

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Vízi Közmű és Környezetmérnöki Tanszék

# Debrecen város rövidtávú ivóvízfogyasztás előrejelzése

Tudományos Diákköri Dolgozat

Készítette: Bíbok Attila

Témavezető: Dr. Buzás Kálmán, egyetemi docens  
Dr. Darabos Péter, adjunktus

Budapest, 2012

## Tartalomjegyzék

1. Bevezetés.....	3
2. Felhasznált adatok .....	4
2.1. Rendelkezésre álló adatsorok .....	4
2.2. Ivóvízellátó rendszer leírása.....	5
3. Adatok feldolgozása.....	7
3.1. Hibás adatok kiszűrése.....	8
3.2. Archivált értékek interpolálása .....	9
3.3. Félórás fogyasztási adatok számítása .....	10
3.3.1. Szivattyúk fogyasztásának számítása .....	12
3.3.2. Tározók látszólagos fogyasztás számítása .....	13
3.4. Napi fogyasztási adatok számítása.....	14
3.5. Számított adatsorok ellenőrzése.....	14
4. Napi fogyasztási adatok előrejelzése .....	16
4.1. Vizsgált paraméterek .....	17
4.2. Előrejelzés készítése.....	17
4.3. Összefüggések vizsgálata .....	23
4.4. Modell pontossága.....	24
5. Napon belüli előrejelzés.....	25
5.1. Alkalmazott módszer .....	25
5.2. Vizsgált paraméterek .....	25
5.3. Előrejelzés készítése.....	27
5.4. Összefüggések vizsgálata .....	27
5.4.1. Napon belüli, időpont szerinti eloszlás .....	29
5.4.2. Meteorológiai tényezők hatása.....	31
5.4.3. Munkarendtől való függés .....	35
5.5. Modell értékelése .....	37
6. Összefoglalás.....	43
6.1. Eredmények .....	43
6.2. Alkalmazási lehetőségek.....	44
7. Irodalomjegyzék.....	46

## 1. Bevezetés

A TDK munka célja olyan eljárás keresése és tesztelése, amely alkalmas az ivóvíz fogyasztás rövidtávú előrejelzésére félórás időközönként. Rövidtáv alatt egy napos, illetve egy hetes intervallumot kell érteni, a fogyasztás változásának periodicitásától függően.

Az előrejelzési modellek kalibrálásának előfeltétele, hogy a bemeneti adatokat minél kevesebb hiba terhelje, illetve rendelkezésre álljon egy tanításra alkalmas félórás fogyasztási idősor. Az archivált nyers üzemeltetési adatok éppen ezért feldolgozásra szorulnak. A hibás archivált adatokat ki kell szűrni, a szivattyúk és a víztorony üzemeltetési adataiból pedig elő kell állítani a fogyasztási idősort. A napi részletességű meteorológiai adatokat csak ezután lehet összefésülni a félórás részletességű fogyasztási idősorral. A fenti feladatrészeket egy saját készítésű adatfeldolgozó program végzi.

Az adatok feldolgozása után két eltérő megközelítést vizsgálom a fogyasztás előrejelzésének. Az egyik a napi szintű előrejelzés, melyből általános fogyasztási menetgörbékkel lehet félórás idősorokat előállítani. A másik eljárás pont az ellenkező irányból közelíti meg a problémát: ebben az esetben a napon belüli félórás értékek előrejelzése után, azok napi szintű összegzésével lehet megkapni a napi becsült fogyasztási értékeket. Mindkét esetben elsődleges szempont a meteorológiai adatoktól való függés vizsgálata. A történeti értékek vizsgálatát a tanítási adatsor amúgy is alacsony elemszámának további csökkenése miatt elvettem.

Az eltérő gondolatmenetekhez eltérő statisztikai módszereket alkalmaztam. Az első esetben regresszióanalízissel kerestem az összefüggést a bemeneti adatok és a napi fogyasztás mértéke között. A napon belüli értékek klasszikus statisztikai módszerekkel történő feldolgozására pedig a jóval modernebb fuzzy következtető rendszereket és mesterséges neurális hálózatokat egyesítő eljárást alkalmaztam. Utóbbi eljárás bár jóval kevesebb manuális adatfeldolgozást igényel, a modell kalibrálása és kiértékelése időigényes folyamat. A modellek előnyeit, gyenge pontjait, illetve a kalibrálás nehézségeit részletesen tárgyalom a dolgozat során. Az eltérő modellek összehasonlítására a négyzetes középhibát (RMSE) használom fel, illetve ennek az átlagos fogyasztáshoz viszonyított mértékét.

Az eredmények értékelése után a kész előrejelzési modell lehetséges alkalmazási területeit vázolom fel. Kiemelendő az ivóvízhálózatok szivattyú menetrendjének kalibrálása a várható fogyasztás függvényében. Ennek a lényege a magastározók térfogatát kihasználva az energiaköltség minimalizálása.

## 2. Felhasznált adatok

A Debreceni Vízmű Zrt. által szolgáltatott adatok adják a kutatás alapját. Fontos megkülönböztetni a modern üzemirányító rendszerből nyert 5 perces időközönként rögzített adatokat a naponta archiváltaktól. Ez utóbbi 1980-tól, míg az 5 perces adatsorok 2007-től állnak rendelkezésre nagyobb megszakításokkal. Továbbá ismert Debrecen ivóvíz hálózatának kialakítása, mely a fekvéséből és méretéből adódóan csupán egy nyomászónából áll.

### 2.1. Rendelkezésre álló adatsorok

Az üzemirányítási rendszerből nyert adatsorok közül a vízmű telepekhez tartozó szivattyúk vízhozamaira, továbbá a város víztornyához tartozó tárolt térfogat értékekre, vagy előjeles vízforgalomra van szükség a vízfogyasztás számításához. A vízhozam értékek minden esetben l/s mértékegységgel lettek archiválva 5 perces időközönként. A 2.1.1-es táblázatban foglaltam össze a számítás szempontjából szükséges adatsorokat.

Adatsor neve	Archivált adatkód	Mértékegység
I. telep - 1. hálózati sziv. mennyiség pillanat érték	10	l/s
I. telep - 1. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	11	m <sup>3</sup>
I. telep - 2. hálózati sziv. mennyiség pillanat érték	17	l/s
I. telep - 2. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	18	m <sup>3</sup>
I. telep - 3. hálózati sziv. mennyiség pillanat érték	24	l/s
I. telep - 3. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	25	m <sup>3</sup>
I. telep - 5. hálózati sziv. mennyiség pillanat érték	32	l/s
I. telep - 5. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	33	m <sup>3</sup>
I. telep - 6. hálózati sziv. mennyiség pillanat érték	39	l/s
I. telep - 6. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	40	m <sup>3</sup>
II. telep - 2. hálózati sziv. mennyiség pillanat érték	55	l/s
II. telep - 2. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	56	m <sup>3</sup>
II. telep - 3. hálózati sziv. mennyiség pillanat érték	62	l/s
II. telep - 3. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	63	m <sup>3</sup>
II. telep - 4. hálózati sziv. mennyiség pillanat érték	70	l/s
II. telep - 4. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	71	m <sup>3</sup>
II. telep - 6. hálózati sziv. mennyiség pillanat érték	77	l/s
II. telep - 6. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	78	m <sup>3</sup>
II. telep - 7. hálózati sziv. mennyiség pillanat érték	85	l/s
II. telep - 7. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	86	m <sup>3</sup>
III. telep - 1. hálózati sziv. mennyiség pillanat érték	100	l/s
III. telep - 1. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	101	m <sup>3</sup>
III. telep - 2. hálózati sziv. mennyiség pillanat érték	107	l/s
III. telep - 2. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	108	m <sup>3</sup>
III. telep - 3. hálózati sziv. mennyiség pillanat érték	114	l/s
III. telep - 3. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	115	m <sup>3</sup>
III. telep - 4. hálózati sziv. mennyiség pillanat érték	121	l/s
III. telep - 4. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	122	m <sup>3</sup>
III. telep - 5. hálózati sziv. mennyiség pillanat érték	128	l/s
III. telep - 5. hálózati sziv. összegzett mennyiség (07:00-kor nullázva)	129	m <sup>3</sup>
Integrált összes termelés	161	m <sup>3</sup>
Víztoronytöltődés (negatív - ürül)	135	l/s
Víztorony tárolt térfogat	138	m <sup>3</sup>

2.1.1-es táblázat - Felhasznált archivált adatsorok

A napi víztermelési adatsor vízmű telepekre bontva, már 1980-tól kezdve rendelkezésünkre áll. Ezeket az értékeket a hálózati veszteség is terheli, tekintettel arra, hogy a betáplálási pontokon mérik a vízhozam értékeket. Gyakorlati megfontolások alapján azonban, a hálózati veszteség változékonyságát közel állandónak feltételezve, a tendencia megegyezőnek tekinthető a fogyasztás tendenciájával. A továbbiakban a termelési adatok alapján számított zóna vízmérleget az egyszerűbb fogalmazás érdekében vízfogyasztásként fogom említeni.

Minden szivattyú vízhozamán túl van egy összegzett vízmennyiségi adatsor is, amely 24 óránkénti integrálja a vízhozamnak. Ez minden nap 07:00-kor kerül nullázásra. Habár a számításokat e nélkül is el lehet végezni, a hibaszűrés, illetve a hihetőség vizsgálat szempontjából hasznos az összegzett vízmennyiségből számított vízfogyasztás is, mivel mind a két adatsort terhelhetik – és terhelik is – archiválási hibák.

A rendszer egyetlen víztornyához is tartoznak adatsorok. A medence töltődéséből/ürüléséből keletkező látszólagos fogyasztás számítása elengedhetetlen a pontos vízmérleg felállításához. Ezt a látszólagos fogyasztást mind a víztorny előjeles vízhozamából, mind a tárolt térfogat értékek különbségéből elő lehet állítani.

A napi vízfogyasztási adatokkal párhuzamosan a meteorológiai adatsorok is rendelkezésre állnak. Ezek közül a középhőmérséklet csak 1981-től, azonban ez a számítást nem befolyásolja, hiszen az 1980-as évek óta bekövetkezett változások ezt az időszakot értékelhetetlenné teszik az ilyen rövidtávú előrejelzés szempontjából. A napi részletességű adatokat a 2.1.2-es táblázatban ábrázoltam.

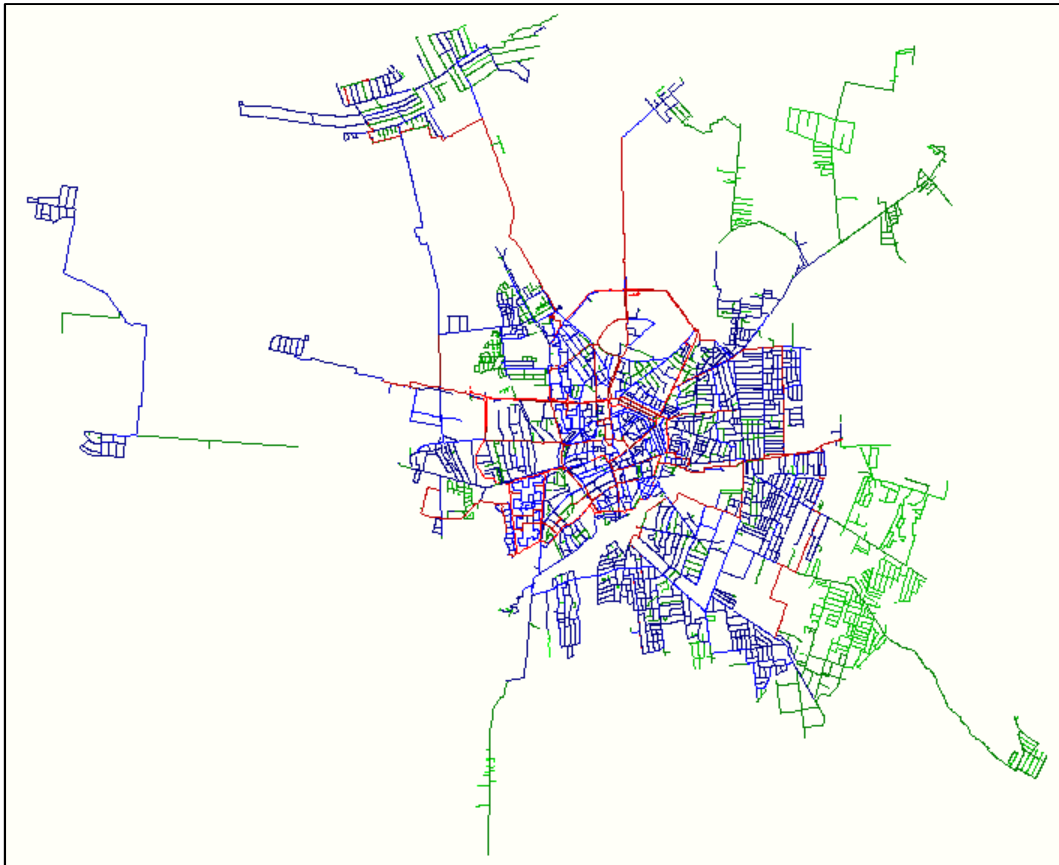
Adat típusa	Adatsor neve	Mértékegység
Meteorológiai	Minimális napi hőmérséklet	°C
	Napi középhőmérséklet	°C
	Maximális napi hőmérséklet	°C
	Napi csapadékmagasság	mm
Üzemeltetési	Tárolt térfogat (szívómedencék)	m <sup>3</sup>
	Napi víztermelés	m <sup>3</sup> /nap

2.1.2-es táblázat - Napi vízfogyasztási és meteorológiai adatok

Továbbá elérhetőek még a Debrecen Vízmű Zrt. adatai a vízárak változásáról, bekötések számáról, illetve az ellátott lakosság számáról és arányáról. Az adatok alapján megállapítható, hogy 2003-tól lezárultak a nagyobb fejlesztések és az ivóvíz ellátottság 99,89% lett. Ettől fogva a hálózat méretéből fakadó változások elhanyagolhatóak, illetve a lakosság is alig 1%-kal nőtt, ami érdemi változást nem jelent egy több mint 200 ezres városnál.

