsz (*Alburnus alburnus*) gyakorisági térképe nyáron, apadó vízjárásnál a Radványi-szigetnél**

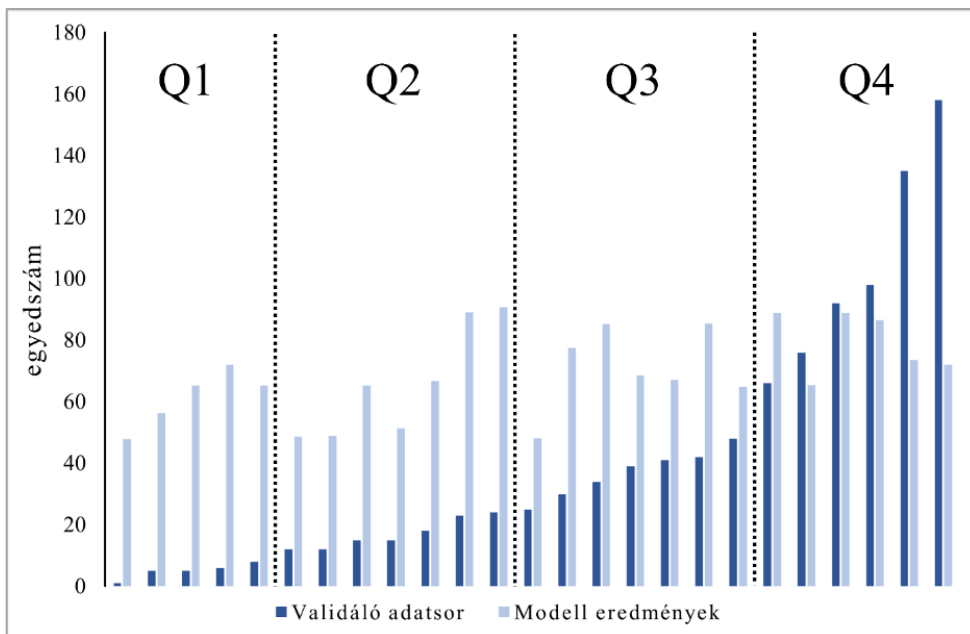
A szélhajtó kűsz a Pireneusi- és az Appennini-félszigeten kívül szinte egész Európában általános elterjedt

pontyfélé (*Kottelat és Freyhof 2007*). Társas életmódú halfaj, amely minden, nem túlságosan kicsi álló- és folyóvízben megtalálható. Legnagyobb állományai a folyók síkvidéki, lassabb folyású régióiban élnek (*Harka és Sallai 2004*).



5. ábra. A szélhajtó kűsz (*Alburnus alburnus*) becslött abundancia (sűrűsödési) térképe a hazai Duna felső szakaszán (A) és a Radványi sziget környékén (B) nyáron, apadó vízjárásnál, összehasonlítva a sebességgel és a vízmélységgel, valamint az RF modell által megadott súlyok eloszlása a szélhajtó kűszre nézve (C)

Figure 5. Predicted macro (A) and micro (B) abundance (optimal habitat) map of the bleak (*Alburnus alburnus*) with water depth and flow velocity in summer, in receding flow regime, in the upper section of the Hungarian Danube and feature importance given for the bleak (C) given by the RF model



6. ábra. A szélhajtó kűsz (*Alburnus alburnus*) fogott egyedszámok növekvő sorrendben, a kvartilisek jelölésével, és a hozzájuk tartozó becslött egyedszámok

Figure 6. Sampled individual numbers of bleak (*Alburnus alburnus*) from the validating dataset (dark blue) in increasing order with quartiles and estimated individual values belonging to them (light blue)

Az RF által adott súlyok alapján a sodrási sebesség a leginkább limitáló tényező az előfordulásában (5. ábra). Azonban sok más fajjal ellentétben a vízmélyesre kevésbé érzékeny, köszönhetően életmódjának, ugyanis leggyakrabban a felszín közelében táplálkozik planktonszervezetekkel, vízre hulló rovarokkal, növényi részekkel, vagy szerves törmelékekkel (Harka és Sallai 2004). A Radványi-szigetnél (Neszmély) készített 5. ábrán is jól látszik, hogy a gyakorlatilag állóvíz mellékágban vannak számára a legmegfelelőbb élőhelyek, annak ellenére, hogy a vízmélység itt elég nagy.

