

MÉLYTANULÓ ALGORITMUSOK ALKALMAZÁSA A VÍZGAZDÁLKODÁSBAN

**Dr. Kozák Péter Ph.D.¹, dr. Vizi Zsolt², Fehérváry István³,
dr. Benyhe Balázs³, Fiala Károly³, Lázár Miklós³**

¹Nemzeti Közszolgálati Egyetem Víz tudomány Kar, egyetemi docens

²Szegedi Tudományegyetem TTIK Bolyai Intézet, egyetemi adjunktus

³Alsó-Tisza vidéki Vízügyi Igazgatóság Vízrajzi és Adattári Osztály

KIVONAT

A vízgazdálkodási feladatok végrehajtásával kapcsolatosan a mélytanuló algoritmusok új távlatokat nyithatnak az adatsorok analízise, az előrejelzési feladatok végrehajtása során. A korábbi gyakorlatban szereplő korrelációs statikus jellegű összefüggések mellett a rekurrens neurális hálózatok alkalmazásával lehetőség nyílik a vízgyűjtő egészében végbemenő változások dinamikus figyelembe vételére. A dolgozat keretében az Alsó-Tisza vidéki Vízügyi Igazgatóság működési területén a mélytanuló algoritmusok alkalmazásával kapcsolatos eddigi tapasztalatok kerülnek ismertetésre.

KULCSSZAVAK: mélytanuló algoritmusok, mintázat felismerés, rekurrens neurális hálózat

1. Bevezetés

A vízgazdálkodás valamennyi szakterületéhez kapcsolódóan egyre több és több adat áll rendelkezésre. Napjaink korszerű vízrajzi műszerei rendkívüli gyakorisággal állítják elő azon adatokat, amelyekkel a hidrológiai elemek változása jellemezhető. A hatékony vízgazdálkodási döntések meghozatalához nélkülözhetetlen ezen adattömeg gyors és eredmény-orientált feldolgozása. A hagyományos hidrológiai feldolgozási módszerek és eljárások a hidrológiai folyamatokhoz kapcsolódó feldolgozási módszerekkel biztosítanak átfogó képet a változásokról. A feldolgozások jellemzően napi egy adatból alkotott idősorok vizsgálata alapján vonnak le következtetéseket, határoznak meg mérőszámokat. Azonban a vízrajzi adatbázisokban a vizsgálatokba bevont adatok többszöröse található, melyek még nem kerültek feldolgozásra. Felvetődik a kérdés, hogy a kiterjesztett adat-tartalmú elemzések eredményezhetnek-e új összefüggéseket a vizsgált hidrológiai folyamatok vonatkozásában? Illetve az adatsorokban azonosítható mintázatok alapján növelhető-e az adatsorok vizsgálatának hatékonysága, például a hidrológiai előrejelzések pontossága, gyorsasága?

Jelen dolgozatban az Alsó-Tisza vidéki Vízügyi Igazgatóság gyakorlati feladataihoz kapcsolódó, néhány szakterületet érintő minta-vizsgálati alkalmazás kerül ismertetésre. A vizsgálatok matematikai-módszertani megalapozását a Szegedi Tudományegyetem Természettudományi és Informatikai Karának Bolyai Intézetében Dr. Vizi Zsolt egyetemi adjunktus által irányított kutatás-fejlesztési csoportja végezte.

2. A mélytanuló algoritmusok alkalmazási területei a vízgazdálkodásban

Napjainkban a nagytömegű adatok gyors feldolgozása az élet valamennyi területén nélkülözhetetlen. A tőzsde, a forgalomirányítás, az adatbányászat, az e-közigazgatás, a bűnüldözés, az arc-felismerés, az automatizált képfeldolgozás mind-mind nagytömegű adat mintázat felismerésére és feldolgozására alapozva biztosít olyan döntéstámogatói háttérrel, amellyel az adatok változásához illeszkedő, gyors vizsgálatok hajthatók végre.

A hagyományos hidrológiai feldolgozások végrehajtása során az adatsorok statikus vizsgálatával kerültek/kerülnek megállapításra azon összefüggések amelyek alapján az adott

hidrológiai jellemző előrejelzése történik. Ez feltételezi, hogy az adott hidrológiai folyamatot meghatározó környezeti jellemzők nem változnak/változtak meg. Amennyiben ezek változnak, úgy a hidrológiai vizsgálatokat újra el kell végezni, Amennyiben nem ismerjük fel kellő időben a környezeti tényezők következtében lezajló változásokat (pl: medrek, csapadékvizonyok, talajvízszintek, stb.) úgy a korábbi összefüggések alkalmazásával növekedni fog az előrejelzés pontatlansága, amely akár az operatív vízgazdálkodási döntések hatékonyságát is csökkentheti.

A vízgazdálkodásban igényelt döntések környezeti feltételei a klímaváltozás következtében jelentős mértékben megváltoztak, ennek következtében a korábbi éghajlati-környezeti viszonyokhoz igazodóan felírt regressziós kapcsolatok felbomlottak, azok alapján kidolgozott méretezési elvek, módszerek felülvizsgálata időszerű. A mérési technológia fejlődése olyan többlet adat mennyiséget tett elérhetővé, amely alapján a korábban alkalmazott, statikus összefüggések helyett az adatsorokra jellemző változásokat megtanuló adatvezérelt algoritmusok segítségével végezhetünk előrejelzéseket (pl. a felszíni lefolyások, - vízállások tekintetében) vagy térbeli mintázatfelismerést (pl. a felszínfedettség, belvízi elöntés kapcsán). Ezen egyre növekvő adathalmazok hatékony feldolgozása a korábbi statikus megközelítéssel már nem biztosítható. Az adatok dinamikus megközelítésére alapozott módszerek kifejlesztésére és alkalmazására van szükség. Habár a dolgozatban alkalmazott vizsgálatok elsődleges célja a mélytanuló algoritmusok hidrológiai alkalmazásainak feltárása volt és a vizsgálatok jelen szakaszában csak a múltban bekövetkezett változások figyelembe vételére volt lehetőség, azonban a vizsgált algoritmusok a rendkívül gyors tanítási időszükséglet miatt alkalmasak a változások dinamikus követésére is újabb tanítási időszakot követően.

