

Az ANFIS mesterséges neurális hálózat lehetséges bemeneti adatainak vizsgálata csapadék-lefolyás modellezés esetén

Négyesi Klaudia¹, Nagy Eszter Dóra²

¹ Budapesti Műszaki és Gazdaságtudományi Egyetem, Vízépítési és Vízgazdálkodási Tanszék, doktorandusz (e-mail: negyesiklaudia@edu.bme.hu)

² Budapesti Műszaki és Gazdaságtudományi Egyetem, Vízépítési és Vízgazdálkodási Tanszék, adjunktus (e-mail: nagy.eszter@emk.bme.hu)

DOI: 10.59258/hk.18334



Kivonat

Napjainkban a mesterséges neurális hálózatok használata széleskörűen elterjedt számos tudományterületen, így a hidrológia területén is egyre több kutatás készül az alkalmazhatóságuk vizsgálatára. Csapadék-lefolyás modellezés esetén az egyik legígéretesebb hálózat az úgynevezett adaptív neuro-fuzzy következtető rendszer, avagy az ANFIS (*Adaptive Neuro Fuzzy Inference System*), melynek előnye, hogy hatékonyan ötvözi mesterséges neurális hálózatok taníthatóságát a fuzzy rendszerek rugalmas szerkezetével. Az ANFIS alkalmazása lehetővé teszi a nem lineáris környezeti folyamatok modellezését és a nem lineáris komponensek összefüggéseinek azonosítását, többek között az általunk alkalmazott Takagi-Sugeno típusú rendszer segítségével. Jelen kutatás során kilenc-kilenc eltérő input adatsorral kalibrált ANFIS alapú csapadék-lefolyás modell eredményeit hasonlítottuk össze a Torna-patak és az Arany-patak vízgyűjtőire. A modelleket MATLAB környezetben építettük fel az „anfis” függvény használatával. A vizsgált input adatsorok között szerepelt a csapadék, a megelőző napi vízhozam, a megelőző csapadék index, a hőmérséklet és a potenciális párolgás. A bemeneti adatok előfeldolgozása során megtörtént az adatsorok ellenőrzése, illetve standardizálása. A felépített modellek két csoportra oszthatóak: hat-hat modell esetében figyelembe vettük, további három-három modell esetében elhagytuk a vízhozam adatok bemeneti adatsorként történő felhasználását. A modellek érzékenységvizsgálata alapján a kezdeti kapcsolatok optimális értéke 2, amíg a maximális tanítási körök értékére 500 elegendőnek bizonyult. A mintaterületekre futtatott kilenc-kilenc modell eredményei alapján megállapítható, hogy az ANFIS alapú csapadék-lefolyás modellek megfelelő modellhatékonysággal rendelkeznek a kalibráció során, azonban a modellek teljesítménye a validáció során ettől lényegesen elmarad. A vízhozam adatokat inputként alkalmazó, illetve nem alkalmazó modellek esetén is azon modellek teljesítettek a legjobban, melyek az összes vizsgált input adatsort tartalmazták. Azonban a modellek bizonyos időpillanatokban kiugró értékeket eredményeztek, melyre a lehetséges magyarázat a mesterséges neurális hálózat számítási módszerében vagy szerkezetében keresendő.

Kulcsszavak

Adaptív Neuro-Fuzzy Következtető Rendszer, ANFIS, csapadék-lefolyás modellezés, mesterséges neurális hálózat, Arany-patak, Torna-patak.

Analysis of input combinations used by the ANFIS artificial neural network for rainfall-runoff modeling

Abstract

Artificial neural networks (ANNs) are gaining popularity across various scientific fields. Thus, studies to analyze the applicability of ANNs have also been appearing within the field of hydrology. In the case of rainfall-runoff modeling, one of the most promising networks is the Adaptive Neuro-Fuzzy Inference System (ANFIS), which effectively combines the learning capability of neural networks and the flexible structure of fuzzy systems. ANFIS can model nonlinear phenomena and identify nonlinear components using, for example, the currently applied Takagi-Sugeno type system. In this study, nine-nine ANFIS-based rainfall-runoff models with different input datasets were compared for the catchments of Torna- and Arany-creek. The models were built using the MATLAB software and the "anfis" function. The input datasets included precipitation, antecedent discharge, antecedent precipitation index, temperature, and potential evaporation. The preprocessing of the data included the examinations of quality and standardization. The models can be divided into two groups: six-six models included antecedent discharge as an input, while three-three models did not use the antecedent discharge data. The sensitivity analysis of the models revealed that an optimal number of initial FIS was 2, and 500 training epochs were sufficient. Based on the results from the nine-nine models, ANFIS-based rainfall-runoff models demonstrated adequate model efficiency during calibration, but their performance decreased during validation. Whether the models used the discharge time-series as an input or not, the best-performing models were those that included all the examined input datasets. However, some models produced outlier values at certain time steps, which could be a result of the computational methods or the structure of the neural network.

Keywords

Adaptive Neuro Fuzzy Inference System, ANFIS, rainfall-runoff modeling, artificial neural network, Arany-creek, Torna-creek.

BEVEZETÉS

Napjainkban a mesterséges neurális hálózatok alkalmazása széleskörűen elterjedt minden tudományterületen, így

a hidrológiai alkalmazásukra is egyre több példát láthatunk. Ezen hálózatok előnye a nemlineáris folyamatok leírása, a gyors működés, a bemeneti adatok megválasztása-

nak lehetősége, illetve a modellezett idősor leképezésének folytonossága és differenciálhatósága. Azonban hátrány az adatigény, hiszen a kapcsolat pontos jellemzéséhez a lehető leghosszabb idősorokra van szükség, illetve a black-box működési elv, ami nem teszi lehetővé a folyamatok teljes megértését. További hátrány a korlátozott extrapolációs képesség, ugyanis amennyiben a tanítási idősor nem rendelkezik ritkán előforduló, de kiugró értékekkel, a hálózat nem fogja tudni leírni ezen eseteket a tesztelés során (*Fülöp és Józsa 1998*). A hidrológiai modellezésben az egyik legígéretesebb mesterséges neurális hálózat az úgynevezett adaptív neuro-fuzzy következtető rendszer, avagy az ANFIS (*Adaptive Neuro Fuzzy Inference System*) (*Chang és társai 2018*). Az ANFIS a mesterséges neurális hálózatok hatékony taníthatóságát ötvözi a rugalmas szerkezettel. Mindemellett képes a nem lineáris függvények modellezésére és a nem lineáris komponensek azonosítására a Takagi-Sugeno típusú rendszer segítségével. A tanítási algoritmus globális, melynek alapja a legkisebb négyzetek módszere (*Ang és társai 2023*). Előnye, hogy amellett, hogy rendelkezik a neurális hálózatok taníthatósági képességeivel, képes egy átláthatóbb struktúrát biztosítani, amely bizonyos mértékig lehetővé teszi a fizikai folyamatok interpretálását is (*Ang és társai 2023*).