#### Modell tesztelése a szélhajtó küsz (*Alburnus alburnus*) példáján

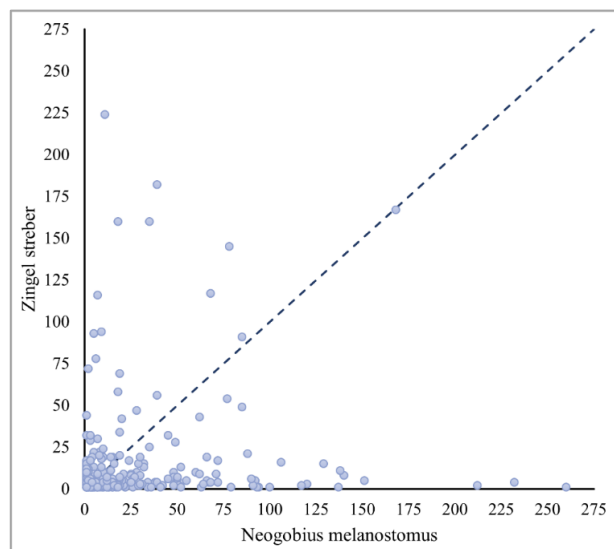
A modell pontosságának és hatékonyságának tesztelését a szélhajtó küsz egyedszám adatain hajtottuk végre, ugyanis ez a faj a modell által adott súlyok alapján kiemelkedően a sodrási sebességre mutatkozott érzékenynek (3. ábra), ezenkívül a validáló adatsorban ebből a fajból állt rendelkezésre a legtöbb, nullától különböző rekord. A mintavételek szelektálása a sebességet befolyásoló változók szerint történt, így végül 26 db középvízi, apadó vízjárásnál, júliusban – a hónap azért lett megadva, mert bár nem szerepel nagy súllyal ennél a fajnál, azonban a térbeli eloszlásban meghatározó – vett minta jelentette a valós adatokat, amelyekhez hozzá lettek rendelve a modell becslésének mediánjai.

A modell a megadott körülmények mellett minden mintához hasonló értékeket adott (szórás:  $\sigma = 14,25$ ; terjedeleme = 42,81). Ezzel szemben a valódi egyedszámok sokkal nagyobb terjedelemben vettek fel értékeket (szórás:  $\sigma = 41,53$ ; terjedeleme = 157). A legtöbb esetben a modell eléggé túlbecsül, azonban a valódi és a becsült értékek közötti eltérés kvartilisenként csökken és a 4. negyedben a legkisebb. Ezenkívül a 6. ábrán látszik, hogy van egy szűk intervallum, ahol elfogadhatóan kicsi a hiba. Ez a tartomány egyfajta hidromorfológiai optimum, ahol ezek a változók határozzák meg az egyedszámot. A többi esetben feltehetőleg kis részben más abiotikus változók és nagyobb részben biotikus változók magyarázzák a tapasztalhatóan nagy eltérést, ami legtöbbször negatív. Ebből következtethetünk arra, hogy a biotikus tényezők nagyobb limitáló erővel bírnak az általunk használt hidromorfológiai változóknál. Ez a feltételezés jól összecseng az úgynevezett biotikus-abiotikus korlátozó hipotézissel, mely szerint a biotikus tényezők, mint a potenciális predátorok és kompetitorok abundanciája – kisebb térléptékben vizsgálva – az abiotikus paramétereknél jobban befolyásolják, vagy felülírják a halközösségek mennyiségi összetételét, vagy egy adott halfaj abundanciáját (Jackson és társai 2001, Quist és Hubert 2005, Chevalier és társai 2022).

#### Fajpárok elemzésének eredményei *Neogobius melanostomus* – *Zingel streber* (idegenhonos generalista – őshonos specialista)

A kerekfejű géb generalista faj, amit alátámasztanak az RF modell súlyai is, hiszen ennél a fajnál gyakorlatilag nincsen kiemelkedő, limitáló változó (3. ábra). Tulajdonképpen minden folyami élőhelyen jelen van, azonban a nagy sodrási sebességű (~ 46 cm/s <) és durva aljzatú élőhelyeken szignifikánsan kisebb az abundanciája (Szalóky és társai 2021). A Dunában előforduló ponto-kaszpikus

géb fajok közül ez a leggyakoribb (Szalóky és társai 2015). A többi gébfajhoz hasonlóan nehéz megítélni, hogy betelepüléséhez mennyire közvetlenül járult hozzá az emberi tevékenység (Erős és Vörös 2017, Takács és társai 2017, Tavares és társai 2020). Ezzel szemben a német bucó a Duna vízgyűjtőjén őshonos, fokozottan védett halfaj, amely elsősorban áramláskedvelő, ezáltal nagyobb egyedszámban az erősebb sodrású folyószakaszokon, a meder mélyebb részein fordul elő, ahol a mederanyag-összetételben a nagyobb kavicsok és kövek dominálnak (Füstös és társai 2021, Szalóky és társai 2021).