## 2.2. Ivóvíz ellátó rendszer leírása

A Debrecen Vízellátó Rendszert három vízműtelep táplálja. A telepek I-es, II-es, és IV-es telep néven szerepelnek az adatbázisban. A vízmű telepekhez 5-7 szivattyú tartozik, melyek közül az I/6-os, II/3-as és II/6-os szivattyú fordulatszám-szabályzós elven működik. Debrecen városban egyetlen víztorny található, mely 3000 m<sup>3</sup> térfogatú, azonban a szerkezetileg 2 kamrára osztott tárolótérből, szigetelési hibák miatt csupán az egyik kamra használható. A ténylegesen rendelkezésre álló térfogat 1300 m<sup>3</sup>. Ez a rendszer egyetlen zónájához tartozik, értelemszerűen a fent említett szivattyúk betáplálásával együtt.



2.2.1-es ábra: Debrecen város ivóvíz hálózata

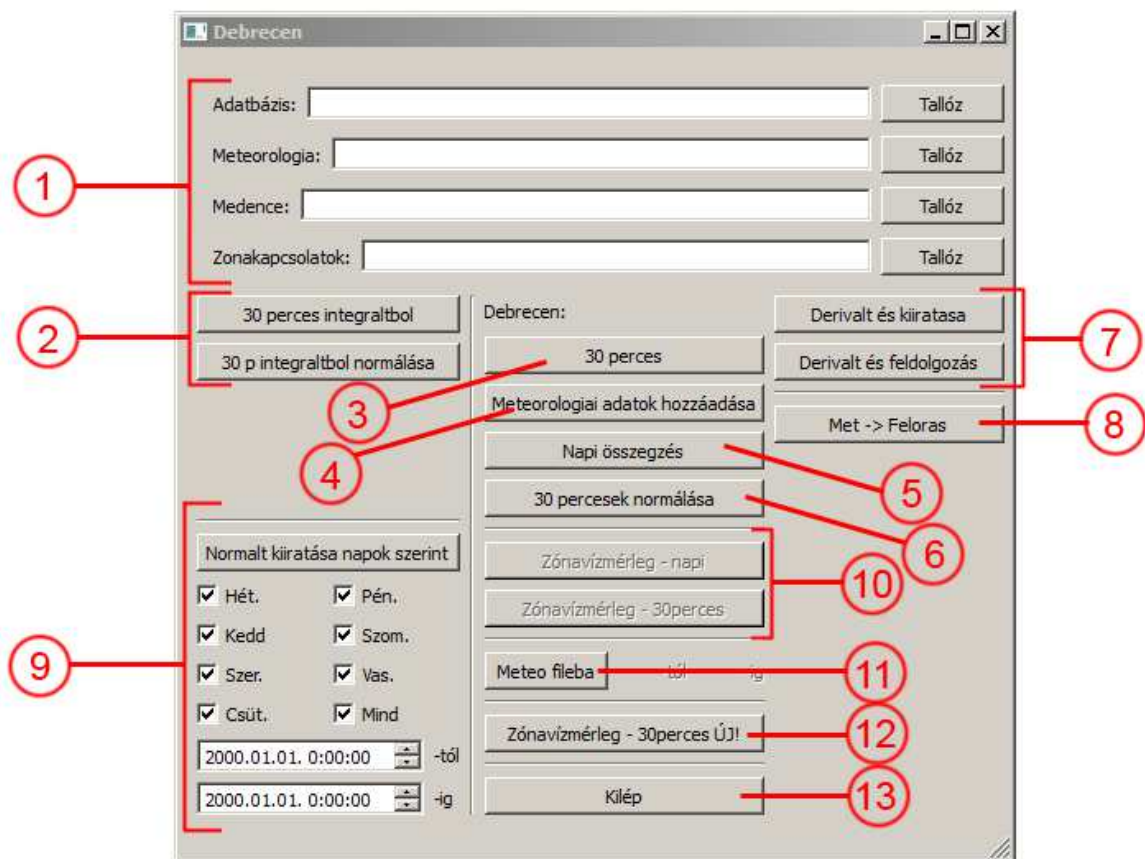
### 3. Adatok feldolgozása

A vízmérleg felállításánál a félórás fogyasztási idősor előállítás a cél. Mivel a mintavételi időpontok nem 5 percre kerekednek, minden kerek félóra interpolálni kell az archivált értékeket a számítás elvégzéséhez. Az adatsorok folytonossági hibái miatt ezt még a tényleges számítás előtt végre kell hajtani.

Ezután következhet a félórás vízfogyasztás számítása, mely a szivattyúk integrált vízhozamából és a víztorony térfogatváltozásából tevődik össze. Itt fontos felhívni a figyelmet arra, hogy az archivált adatok Debrecen esetében l/s mértékegységben vannak, míg a félórás idősorokat következetesen m<sup>3</sup>/félóra mértékegységgel számítjuk.

A napi vízfogyasztási értékeket az aktuális napra eső 48 db félórás adat összegzéséből kapjuk. Ez mind a napon belüli fogyasztási adatok normálásához, mind az archivált napi fogyasztási adatokkal történő összevetéshez szükséges. Előbbi a napi fogyasztási adatok alapján történő előrejelzés szempontjából lényeges, utóbbi pedig a hibaszűrés miatt.

A nyers adatok feldolgozását az általam erre a célra készített program végzi. A program Qt fejlesztői környezetben készült, aminek egyik nagy előnye, hogy ingyenesen hozzáférhető. Az alkalmazás megírása a TDK dolgozat szempontjából nem túl lényeges, bár a programot alkotó 4242 sornyi kód mennyisége szemlélteti, hogy a munka egy jelentős részét a program megírása tette ki. Ez a kódtömeg tartalmazza azokat a funkciókat is, melyek a munka során alapjaiktól lettek újraírva – és így használaton kívül kerültek – a későbbi felhasználás szempontjait szem előtt tartva. Az ilyen, inaktív függvények kb. 1200 sort tesznek ki a 4242-ből.



3.1-es ábra – Az adatfeldolgozó program kezelőfelülete

- 1. Adatbázisok elérési útvonalai:** A különböző feladatokhoz szükséges adatbázisok megadásához szükséges felület. Az **adatbázis**nál az üzemeltetési adatbázist kell megadni, a **meteorológia** mezőnél értelemszerűen a meteorológiai adatokat tartalmazó adatbázist, a **medencénél** a tározók „vízszint - tárolt térfogat” adatpárjait tartalmazót, míg a **zónakapcsolatoknál** a fogyasztási zónák kapcsolatait leíró adatbázist lehet megadni.
- 2. Számítások az integrált értékekből:** az integrált termelési adatsorból számítja ki a félórás fogyasztási idősort, numerikusan deriválva azt, majd a víztorony tárolt térfogatának változásával korrigálja. A normálás esetén a napi fogyasztási menetgörbét állítja elő % mértékegységben az integrált termelésből számított idősből. Utóbbi csak a hiánytalan napon belüli adatsorok esetében lehetséges
- 3. 30 perces:** félórás idősorok generálása a szivattyúk és a víztorony archivált adatai alapján. Tartalmazza az adatok előzetes feldolgozását is.
- 4. Meteorológiai adatok hozzáadása:** a napi részletességű meteorológiai adatokkal összefésüli a félórás számított adatsort a Debreceni rendszerben.
- 5. Napi összegzés:** a félórás adatsorok alapján, napi szinten történő vízfogyasztás összegzése  $m^3$  mértékegységben. Ha elérhető az integráltból számított 30 perces idősor, akkor azt is összegzi.
- 6. 30 percesek normálása:** a szivattyúk vízhozamából számított félórás fogyasztási adatokat elosztja az aznapi összes fogyasztással, így a napi fogyasztási menetgörbét adja, mértékegység nélkül [%].
- 7. Deriváltak számítása:** a félórás adatsorok első- és másodrendű numerikus deriváltjainak számítása. A félórás adatok feldolgozásában ennek az alkalmazását végül elvettem, de a funkció működik.
- 8. Met → Félórás:** a már korábban kiszámolt félórás adatsort összefésüli a meteorológiai adatokkal. Funkciójában hasonló a '4'-es pontban jelölthöz, azonban itt a munkarendtől való függést is automatikusan a program végzi a megadott munkaszüneti, pihenő- és ledolgozott napok alapján.
- 9. Normált értékek kiírása fileba:** a '6'-os, vagy '2'-es pontban előállított értékek kiírása egy .csv fileba.
- 10. Zónavíz mérleg (rég):** a célja azonos a '12'-es pontjával, azonban csak a debreceni rendszerre megírva. Bár gyors, de a rendszerkialakítás szempontjából merev volt, így elvettem az alkalmazását a rugalmasabb felhasználás miatt.
- 11. Meteo fileba:** A napi fogyasztási adatok feldolgozásához szükséges határhőmérsékleti adatok alapján nyári, illetve téli időszakokra bontott adatsorok kiírása .csv fájlalba.
- 12. „Zónavíz mérleg – 30perces ÚJ!”:** zónavíz mérleg számítást végez a zónakapcsolati adatbázisban megadott kapcsolatok alapján. Ezzel teljes mértékben helyettesíthető a '3'-as ponttal jelölt funkció a megfelelő zónakapcsolati adatbázis elkészítése után.
- 13. Kilépés:** a program bezárása.



### 3.1. Hibás adatok szűrése

A SCADA rendszer által gyűjtött adatokban óhatatlanul jelentkeznek mérési hibák. Ezek következhetnek a mérőműszerek, a távközlési eszközök meghibásodásából, vagy az archiváló program hibás működéséből. A debreceni adatok esetében főleg az irreális értékek és az adat duplikációk azok, melyek leggyakrabban jelentkeztek. A hibás adatok értelemszerűen használhatatlanok, így törölni kell őket, mivel a modell eredményeit torzítanák.

Érdemes még ide sorolni az adathiányokat is, annak ellenére, hogy nem hibás adatokról, hanem a mérések, avagy azok archiválásának a hiányáról van szó. A félórás fogyasztási idősorok előállításánál fontos szempont, hogy hány mintavétel marad ki. A mintavételek kis száma miatt igen nagyvonalú és helytelen megközelítés lenne a hibás méréseket tartalmazó adatsorok figyelmen kívül hagyása, hiszen indokolatlan néhány hibás érték miatt egy tízezres nagyságrendű idősort kidobni.

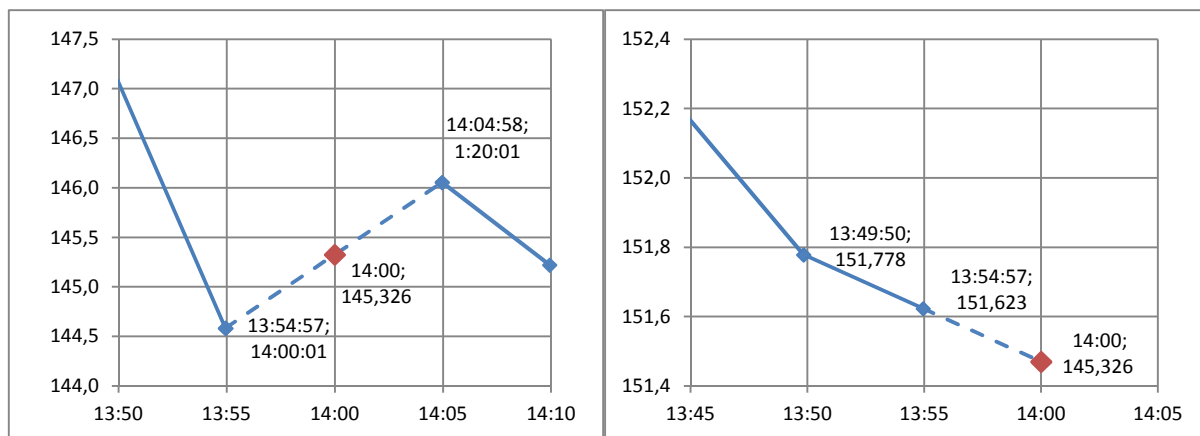
Az ilyen jellegű hibák egy speciális esete, az "Integrált összes termelés" – 161-es adatkód – minden nap 7:00-kor történő nullázásánál fellépő hibás értékek. Ha a mintavételi helyeken nem azonos időpontban történik a mintavétel, előfordulhat és elő is fordult olyan eset, hogy 6:55 és 7:00 között volt egy köztes érték, amely kisebb volt a 6:55-ösnél. Ez az integrál görbe numerikus deriválásánál negatív előjelű fogyasztásként jelentkezett, ami ennél a rendszernél teljes mértékben lehetetlen.