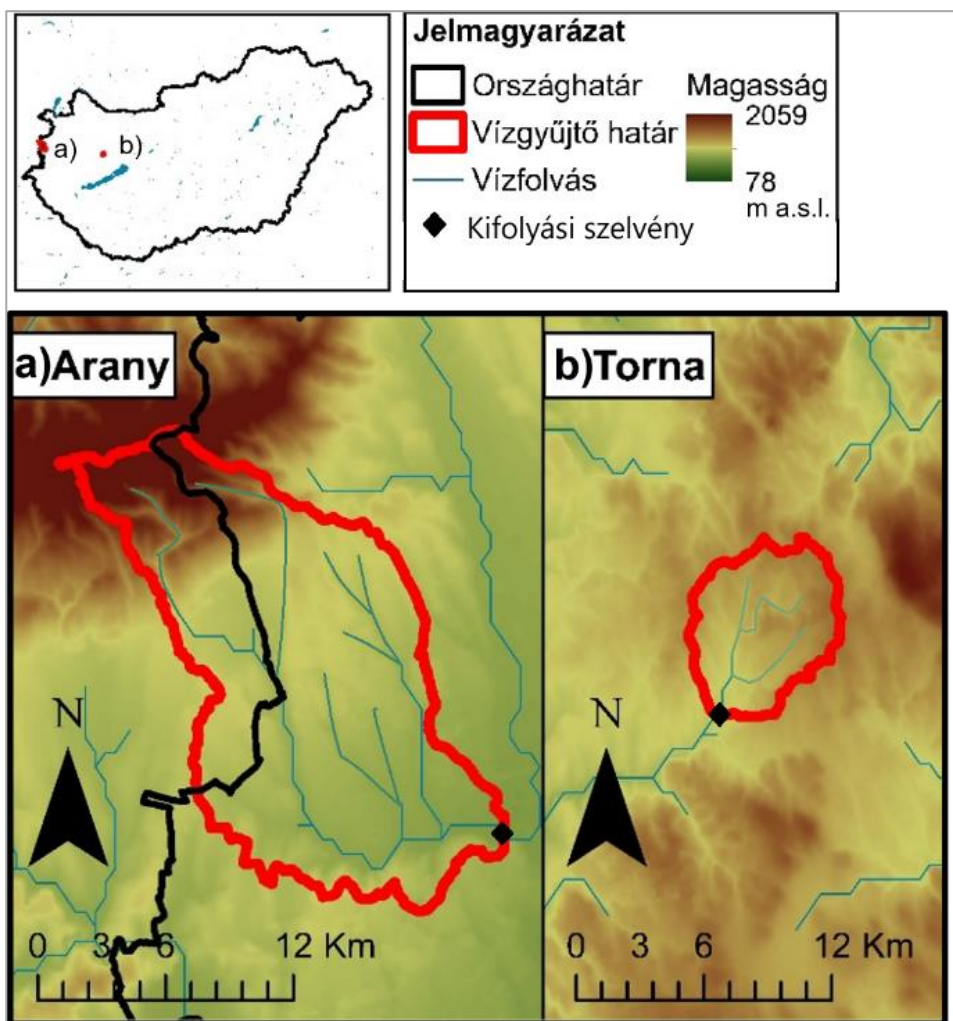
Számos nemzetközi publikáció alapján az ANFIS alapú hidrológiai modellek eredményei összevethetőek a regressziós (*Nayak és társai 2004, Ang és társai 2023, Rathnayake és társai 2023*), konceptuális (*Chang és társai 2016, Chang és társai 2018, Liptay 2022*) vagy akár fizikai (*Talei és társai 2010, Talei és Chua 2012, Talei és társai 2013*) modellek hatékonyságával, avagy képesek azonos vagy akár jobb hatékonyságot elérni. Az ANFIS alapú csapadék-lefolyás modellek alkalmasabbnak bizonyultak a tetőző vízhozamok becslésére (*Nayak és társai 2004, Tayfur és társai 2006, Wang és társai 2009, Talei és társai 2010*) az említett modellekhez képest, azonban ezen mesterséges neurális hálózat alkalmazása számos kihívást is jelent. Ezen kihívások közé tartozik, hogy a modellezés során alapvető fontosságú az adatok előzetes feldolgozása, az input adatok megválasztása, a tanítási, validálási és tesztelési idősor kijelölése, a hosszú-idejű változások figyelembevétele, a modellezett folyamatok értelmezhetősége és a paraméterek meghatározása (*Ang és társai 2023*).

A vizsgálat célja az ANFIS mesterséges neurális hálózat alkalmazhatóságának elemzése két magyarországi

vízgyűjtő csapadék-lefolyás modellezése során, melynek részét képezik az ANFIS korábbiakban ismertetett kihívásai különös tekintettel az input adatok megválasztására. Mivel az input adatok megválasztása és feldolgozása alapvetően meghatározza a modell hatékonyságát (*Talei és társai 2010*), a kutatás során mintaterületenként öt lehetséges bemenő adat 9 alternatív kombinációját vizsgáltuk mintaterületenként, ANFIS alapú csapadék-lefolyás modellek segítségével. A modellezéshez a MATLAB szoftver „anfis” függvénye kerül alkalmazásra. Az eredmények alapján meghatároztuk a javasolt input adatokat, illetve a neurális hálózat segítségével előállított modellek hatékonyságát.

VIZSGÁLT VÍZGYŰJTŐK

A modellezésre két kisvízgyűjtőt választottunk ki, melyek az Arany-patakhoz (Olad) és Torna-patakhoz (Városlőd) tartoznak. Ennek oka, hogy a kisvízgyűjtők hidrológiai modellezése kihívást jelenthet fizikai vagy konceptuális modellekkel, így egy mesterséges neurális hálózattal történő modell megoldást nyújthat, amennyiben rendelkezésre állnak a szükséges idősorok. A választott vízgyűjtők kellő adatellátottsággal rendelkeztek az ANFIS alapú csapadék-lefolyás modellek vizsgálatához, mely a szükséges vízhozam, csapadék, hőmérséklet és potenciális párolgás idősorokat jelentette. A vízgyűjtők és elhelyezkedésük az *1. ábrán* látható. A vízgyűjtők az ország nyugati részén találhatóak, az Arany-patak a Nyugat-dunántúli Vízügyi Igazgatóság területéhez, amíg a Torna-patak a Közép-dunántúli Vízügyi Igazgatóság területéhez tartozik. Az Arany-patak 106 km²-es vízgyűjtőterülettel, amíg a Torna-patak 20 km²-es vízgyűjtőterülettel rendelkezik. A vízgyűjtőkön jellemző átlagos esés 2,0% a Torna-patak, illetve 2,45% az Arany-patak esetén. A leghosszabb lefolyási hossz 9 km a Torna-patak vízgyűjtőjén, amíg 42 km az Arany-patak vízgyűjtőjén. A jellemző éves csapadék-összeg 650-750 mindkét vízgyűjtő esetén. A vizsgált vízgyűjtők területén a vályog a legmeghatározóbb talajképző kőzet, azonban Torna-patak esetén homokos vályog is nagymértékben található az AGROTOPO adatbázis alapján (*AGROTOPO 2019*). A területhasználatot tekintve az erdős területek aránya mindkét vízgyűjtő esetén jelentős, azonban az Arany-patak esetén számottevőek a mesterséges felületek is, amíg a Torna-patak esetén a mezőgazdasági területek le lehetők fel nagyobb mértékben (*CORINE 2019*).