A mélytanuló algoritmusok nemzetközi szakirodalmi megjelenése már a 2000-es évek elejétől datálható [3]. Az algoritmusok alkalmazásra egyrészt a hidrológiai jellemzők adatsorainak feldolgozása, másrészt a távérzékelési folyamatok eredményeként kapott nagyfelbontású felvételek elemzése biztosít feladatokat.

Az adatsorok analízisével kapcsolatosan egyre több szakterületen kerültek alkalmazásra ezen eljárások, módszerek. A felszíni hidrológia, a csapadék analízis, az árvízi előrejelzés, a felszín alatti vizek változásával kapcsolatos vizsgálatok közös jellemzője, hogy nagytömegű adat vizsgálataira alapozva a múltban bekövetkező események összefüggései alapján határozzunk meg egy jövőbeni esemény során kialakuló értéket.

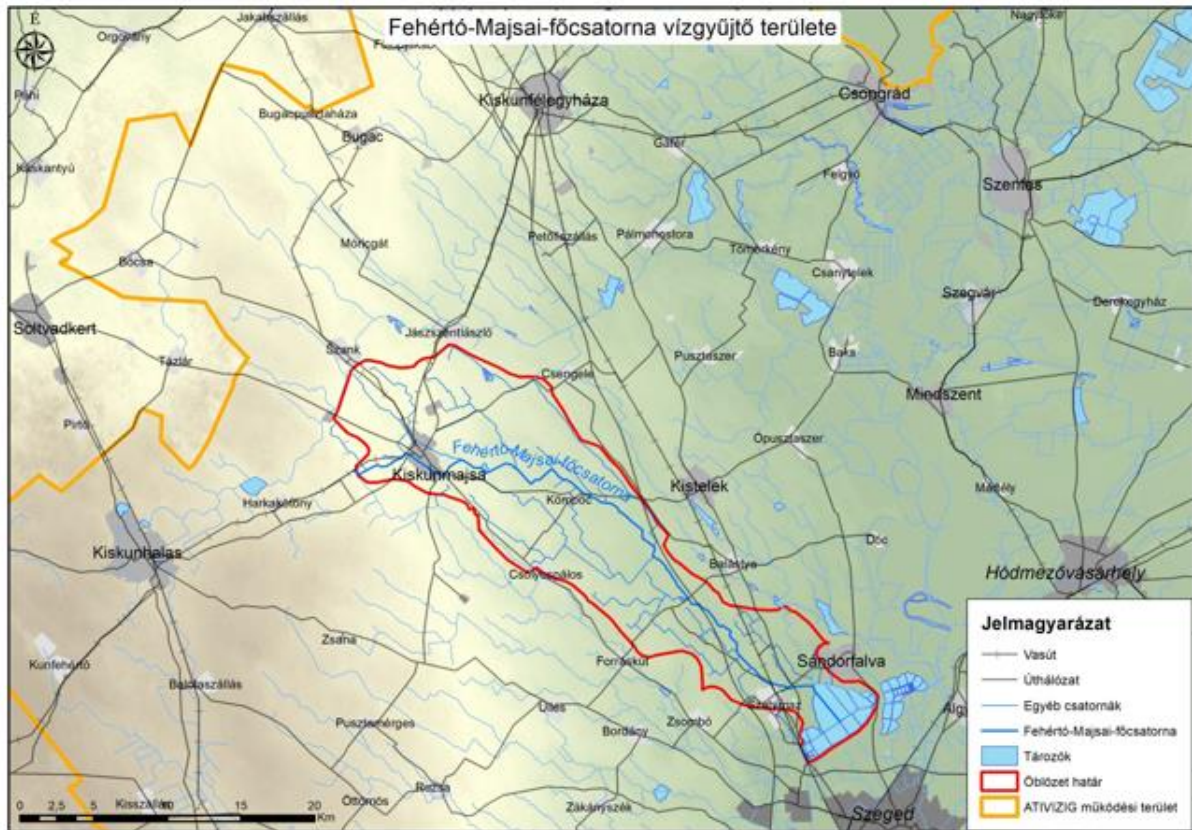
3. A felszín alatt vizek és a felszíni lefolyás vizsgálata

Az Alsó-Tisza vidéki Vízügyi Igazgatóság munkatársai vizsgálatokat végeztek a Duna-Tisza közti Homokhátság lefolyási viszonyainak elemzésére. A vizsgálatok során elsődleges cél az volt, hogy a vizsgálati terület (1. ábra) vonatkozásában a talajvíz-csapadék-felszíni lefolyás értékek összefüggései feltárásra kerüljenek.