Az adatok feldolgozása során a 7:00-s nullázás problémáját kiküszöbölendő, az adatfeldolgozó program minden nap 7:00:00-kor beilleszt az idősorba egy zérus fogyasztási értéket. Ennek köszönhetően, ha nem pontosan 7:00-kor történne az archiválás, nem fog "elveszni" az első archivált fogyasztási érték. Ennek a 7:00-s zérus értéknek a beszúrása az interpolálással azonos folyamatban történik, de azt nem befolyásolja, hiszen a rendezett adatsoron idő szerint előre haladva megyünk, így a 7:00-nál interpolált érték az előző napi integrált fogyasztás utolsó értéke lesz.

### 3.2. Archivált értékek interpolálása

Az idősorok a gyakorlatban csak a legritkább esetben folytonosak. Az adathiányok véletlenszerűen jelentkeznek. Mivel a vízhozam értékeket integrálni kell a félórás időintervallumokon, fontos meghatározni, hogy mely félórás fogyasztási értékek számítanak a feldolgozás szempontjából megbízhatónak.

A jelenlegi számítási módszer szerint azok a félórák tekinthetőek felhasználhatónak, melyek kezdeti és végpontjától fél órán belül található archivált érték. Ebből a szempontból a program rugalmasan kezeli az adatsort, hiszen akár 20 perces adathiányt is megenged. Ezt, ha az eredmények megkívánják, tetszőlegesen lehet szigorítani, de 5 percnél kevesebbet nem érdemes megadni – mivel ez az archiválási időköz is. Amennyiben ennek a feltételnek megfelel az adott, félórára kerek időpont, viszont nincsen ilyen időpontban archivált érték, akkor alkalmazom az inter-, vagy extrapolációt. Jelen esetben közönséges lineáris interpolációval állítjuk elő a félórára kerek időpontokhoz tartozó értékeket.



3.2.1-es ábra – Interpoláció és extrapoláció esete

### 3.3. Félórás fogyasztási adatok számítása

Amennyiben rendelkezésre állnak a szükséges adatok, elkészíthetjük a félórás fogyasztás értékek számítását, illetve összegzését. Mivel két független mérésből is tudunk fogyasztást számolni, lehetőségünk nyílik a számított adatok ellenőrzésére is. Az egyik eljárás során az "Integrált összes termelés" adatsorból, míg a másik esetben az egyes szivattyúk vízhozam értékeiből kapjuk meg a szükséges vízfogyasztási adatot. A félórás fogyasztásoknál mindig a kezdeti időponttal jelöljük a félórás intervallumot.

Az "Integrált összes termelés" adatsor mértékegysége  $m^3$ , és minden nap 7:00:00-kor nullázódik. Fontos megjegyezni, hogy a félórás adatok feldolgozásánál minden nap 0:00:00-kor váltunk napot. A vízfogyasztási értékeket úgy kapjuk meg, hogy kivonjuk a vizsgált félóra végpontjához tartozó értékből a kezdeti időponthoz tartozó értéket. Úgy is mondhatjuk, hogy az idő szerint numerikusan deriváljuk az összegzett termelésfüggvényt. Ehhez elengedhetetlen, hogy interpolálva legyenek az adott időpontokra az archivált adatsor értékei. Ez alól nem képez kivételt a reggel 6:30 és 7:00 közötti intervallum sem, viszont itt két érték is tartozik az adott nap 7:00-s időpontjához. Lesz egy, a 7:00 előtti utolsó két értékből extrapolált érték, és egy szigorúan zérus érték. A számítás során itt is érvényesül az, hogy csak azok a félórás idősorok használhatóak fel, ahol a kezdeti, és a végponthoz is tartozik érték (3.2.1-es ábra).

$$Q_j [m^3 / \text{félóra}] = V_{i+1} [m^3] - V_i [m^3]$$

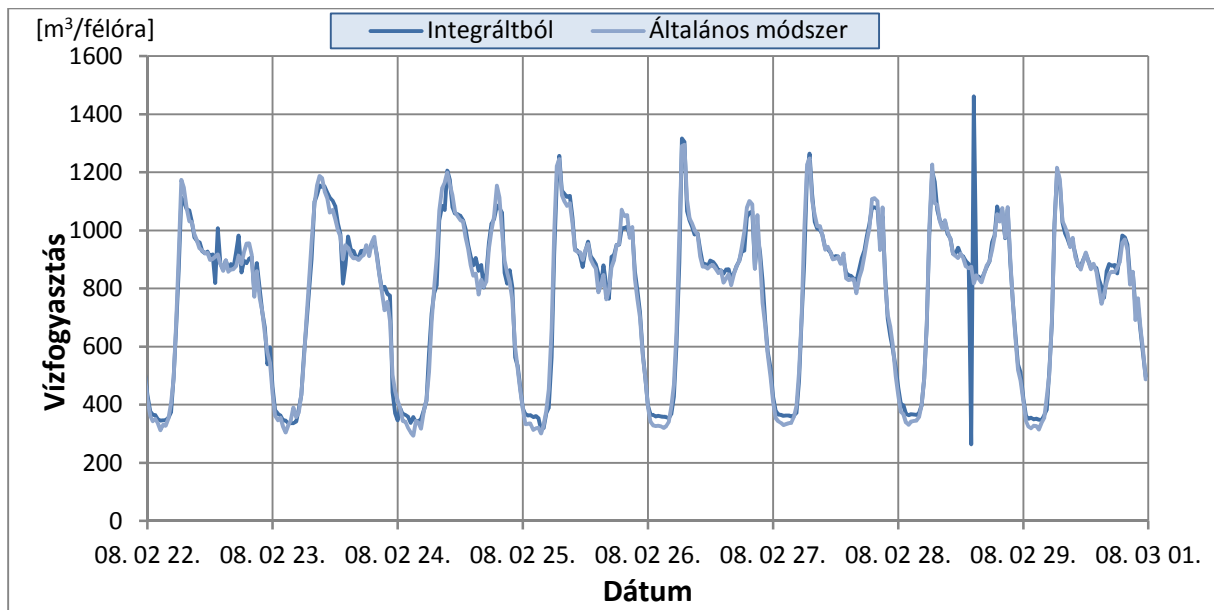
Ahol:

- $Q_j$  - félórás fogyasztási érték
- $V_{i+1}$  - félórás intervallum végponti időpontjára interpolált integrált fogyasztási érték
- $V_i$  - félórás intervallum kezdőponti időpontjára interpolált integrált fogyasztási érték

A másik lehetőség, hogy kiszámoljuk egyenként a szivattyúhoz tartozó vízfogyasztási adatsorokat, majd ezeket összegezzük félóránként. Ebben az esetben is figyelembe kell venni a víztorony látszólagos fogyasztását is, melyet jelen esetben a tárolt víztérfogatok előjeles különbségéből, vagy közvetlenül a víztoronynál mért előjeles vízhozam időintervallumon történő integrálásából kaphatunk.

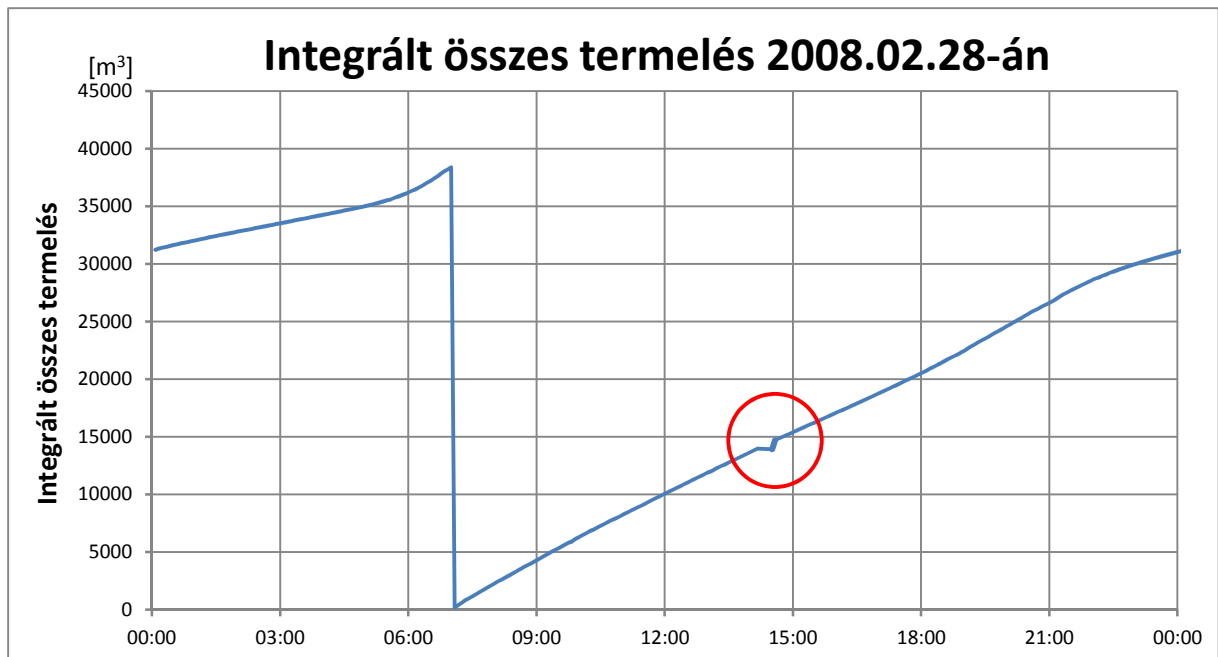
$$Q_{nyomászóna} = \sum_{i=1}^n Q_{szivattyú,i} + \sum_{j=1}^m Q_{viztorony,j}$$

Ez utóbbi eljárás jóval számításigényesebb. Kiterjedtebb, és összetettebb rendszerek esetében azonban sokkal kézenfekvőbb és egyszerűbb megoldásnak számít. Továbbá egy több zónából álló rendszer esetében kicsi annak az esélye, hogy rendelkezésre álljon egy összegzett adatsor. Utóbbi esetre jó példa a MIVIZ Miskolci Vízmű Kft. rendszere, mely 18 zónára bontható, meglehetősen bonyolult kapcsolatokkal. Az adatfeldolgozó program a kapcsolatokat tartalmazó adatbázis betöltése után képes a zónák különálló, illetve összegzett vízmérlegét is előállítani. Lehetőség van továbbá összevont vagy fiktív zónák bevitelére, és vízmérleg számítására is.



3.3.2-es táblázat: Félórás fogyasztási adatsorok, eltérő számítási eljárással

A 3.3.2-es táblázaton jól látható, hogy a két eltérő számítási eljárásból adódó fogyasztási adatsorok gyakorlatilag megegyeznek. Az apróbb eltérések az interpolációból erednek. Míg az általános módszer pillanatnyi vízhozam értékekből számítja a fogyasztást, addig az integrált termelés az átfolyt térfogat alapján számol. Utóbbit nem terheli a mérési pontok közötti vízhozam-ingadozás (lásd:3.3.4-es ábra). Az integrált termelési adatsornál azonban jól láthatóak a február 28-ai hibás archivált értékekből fakadó irreális értékek.