1. ábra. A vizsgált vízgyűjtőterületek
Figure 1. The examined catchments

FELHASZNÁLT ADATOK

A kutatás során a modellek output adatsora a vízhozam (Q), lehetséges input adatsorai pedig az egy időlépéssel korábbi vízhozam ($Q(t-1)$), a csapadék (P), a hőmérséklet (T), a potenciális párolgás (PET) és a megelőző csapadék index (*Antecedent Precipitation Index, API*). A tanulmányhoz felhasználtuk az Arany-patak és a Torna-patak vízhozam idősorait 2001. január 1-től 2017. december 31-ig Olad és Városlőd állomásokra. Az adatsorok változó (átlagosan 15 perces) időbeli felbontással rendelkeztek, azonban a vizsgálathoz napi időbeli felbontással kerültek feldolgozásra hibaszűrő természete miatt. Jelen tanulmányban az ECMWF (*European Centre for Medium-Range Weather Forecasts*) reanalízis (*C3S 2019*) adatbázist használtuk fel. Az egyes szcenáriókban felhasznált hőmérséklet és a potenciális párolgás idősorai a felszíni mérőállomásokról származtak (*ODP 2023*). A vizsgálatok részét képezte a megelőző csapadék index alkalmazása, melynek számítása az (1) egyenlet alapján történt:

$$API = \sum_{t=1}^M P_t k^t, \quad (1)$$

ahol P_t a csapadék t választott időablak esetén, k pedig egy tapasztalati érték, mely 0,8-0,98 (*Koehler és Linsley 1951*) közötti értéket vehet fel. A megelőző csapadék index információt szolgáltat a vízgyűjtőn található talaj telítettség állapotáról. Jelen kutatás során az index 7 napos időablakra

vonatkozó értékét alkalmaztuk, $k = 0,98$ konstans esetén szakirodalom alapján (*Kontur és társai 2003*). A mesterséges neurális hálózatok alkalmazása során elengedhetetlen lépés az adatok standardizálása, mely a (2) formula alkalmazásával történt (*Van Ooyen és Nienhuis 1992*):

$$x_n = f_{min} + \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \cdot (f_{max} - f_{min}), \quad (2)$$

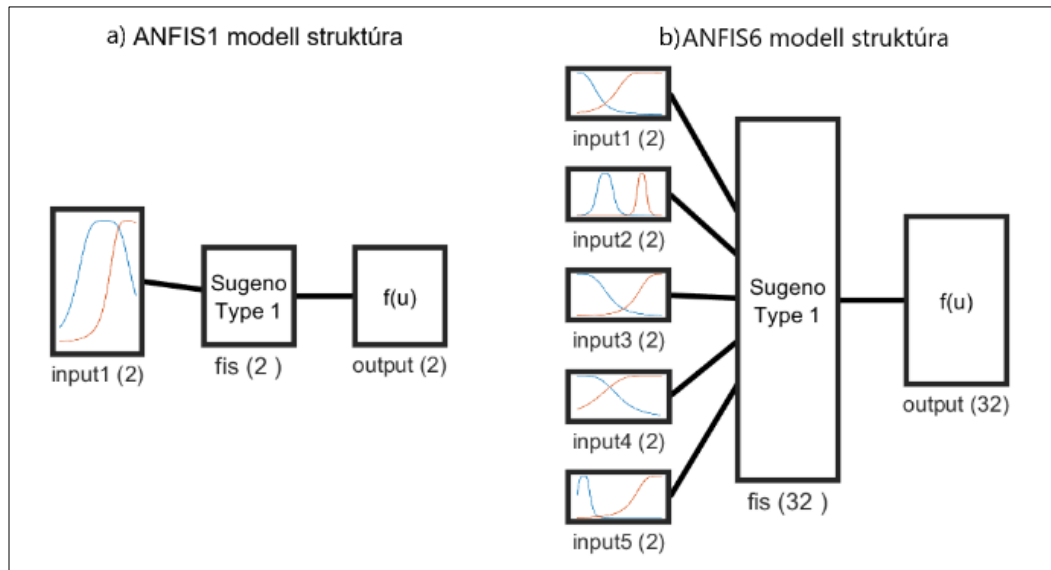
ahol a x_n standardizált adat, x_i a mért adat, f_{max} és f_{min} a kívánt értéktartomány, x_{min} és x_{max} a minimum és maximum értékei az adatsornak. Az értéktartományt $f_{min} = 0,1$ és $f_{max} = 0,9$ között vettük fel. A mesterséges neurális hálózatok esetén szintén gyakran alkalmazott adatelőkészítési eljárás a különböző szűrők alkalmazása a zajjal terhelt adatsorok esetén. Jelen kutatásban a napi időbeli felbontással rendelkező adatsorok esetén nem alkalmaztunk adatszűrést, ugyanis a kezdeti vizsgálatok során az adatok szűrése nem befolyásolta jelentős mértékben a modellek teljesítményét.

MÓDSZERTAN

A kutatás során az ANFIS modellek felépítése és alkalmazása MATLAB szoftver segítségével történt. Az alkalmazott függvény az „anfis”, melynek tanítási algoritmus hibrid, a legkisebb négyzetek módszerét és a hibavisszaterjesztéses gradiens módszert ötvözi. A fuzzy következtető

rendszer (FIS) automatikusan generált a függvény alkalmazása során az úgynevezett „grid partition” módszer segítségével (MATLAB 2024). A függvény futtatása során a bemeneti és kimeneti adatok alapján automatikusan történik a szabályok detektálása és a lineáris, il-

letve nem lineáris paraméterek hangolása. A 2. ábrán két generált szerkezet látható az ANFIS1 modell és az ANFIS6 modell példáján, előbbi esetben 1 input alapján 2 szabállyal, utóbbi esetben pedig 5 input alapján 32 szabállyal.