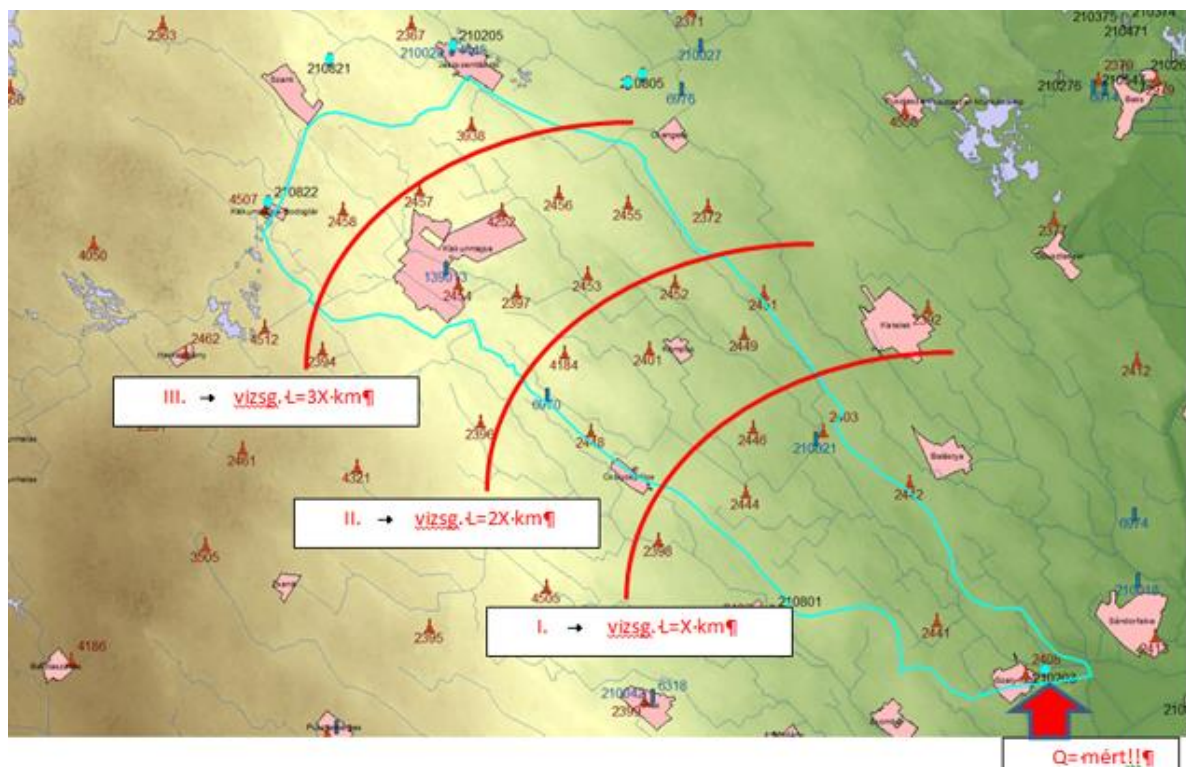
A vizsgálati területként a Fehértó-Majsai főcsatorna vízgyűjtője került kiválasztásra, amely a Duna-Tisza közti Homokhátság délkeleti lejtőjén helyezkedik el. A vizsgálatok célja a főcsatornán működő mérőszelvényben regisztrált vízhozamok (1962.01.01.-2019.12.31) és a vízgyűjtő területen mért csapadék összegek, valamint a talajvízszintek közötti összefüggésekre alapozott lefolyás előrejelzése volt (2. ábra).

A vizsgálatok végrehajtása során kiemelt figyelmet kellett fordítani az adatok előkészítésére. Az adatsorok mélytanuló algoritmusokkal való vizsgálata során kiemelt figyelmet kell fordítani a szélsőértékek és a szezonális vizsgálatára. Az alkalmazható vizsgálati módszerek – bár

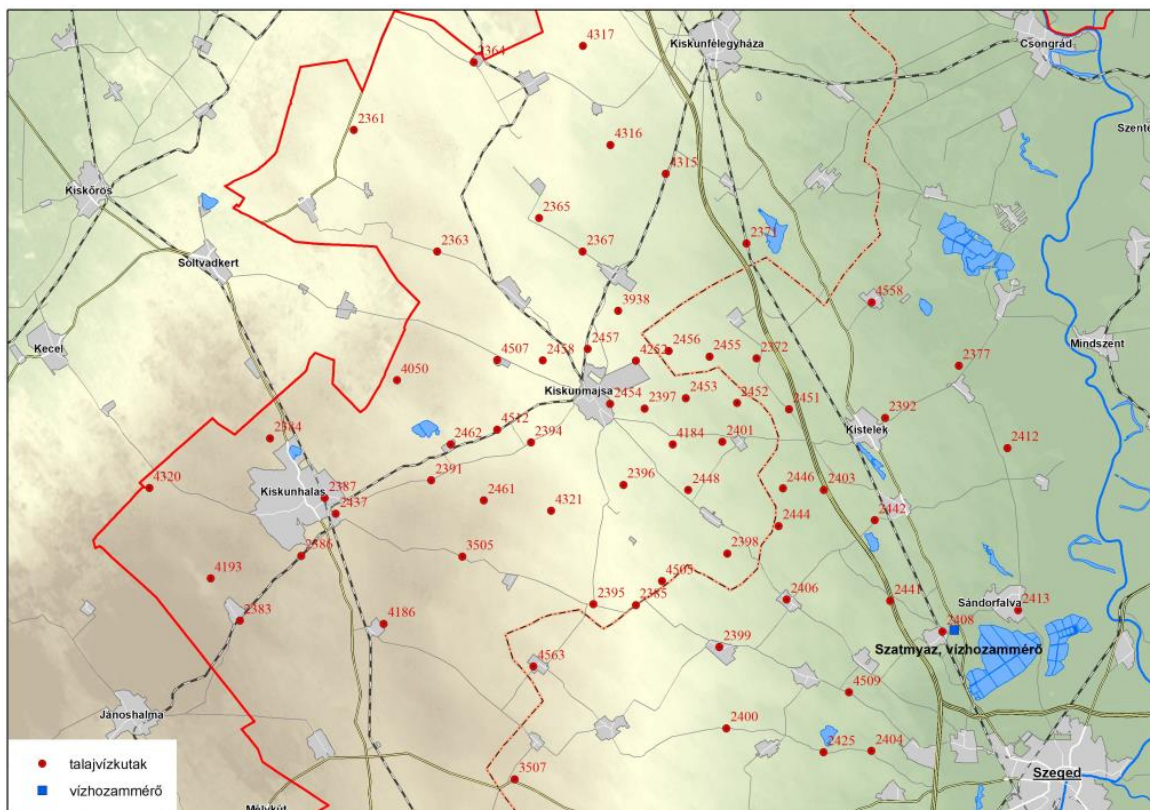
szabad forráskóddal felhasználhatók – de érzékenyek az adatsorok inhomogenitására, ilyen esetekben további adat előkészítési eljárások szükségesek az eredményes felhasználáshoz.



1. ábra. a Vizsgálati terület elhelyezkedése Forrás: ATIVIZIG 2022.