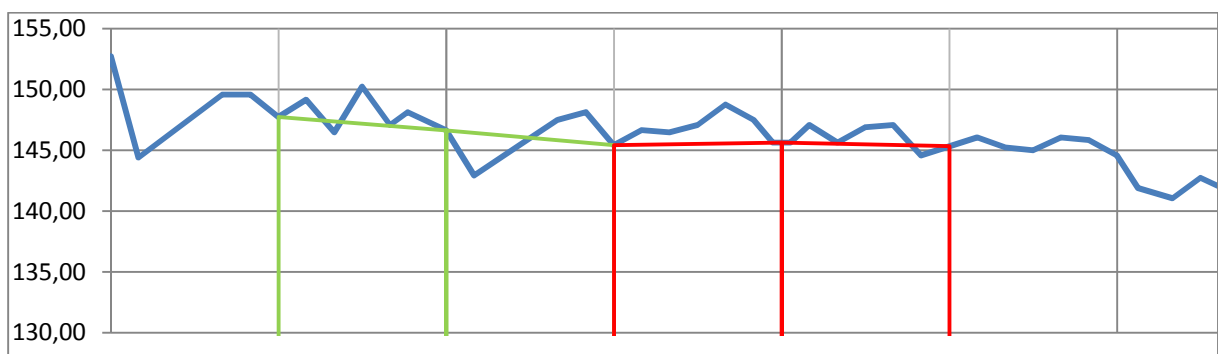


3.3.3-as ábra: Az integrált víztermelés adatsor jellemző hibája

A SCADA rendszerben az „Integrált összes termelés” minden nap 07:00-kor kerül nullázásra. Ennek köszönhető a szembeűnő ugrás a grafikonon ebben az időpontban. Mivel ennek a függvénynek a deriválásával kapjuk a rendszerbe betáplált vízhozamot, a nullázás időpontjától független ez az eredmény. A 3.3.3-as ábrán pirossal bekarikázott szakasznál jól látható, hogy a függvényben két törés található. Ez a fogyasztási görbén (3.3.2-es ábra) két tusként jelenik meg. Ilyen üzemállapot a valóságban nem fordulhat elő, főleg mivel ugyanilyen hibák miatt jelentkeztek 2500 m<sup>3</sup>/h körüli fogyasztási tuskék is, mely az átlagos fogyasztás 4-5-szöröse.

### 3.3.1. Szivattyúk vízszállítás idősorainak kezelése

A szivattyúk vízszállításának kiintegrálásánál trapézokkal közelítettem a vízszállítás idősort. A trapézokkal történő közelítés az adatsorok szempontjából azt jelenti, hogy a mérési pontok között lineárisan változik a vizsgált érték. Ettől a valóság eltérhet, viszont az 5 perces mintavételek, a magas lakosság (~200e fő) és a nagyfogyasztók hiánya miatt a mintavételek közötti ingadozást elhanyagolhatjuk, így kielégítőnek tekinthető a lineáris interpoláció. Mivel a mintavételi időköz 5 perc, ez a legnagyobb részletesség, amellyel a szivattyúk fogyasztása közelíthető. Alább egy olyan fiktív esetet szemléltettem, ahol látható a trapézzal közelítésnél kiegyenlítő eltérések (zöld), illetve a pillanatnyi mintavételekből fakadó hiba (piros) is.



3.3.4-es ábra - Vízfogyasztás közelítése trapézzal

Itt is igaz a fenti állítás, hogy csak azok az intervallumok értékelhetőek, ahol mind a kezdeti, mind a végpontban van archivált érték.

### 3.3.2. A víztorony vízforgalmának számítása

A tározók vízforgalmának számítása kétféle módon történhet. Abban az esetben, ha van távadó felszerelve a tározó vízhozam mérőjére, akkor a tározó relatív vízszintjére és geometriai adataira nincs szükség. Ilyenkor az archivált vízhozam közvetlenül a tározó vízforgalmát adja meg. Ezt átszámítva  $\text{m}^3/\text{félóra}$  mértékegységre, előjelesen összegezhető a szivattyúk vízszállítás értékeivel. Debrecen város hálózatánál ez az eset áll fenn.

A másik eset, például Miskolcon, hogy a relatív vízszintek, illetve a tározók geometriai adatai állnak rendelkezésre. Ahhoz, hogy  $\text{m}^3/\text{félóra}$  mértékegységben megkapjuk a látszólagos fogyasztást, a vizsgált félórás időszak vég- és kezdőpontjában mért tárolt térfogat különbségét kell kiszámítani. A nem hasáb alakú tározók esetében több „tárolt térfogat – relatív vízszint” értékpárt lehet megadni. A megadott pontok közötti vízszint értékek esetén lineáris interpolációt alkalmazunk.

$$V_{(h)} = V_i + \frac{V_{i+1} - V_i}{h_{i+1} - h_i} * (h - h_i)$$

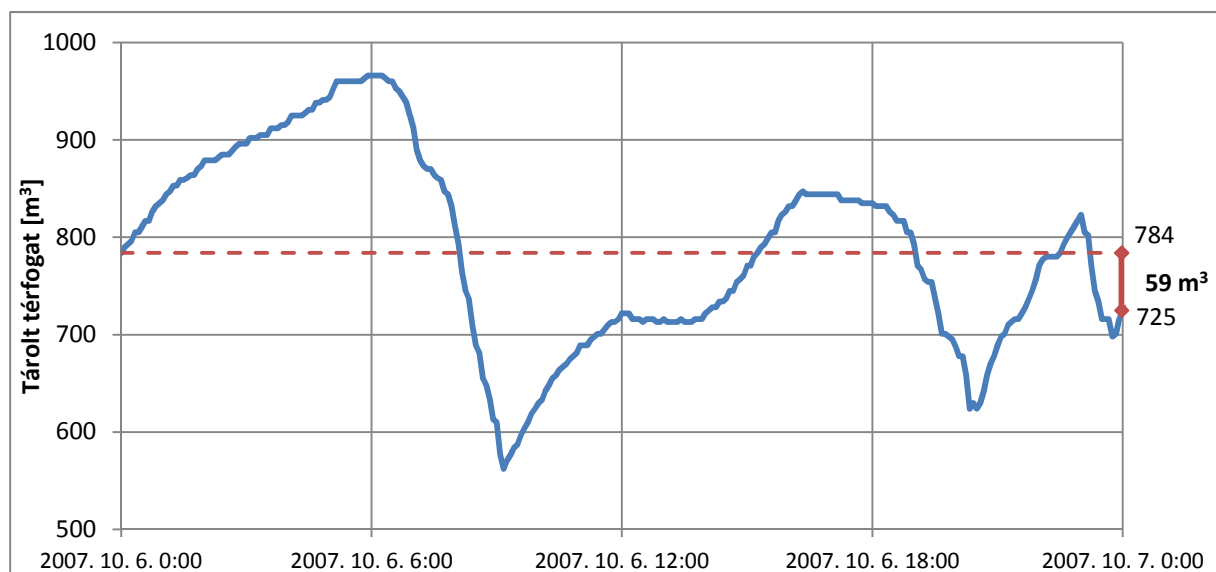
Ahol:

- $h$ : tározó vízszintje
- $V_{(h)}$ :  $h$  vízszinthez tartozó tárolt térfogat
- $h_i, h_{i+1}$ : a tárolt térfogat – vízszint görbe  $h$ -val szomszédos vízszint értékei
- $V_i, V_{i+1}$ : a tárolt térfogat – vízszint görbe  $h$ -val szomszédos tárolt térfogat értékei.

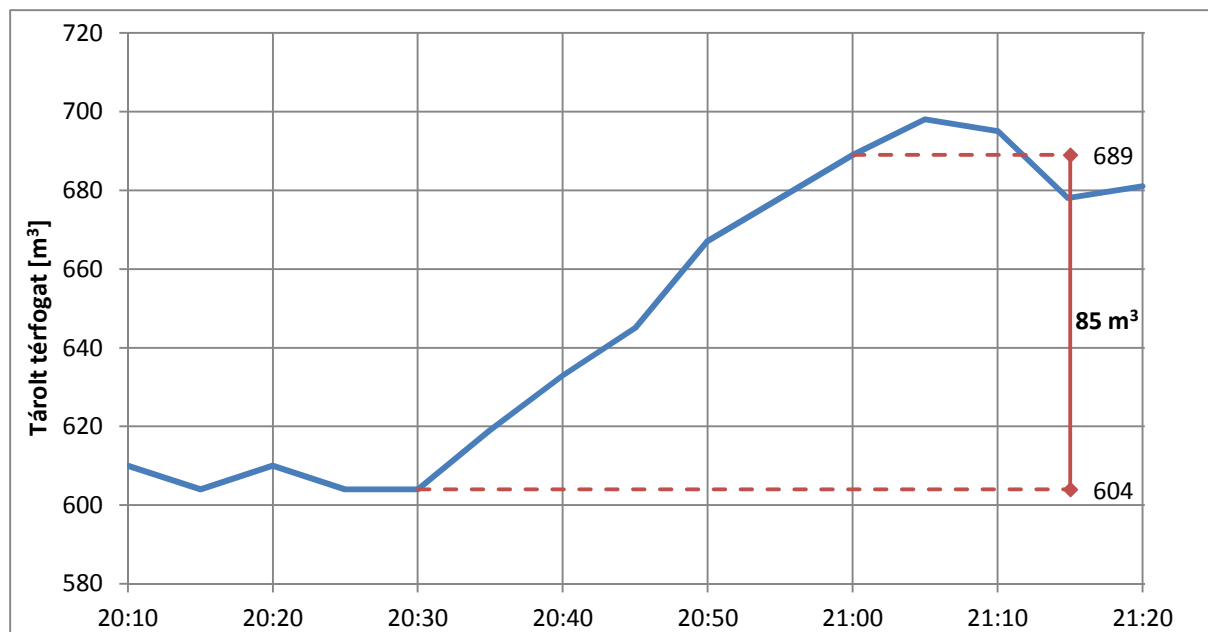
$$Q_{\text{tározó},i} = V_i - V_{i+1}$$

Ahol:

- $V_i$ : kezdeti időpontban a tárolt térfogat
- $V_{i+1}$ : az intervallum végpontjában a tárolt térfogat



3.3.5-ös ábra – Napi vízfogyasztás számítása tárolt térfogattól



3.3.6-ös ábra – Félórás vízfogyasztás számítása tárolt térfogatból

### 3.4. Napi fogyasztási adatok számítása

A napi szinten összegzett vízfogyasztás értékek előállításához meg kell vizsgálni, hogy rendelkezésre áll-e elég adat az összegzés elvégzéséhez. Ha ezt nem tennénk meg, az olyan napokon, ahol nem áll rendelkezésre mind a 48 félórás érték, a valószínűleg kisebb értékek adódnának, melyek az előrejelzés készítése során torzítanák a modellt.

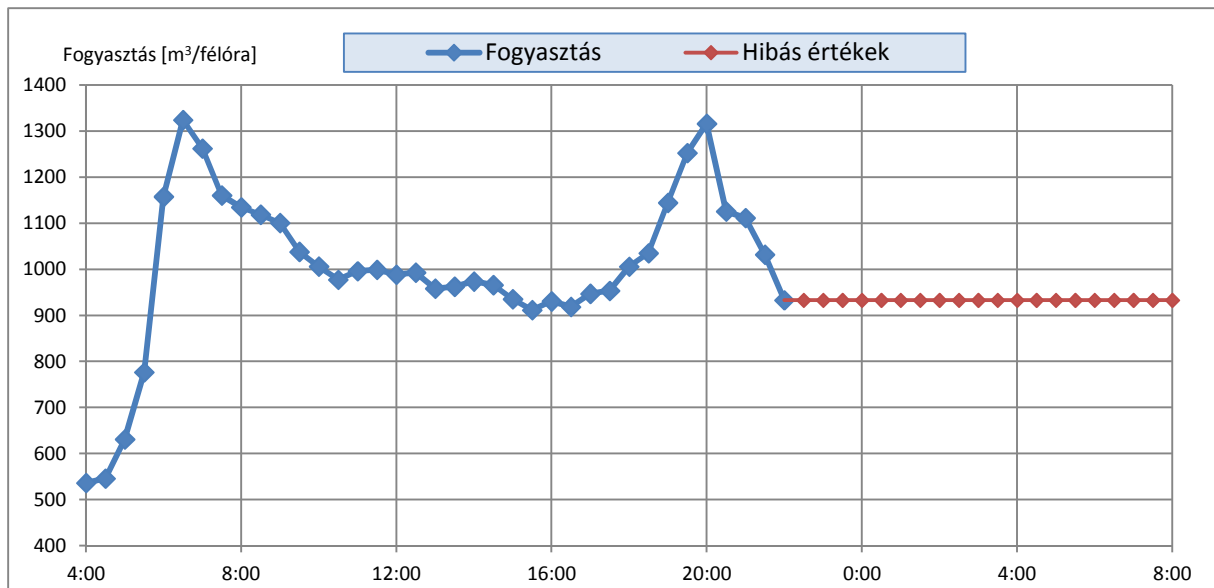
Az alapul véve, hogy a napi fogyasztási adatokat csak közvetve, a félórás idősorokból állítjuk elő, megengedünk apróbb adathiányokat, melyek érdemben nem befolyásolják a napi, sőt még a félórás értékeket sem (lásd: 3.2-es fejezet). Ha az adott napon mind a 48 félórásra kerek időpontban van archivált adat, vagy még lehet oda interpolálni adatot, lehetséges az aznapi fogyasztás összegzése is.

$$Q_{a,i} = \sum_{j=1}^{48} Q_{30min,j}$$

### 3.5. Számított adatsorok ellenőrzése

A feldolgozás során végzett hihetőség vizsgálaton kívül a már kiszámolt fogyasztási idősort is érdemes ellenőrizni. Főleg az archivált idősorban jelentkező szakadások, adathiányok széleinél jelentkezhetnek hamis, vagy irreális értékek.