7. ábra. A német bucó (*Zingel streber*) és a kerekfejű géb (*Neogobius melanostomus*) együttes szórásdiagramja origón átmenő egyenessel ( $n = 291$ )

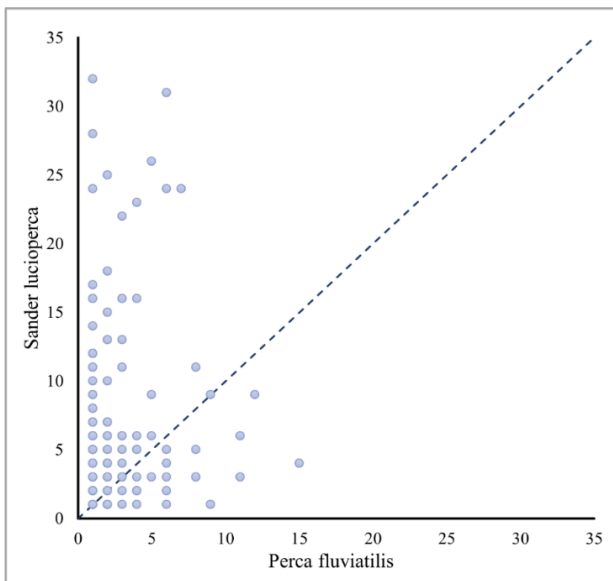
Figure 7. Joint scatter plot of the native, specialist Danube streber (*Zingel streber*) and the invasive, generalist round goby (*Neogobius melanostomus*) with a  $f(x) = x$  line ( $n = 291$ )

A két faj mindössze 291 esetben fordult elő egy mintában. Az egyedszámok szóródási grafikonja nem mutat semmilyen határozott kapcsolatot a két változó között (7. ábr), a pontok enyhén húznak a referenciaegyenes jobb oldalára, tehát a kerekfejű géb felé. Ezt alátámasztja a páros t-próba eredménye is, mely szerint a rész-arány változók különbségének átlaga 0,20 ( $p < 0,0001$ ), tehát az együttes előfordulás esetén átlagosan 20%-kal több kerekfejű géb van, mint német bucó. Ez komoly táplálékkonkurenciát jelenthet a két faj között és a kapott értékek aggasztóak a német bucó állományára nézve. Azonban a szórásdiagramról leolvasható, hogy azokon a területeken, ahol a német bucó nagy mennyiségben fordul elő, ott a kerekfejű géb csak kis egyedszámban van jelen (7. ábra), tehát a specialista faj a számára optimális élőhelyen versenytársa az agresszívan terjeszkedő invazív fajnak. Ez a megállapítás összhangban van korábbi eredményeinkkel, melyek alapján a kerekfejű géb, az egyébként generalista élőhelyhasználata mellett, előnyben részesíti a kisebb sodrási sebességű és kevésbé durva mederanyag összetételű (a kettő összefügg) élőhelyeket (Szalóky és társai 2021). A nagyon mély és nagy sodrású területeken azonban szignifikánsan kisebb az abundanciája, annak ellenére, hogy a hazai ponto-kaszpikus gébfajok közül ez az egyetlen, amely igazán nagy számban van jelen nyílt vízi élőhelyeken (Szalóky és társai 2021).



### *Sander lucioperca* – *Perca fluviatilis* (két táplálékkonkurens)

A fogassüllő és a csapósügér bizonyos korban lehetnek egymás táplálékkonkurensai, ugyanis a süllők és az idősebb sügerek egyaránt kis halakat fogyasztanak. Így érdekes vizsgálni, hogy a két faj hogyan osztja fel egymás között a rendelkezésre álló élőhelyeket. Azt azonban fontos megjegyezni, hogy sem a fajpárok elemzésénél, sem pedig a modell betanításánál nem lettek figyelembe véve az egyes korcsoportok, mint függő változók. Így nem vonhatók le messzemenő következtetések a két faj közös élőhelyhasználatáról, a különböző korcsoportok eltérő habitat- és táplálékpreferenciái miatt. Ez a fajpár tehát jó példa a modell korlátaira.