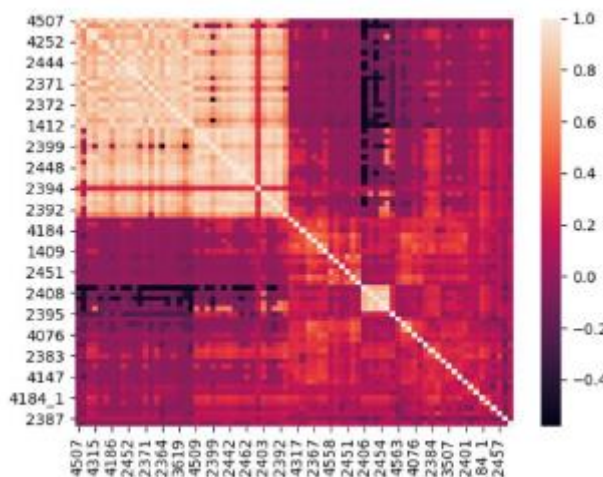
2. ábra. A mért vízhozamok és a vizsgálatba vont talajkutak elhelyezkedése Forrás: ATIVIZIG 2022.



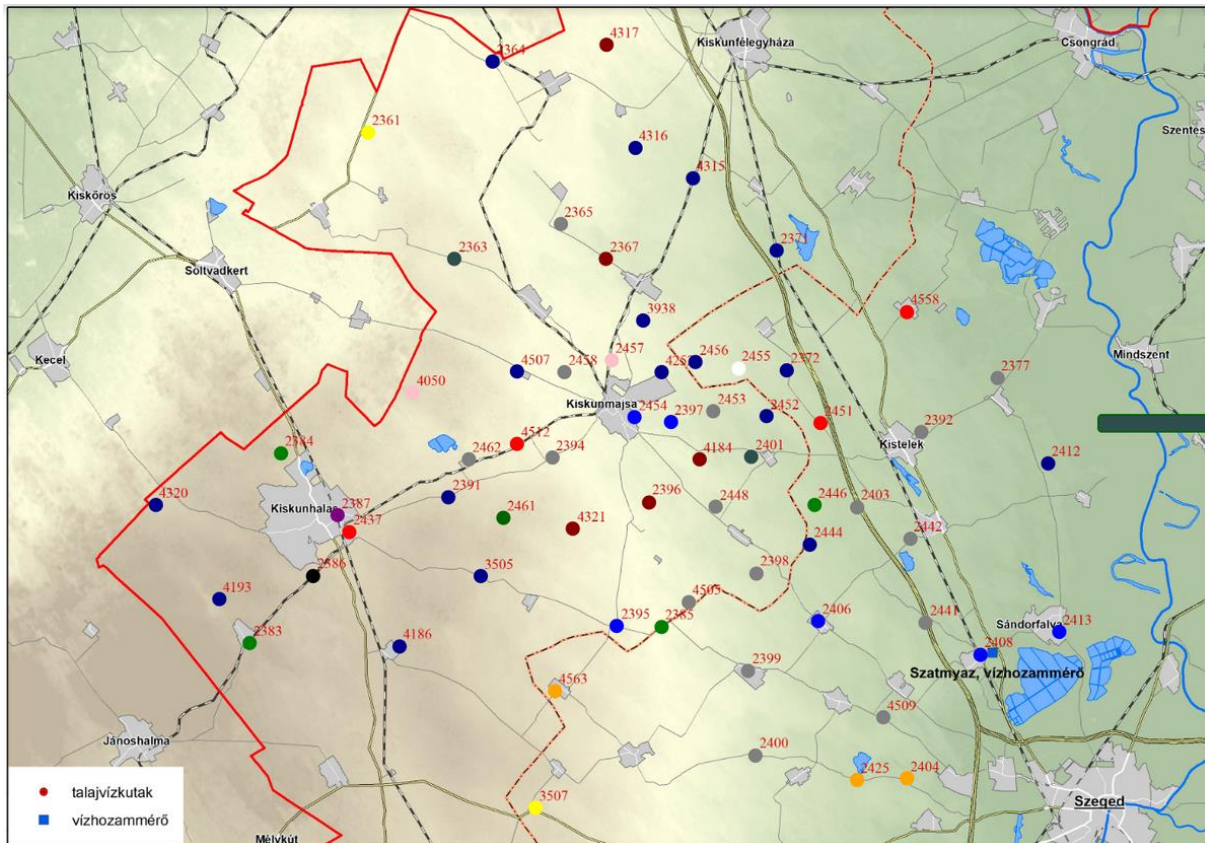
3. ábra. A vizsgálatokba vont talajkutak elhelyezkedése Forrás: ATIVIZIG 2021.

A vizsgálatok első fázisában az adatsorok előkészítését kellett elvégezni az algoritmusok hatékony alkalmazásához. A talajvizek tekintetében az alábbi korrelációs analízis lépések kerültek végrehajtásra azon célból, hogy a mérési szelvényben mért vízhozam szempontjából releváns talajvízkutak leválogatásra kerülhessenek:

- A. Szezonális tesztelése
 - a. autokorrelációs függvény alkalmazása
- B. Stacionaritás tesztelése (ACF)
 - a. augmented Dickey-Fuller teszt (ADF)
 - b. differenciaképzés
- C. Korrelációs mátrix képzése
 - a. ACF és CCF (kereszt-korreláció)



4. ábra. Keresztkorreláció mátrix Forrás: Vízy Zs. 2021.



5. ábra. A legszorosabb korrelációval rendelkező talajvízkutak térképi megjelenítése

Forrás: Víz Zs. 2022.