Mivel a vízfogyasztás – mérési ponttól függően – tized vagy század pontosságú, elhanyagolható az esélye, hogy két egymás utáni félórában századra azonos fogyasztási érték jelentkezzen. Ha ilyen ismétlődés adódik, akkor az időben első ilyen adatot – amennyiben a többi feltétlenül megfelel – leszámítva törli az idősorból a program. Az ilyen ismétlések, még ha hihető tartományban is vannak, nagymértékben torzíthatják a későbbi előrejelzési modellt mind napi, mind napon belüli szinten.



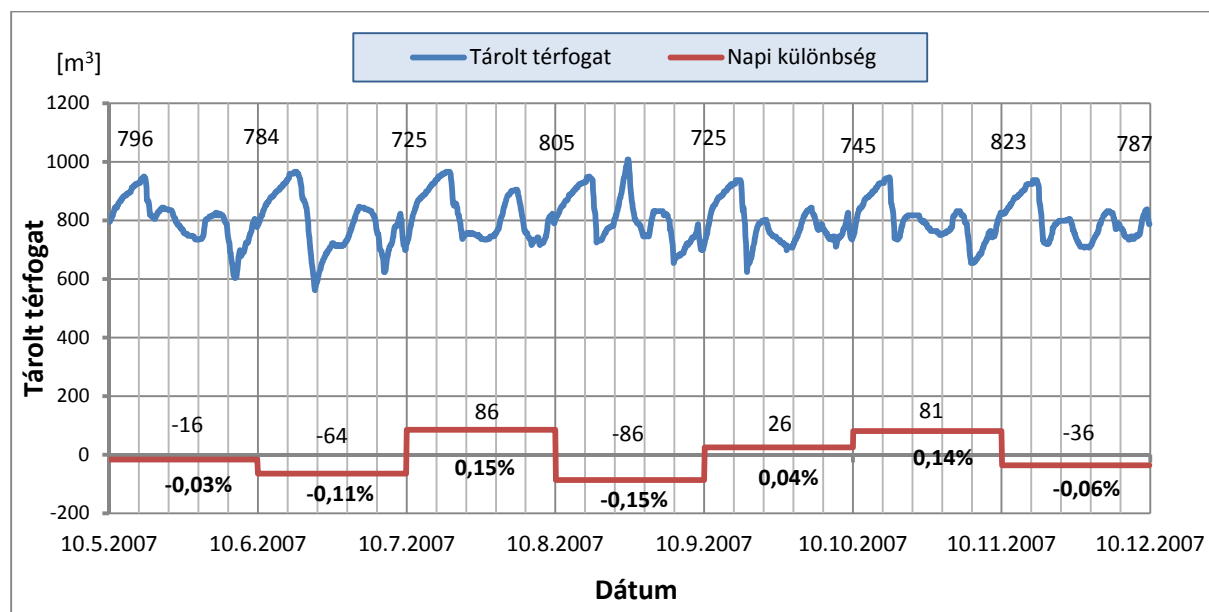
3.5.1-es ábra: Hibás, ismétlődő fogyasztási értékek szűrése

A 3.5.1-es ábra ilyen hibás, ismétlődő értékekre mutat példát. A pirossal jelölt félórás értékek törlésre kerülnek, hogy ne befolyásolják az előrejelzési modelleket. Az ilyen jellegű hibákat, mivel reális értéktartományban van, érdemes a fogyasztás számítása után szűrni a szélsőséges, vagy zérus értékekkel ellentétben, melyeket a fogyasztás-számítás előtt szűr az adatfeldolgozó program.

## 4. Napi fogyasztási adatok előrejelzése

Egészen 1980-ig visszamenőleg állnak rendelkezésre Debrecen város lakossági vízfogyasztásának napi részletességű adatai. Ezek az adatok még manuális leolvasással lettek rögzítve. Mind a három vízműtelepre szétbontva vannak termelési adatok, melyeket összegezve megkapjuk a napi lakossági vízfogyasztást.

A víztoronyban tárolt víz térfogatával ez az érték nincs korrigálva, azonban a víztoronyban maximálisan tárolható  $3000\text{m}^3$  (üzemeltetési okok miatt jelenleg csak  $1300\text{m}^3$ ) a napi fogyasztási értékekhez viszonyítva csupán 1,2-2,0%-ot jelent. Ez az 1 % körüli eltérés is csak olyan szélsőséges esetben jelentkezhet, ha a vizsgált nap kezdeti időpontjában a víztorony üres volt, míg az utolsó időpontban a túlfolyó szintet érte el. Egy üzemszerűen működő rendszer esetében – a 24 órás üzemeltetési periódusok miatt – a tárolt térfogatok napi szinten mért eltérése ennél jóval alacsonyabb. Az így keletkező termelés és fogyasztás közötti eltérés nem „elveszik” csupán a vizsgált nap előtti vagy utáni naphoz fog adódni a tárolt térfogatok különbségének előjelétől függően. A fenti probléma átlagosan  $56,42\text{ m}^3/\text{nap}$  eltérést okoz. Ez az eltérés 0,96%-e a sokéves napi átlagfogyasztásnak, mely elenyésző (4.1.1-es ábra). Ráadásul ezt az értéket csak azokon a napokon tudjuk meghatározni, amikor félórás archivált adatok állnak rendelkezésre. Az 1980-tól regisztrált adatok között csak a szívómedencékben tárolt víztérfogat van feltüntetve, ami a számításaimtól független, azokat nem befolyásolja.



4.1.1-es ábra: Víztorony által tárolt víztérfogat ingadozása napi szinten

Rendelkezésünkre áll továbbá a napi minimális, közép, illetve maximális hőmérsékleten kívül az aznapi csapadékmennyiség milliméter mértékegységben. A napi fogyasztási értékek hitelessége a folyamatirányító rendszer által szolgáltatott adatokból előállított napi fogyasztási értékekkel ellenőrizhető. Reményeink szerint, ha a napi termelési értékeket kellő pontossággal sikerül előre jelezni, ebből a százalékos napon belüli fogyasztási adatokkal akár napon belüli előrejelzés is készíthető.



#### 4.1. Vizsgált paraméterek

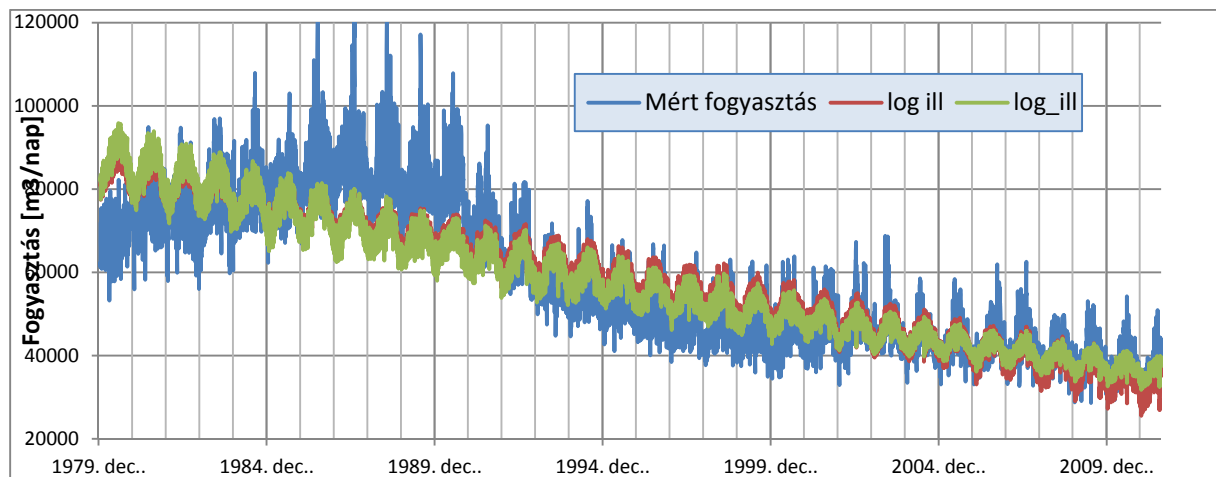
A vizsgált paramétereknél a gyors számítási eljárás miatt érdemes az összes rendelkezésre álló adatot megvizsgálni, ellenben a napon belüli (5. fejezet), ahol a bemeneti adatsorok számát a lehető legkisebbre kellett csökkenteni a modell torzítása nélkül.

- **Datum:** Az adott nap Excel-féle Timestamptel megadott értéke.
- **HetNapja:** [0;1] értéket felvevő úgynevezett *bulk* változó. 1 értéket a munkanapokon rendeltünk hozzá. 0 értéket pedig a szombat, vagy munkaszüneti napok során. Itt nem célravezető a napok számokkal történő megjelölése (a napon belüli ellentétben 5.2. fejezet) a klasszikus statisztikai eljárások korlátai miatt.
- **Tmax:** az adott napra eső maximális hőmérséklet °C mértékegységben megadva.
- **Tkozep:** az adott napra eső középhőmérséklet °C mértékegységben megadva.
- **Csapadék:** az adott nap leesett csapadék milliméter mértékegységben megadva.

A vízfogyasztási értékek vizsgálatánál a víztorony tárolt térfogatai nem állnak rendelkezésre (4. fejezet). Éppen ezért vízfogyasztás helyett a pontos megnevezés a víztermelés lenne. A számokban mérhető különbség kis mértéke miatt ezt elhanyagoljuk és a továbbiakban fogyasztásként hivatkozunk rá.

#### 4.2. Összefüggések vizsgálata

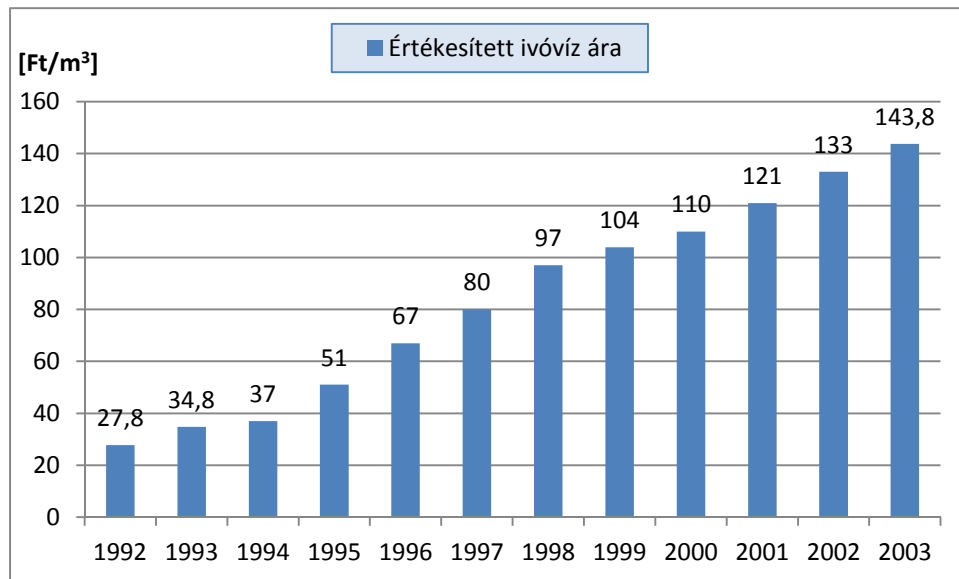
Első lépésben a teljes rendelkezésre álló időszakra vizsgáltuk a lineáris, illetve logaritmikus illeszkedést. Nem tettünk különbséget nyári és téli időszakok között. A közép-, és maximális napi hőmérsékletet külön vizsgáltuk, hogy lássuk, melyiktől függ nagyobb mértékben a vízfogyasztás.



4.2.1-es ábra – 1980-tól számított lineáris és logaritmikus regresszió

Ahogy a 4.2.1-es ábrán is látszik, mindkét eljárás alábecsüli a fogyasztási értékeket, ha az utolsó 5-7 éves intervallumot nézzük. Ennek legfőbb oka a 80-as évektől a 90-es évek elejéig tartó – ma már igen magasnak tekinthető – fogyasztási csúcs a görbén.