8. ábra. A fogassüllő (*Sander lucioperca*) és a csapósügér (*Perca fluviatilis*) együttes szórásdiagramja origón átmenő egyenessel ( $n = 172$ )

Figure 8. Joint scatter plot of pike perch (*Sander lucioperca*) and European perch (*Perca fluviatilis*) with  $f(x) = x$  line ( $n = 172$ )

Az egyedszámok együttes szórásdiagramján látszik, hogy a pontok szinte kizárólag a bal oldalon, azaz a fogassüllőnél sűrűsödnek (8. ábra). Ezt támasztják alá a részarány változókból kapott eredmények is, ugyanis a különbségek átlagának becslést értéke 0,29-nek adódott, vagyis  $p < 0,0001$  (páros t-próba). A két faj együttes előfordulása esetén tehát átlagosan 29%-kal több a fogassüllő, mint a csapósügér. A szórásdiagramról ezenkívül leolvasható, hogy kis egyedszámok esetén a két fajból jellemzően ugyanannyi van (8. ábra), azonban, ahogy az eredmények is mutatják, a fogassüllő uralkodóbb faj a Dunán, mint róna. Ezt az RF modell eredményei is alátámasztják, ugyanis a csapósügér jelenlétének nagyon erős limitáló tényezője a víz sodrásának sebessége (3. ábra), tehát valamilyen sodrás-alapú felosztás sejtethető a két faj között, amely alapján a fogassüllő a nagyobb sodrású területeken fordul elő nagy számban, míg a csapósügér inkább a lassú folyású, vagy állóvizet kedveli. Habár a két faj nem közeli rokonai egymásnak (Sloss és társai 2004), lehetséges, hogy egyfajta vikarizmus (fajhelyettesítés) áll a jelenség háttérben, hasonlóan más fajok kialakulásához a sügérfélek (*Percidae*) családjában (Near és társai 2001).

### ÖSSZEGZÉS ÉS KÖVETKEZTETÉSEK

Kutatásunkban a dunai halállomány felmérések hosszú távú adatsorát (~15 év; 2004-2022) használtuk egy RF modell betanítására. A független változóként használt 9 paraméter magában foglalja a legfőbb hidromorfológiai jellemzőket és az abiotikus tényezőket, a függő változóként megadott 20 halfaj pedig igen változatos csoportot alkot ökológiai igények és természetvédelmi jelentőség terén egyaránt. A modell újszerűsége a halak elterjedését prediktáló modellek között a gépi tanulási módszer, illetve a nagyszámú függő és független változó. Jelen munka eredményei megerősítik a Dunán végzett korábbi élőhelymodellezések eredményeit, amelyek más módszerekkel és függő változóként kevesebb halfaj adatait felhasználva készültek (Füstös és társai 2019, Szalóky és társai 2021).

A hal-élőhely kapcsolatok közvetlen modellezésével korábban a szerzőtársak is foglalkoztak, amely során főkomponens-analízissel és regressziós fákkal elemezték a bentikus halközösségek fajösszetételét és az egyes fajok élőhelyhasználatát, valamint a bentikus élőhelyek homogenitását (Szalóky és társai 2021). Az ott használt független változók nagyjából egyeznek a jelen kutatásban használtakkal (2. ábra), azonban többféle mintavételt különböztettek meg, határozottan finomabb mederanyag osztályozást használtak (6 db osztály), és az abiotikus változók mintavételezését nagyobb léptékben végezték, mint az itt használt modellek rácsléptéke (Szalóky és társai 2021). Az ott vizsgált fajok közül a karikakeszeg és a kerekfejű géb nagyon általános élőhelyhasználatot mutatott, ami összhangban van a modellünk által adott súlyok egyenletes eloszlásával erre a két fajra nézve (3. ábra). Az általuk szignifikánsan kimutatható élőhelypreferenciákkal rendelkező fajok közül a német bucó, a magyar bucó és a halványfoltú küllő jelen modell súlyai alapján is specialistának mutatkoztak (3. ábra). A selymes durbincs esetében azonban az észlelhető élőhelypreferencia leginkább a mederanyag-összetételhez kötődött, amit a modellünk kevésbé látványosan mutatott. Hasonló a helyzet a két bucófaj eltérő élőhelypreferenciáinak kimutatásával is, ugyanis a modellünk a korábbiakban tapasztalt eltérő mederanyag-összetétel preferenciát nem igazán jelezte, azonban jól megfigyelhető a magyar bucó esetében tapasztalható nagyobb érzékenység a sodrási sebességre és a vízmélységre (3. ábra). A két kutatás eredményei tehát konzisztensnek bizonyultak. Ezen túlmenően az idézett munkában kimutatható volt, hogy az egyes független (hidromorfológiai) változók között is van korreláció. Erre alapozva egy következő lépésben érdemes lehet megfontolni egyes változók összevonását, a redundancia csökkentését, a helyükre pedig új, független változókat foglalni a modellbe.