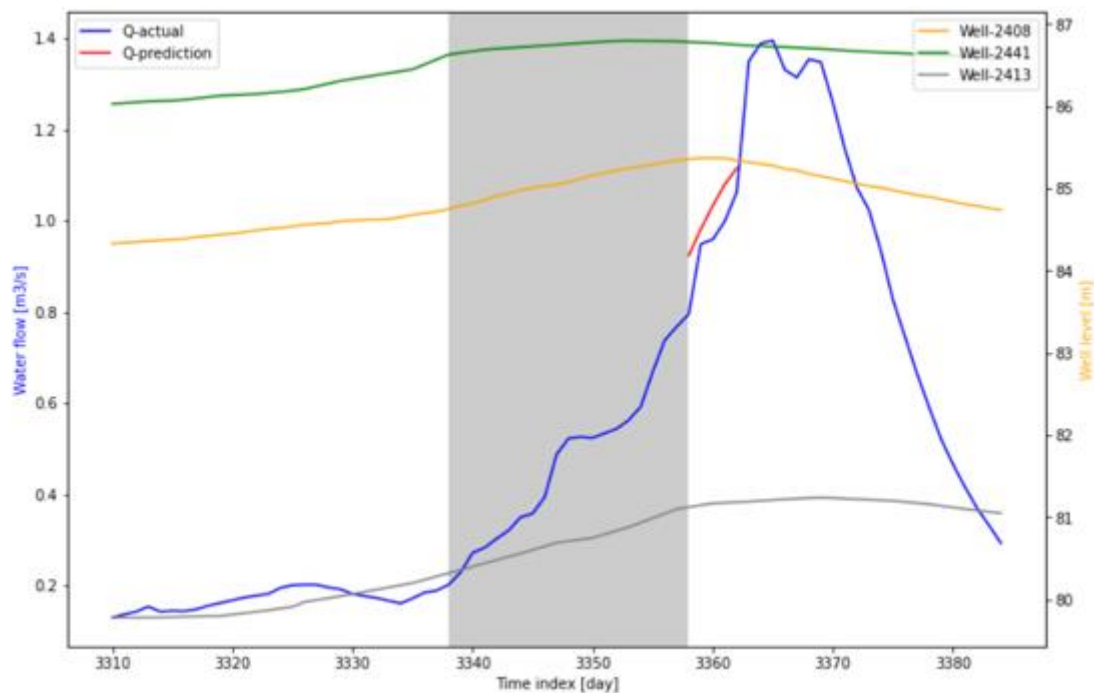
Az előkészítő vizsgálatok eredményeként kiválasztásra kerültek azon talajvízkutak amelyek vízjárása a legjobb egyezést mutatta a mérési szelvény vízhozam adataival.

A vizsgálatok további szakaszában megalkotásra kerül az a rekurrens mélytanuló algoritmus mely segítségével a lefolyási szelvény vízhozamai előrejelzésre kerültek a talajvízszintek segítségével. Az algoritmus 1962-2019 közötti időszakra vonatkozó lefolyási és talajvízadatok vizsgálatára alapozva került betanításra, majd 30 napos múltbeli adatokra történő visszatekintés alapján 5 napra vonatkozó lefolyás előrejelzést határozott meg. A modellfuttatás vizsgálati szakaszát az 5. ábra szemlélteti. A ábrán kék vonal a vízállás tény adatait, míg a piros vonal az öt napra vonatkozó előrejelzést mutatja.

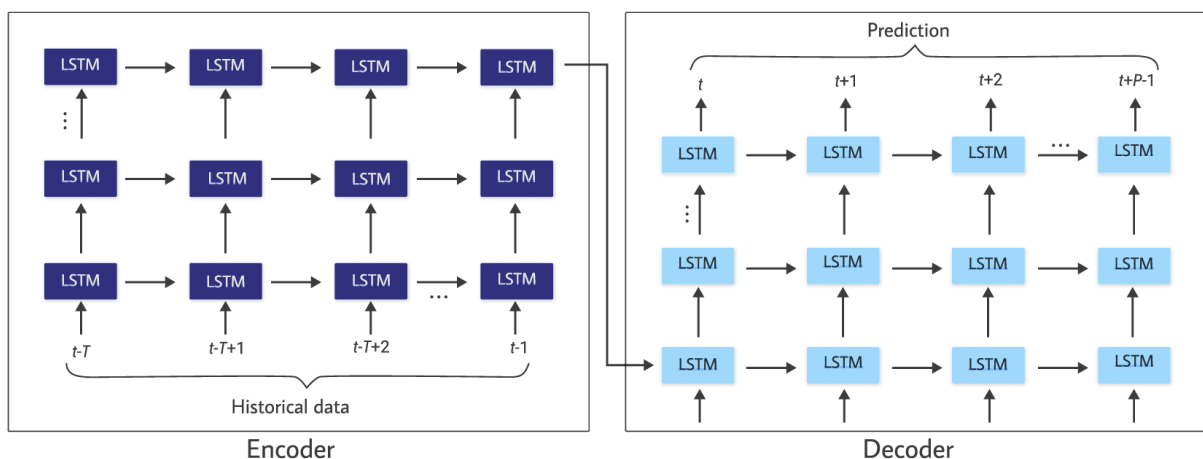
Az ábra alapján megállapítható, hogy az alkalmazott mélytanuló algoritmus alkalmas a lefolyások talajvízjárásból történő előre jelzésére. Az eljárás az ATIVIZIG-nél jelenleg tesztelési folyamatban van a későbbi napi alkalmazás feltételeinek meghatározására.

4. Folyami vízállások előrejelzése

A folyami vízállások előrejelzése során a neurális hálózatok területéről egy, az idősorok predikciójában széles körben alkalmazott családot, az LSTM-cellát [1] alkalmazó hálózattípust választottunk. Mivel egy bemeneti szekvenciából (múltbeli adatokon tekintett időablak) egy másik idősort (predikció) szeretnénk előállítani, ezért ezen a területen klasszikus encoder-decoder struktúrát ([2], 7. ábra) alkalmaztunk. Ebben az architektúrában a számolások első fázisa a múltbeli adatok feldolgozása, az információk tömörítése és az előrejelzés szempontjából releváns mintázatok kinyerése, a második részben pedig az előbb kiszámolt "tulajdonságok" alapján történik a következő időpontokhoz tartozó értékek becslése.