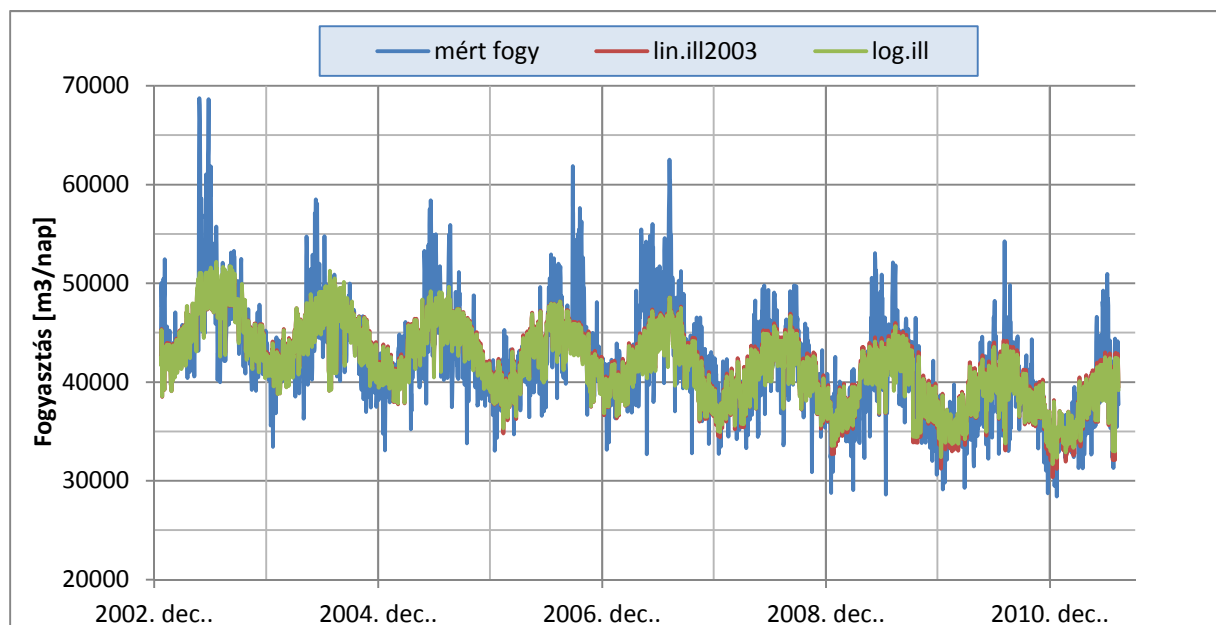
Abban az időben a fogyasztói szokások jócskán eltértek a jelenlegitől. A víz és az egyéb közművek ára jelentősen alacsonyabb volt, mint ma. A rendszerváltás utáni meredek csökkenés a vezetékes ivóvíz árának növekedésével magyarázható. Ez a csökkenés azonban nem lineáris, hiszen tartani fog a minimálisan szükséges vízfogyasztáshoz, melyet még akkor is kénytelen a fogyasztó megvenni, ha az nagyobb anyagi teherrel jár, mivel a mindennapi élethez elengedhetetlen.



4.2.3-as ábra Debreceni vízművek árváltozás grafikonja (forrás: Debrecen vízművek)

A fogyasztói szokásokban való eltérést jól mutatja, hogy a rendszerváltás környékén a nyári időszakokban jóval nagyobb a vízfelhasználás mértéke a téli időszakhoz viszonyítva. Ebből arra következtethetünk, hogy a vízfelhasználás hőmérséklettől való függése is más mértékű, mint a kétezres években.

Annak ellenére, hogy rengeteg adat áll rendelkezésünkre, a továbbiakban a napi szinten történő előrejelzésnél a 80-as évek végén jelentkező fogyasztási csúcsot kihagyjuk az előrejelzésből. 1995-től a fogyasztási trend nem változik, gyakorlatilag lineáris a vízfogyasztás csökkenése. Azonban ennek ellenére érdemes 2003-tól vizsgálni az adatsort, hiszen akkor fejeződtek be az utolsó nagyobb hálózatbővítési munkálatok és a lakosság számának érdemi növekedése (2.1-es fejezet). Ez megkönnyíti a munkánkat, mivel nem szükséges fajlagos fogyasztási értékekkel számolni, ha állandónak vesszük az ivóvízzel ellátott lakosság számát.



4.2.4-es ábra –2003-tól számított lineáris és logaritmus regresszió

		1980-tól		2003-tól	
		RMSE [m <sup>3</sup> /nap]	Hiba[%]	RMSE [m <sup>3</sup> /nap]	Hiba[%]
Tköz	lin. Reg.	8679,08	14,68%	4166,13	9,93%
	Log. Reg.	9421,85	15,94%	5019,48	11,96%
Tmax	lin. Reg.	8882,81	15,03%	3362,55	8,01%
	<b>Log. Reg.</b>	<b>9132,51</b>	<b>15,45%</b>	<b>2364,19</b>	<b>5,63%</b>

4.2.2.-es táblázat: Négyzetes középhiba nagysága, átlagfogyasztáshoz viszonyított mértéke

Az így lerövidített adatsorból számított regressziós görbe sokkal jobban közelíti a valós értékeket. Itt már érdemes a hőmérséklettől, csapadéktól, és munkarendtől való függést is vizsgálni, nem csupán az átlagos hibát, vagy a korrelációt. Itt a T-próba segítségével azt vizsgáltam, hogy szükségesek-e az együttthatók a fogyasztás becsléséhez.

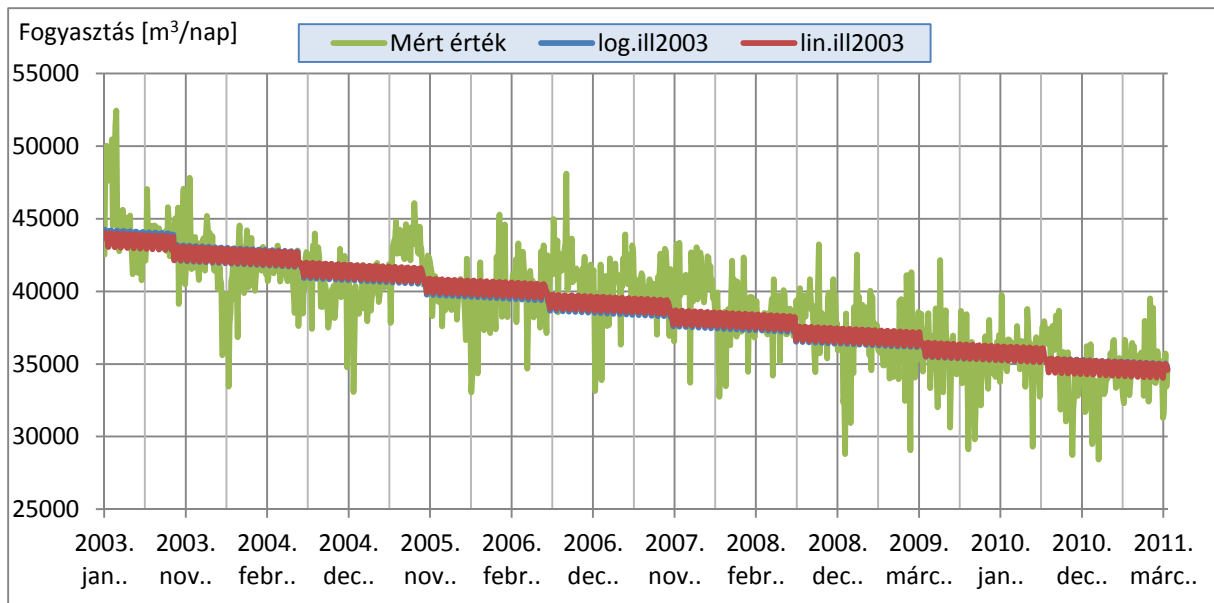
		Csapadék	Hőmérséklet		Munkarend	Dátum	Konstans
			Tmax	Tköz			
1980-tól	lin. Reg.	-6,672532		23,85472	13,07802	-184,505	242,1352
	Log. Reg.	-8,006378		27,97369	14,58455	-210,564	991,5316
	lin. Reg.	-4,043697	35,73522		13,17423	-181,941	238,5211
	Log. Reg.	-5,193941	41,56549		14,74238	-209,924	1024,585
2003-tól	lin. Reg.	-6,992924		17,91979	12,25083	-33,3138	45,51612
	Log. Reg.	-12,1722		32,70977	20,93915	-59,5553	300,8797
	lin. Reg.	-7,961453	46,4955		14,96087	-41,0578	54,81484
	Log. Reg.	-8,217349	47,61019		15,20035	-43,6481	219,7695

4.2.6-os táblázat: A vizsgált paraméterek T-próba értékei

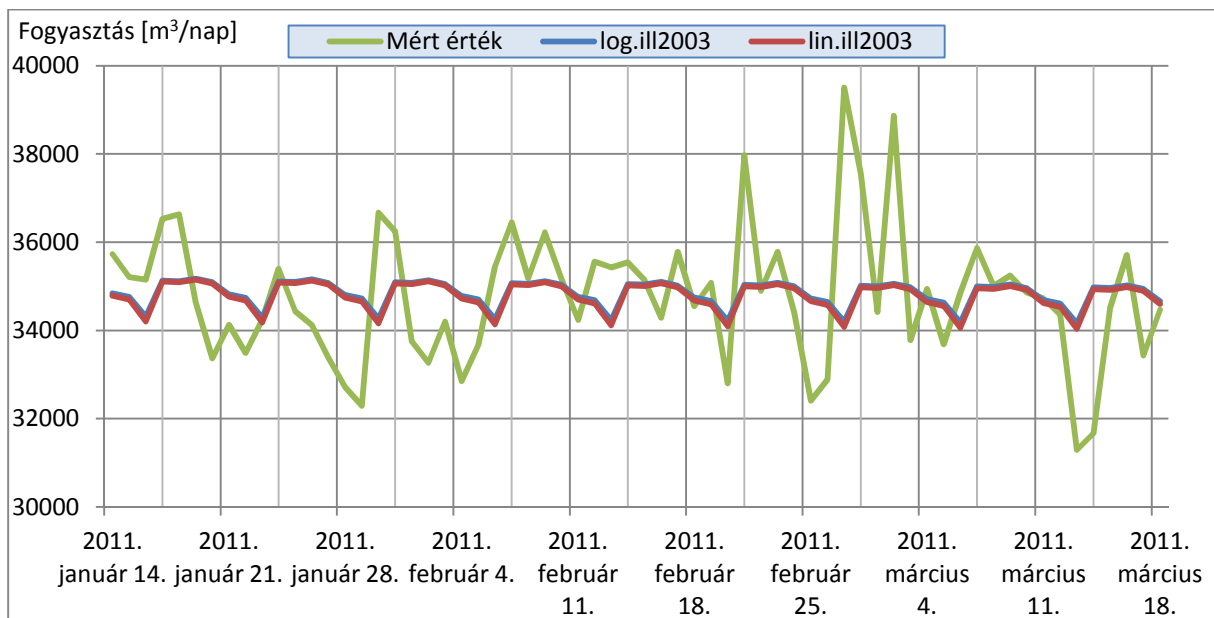
Az így kapott értékeket összehasonlítva látható, hogy a legnagyobb mértékben a dátumtól való függés érvényesül. Ez az eredmény várható volt, hiszen ha megvizsgáljuk a hőmérséklettől való függést, látható, hogy lineáris, vagy logaritmikus illeszkedéssel nem követhető le. Feltételezésem szerint van egy hőmérsékleti határ, amelytől lejjebb a hőmérséklet már nem befolyásolja a vízfogyasztás mértékét. Például a vízfogyasztás szempontjából mindegy, hogy télen +3, vagy -10 °C van.

Annak a határhőmérsékletnek a megtalálásához, amitől már függ a vízfogyasztás, két időszakra bontottam éves szinten az adatsort. A maximális napi hőmérséklettől való függőségénél a 19-22 °C-ig terjedő tartományt vizsgáltam meg. Minden évben, az első olyan nap a kezdete, amelyen a hőmérséklet elérte vagy meghaladta az adott határhőmérsékletet. A nyári időszak végét az utolsó ilyen nap jelöli. Az ide nem tartozó napokat a téli időszakhoz csatoltam. A nyári időszakba tartozó szakaszok hossza a hőmérséklet növekedésével rövidül, hiszen később lesz várható az első hőmérsékleti feltételt teljesítő nap, az utolsó pedig értelemszerűen hamarabb.

A téli időszakok során a hőmérséklettől és a csapadékmennyiségtől való függést kizárhatjuk. Itt egyedül a munkarendtől való függés az, ami befolyásolja az értékeket az éves szinten csökkenő trenden kívül. Utóbbit, a dátumot timestampmel jelölő adatsorból kapjuk. Ez a legkisebb négyzetek módszere miatt a szélsőségekre érzéketlen becslést ad, ami a mi szempontunkból előnytelen. Az üzemeltetés szempontjából a magas fogyasztási értékek előrejelzése az elsődleges.



4.2.7-es ábra: Téli időszak becült és mért fogyasztási értékei



4.2.8-as ábra: Téli időszak becült és mért fogyasztási értékei

A nyári időszakoknál azzal a feltételezéssel élünk, hogy egy adott napon eső csapadék hatással van az utána következő napokban öntözésre fordított vízmennyiségre, illetve a több napig tartó forrás jobban növeli a vízfogyasztást. Ezt a feltevést a modellbe úgy vittem bele, hogy az adatfeldolgozó program minden egyes napra kiszámolta a vizsgált napot megelőző 1. 2. és 3. napi hőmérséklet és csapadékösszeget (4.2.9-es táblázat).