2. ábra. Példa a modellek szerkezetére az ANFIS1 és az ANFIS6 modellek esetén (Torna-patak)
Figure 2. Example of model structure in the case of ANFIS1 and ANFIS6 models (Torna-creek)

A MATLAB „anfis” függvényéhez megadható az input és output adatok tanítási adatsora (training data); egy validációs adatsor (validation data), mely megakadályozza a rendszer túltanítását; a mesterséges neurális hálózat szerkezetét megalapozó kezdeti kapcsolatok száma (initial FIS), illetve a maximális tanítási körök száma (epoch number). A modell érzékenységvizsgálata során a kezdeti kapcsolatok számára 2, 3, 4, 5 és 6 értékeket vizsgáltunk. Minél több a kapcsolati függvény a modellben, annál jobban növekszik a modell komplexitása, mely gyengébb modellteljesítményre vezethet az igazolás során (Chang és társai 2016, Bartoletti és társai 2017). Megvizsgáltuk, hogy hogyan befolyásolja a

modell teljesítményét, amennyiben a maximális tanítási körök értéke 5, 10, 40, 100, 200, 500, illetve 1000. A tanítási időszámok a teljes értéktartományt le kell fednie, hogy elkerülhető legyen a mesterséges neurális hálózattal történő extrapoláció. Ennek megfelelően a Torna-patak esetén 12 év hosszú adatsort, az Arany-patak esetén 13 év hosszú adatsort alkalmaztunk a hálózat tanítására. A Torna-patak esetében 2 évet, az Arany-patak esetén pedig 1 évet jelöltünk ki validációs adatsorként, amíg a modellek tesztelését mindkét esetben 3,5 éves időszor alapján végeztük el. Az 1. táblázatban a vizsgált modellekhez megválasztott input adatsorok láthatóak.

1. táblázat. Az eltérő ANFIS modellekhez választott input adatsorok.
Table 1. Input data of the different ANFIS models.

Modell	Input adatsorok
ANFIS1	Q(t-1)
ANFIS2	P, Q(t-1)
ANFIS3	P, Q(t-1), API
ANFIS4	P, Q(t-1), PET
ANFIS5	P, Q(t-1), PET, T
ANFIS6	P, Q(t-1), PET, T, API
ANFIS7	P, PET
ANFIS8	P, PET, API
ANFIS9	P, PET, API, T

A modell eredményeit négy különböző hibamutatóval vizsgáltuk. A determinációs együttható (R^2) mellett a hidrológiában gyakran alkalmazott Nash-Sutcliffe hatékonyság mutatószámot alkalmaztuk. A Nash-Sutcliffe hatékonyság mutatószám (NSE) értéke $-\infty$ és 1 közötti tartományban mozoghat. Minél közelebb van a számított érték

1-hez, annál jobbnak tekinthető a modell hatékonysága. Számítása a (3) képlet segítségével történik:

$$NSE = 1 - \frac{\sum(Q_{mért} - Q_{szim})^2}{\sum(Q_{mért} - Q_{átlag})^2} \quad (\text{Nash és Sutcliffe 1970}) \quad (3)$$

A négyzetes középérték hiba (RMSE) 0 és $+\infty$ közötti értékeket vehet fel. Minél jobbnak közelít a (4) egyenlet

alapján számított érték 0-hoz, annál nagyobb a modell hatékonysága.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |Q_{mért} - Q_{szim}|} \quad (\text{Jackson és társai 2019}) \quad (4)$$

Az átlagos abszolút relatív hiba (MARE) egy adatsor esetén a negatív értékeket figyelmen kívül hagyja, illetve egy konstans érték segítségével kerül el a zérus értékkel járó számítási hibákat. Értéke 0 és $+\infty$ közötti tartományon mozoghat. Amennyiben az (5) egyenlettel számított érték közelítőleg 0, a modell teljesítménye kiváló:

$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Q_{szim}^* - Q_{mért}^*}{Q_{mért}^*} \right| \quad (5)$$

2. táblázat. Hatékonyság mutatószámok értékei a különböző ANFIS modellek esetén.

Table 2. Values of model efficiency in the case of the different ANFIS models.

Modell	Folyamat	Arany-patak				Torna-patak			
		NSE	RMSE	R ²	MARE	NSE	RMSE	R ²	MARE
ANFIS1	Kalibráció	0,505	0,009	0,505	0,491	0,749	0,008	0,749	0,106
	Tesztelés	0,300	0,004	0,518	0,357	0,172	0,012	0,279	0,132
ANFIS2	Kalibráció	0,541	0,008	0,541	0,520	0,788	0,007	0,788	0,142
	Tesztelés	0,190	0,004	0,473	0,375	0,034	0,013	0,221	0,150
ANFIS3	Kalibráció	0,672	0,007	0,672	0,530	0,857	0,006	0,857	0,122
	Tesztelés	0,552	0,003	0,594	0,365	-1,563	0,022	0,087	0,145
ANFIS4	Kalibráció	0,532	0,008	0,533	0,317	0,804	0,007	0,804	0,122
	Tesztelés	0,377	0,003	0,607	0,271	0,159	0,012	0,257	0,136
ANFIS5	Kalibráció	0,685	0,007	0,685	0,632	0,843	0,006	0,843	0,109
	Tesztelés	0,360	0,004	0,446	0,501	-9,225	0,043	0,026	0,203
ANFIS6	Kalibráció	0,825	0,005	0,826	0,397	0,916	0,005	0,916	0,104
	Tesztelés	0,420	0,003	0,530	0,343	-3,011	0,027	0,049	0,201
ANFIS7	Kalibráció	0,218	0,011	0,218	1,531	0,105	0,015	0,105	0,771
	Tesztelés	-13,156	0,017	0,001	1,435	0,127	0,013	0,141	0,346
ANFIS8	Kalibráció	0,425	0,009	0,425	1,349	0,379	0,013	0,379	0,692
	Tesztelés	-0,355	0,005	0,113	1,031	0,075	0,013	0,115	0,311
ANFIS9	Kalibráció	0,473	0,009	0,474	1,300	0,439	0,012	0,439	0,651
	Tesztelés	-3,831	0,010	0,040	1,064	0,142	0,013	0,160	0,303

Ahogy az 1. táblázatban is látható volt, az első hat modell esetén vettük figyelembe az egy nappal korábbi vízhozam értékeket, mint bemeneti adat. Az Arany- és a Torna-patak esetén is megfigyelhető, hogy a kalibráció során minden esetben legalább megfelelő (NSE > 0,5), több esetben jó (NSE > 0,7) modellhatékonyságot képesek elérni az ANFIS alapú csapadék-lefolyás modellek. Azonban a tesztelés, avagy igazolás során a modellek többségének hatékonysága jelentősen csökken, ahogyan az a 3. ábrán is látható. Más magyarországi vízgyűjtőn elvégzett kutatás során Liptay (2022) is rávilágított a mesterséges neurális háló alapú csapadék-lefolyás modellek kiemelkedő modellhatékonyságára a tanítási időszak alatt, azonban arra is felhívta a figyelmet, hogy az igazolás során a modellek teljesítménye már sokkal gyengébb.