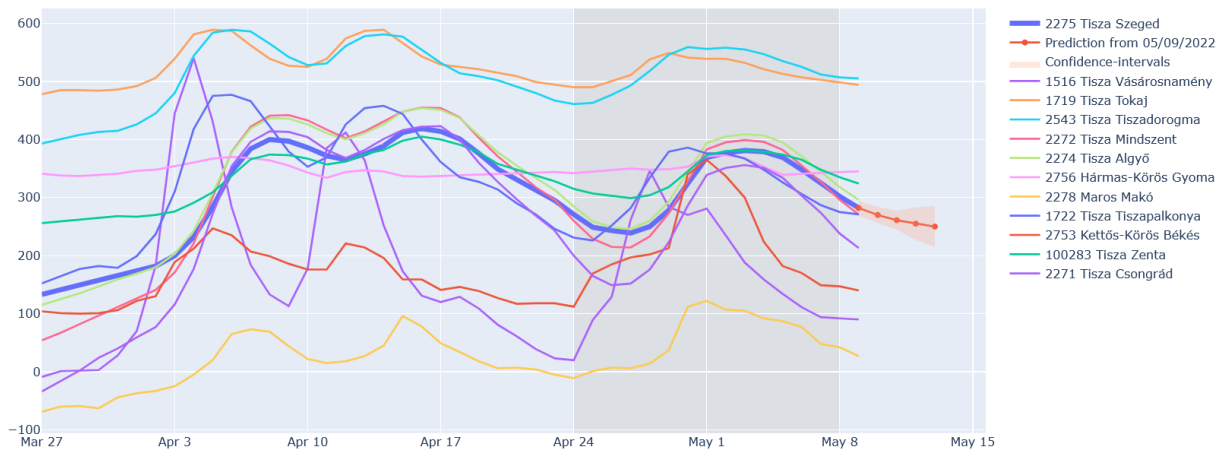
6. ábra. Lefolyás előrejelzés talajkutak idősorai alapján Forrás Vizi Zs. 2021.



7. ábra. Az idősorok előrejelzésére alkalmas encoder-decoder architektúra LSTM cellákkal. Itt a múltbeli adatokon vett időablak T széles, az első becslés a t -edik időpontra történik, a predikció hossza pedig P . Forrás Vizi Zs. 2022.

A modell paramétereit adatokon optimalizálva (másnéven a modellt tanítva), kerültek meghatározásra. A tanítási időszak 1951. január 1-től 2018. december 31-ig terjedt és 12 állomás napi adatsora lett felhasználva a tanítás során, míg 2019. évi adatok tesztelésre/validálásra kerültek bevonásra. Az adatok alapján Csongrádra, Mindszentre, Algyőre és Szegedre lettek definiálva az előrejelzési modellek. A megvalósíthatósági tanulmány elkészítése során a bemeneti idősorok halmazát a vízrajzi osztály szakemberei határozták meg, de számos egyéb matematikai alapú technika is létezik a megfelelő bemeneti halmaz megtalálására. A betanulás és az élés működés során a modell 14 napos múltbeli ablakkal dolgozik és 5 napra előre jelzi a vízállás értékeit. A predikció körüli, úgynevezett konfidenciasávot a megelőző időszakbeli predikciós hibák szórásaként számoltuk ki minden napra, ezzel a

pontszerű becsléshez képest a védekezésben dolgozó szakértő a modell bizonytalanságát is figyelembe veheti az elemzésben (8. ábra).



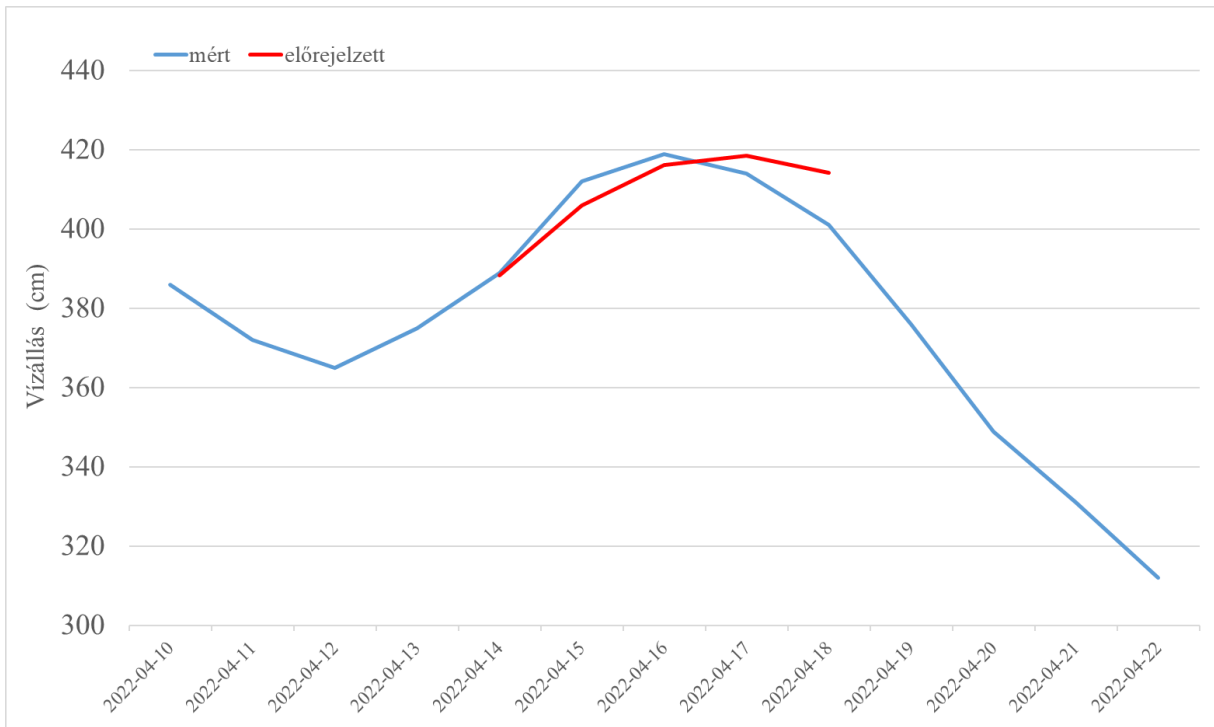
8. ábra. Az előrejelző modell működése: A szürkével jelölt rész a múltbeli időablak, a vastag vonallal ábrázolt idősorra (2275, Szeged) történik a predikció, amelynek első napja 2022. május 9.-e, a becsült adatpontokat pedig megjelenítettük a görbén. Az előrejelzés köré rajzolt konfidencia-sáv a becslés bizonytalanságát fejezi ki (az időben előrehaladva egyre szélesebb tölcserűt látunk). A jelmagyarázaton a többi vízmérce a bemeneti adatsorokat mutatja, a modell ezekből becsülte meg a jövőbeli értékeket. Forrás: Fehérvári I 2022.