		Csapadék		
		1 nap	2 nap	3 nap
Hőmérséklet	1 nap	sh1-sT1	sh2-sT1	sh3-sT1
	2 nap	sh1-sT2	sh2-sT2	sh3-sT2
	3 nap	sh1-sT3	sh2-sT3	sh3-sT3

4.2.9-es táblázat: a történelmi meteorológiai adatok

Lehetséges szempont a csapadékok vizsgálata során, hogy a vizsgált napot megelőző, adott időintervallumon belüli esős napok száma, vagy a csapadékösszeg befolyásolja jobban a fogyasztást <sup>[1]</sup>. Az esős napok számának vizsgálatát Debrecen talajviszonyai, illetve éghajlati adottságai miatt elvettem. Mivel a város jellemzően homokos talajon fekszik, a nedvességet rövidebb ideig tudja tárolni a talaj. Így egy adott csapadék öntözés és vízfogyasztás-csökkentő hatása rövidebb idejű lesz, mint kötött talaj esetén.

A hőmérséklettől való függés esetében egyértelműen a napi maximális hőmérséklet befolyásolja jobban a fogyasztást (4.2.6-os táblázat). Ez logikus is, hiszen a középhőmérséklet alacsonyabb, ha éjszaka jobban lehül a levegő, azonban a fogyasztást az éjszakai hideg nem befolyásolja, hiszen ilyenkor az emberek többsége alszik. Ezzel szemben a maximális hőmérséklet nappal jelentkezik, amit az ember jobban megérez, így több vizet is fogyaszt. Az eredmények is ezt támasztják alá. A továbbiakban középhőmérsékletet nem, csak a napi maximális hőmérsékletet vizsgáljuk.

Az összegzett, vizsgált napot megelőző értékek kombinációira elvégeztem mind a lineáris, mind a logaritmusos illeszkedés számítását. Vizsgáltam még az összegzett hőmérsékletek logaritmusával elvégzett lineáris illeszkedést is, melynek az eredményei az „sTlog” jelű mezőkben találhatóak. Ezt megismételtem a négy vizsgált határhőmérsékletre is. Ez összesen  $3 \times 4 \times 9 = 108$  eltérő esetet jelent. Fontos megjegyezni, hogy ezek az esetek nem tartalmazzák a munkarendtől való függést, csupán a hőmérséklettől való függés pontosabb megértését hivatottak megmutatni. Az alábbi táblázatokba foglaltam össze az áttekinthetőség érdekében színekkel is kiemelve az eredményeket.

„ $r^2$ : A determinációs együttható. A becült és a tényleges  $y$  értéket hasonlítja össze. Értéke 0 és 1 közötti lehet. Ha értéke 1, akkor a minta elemei között teljes korrelációs kapcsolat van, vagyis a becült és a tényleges érték megegyezik. Ha a determinációs együttható értéke 0, akkor a regressziós egyenlet nem alkalmas az  $y$  értékének előrejelzésére.” <sup>[2]</sup>

$r^2$		Hőmérséklet határ (Maximum): 19 °C			Hőmérséklet határ (Maximum): 20 °C			Hőmérséklet határ (Maximum): 21 °C			Hőmérséklet határ (Maximum): 22 °C		
		sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3
lin	sT1	0,444	0,448	0,451	0,446	0,450	0,454	0,391	0,397	0,402	0,433	0,436	0,440
	sT2	0,446	0,450	0,453	0,449	0,453	0,456	0,398	0,402	0,405	0,441	0,443	0,445
	sT3	0,448	0,452	0,455	0,451	0,455	0,458	0,401	0,406	0,409	0,442	0,447	0,448
sTlog (lin)	sT1	0,441	0,446	0,450	0,443	0,448	0,452	0,405	0,410	0,415	0,456	0,460	0,464
	sT2	0,444	0,448	0,451	0,446	0,450	0,453	0,408	0,412	0,416	0,458	0,462	0,465
	sT3	0,446	0,450	0,453	0,449	0,453	0,456	0,410	0,415	0,418	0,460	0,464	0,467
log	sT1	0,473	0,477	0,480	0,476	0,480	0,483	0,418	0,424	0,428	0,465	0,468	0,471
	sT2	0,475	0,479	0,482	0,479	0,482	0,485	0,424	0,429	0,432	0,473	0,475	0,476
	sT3	0,477	0,482	0,484	0,481	0,485	0,488	0,428	0,433	0,435	0,475	0,479	0,480

4.2.11-es táblázat: az  $r^2$  (determinációs együtthatók) értékei

Tmax T-próba		Homerseklet hatar (Maximum): 19 °C			Homerseklet hatar (Maximum): 20 °C			Homerseklet hatar (Maximum): 21 °C			Homerseklet hatar (Maximum): 22 °C		
		sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3
lin	sT1	10,602	11,029	11,263	11,064	11,457	11,681	11,976	12,202	12,352	8,337	8,546	8,636
	sT2	10,189	10,681	11,210	10,625	11,097	11,615	9,724	10,106	10,572	7,603	7,917	8,249
	sT3	10,370	10,581	11,205	10,773	10,956	11,577	9,596	9,731	10,303	8,058	8,164	8,577
sTlog (lin)	sT1	12,506	12,897	13,098	13,026	13,392	13,589	11,708	12,046	12,226	11,907	12,234	12,407
	sT2	11,465	11,910	12,412	11,891	12,330	12,831	10,912	11,322	11,812	11,251	11,650	12,135
	sT3	11,246	11,404	11,993	11,600	11,741	12,343	10,892	11,002	11,593	11,338	11,433	12,012
log	sT1	11,255	11,735	11,981	11,754	12,199	12,438	12,537	12,806	12,978	8,982	9,216	9,316
	sT2	10,788	11,340	11,910	11,256	11,786	12,347	10,200	10,637	11,153	8,188	8,537	8,893
	sT3	10,890	11,146	11,825	11,317	11,541	12,221	10,009	10,181	10,816	8,613	8,746	9,194

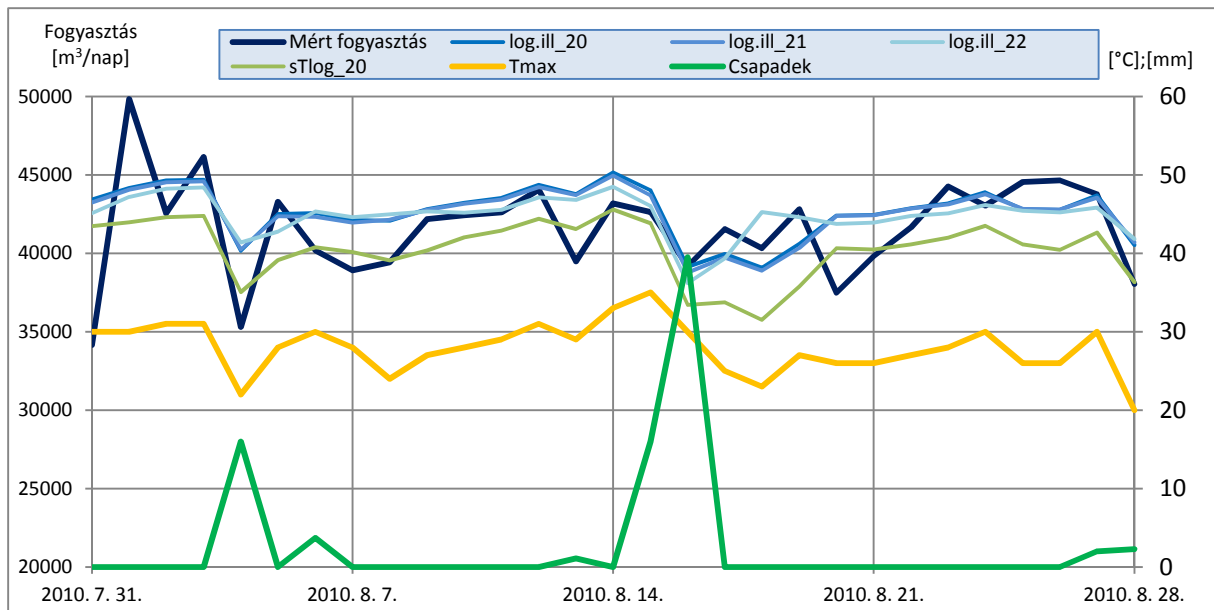
4.2.12-es táblázat: T-próba eredményei a vizsgált napon mért maximális hőmérsékletre

sTmax T-próba		Homerseklet hatar (Maximum): 19 °C			Homerseklet hatar (Maximum): 20 °C			Homerseklet hatar (Maximum): 21 °C			Homerseklet hatar (Maximum): 22 °C		
		sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3
lin	sT1	3,227	2,655	2,365	3,651	3,105	2,804	3,256	2,696	2,396	8,127	7,504	7,232
	sT2	4,247	3,769	3,186	4,718	4,268	3,687	5,183	4,610	3,957	9,389	8,744	8,145
	sT3	4,791	4,622	4,046	5,315	5,174	4,604	5,867	5,585	4,903	9,617	9,246	8,603
sTlog	sT1	1,544	0,999	0,759	2,016	1,481	1,215	2,122	1,464	1,119	2,231	1,605	1,265
	sT2	3,054	2,615	2,075	3,607	3,180	2,624	3,447	2,894	2,229	3,389	2,877	2,229
	sT3	4,001	3,875	3,345	4,621	4,514	3,971	4,193	3,970	3,308	4,025	3,829	3,196
log	sT1	3,117	2,494	2,208	3,571	2,974	2,673	3,206	2,588	2,290	8,374	7,717	7,459
	sT2	4,227	3,692	3,081	4,737	4,232	3,619	5,218	4,576	3,887	9,756	9,067	8,457
	sT3	4,904	4,698	4,079	5,483	5,309	4,692	6,014	5,689	4,953	10,100	9,700	9,033

4.2.13-as táblázat: T-próba eredményei a vizsgált napot megelőzően összegzett Tmax értékekre

sh T-próba		Homerseklet hatar (Maximum): 19 °C			Homerseklet hatar (Maximum): 20 °C			Homerseklet hatar (Maximum): 21 °C			Homerseklet hatar (Maximum): 22 °C		
		sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3	sh1	sh2	sh3
lin	sT1	-3,516	-4,870	-5,970	-3,280	-4,661	-5,837	-3,549	-5,107	-6,254	-3,813	-4,899	-5,885
	sT2	-3,982	-5,175	-6,008	-3,815	-5,020	-5,892	-4,029	-5,243	-6,036	-4,593	-5,368	-5,792
	sT3	-4,211	-5,561	-6,271	-4,078	-5,466	-6,202	-4,293	-5,667	-6,288	-4,885	-6,015	-6,298
sTlog	sT1	-3,463	-5,033	-6,190	-3,217	-4,809	-6,047	-3,070	-4,652	-5,838	-2,863	-4,403	-5,589
	sT2	-3,816	-5,179	-6,137	-3,654	-5,013	-6,003	-3,472	-4,811	-5,748	-3,268	-4,592	-5,530
	sT3	-4,071	-5,510	-6,319	-3,954	-5,419	-6,239	-3,710	-5,155	-5,923	-3,476	-4,918	-5,706
log	sT1	-3,951	-5,309	-6,270	-3,701	-5,106	-6,169	-4,062	-5,665	-6,690	-4,162	-5,234	-6,092
	sT2	-4,415	-5,601	-6,298	-4,239	-5,456	-6,215	-4,546	-5,797	-6,468	-4,974	-5,718	-5,990
	sT3	-4,664	-5,995	-6,555	-4,528	-5,916	-6,522	-4,826	-6,227	-6,713	-5,288	-6,393	-6,509

4.2.13-as táblázat: T-próba eredményei a vizsgált napot megelőző összegzett h értékekre



4.2.14-es ábra: Becsült és mért fogyasztási értékek nyári időszakra, meteorológiai adatokkal

A 4.2.14-es ábrán jól látszik, hogy a logaritmikus regresszióval becsült értékek alig különböznek a határhőmérséklet, illetve a meteorológiai történeti adatok változásával. A halványzölddel jelölt „sTlog” görbe, amely lineáris regresszióval, de a hőmérsékleti történeti értékek logaritmusával lett számítva, elkülönül a másik három becsült görbétől. Alábecsüli a mért értékeket, ami várható is. Az ábrán 2010-es időpontok szerepelnek, melyek a vizsgált intervallum vége felé találhatók. E miatt a lineáris regresszióval vizsgált dátumértékek az átlagos dátumérték alatt vannak, amiből az következik, hogy a lineáris közelítésnek köszönhetően egyre nagyobb hibát fog eredményezni az idő múlásával. A logaritmikus regresszió vizsgálat esetében jobban közelíthető az a valóságnak megfelelő folyamat, melynél a vízfogyasztás konvergál a minimálisan szükséges vízigényhez, a vízárak növekedésével.

### 4.3. Előrejelzés készítése

A klasszikus statisztikai elemzésnél két eltérő regressziós eljárást különböztetünk meg jelen esetben. Az egyik a lineáris, a másik a logaritmikus regresszió.