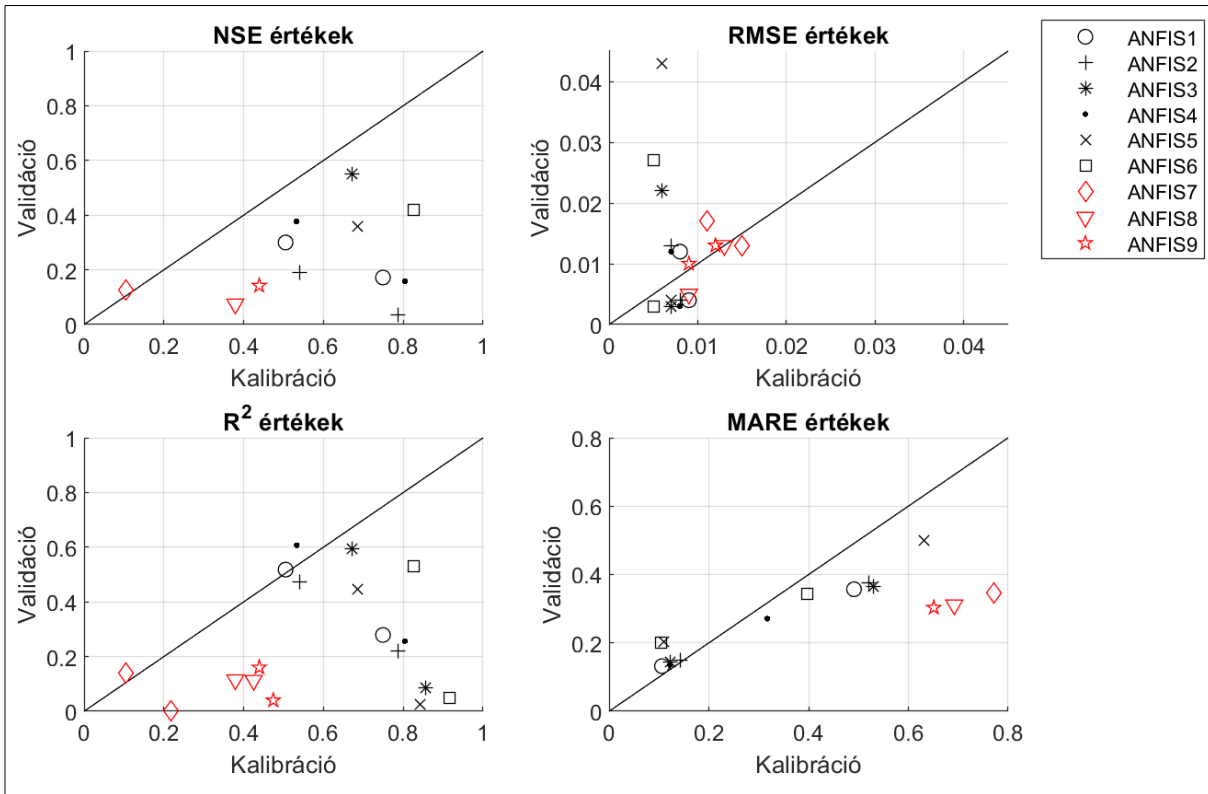
A modellhatékonyság mutatószámai alapján a megelőző vízhozam adatokat tartalmazó modellek közül az

ahol $Q_{szim}^* = Q_{szim} + \frac{Q_{átlag}}{100}$ és $Q_{mért}^* = Q_{mért} + \frac{Q_{átlag}}{100}$ (Pushpalatha és társai 2012).

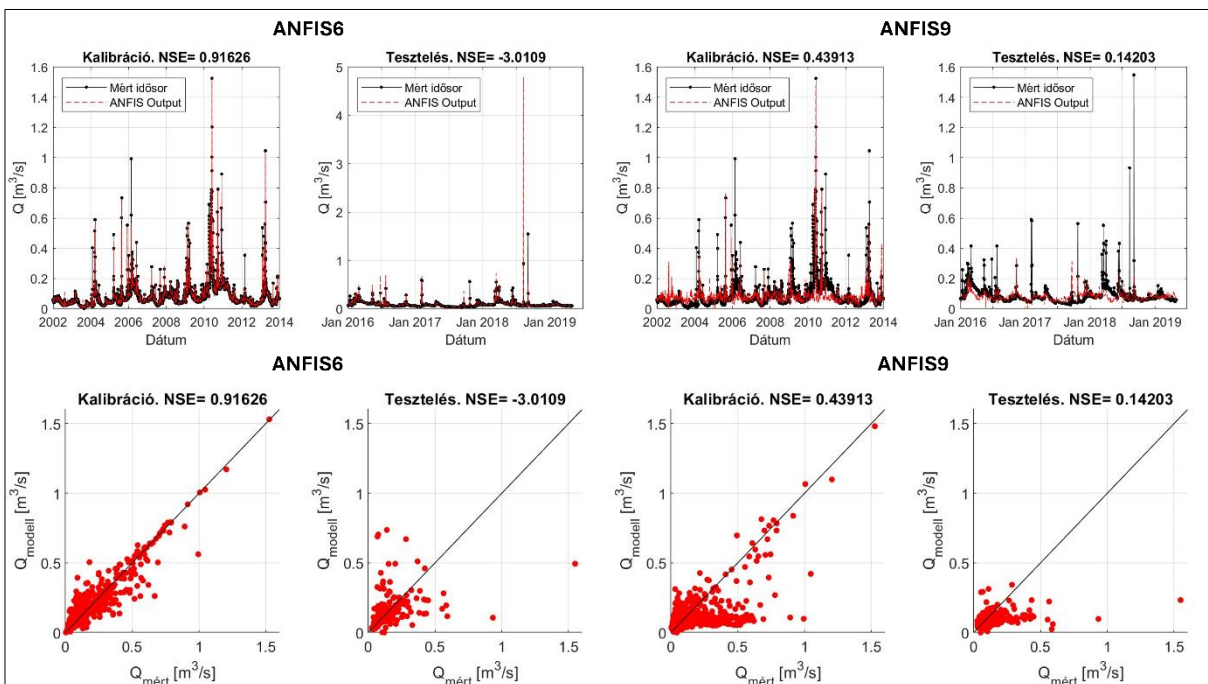
EREDMÉNYEK

Az érzékenységvizsgálat eredményei alapján a kezdeti kapcsolatok optimális száma 2, mivel amikor több kezdeti kapcsolatot alkalmaztunk, az rontotta a modell teljesítményét, valamint növelte a számítási időt is. A maximális tanítási körök értékére 500 elegendőnek bizonyult, biztosítva a modell hatékony konvergenciáját. A 2. táblázatban a kilenc ANFIS modell által elért hatékonyság mutatószámok értékei láthatóak a két vizsgált vízgyűjtő esetén.