Az RF modellben adódott súlyok alapján a 20 leggyakoribb halfajnál meg tudtuk állapítani, hogy az egyes abiotikus változók mekkora szerepet játszanak az előfordulásukban. Fajonként a súlyok értékének megoszlása információt ad a fajok generalista, vagy specialista mivoltáról ezen paraméterekre nézve. Összesítve pedig megállapítottuk, hogy a sodrási sebesség, a vízmélység és a mederanyag-összetétel az a három kiemelkedően fontos változó, ami folyami viszonyok között egyezik a szakirodalmi állásponttal (U.S. Fish and Wildlife Service 1985, Baranya

és társai 2018). Fontos azonban kiemelni, hogy az egyes változókhoz kapott alacsony súlyok nem feltétlenül jelentik az adott változó kisebb jelentőségét, egyszerűen csak az adott paraméterek mellett nem bizonyultak meghatározó körülménynek (pl: mintavételi módszer, napszak).

A modell a 2D hidrodinamikai szimulációkkal előállított eredményezőkre épülő lekérdezések segítségével képes adott körülmények mellett prediktálni az egyes halfajok elterjedését a Duna teljes hazai szakaszán. Az általa adott értékek jól követik a hossz- és keresztirányú összefüggéseket is. A teszteléseink azonban rávilágítottak arra, hogy a modell az esetek legnagyobb részében túlbecsüli az egyedszámokat és a magyarázó változók összetételéből adódóan nyilvánvalóan nem alkalmas pontos egyedszámbebecslésre. Ugyanis a modellben nem szereplő biotikus változók jóval meghatározóbbak és nagyobb limitáló erővel bírnak a halak elterjedésében és számosságában. A számított egyedszámok tehát nem önmagukban, nyersen értelmezendők, hanem a nagyságrendjükből az adott halfajok számára megfelelő, vagy kevésbé megfelelő élőhelyfeltok térbeli eloszlására következtethetünk. Modellünk emellett alkalmas az egyes körülményekben bekövetkező változások hatásvizsgálatára is, kis és nagy léptékben egyaránt. A modell másik fontos korlátja a halak eltérő korcsoportjainak eltérő élőhelyhasználatából fakad. A tanítás során ugyanis, nem lettek figyelembe véve a különböző korcsoportok, így a kapott összefüggések összességében írják le az adott faj mondhatni átlagos habitatpreferenciáit, ami azonban egyes fajok esetén a különböző korcsoportok között eltérhet.

Habár, mint megjegyeztük, a modellbeli egyedszámokat nem az adott körülmények között ténylegesen várható egyedszámokként kell kezelni, volt egy szűk tartomány, a hidromorfológiai optimum, ahol a modell becslései igen közel álltak a valósághoz. Ez további vizsgálatra érdemes, ugyanis kérdéses, hogy halfajonként a terjedelme és a helye mennyire változékony. Azaz, vannak-e olyan halfajok, amelyeknél kevésbé, vagy másként érvényesülnek a biotikus változók.

A fajpárok klasszikus statisztikai módszerekkel történő elemzésében jól látszott, hogy a modell által adott súlyok bizonyos esetekben magyarázatul szolgálhatnak populációs kapcsolatokra és meghatározzák az egyes fajok életterét, illetve fordítva, az egyes populációs kapcsolatok, mint biotikus változók szintén meghatározóak lehetnek az egyedszámok tekintetében.

Kutatásunk során rengeteg további kérdés merült fel. Elsősorban nyilvánvalónak tűnik a modell bővítése további abiotikus változók bevezetésével a független változók sorába. Ezzel egyrészt elméletben javíthatnánk az élőhelybecslések pontosságán, valamint egy abiotikus változót széleskörben lefedő modell segíthet a biotikus tényezők hatásának indirekt vizsgálatában. Szóba jöhet például a pH, amely tavaknál igen meghatározó (*Matuszek és Beggs 1988*), vagy az oldott oxigénmennyiség, melyet vízfolyásos modellezéseknél szoktak is alkalmazni háttérváltozóként, ugyanis jól jelzi az élőhelykülönbséget hosszú élőhelygrádienseknél. Ezen változók mérése vagy modellezése azonban egyáltalán nem könnyű és a tapasztalatok

azt mutatják, hogy a hidromorfológiai változók adják a variancia legnagyobb hányadát (*Szalóky és társai 2021*), így a további abiotikus változók bevezetésének csekély lenne a megtérülése a jelen lépték mellett.

További lehetőségeket hordoz magában a modell alkalmazhatóságának térbeli növelése, ami irányulhat a Duna, vagy mellékfolyói további szakaszaira, melyekhez vannak is rendelkezésre álló adatsorok (pl: Nemzetközi Duna Expedíció – JDS). Ebben kihívást jelentenek, főleg a mellékfolyók esetében, az eltérő hidrodinamikai és hidromorfológiai körülmények. A térbeli extrapolációt elősegíti, hogy már számos technikai megoldás létezik a ML modellek térbeli használhatóságának növelésére (*Meyer és Pebesma 2021*).