A modell futtatási tapasztalatai

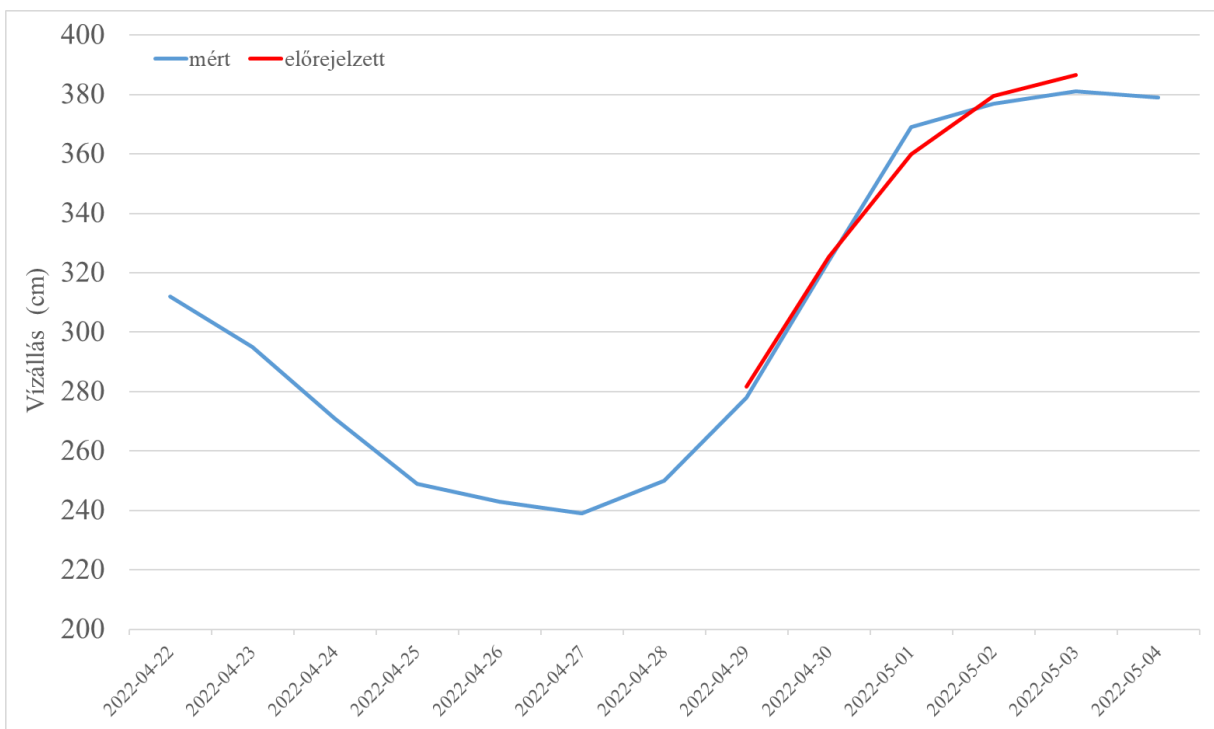
Az elkészült LSTM alapú modellt az ATIVIZIG munkatársai 2022 április-május időszakában tesztelték. A futtatások tapasztalatait az alábbi ábrákon foglaljuk össze. A 9 ábrán egy apadó hullám után egy újabb áradás érkezik a szegedi vízmércéhez. A modell az árhullám emelkedő ágában a következő napi adatot (2022.04.14) -1 cm pontossággal adata meg majd a 2. napon -6 cm-rel alacsonyabb vízállást jelez előre mint a valóságban. A 3. napon -3 cm-re mérséklődik a hiba, míg a 4. napon +4 majd az ötödik napon +13 cm a modell eltérése a valósághoz képest. A tetőzés magasságát a modell 1 cm pontossággal adja meg 1 napos késéssel, ami ez addigi előrejelzési tapasztalatokat figyelembe véve gyakorlati szempontból is használhatónak értékelhető (9. ábra).

Egy hasonló szituációban április 28-án is jól teljesít a modell. Ebben az esetben is egy apadás után kezd el újból áradni a Tisza Szegednél. A 1. napon +4 cm-rel haladja meg ez előrejelzés a mért vízállást, míg a 2. napra +2 cm-re csökken a különbség. A 3. napon -9 cm az eltérés, míg a 4. és 5. napon újból 5 cm alatti hibával számol a modell (10. ábra).