ANFIS6, a vízhozamot input adatként nem alkalmazó modellek közül pedig az ANFIS9 az egyik leghatékonyabb modell struktúra a kalibrálás során. Ez a két modell, tekintve a vízhozamot alkalmazó, illetve nem alkalmazó modellek csoportjait, az összes vizsgált input adatsort tartalmazták. A 3. ábra alapján alacsonyabb hatékonysággal azon modellek rendelkeztek, amelyek nem használják a vízhozamot input adatként. A 4. ábrán láthatóak a két modell tanítási és tesztelési idősorai, illetve a modellezett és mért vízhozamok egymáshoz viszonyított értékei a Torna-patak esetén. Az ANFIS6 modell esetén a tanítási időszak alatt kimagasló a modell hatékonysága. A tesztelés során látható, hogy habár a legtöbb tetőzési értéket megközelíti a modell, van olyan árhullám, mely esetén jelentősen felülbecsli a tetőzés értékét a modell. Az ANFIS9 esetén a modell már több esetben alulbecsülte a mért vízhozamot a kalibráció során is. A tesztelés során a legnagyobb tetőzési vízhozamokat nem tudta a modell közelíteni.



3. ábra. Hatékonyság mutatószámok értékei a kalibráció és a validáció esetén.
Figure 3. Values of model efficiency in the case of calibration and validation.



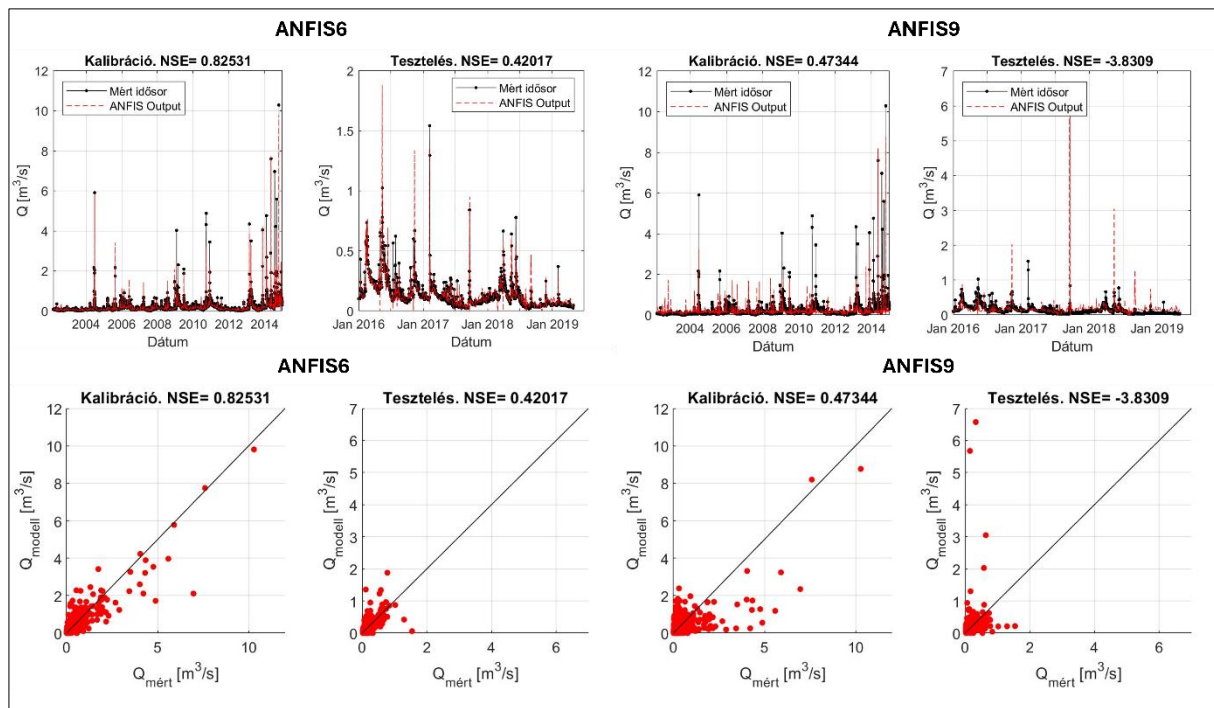
4. ábra. Az ANFIS6 és ANFIS9 modellek eredményei (Torna-patak).
Figure 4. The results of ANFIS6 and ANFIS9 models (Torna-creek).

A 4. ábra a Torna-patak esetén szemlélteti az ANFIS6 és ANFIS9 csapadék-lefolyás modellek eredményeit. Az ANFIS6 a tanítási és a tesztelési időszak során is megfelelő hatékonyságot mutat, amíg az ANFIS9 a kalibráció során több esetben alulbecsli a mért vízhozam értékét, a tesztelés során pedig számos esetben felülbecsli a tetőző vízhozam értékeket. Megfigyelhető,

hogy az ANFIS9 a tesztelés során egy esetben jelentős felülbecslést ad egy árhullám tetőzési vízhozamára. A 4. ábra és az 5. ábra alapján a Torna-patak és az Arany-patak esetén is megfigyelhető, hogy bizonyos esetekben az ANFIS modellek kiugró értékeket eredményeznek. A kiugró értékekre lehetséges magyarázat lehetne az extrapoláció, avagy a tesztelési adatsor nagyobb értékeket

tartalmaz, mint a tanítási idősor. Jelen kutatásban nem történt extrapoláció, így további lehetséges magyarázat

a neurális hálózat számítási módszerében vagy szerkezetében keresendő.



5. ábra. Az ANFIS6 és ANFIS9 modellek eredményei (Arany-patak).
Figure 5. The results of ANFIS6 and ANFIS9 models (Arany-creek).

DISZKUSSZIÓ

Az ANFIS típusú mesterséges neurális hálózat jelen kutatás során is biztató eredményeket adott csapadék-lefolyás modellezés esetén, azonban az ismert kihívások mellett számos újabb kérdés felmerült az alkalmazása során. Ahogyan az a korábbiakban ismertetésre került, az egyik elsődleges fontosságú feladat neurális hálózatok alkalmazása előtt az adatok előkészítése, melyhez elengedhetetlen az adatok gondos megválasztása és minőségének ellenőrzése, illetve az adatsorok standardizálása. További megfontolandó lehetőség az adatok előfeldolgozásához az adatok transzformálása vagy szűrő alkalmazása az adatsoron.