Ahhoz, hogy a modell alkalmas legyen pontos egyedszámbebecslésre, abiotikus változók mellett rengeteg biotikus változót is hozzá kellene venni. Ezek közül soknak, például a trofikus kapcsolatoknak a mérése, vagy értékelése rendkívül nehéz, például egy faj egyedei különböző korban más-más trofikus szinten helyezkedhetnek el. Ezenkívül igen nehéz megállapítani, hogy egy faj milyen hatást fejt ki az ökoszisztémára, tehát nehéz megmondani, hogy mely biotikus változók figyelembevétele szükséges az egyedszám pontos becsléséhez (*Franco és társai 2020*). A biotikus változók hozzáadásához tehát szükség lenne az egyes fajok populációs kapcsolatrendszerében történő elemlyedésre.

Kutatásunk újabb előrelépést jelent a mesterséges intelligencia által támogatott ökológiában azzal, hogy bemutattuk a Random Forest modellek alkalmazhatóságát a hal-élőhely kapcsolatok értékelésében. Ezenkívül az eredményeink rávilágítottak arra, hogy a gépi tanulás megközelítések ígéretesnek bizonyulnak a halak élőhelyválasztását befolyásoló paraméterek jelentőségének számszerűsítésében, és így segítségünkre lehetnek a jövőben a mintavételek hatékonyságának növelésében. Másrészt, ismét bebizonyosodott a részletes, nagy tér- és időintervallumot átölelő adatbázisok jelentősége, mint a gépi tanulás módszerek esszenciális tanító adatsorai.

A közlemény alapját Jandó Benedek dolgozata képezi, melyet a 36. Országos Tudományos Diákköri Konferencián és a 2023. évi Stockholm Junior Water Priceon (Stockholmi Ifjúsági Víz Díjon) mutatott be Magyarország képviselőjében, a Global Water Partnership támogatásával.

## IRODALOMJEGYZÉK

*A Víz keretirányelv hazai megvalósítása (2009). Vízgyűjtő-gazdálkodási terv. Budapest: Vízügyi és Környezetvédelmi Központi Igazgatóság.*

*Baranya S., Fleit G., Józsa J., Szalóky Z., Tóth B., Czeglédi I., Erős T. (2018). Habitat mapping of riverine fish by means of hydromorphological tools. Ecohydrology, 11(7). <https://doi.org/10.1002/eco.2009>*

*Bergström, P., Gonzalez-Mirelis, G., Lindgarth, M. (2011). Interaction between classification detail and prediction of community types: implications for predictive modelling of benthic biotopes. Marine Ecology Progress Series, 432. pp. 31-44. <https://doi.org/10.3354/meps09129>*