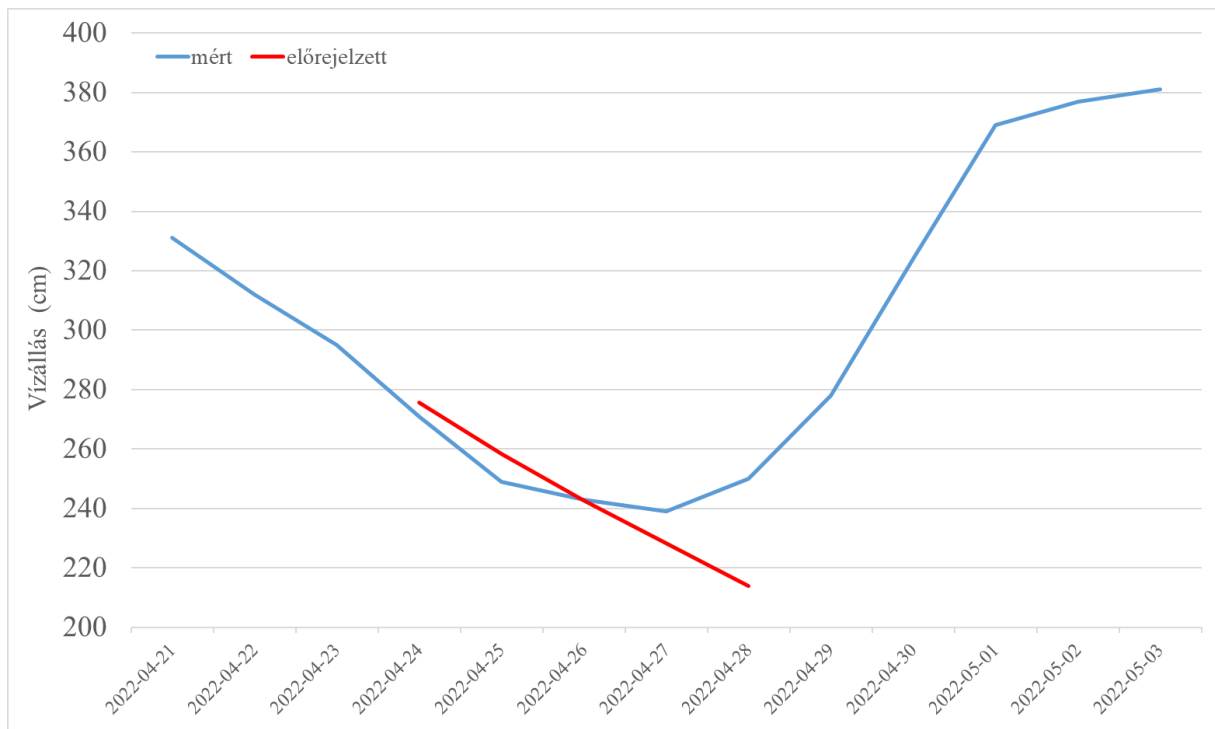
Április 23-án az apadó ágban a modell 1 napos előrejelzése 5 cm adódik magasabbnak, mint a valóság majd ez a különbség a második napra +10-cm-re nő. A harmadik napra a modell cm-re pontosan jelzi előre a vízállást, azonban a negyedik és ötödik napon nem látja előre a trend megfordulását és 10 cm meghaladó hibával becsüli alul a vízállást. Itt jegyeznénk meg azt is, hogy a Közép-Tiszáról és a Marosról történő állomások további bevonásával a modellezés pontossága előreláthatólag tovább javítható.



9. ábra. A modell előrejelzése április 13-án az elkövetkező 5 napra a szegedi vízmércére
 Forrás: Fehérvár I 2022.



10. ábra. A modell előrejelzése április 28-án az elkövetkező 5 napra a szegedi vízmércére
 Forrás: Fehérvár I 2022.



11. ábra. A modell előrejelzése április 23-án az elkövetkező 5 napra a szegedi vízmércére
Forrás Fehérváry I 2022.

A bemutatott példákon kívül több futtást végeztünk több vízmércére (Mindszent, Algyő, Csongrád) és a tapasztalatok alapján a napi gyakorlatban a többi előrejelzési módszert kiegészítve egy jól használható előrejelzés modellnek ítéltük meg a neurális hálót, hiszen pontossága ± 10 cm-en belül volt a vizsgált időszak nagy részében. A jövőben az adatok bővítése és a neurális háló struktúrájának tökéletesítése, illetve modernebb architektúrák alkalmazása [4] a pontosság növekedéséhez vezethet, ami igazolhatja a neurális háló létjogosultságát a hidrológiai modellezésben.

5. Következtetések, javaslatok

Jelen dolgozat keretein belül –a teljesség igénye nélkül- kerültek bemutatásra azon jelenleg is alkalmazás alatt álló vizsgálati eljárások, amelyekkel a mélytanuló algoritmusok vízgazdálkodási alkalmazása szemléltethető. A vízgazdálkodási folyamatok monitoringjának jelenlegi kiépítési szintje alapján a jelenlegi ismereteinket messzemenőig meghaladó új ismeretek szerezhetők a rendelkezésünkre álló nagytömegű adatok újszerű feldolgozásával. Ezen korábbiakban nem alkalmazott eljárásokkal -megfelelő előkészítettségű adatok rendelkezésre állása esetén- a vízgazdálkodás számos szakterületét érintően alapozhatók meg hatékony döntések. A hangsúly ezen eljárások esetében a rendelkezésre álló nagytömegű adat teljes körű –nem pedig ritkított- felhasználásán van. Az adatokban meglévő –és eddigiekben nem vizsgált- mintázatok és összefüggések felhasználásával növelhető a vízgazdálkodási döntések hatékonysága. A mélytanuló algoritmusok használata nem igényel speciális számítási erőforrásokat, ezáltal széleskörben alkalmazhatók a valamennyi vízgazdálkodási döntési szinten kapcsolatosan. A rekurrens típusú neurális hálózatok erőssége, hogy kifejezetten az előrejelzési feladatok számára kerültek kifejlesztésre, illetve ezen a területen több évtizedes alkalmazási gyakorlattal, és külföldi szakirodalmi háttérrel rendelkeznek.

A dolgozatban az Alsó-Tisza vidéki Vízügyi Igazgatóság és a Szegedi Tudományegyetem közös kutató-fejlesztő csoportja által elkészített alkalmazások eddigi tapasztalatai kerültek bemutatásra. Az alkalmazások minta jelleggel készültek azon célból, hogy megalapozzák egy későbbi komplex algoritmus kifejlesztését, illetve azokkal álljanak rendelkezésre tapasztalatok a mélytanuló algoritmusok vízgazdálkodási alkalmazási lehetőségeiről. Az eddigi tapasztalatok kedvezőek, mind a talajvíz-lefolyás, mind pedig a folyami vízállás előre jelzésével kapcsolatban, melyek alapján az alkalmazások használata és szükség szerinti finomítása folyamatban van.

Szakirodalom

- [1] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, 9(8), pp.1735-1780.
- [2] Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.
- [3] Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep learning*. MIT press.
- [4] Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), pp.1748-1764.