Az adatsorok feldolgozása mellett a tanítási és tesztelési időszakok megválasztása is alapvetően meghatározza a modell hatékonyságát. Az adatsorok megválasztása során az extrém értékekre is figyelemmel kell lenni, elkerülve az extrapolációt, illetve a kevés előfordulásból adódó alulbecsléseket a neurális hálózat által. Mivel a matematikai modellek alapvetően érzékenyek a kiugró értékekre, jelen kutatásban is az extrém értékek jelentették a legnagyobb kihívást az ANFIS számára. Ennek oka, hogy a modell reprezentativitása csökken olyan helyzetekben, melyek ritkán fordulnak elő. Kulcsfontosságú feladat a neurális hálózatok megtanítása a lehetséges extrém környezeti jelenségekre anélkül, hogy hibás extrém értékeket jelezzenek elő vagy hibás megfigyeléseket vegyenek alapul.

A neurális hálózatok használatának további lényeges kihívása a bemeneti adatok megválasztása, mely jelen kutatás fő témaköre. A bemeneti adatok megválasztása kapcsán nem elegendő a rendelkezésre álló adatsorok minőségi ellenőrzése. A modellezett vízgyűjtő meghatározó

hidrológiai folyamatai alapján szükséges átgondolni a lehetséges bemeneti adatsorokat. Ennek megfelelően a kutatásban vizsgált adatsorok mellett akár szél, hóolvadási vagy talajnedvességi adatsorok is javíthatják a modellek teljesítményét, amennyiben rendelkezésre állnak. Jelen két vizsgált vízgyűjtő alapján az elvárásoknak megfelelően a megelőző vízhozam jelentősen javítja a modellek teljesítményét. Emellett a megelőző csapadék index alkalmazása mutat javulást a hatékonyságot illetően, mely jellemzést adhat a vízgyűjtő talajnedvességéről, lefolyási tulajdonságairól.

A kutatást számtalan irányban lehet továbbfejleszteni, illetve kiegészíteni. Egy meghatározó irány lehet az ANFIS hálózat felépítésének elemzése, eltérő számítási módszer alkalmazása. Mindemellett napjainkig is csekély tanulmány található az eltérő mesterséges neurális hálózatok vizsgálatára hazánkban, annak ellenére, hogy rengeteg más modell alkalmazására lenne lehetőség, akár LSTM (Long Short Term Memory), SVM (Support Vector Machine), GP (Genetic Programming) vagy mély tanulás (Deep Learning). A mesterséges intelligencián alapuló modellek mellett elemezhető irány lehetne a mesterséges neurális hálózatok és az egyéb modellek, például autoregressziós modellek ötvözése.

ÖSSZEFOGLALÁS

A mesterséges neurális hálózatokat napjainkban már minden tudományterületen alkalmazzák. A hidrológia területén is egyre több kutatás készül a mesterséges neurális hálózatok alkalmazhatóságának vizsgálatára, melyek közül csapadék-lefolyás modellezés esetén az egyik legígéretesebb hálózat az úgynevezett adaptív neuro-fuzzy következtető rendszer, avagy az ANFIS (*Adaptive Neuro Fuzzy*

Inference System). Az ANFIS előnye, hogy hatékonyan ötvözi mesterséges neurális hálózatok taníthatóságát a fuzzy rendszerek rugalmas szerkezetével. Az ANFIS képes a nem lineáris függvények modellezésére és a nem lineáris komponensek azonosítására a Takagi-Sugeno típusú rendszer segítségével. Az ehhez alkalmazott tanítási algoritmus globális, melynek alapja a legkisebb négyzetek módszere (Ang és társai 2023).

Jelen tanulmányban összehasonlítottunk kilenc eltérő input adatsorral rendelkező ANFIS mesterséges neurális hálózat alapú csapadék-lefolyás modellt a Torna-patak és az Arany-patak vízgyűjtőire. A modelleket MATLAB szoftver segítségével építettük fel az „anfis” függvény használatával. A vizsgált input adatsorok között szerepelt a csapadék, a megelőző napi vízhozam, a megelőző csapadék index, a hőmérséklet és a potenciális párolgás, melyek előfeldolgozása során megtörtént az adatsorok ellenőrzése, illetve standardizálása. A modellek két csoportra bonthatóak: hat modell tartalmazza a megelőző vízhozamot input adatként, amíg három modell nem használ vízhozam adatokat bemeneti adatsorként. A modellek érzékenységvizsgálata alapján a kezdeti kapcsolatok optimális értéke 2, mivel a több kezdeti kapcsolat alkalmazása rontotta a modell teljesítményét, valamint növelte a számítási időt. A maximális tanítási körök értékére 500 elegendőnek bizonyult, biztosítva a modellek hatékony konvergenciáját. A vizsgált kilenc modell eredményei alapján megállapítható, hogy az ANFIS alapú csapadék-lefolyás modellek megfelelő modell hatékonysággal rendelkeznek a kalibráció során, azonban a modellek teljesítménye csökken az igazolás során. Mindkét csoport esetén azon modellek teljesítettek a legjobban, melyek az összes vizsgált input adatsort tartalmazzák. Azonban a Torna-patak és az Arany-patak vizsgálata során az ANFIS modellek bizonyos időpillanatokban kiugró értékeket eredményeztek. A kiugró értékekre lehetséges magyarázat lehetne az extrapoláció, azonban jelen kutatásban a tesztelési idősor értéktartománya nem lépte túl a tanítási idősor értéktartományát. A további lehetséges magyarázat a neurális hálózat számítási módszerében vagy szerkezetében keresendő.

A vizsgált ANFIS alapú csapadék-lefolyás modellek fejlesztéséhez további elemzések szükségesek. A kutatás továbbfejlesztése során az adatok előfeldolgozási lehetőségeinek széleskörű feltárása segítheti a modellek hatékonyságának növelését. A kiugró értékek kezelésére a jövőbeli kutatás során elengedhetetlen a számítási módszer és a hálózat szerkezetének felülvizsgálata. A jelen tanulmányban alkalmazott Takagi-Sugeno rendszer helyett vizsgálható lehetne a Mandami-típusú rendszerek alkalmazása, illetve az optimalizációs módszerek hibrid alkalmazása helyett csak a hiba visszaterjesztéses gradiens módszer használata. Mindemellett vizsgálati lehetőség az eltérő módszerek alkalmazása a FIS generálására, hogy elérhető legyen a lehető legjobb modellelhatékonyságot eredményező modell struktúra.

KÖSZÖNETNYILVÁNÍTÁS

A közlemény a Kulturális és Innovációs Minisztérium ÚNKP-23-3-I-BME-167 kódszámú Új Nemzeti Kiválóság

Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

IRODALOMJEGYZÉK

AGROTOPO (2019). <https://www.mta-taki.hu/hu/osztalyok/kornyezetiinformatikai-osztaly/agrotopo> (Letöltve: 2019.10.25.)