- Chevalier, M., Tedesco, P., Grenouillet, G. (2022). Spatial patterns in the contribution of biotic and abiotic factors to the population dynamics of three freshwater fish species. *PeerJ Life & Environment*, 10:e12857. <https://doi.org/10.7717/peerj.12857>
- Copp, G.H. (1990). Shifts in the microhabitat of larval and juvenile roach, *Rutilus rutilus* (L.), in a floodplain channel. *Journal of Fish Biology*, 36. pp. 683-692. <https://doi.org/10.1111/j.1095-8649.1990.tb04322.x>
- Elmahdy, S., Ali, T., Mohamed, M., Howari, F., Abouleish, M., Simonet, D. (2020). Spatiotemporal Mapping and Monitoring of Mangrove Forests Changes From 1990 to 2019 in the Northern Emirates, UAE Using Random Forest, Kernel Logistic Regression and Naive Bayes Tree Models. *Frontiers in Environmental Science*, 8. pp. 102-125. <https://doi.org/10.3389/fenvs.2020.00102>
- Erős T., Vörös J. (2017). Áttekintés a hazai idegenhonos inváziós halak, kétélűek és hullók jelenlegi helyzetéről. *Magyar Tudomány*, 4. pp. 426-428.
- Farkas-Iványi K., Trájer A. (2015). The influence of the river regulations on the aquatic habitats in river Danube, at the Bodak branch-system, Hungary and Slovakia. *Carpathian Journal of Earth and Environmental Sciences*, 10(3). pp. 235-245.
- Franco, A., Garcia-Berthou, E., dos Santos, L. (2020). Ecological impacts of an invasive top predator fish across South America. *Science of The Total Environment*, 761., 143296. <https://doi.org/10.1016/j.scitotenv.2020.143296> PMID:33187704
- Füstös V., Baranya S., Fleit G., Erős T., Szalóky Z., Tóth B., Józsa J. (2019). A felső-magyarországi Duna élőhelyszempontú hidrodinamikai vizsgálata. *Pisces Hungarici*, 13. pp. 81-90.
- Füstös V., Erős T., Józsa J. (2021). 2D vs. 3D Numerical Approaches for Fish Habitat Evaluation of a Large River-Is 2D Modeling Sufficient? *Periodica Polytechnica Civil Engineering*, 65(4). pp. 1114-1125. <https://doi.org/10.3311/PPci.17788>
- Garner, P. (1995). Suitability indices for juvenile 0+ roach [*rutilus rutilus* (L.)] using point abundance sampling data. *Regulated Rivers: Research & Management*, 10(2-4). pp. 99-104. <https://doi.org/10.1002/rrr.3450100206>
- Harka Á., Sallai Z. (2004). Magyarország halfaunája. Szarvas: Nimfea Természetvédelmi Egyesület., p. 269.
- Jackson, D.A., Peres-Neto, P.R., Olden, J.D. (2001). What controls who is where in freshwater fish communities – the roles of biotic, abiotic and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(1), pp. 157-170. <https://doi.org/10.1139/f00-23>
- Kottelat, M., Freyhof, J. (2007). *Handbook of European Freshwater Fishes*. Cornol, Switzerland and Berlin, Germany: Maurice Kottelat and Jörg Freyhof. p. 646.
- Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., Kneeshaw, D. (2018). Application of machine learning methods in forest ecology: recent progress and future challenges. *Environmental Reviews*, 26(10). pp. 339-350. <https://doi.org/10.1139/er-2018-0034>
- Matuszek, J.E., Beggs, G. (1988). Fish Species Richness in Relation to Lake Area, pH, and Other Abiotic Factors in Ontario Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 45(11). pp. 1931-1941. <https://doi.org/10.1139/f88-225>
- Meyer, H., Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9). pp. 1620-1633. <https://doi.org/10.1111/2041-210X.13650>
- Near, T., Page, L., Mayden, R. (2001). Intraspecific phylogeography of Percina evides (Percidae: Etheostominae): an additional test of the Central Highlands pre-Pleistocene vicariance hypothesis. *Molecular Ecology*, 10. pp. 2235-2240. <https://doi.org/10.1046/j.1365-294X.2001.01362.x> PMID:11555265
- Olden, J., Lawler, J., Poff, N. (2008). Machine learning methods without tears: A primer for ecologists. *The Quarterly Review of Biology*, 83(2). pp. 171-193. <https://doi.org/10.1086/587826> PMID:18605534
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python, *JMLR*, 12. pp. 2825-2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- Quist, M.C., Hubert, W.A. (2005). Relative Effects of Biotic and Abiotic Processes: A Test of the Biotic–Abiotic Constraining Hypothesis as Applied to Cutthroat Trout. *Transactions of the American Fisheries Society*, 134(3). pp. 676-686. <http://dx.doi.org/10.1577/T04-112.1>
- Reiczigel J., Harnos A., Solymosi N. (2010). Biostatistika nem statisztikusoknak. Nagykovácsi: Pars Kft.
- Rossier, O., Castella, E., Lachavanne, J.-B. (1996). Influence of submerged aquatic vegetation on size class distribution of perch (*Perca fluviatilis*) and roach (*Rutilus rutilus*) in the littoral zone of Lake Geneva (Switzerland). *Aquatic Sciences*, 58(1), pp. 1-14. <https://doi.org/10.1007/BF00877636>
- Sloss, B., Billington, N., Burr, B. (2004). A molecular phylogeny of the Percidae (Teleostei, Perciformes) based on mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution*, 32. pp. 545-562. <https://doi.org/10.1016/j.ympev.2004.01.011> PMID:15223037
- Sylvain, C., Hervet, É., Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10. pp. 1632-1644. <https://doi.org/10.1111/2041-210X.13256>
- Szalóky Z., Bammer, V., György Á.I., Pehlivanov, L., Schabuss, M., Zornig, H., Weiperth A., Erős T. (2015). Offshore distribution of invasive gobies (Pisces: Gobiidae) along the longitudinal profile of the Danube River. *Fundamental Applied Limnology*, 187(2), pp. 127-133. <http://dx.doi.org/10.1127/fal/2015/0768>
- Szalóky Z., Füstös V., Tóth B., Erős T. (2021). Environmental drivers of benthic fish assemblages and fish-habitat associations in offshore areas of a very large river. *River*

Research and Applications, 37(5). pp. 712-721. <https://doi.org/10.1002/rra.3793>

Szalóky Z., György, Á.I., Tóth B., Sevcsik A., Specziár A., Csányi B., Szekeres J., Erős T. (2014). Application of an electrified benthic frame trawl for sampling fish in a very large European river (the Danube River) - Is offshore monitoring necessary? Fisheries Research, 151. pp. 12-19. <https://doi.org/10.1016/j.fishres.2013.12.004>