Ang, Y., Talei, A., Zahidi, I., Rashidi, A. (2023). Past, Present, and Future of Using Neuro-Fuzzy Systems for Hydrological Modeling and Forecasting. *Hydrology*. 10. 36. <https://doi.org/10.3390/hydrology10020036>

Bartoletti, N., Casagli, F., Marsili-Libelli, S., Nardi, A., Palandri, L. (2017). Data-driven rainfall/runoff modelling based on a neuro-fuzzy inference system. *Environmental Modelling & Software*. <https://doi.org/10.1016/j.envsoft.2017.11.026>

Chang, T.K., Talei, A., Alaghmand, S., Ooi, M. (2016). Choice of Rainfall Inputs for Event-based Rainfall-Runoff Modeling in a Catchment with Multiple Rainfall Stations Using Data-driven Techniques. *Journal of Hydrology*. 545. <https://doi.org/10.1016/j.jhydrol.2016.12.024>

Chang, T.K., Talei, A., Chua, L., Alaghmand, S. (2018). The Impact of Training Data Sequence on the Performance of Neuro-Fuzzy Rainfall-Runoff Models with Online Learning. *Water*. 11. 52. <https://doi.org/10.3390/w11010052>

C3S (2019). *Copernicus Climate Change Service ERA5-Land reanalysis*. Copernicus Climate Change Service, 15/09/2019. <https://cds.climate.copernicus.eu/cdsapp#!/home> (Letöltve: 2019.09.18.)

CORINE (2019). *Land Cover-Copernicus* <https://land.copernicus.eu/pan-european/corine-land-cover> (Letöltve: 2019.10.25.)

Fülöp, I., Józsa, J. (1998). A neurális hálózatok világa. *Hidrológiai Közlöny*, 78. évfolyam, 4. szám, p. 250.

Jackson, E., Roberts, W., Nelsen, B., Williams, G., Nelson, E., Ames, D. (2019). Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environmental Modelling & Software*. 119. <https://doi.org/10.1016/j.envsoft.2019.05.001>

Kohler, M.A., Linsley, R.K. (1951). Predicting the runoff from storm rainfall. *Weather Bureau Research Papers*. Washington.

Kontur, I., Koris, K., Winter, J. (2003). Hidrológiai számítások. Lanograf Kft., Gödöllő.

Liptay, Z. (2022). Neurohydrological prediction of water temperature and runoff time series. *Acta Hydrologica Slovaca*. 23. pp. 190-196. <https://doi.org/10.31577/ahs-2022-0023.02.0021>.

MATLAB (2024). *MATLAB Documentation – anfis*. <https://www.mathworks.com/help/fuzzy/anfis.html>

ODP (2023). *Meteorológiai Adattár* <https://odp.met.hu/> (Letöltve: 2023.09.25.)

Nash, J.E., Sutcliffe, J.V. (1970). River flow forecasting through conceptual models. Part I – A discussion of principles. *Journal of Hydrology* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)

Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., (2004). A neuro-fuzzy computing technique for modeling hydrological time series. *J. Hydrol.* 291 (1–2), pp. 52–66. <https://doi.org/10.1016/j.jhydrol.2003.12.010>

Pushpalatha, R., Perrin, C., le Moine, N. Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology.* 420-421, 171-182. <https://doi.org/10.1016/j.jhydrol.2011.11.055>

Rathnayake, N., Rathnayake, U., Chathuranika, I., Dang, L., Hoshino, Y. (2023). Cascaded-ANFIS to simulate nonlinear rainfall-runoff relationship. *Applied Soft Computing.* 147. <https://doi.org/10.1016/j.asoc.2023.110722>

Talei, A., Chua, L. (2012). Influence of lag time on event-based rainfall–runoff modeling using the data driven approach. *Journal of Hydrology.* s 438–439. pp. 223–233. <https://doi.org/10.1016/j.jhydrol.2012.03.027>

Talei, A., Chua, L., Quek, C. (2010). A novel application of a neuro-fuzzy computational technique in event-based rainfall–runoff modeling. *Expert Systems with Applications.* 37. pp. 7456-7468. <https://doi.org/10.1016/j.eswa.2010.04.015>

Talei, A., Chua, L., Quek, C., Jansson, P. (2013). Runoff forecasting using a Takagi–Sugeno neuro-fuzzy model with online learning. *Journal of Hydrology.* 488. 17–32. <https://doi.org/10.1016/j.jhydrol.2013.02.022>

Tayfur, G., Singh, V., Asce, F. (2006). ANN and Fuzzy Logic Models for Simulating Event-Based Rainfall-Runoff. *Journal of Hydraulic Engineering.* [https://doi.org/10.1061/\(ASCE\)0733-9429\(2006\)132:12\(1321\)](https://doi.org/10.1061/(ASCE)0733-9429(2006)132:12(1321))

Van Ooyen, A., Nienhuis, B. (1992). Improving the convergence of the backpropagation algorithm. *Neural Networks* 5 (3), pp. 465–471. [https://doi.org/10.1016/0893-6080\(92\)90008-7](https://doi.org/10.1016/0893-6080(92)90008-7)

Wang, W., Chau, K., Cheng, C., Qiu, L. (2009). A Comparison of Performance of Several Artificial Intelligence Methods for Forecasting Monthly Discharge Time Series. *Journal of Hydrology.* 374. pp. 294-306. <https://doi.org/10.1016/j.jhydrol.2009.06.019>

SZERZŐK



NÉGYESI KLAUDIA 2022-ben szerzett okleveles infrastruktúra-építőmérnök diplomát a Budapesti Műszaki és Gazdaságtudományi Egyetem Építőmérnöki Karán, jelenleg a Vízépítési és Vízgazdálkodási Tanszék harmadéves doktorandusz hallgatója. Doktori témájának címe „Kisvízgyűjtők csapadék-lefolyás kapcsolatának modellezése”. 2020 óta a Magyar Hidrológiai Társaság tagja.



NAGY ESZTER DÓRA okleveles infrastruktúra-építőmérnök (Budapesti Műszaki és Gazdaságtudományi Egyetem, 2017), 2023-ban szerzett PhD fokozatot a Budapesti Műszaki és Gazdaságtudományi Egyetemen a földtudományok területén, építőmérnöki tudományágból. 2022 óta főállású oktató, jelenleg a Budapesti Műszaki és Gazdaságtudományi Egyetem Vízépítési és Vízgazdálkodási Tanszékén adjunktus. Túlnyomórészt hidrológiával, vízkészlet-gazdálkodással kapcsolatos tárgyakat oktat, fő kutatási területe a kisvízgyűjtők hidrológiája és a csapadék-lefolyás modellezés. 2016 óta a Magyar Hidrológiai Társaság tagja.