Takács P., Czeglédi I., Ferincz Á., Sály P., Specziár A., Vítál Z., Weiperth A., Erős T. (2017). Non-native fish species in Hungarian waters: historical overview, potential sources and recent trends in their distribution. Hydrobiologia 795. pp. 1-22. <https://doi.org/10.1007/s10750-017-3147-x>

Tavares, C., Brauns, M., Hille, S., Krenek, S., Borcherding, J., Weitere, M. (2020). Tracing the coloniza-

tion process of non-native gobies into a large river: the relevance of different dispersal modes. Biological Invasions, 22. pp. 2421-2429. <https://doi.org/10.1007/s10530-020-02281-x>

Tózsza I. (1998). Tájképi homogenitás Magyarországon. Földrajzi Értesítő, 47(3), pp. 432-445. [https://www.mtafki.hu/konyvtar/kiadv/FE1998/FE19983\\_432-445.pdf](https://www.mtafki.hu/konyvtar/kiadv/FE1998/FE19983_432-445.pdf)

U.S. Fish and Wildlife Service (1985). Habitat suitability models and instream flow suitability curves: pink salmon. Biological report, 10.109. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a322912.pdf>

Zajicek, P., Wolter, C. (2018). The gain of additional sampling methods for the fish-based assessment of large rivers. Fisheries Research, 197. pp. 15-24. <https://doi.org/10.1016/j.fishres.2017.09.018>

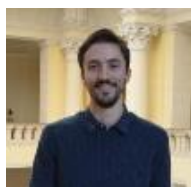
## A SZERZŐK



**JANDÓ BENEDEK** az Állatorvostudományi Egyetem hallgatója angol nyelvű kutató zoológus szakon. Különösen érdeklődik az ökológia és határterületei iránt. 2020 óta tagja a Magyar Biodiverzitás-kutató Társaságnak, továbbá aktív önkéntese és adatgyűjtője a Magyar Madártani és Természetvédelmi Egyesületnek és az Ócsai Madárvárta Egyesületnek.



**FÜSTÖS VIVIEN** okleveles infrastruktúra-építőmérnök, jelenleg PhD-tanulmányait folytatja a Budapesti Műszaki és Gazdaságtudományi Egyetem Vásárhelyi Pál Építőmérnöki és Földtudományi Doktori Iskolában. Kutatási témája a folyami élőhelyek hidromorfológiai vizsgálata mikro- és mezoléptékben. 2017 óta a Magyar Hidrológiai Társaság tagja.



**ERMILOV ALEXANDER ANATOL** okleveles építőmérnök, doktorandusz a Budapesti Műszaki és Gazdaságtudományi Egyetem Vízépítési és Vízgazdálkodási Tanszékén. Doktori kutatásának témája a folyami áramlás és a mederfenék kölcsönhatásának vizsgálata. 2015-ben TDK II. helyezést kapott a „Balaton szélkeltette vízcseré folyamatának numerikus modell alapú vizsgálata” című dolgozattal, majd ezen értekezéssel I. helyezést ért el a Magyar Hidrológiai Társaság Lászlóffy Woldemár diplomamunka pályázatán. MSc diplomamunkáját a Norwegian University of Science and Technology-n védte meg 2017-ben.



**SZALÓKY ZOLTÁN** biológus (ELTE 2002), 2017-ben szerzett PhD fokozatot az Eötvös Loránd Tudományegyetemen. Jelenleg intézeti mérnök a HUN-REN Ökológiai Kutatóközpontban. Kutatási területe álló- és folyóvizek hidro- és halbiológiai vizsgálatai, különös tekintettel nagy folyók halközösségeire.



**ERŐS TIBOR** biológus, 2005-ben szerzett PhD fokozatot az Eötvös Loránd Tudományegyetemen. Jelenleg a HUN-REN Balatoni Limnológiai Kutatóintézet igazgatója. Kutatási területe: halegyüttesek szerveződése édesvizekben, biológiai sokféleség és a környezeti tényezők kapcsolata édesvizekben, mintavétel reprezentativitása, monitorozás rendszerek fejlesztése, természetvédelmi területek kijelölése édesvizek természeti értékei alapján. 1999 óta a Magyar Hidrológiai Társaság tagja.



**BARANYA SÁNDOR** építőmérnöki oklevelét a Budapesti Műszaki és Gazdaságtudományi Egyetemen szerezte 2003-ban, PhD fokozatát ugyanitt 2010-ben. Jelenleg a BME Vízépítési és Vízgazdálkodási Tanszékének egyetemi docense. Kutatási területe folyók medermorfológiai, áramlástan és hordalékvándorlási vizsgálata terpi eljárásokkal és számítógépes modellezéssel. A Magyar Hidrológiai Társaságnak 2003 óta tagja.