Lineáris regresszió esetén a becsült érték képlete:  $y = m_1x_1 + m_2x_2 + \dots +$

$$y = m_1 * x_1 + m_2 * x_2 + \dots + c$$

Behelyettesítve a paramétereinket:

$$y = m_1 * A + m_2 * B_j + m_3 * C + m_4 * D_k + m_5 * E + m_6 * F + c$$

Logaritmikus regresszió esetén a becsült érték képlete:

$$y = c * (m_1^{x_1} * m_2^{x_2} * \dots)$$

Behelyettesítve a paramétereinket:

$$y = c * (m_1^A * m_2^{Bj} * m_2^C * m_2^{Dk} * m_2^E * m_2^F)$$

Ahol:

- y: A becsült napi vízfogyasztás
- m<sub>i</sub>: A változóhoz tartozó súlyok, melyeket a regresszió számítása során kapunk
- A: Dátum excel típusú timestamppal megadva
- B: Összegzett maximális hőmérsékleti értékek a vizsgált napot megelőző 'j' napra vonatkozólag °C mértékegységgel
- C: A vizsgált napon mért maximális hőmérséklet °C mértékegységgel
- D: Összegzett csapadékmennyiség a vizsgált napot megelőző 'k' napra vonatkozólag mm mértékegységgel
- E: A vizsgált napon lehullott csapadékmennyiség mm mértékegységgel
- F: A munkarendre vonatkozó bulk változó. 0 vagy 1 értéket vehet fel
- c: Konstans értéke, melyet a regresszió számítása során kapunk

Logaritmikus regresszió esetén a becsült érték képlete:

$$y = c * (m_1^{x_1} * m_2^{x_2} * ...)$$

$$y = c * (m_1^A * m_2^{Bj} * m_2^C * m_2^{Dk} * m_2^E * m_2^F)$$

#### 4.4. Modell pontossága

A modellek összehasonlítására az elterjedten használt RMSE (Root Mean Square Error) hibajellemzőt alkalmaztam. Magyarul négyzetes középhiba néven ismert. A RMSE értéket az alábbi összefüggés alapján kapjuk:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n}}$$

Ahol:

- RMSE: négyzetes középhiba
- n: x független változó elemeinek száma

Az előrejelzés szempontjából történő összehasonlítást megkönnyítendő, pontossággal is jellemeztem a modelleket. Az RMSE értéket elosztottam az ismert fogyasztási adatsor értékeinek átlagával. Az azonos mértékegységek miatt %-ban kapjuk meg az eredményt.

Változó paraméterek				r <sup>2</sup> [-]	RMSE [m <sup>3</sup> /nap]	Pontosság [%]
Regresszió típusa	Határhőmérséklet Tmax [°C]	Csapadékösszeg [mm]	Hőmérsékletösszeg [°C]			
log	21	sh3	sT3	0,435	4103,28	90,80%
log	22	sh1	sT3	0,475	4201,75	90,53%
sTlog (lin)	20	sh3	sT1	0,452	4668,81	89,42%
<b>log</b>	<b>20</b>	<b>sh3</b>	<b>sT3</b>	<b>0,488</b>	<b>3950,05</b>	<b>91,05%</b>

4.4.1-es táblázat: Nyári regressziós modellek változatainak pontossága a tanítási időszorra



## 5. Napon belüli előrejelzés

A napi szintű előrejelzés esetében a becsült napi fogyasztásból fogyasztási menetgörbékkel lehet visszszámolni a félórás fogyasztási idősorokat. A napon belüli előrejelzés esetében ennek a gondolatmenetnek az ellenkezőjét alkalmazom. A becsült félórás fogyasztási értékeket összegezve kaphatjuk meg a napi fogyasztási adatokat. Ennek több előnye is van: a tanítási adatsor elemszáma jóval (48-szor) nagyobb, így azonos időintervallumon részletesebb adatok állnak rendelkezésre. A becsült félórás értékek hibái részben kiegyenlítik egymást napi szinten, amennyiben a mért érték körül ingadozik a becsült érték. Ez azt jelenti, hogy azonos pontosság mellett a félórás értékekből számított napi fogyasztás becslése pontosabb lesz.

A napon belüli előrejelzés esetében olyan összefüggéseket is könnyedén figyelembe lehet venni, mint például a napi fogyasztási csúcsok változása a meteorológiai adatok függvényében. Ennek köszönhetően az üzemeltetés szempontjából igen fontos fogyasztási csúcsokat és azok mértékét pontosabban lehet előre jelezni. Továbbá a neurális hálózatok tanítási képességét felhasználva olyan, nem lineáris összefüggésekre is fény derülhet, melyeket klasszikus statisztikai módszerekkel nem lehet figyelembe venni, vagy csak igen nehézkesen, sok emberi munkaóra befektetésével.

### 5.1. Alkalmazott módszer

A módszer kiválasztásánál figyelembe vettem, hogy az összefüggések, illetve a napi fogyasztási idősorok elemzése során a klasszikus statisztikai elemzés alapján számított középhiba értéke 10-15% körül volt. A napi előrejelzésből félórás idősorok előállításához általános fogyasztási menetgörbék alkalmazása szükséges, melyek tovább növelnék az amúgy is jelentős középhibát. Így a klasszikus statisztika alkalmazását elvettem a napon belüli előrejelzés készítése során.

Haladva a korral, illetve a jelenlegi trendeknek megfelelően, a választás a Fuzzy rendszerekre, illetve a mesterséges neurális hálók tanításán alapuló eljárásokra esett. Ezek közül a fuzzy rendszerek és a neurális hálózatok előnyeit egyesítő hibrid, ún. ANFIS (Adaptive Neuro Fuzzy Inference System) rendszer alkalmazhatóságát vizsgáltam.

### 5.2. Vizsgált paraméterek

A vízfogyasztás előrejelzéséhez hasonló probléma az elektromos hálózatok terhelésbecslése. A vízellátó hálózatok esetében ilyen nagymértékű pontosságra és részletességre nincs szükség, hála a jelenlegi fogyasztáshoz képest túlméretezettnek számító tározótérfogatnak. Szándékosan kerestem magyar példát, mivel mind az áram-, mind a vízfogyasztás trendjét az éghajlati, és gazdasági tényezők jelentősen befolyásolják. Jól látható, hogy az elektromos hálózatok terhelésbecslésénél a vizsgált időpontot megelőző értékektől való függőséget helyezték előtérbe (5.2.1-es táblázat) <sup>[3]</sup>

A TDK munka során a meteorológiai adatokra fektettem a hangsúlyt. Az adathiányok miatt az amúgy sem méretes tanítási adatsor a történeti értékek függőségvizsgálatától még kisebbre zsugorodna. Hiszen, ha hiányzik egy adott időpontban a fogyasztási érték, minden olyan időpont használhatatlanná válik, melynél a paraméterek között szerepel a hiányzó időponthoz tartozó érték. A tanítási idősor szűkülésének a mértékét jól jelzi, hogy a 20668 elemből álló tanítási idősor 10633 elemesre csökken, ha figyelembe vesszük az előző napi összegzett, illetve az egy héttel korábbi összegzett napi fogyasztást. Ez gyakorlatilag a felére csökkentené a felhasználható idősor nagyságát.

Sorszám	Leírás
1	ÉMÁSZ aktuális összerhelés
2	ÉMÁSZ félórás összerhelés
3	ÉMÁSZ összerhelés (1 órával korábbi)
4	EMASZ összerhelés (1 nappal korábbi)
5	ÉMÁSZ összerhelés (1 héttel korábbi)
6	ÉMÁSZ összerhelés (2 héttel korábbi)
7	ÉMÁSZ összerhelés utolsó differenciája
8	EMASZ összerhelés utolsó előtti differenciája
9	ÉMÁSZ összerhelés egy órás átlaga
10	Naptípus
11	Évszaktípus
12	Félóra (1 –48)
13	Nógrádkövesdi átlaghőmérséklet
14	Jászberényi átlaghőmérséklet

5.2.1-es táblázat: Paraméterek az ÉMÁSZ terhelésbecslő modelljében

E mellett azt is szem előtt tartottam, hogy az információtartam megmaradjon. Ennek megfelelően az alábbi bemeneti adatokat használtam fel:

- **Ido:** Minden nap 0:00:00-tól eltelt órák száma (pl: 0, 10.5, 13, 23.5). A nem egész óránál nem volt értelme az eltelt perceket is külön oszlopba venni, mivel többletinformációval nem bír, az összevonás segítségével ráadásul a napon belüli periódus sokkal átláthatóbb a fuzzy felületek esetében (5.5.7-es ábra). A fél órás idősor miatt minden hiánytalan napra 48 db érték jut.
- **Nap:** 1-9 ig terjedő diszkrét eloszlású változó. 1-7-ig a hét napjait jelöli (, ahol 1: Hétfő), míg 8 az olyan, munkarendet befolyásoló napokat, mint például a nemzeti, egyházi vagy állami ünnepek, pihenőnapok (áthelyezett szombat). Az áthelyezett munkanapokat (szombatra) 9-es számmal jelöltem. Ezeket egymást kizáró eseménynek tekintem, azzal az alapfeltevéssel, hogy a munkanapra és a hétvégére eső ünnepek között nincs különbség.
- **Tmax:** az adott napra eső maximális hőmérséklet °C mértékegységben megadva.
- **Csapadék:** az adott nap leesett csapadék milliméter mértékegységben megadva.

Egyből szembetűnő, hogy az vízfogyasztás éves periódusának figyelembe vételéhez hiányzik az aktuális napig január 01-től eltelt napok száma. Ez tudatosan lett kihagyva, hiszen a nyáron megnövekvő vízfogyasztás a hőmérséklet növekedésével van összefüggésben, aminek szintén egy éves a periodicitása. E miatt többletinformációhoz nem jutnánk vele, viszont lassítaná a tanítási folyamatot. Továbbá a hét napjainak eltolódása miatt a hétnek nem ugyanarra a napjára esnének (A Gergely „öröknapár” 14 különböző évet tartalmaz). Ez akár torzíthatja is az eredményeket.

A meteorológiai adatok csupán napi részletességgel állnak rendelkezésre, így ezeket össze kell fésülni a félórás fogyasztási idősorokkal. Erre szintén az adatfeldolgozó program ad lehetőséget a meteorológiai, és a fogyasztási adatbázis betöltése után.

A csapadékmennyiség esetében bőven elegendőnek tűnik a napi részletesség, viszont a napi maximális hőmérséklet helyett érdemes lenne a pillanatnyi hőmérsékleti adatokat hozzárendelni a félórás fogyasztási értékekhez. Erre a debreceni rendszer esetében jelenleg nincs lehetőség.

A 7-8 as paraméter a táblázatban a deriváltakat jelöli, melyek számítására képes az általam írt adatfeldolgozó program is. Ha később a modell további pontosítására lenne szükség, így könnyedén importálható a neurális hálózat input értékei közé (3.1-es ábra).

A bemeneti, input értékek meghatározásánál figyelembe kellett venni, hogy a tanítási eljárás számítási igénye a bemeneti értékek számának növekedésével nem lineárisan nő. Ennek megfelelően a bemeneti adatsorok számát lehetőleg minél alacsonyabb értéken próbáltam tartani.

Megjegyzendő, hogy például a MIVIZ-nél (Miskolci Vízművek) három ponton is mérnek pillanatnyi hőmérsékletet. Ezek mintavételi időköze megegyezik a szivattyúk, és tározókéval, továbbá mivel azonos adatbázisban foglalnak helyet, a feldolgozást is megkönnyítené. Sajnos ennek a vizsgálatát túlmutat a TDK dolgozat terjedelmén.

### 5.3. Előrejelzés készítése

A számítások elvégzéséhez a MatLab 2011b verzióját használtam. Ezen belül is az ANFIS toolboxot alkalmaztam. Az ANFIS (Adaptive Neuro Fuzzy Inference System) a hybrid rendszerek közé tartozik, mivel egyesíti a fuzzy és a mesterséges neurális hálókat. Az eljárás lényege a következő:

*„A felhasznált jellemző az idősor fraktál dimenziója, amelyet megpróbálnak regressziós módszerrel közelíteni, ezáltal felépítve egy Sugeno típusú fuzzy következtető rendszert. A fraktál dimenzióját az idősornak a BOXCOUNT algoritmussal határozzák meg, amely működése szerint egyre kisebb oldalhosszúságú négyzetekkel próbálja lefedni az idősort megjelenítő görbét, majd egy regressziót próbál illeszteni a négyzetek számára az oldalhosszúság függvényében. Ezt a fraktáldimenzió következő egyenlet szerinti definíciója alapján teszi:*

$$d = \lim_{r \rightarrow 0} \frac{\ln N(r)}{\ln \frac{1}{r}}$$

*Ahol  $r$  az oldalhosszúság és  $N(r)$  a lefedéshez szükséges négyzetek száma. Az így meghatározott jellemző információ már belefoglalható az ANFIS antecendensébe és taníthatóvá teszi rendszert.”<sup>[4][5]</sup>*

### 5.4. Összefüggések vizsgálata

A betanított modell becsült értékei annak ellenére, hogy a tanításra használt adatsorral összehasonlítva elfogadható középhibát generálnak, nem feltétlenül tükröznek reális, és valós összefüggéseket. E miatt elengedhetetlen egy független adatsorra tesztelni a betanított fuzzy rendszert, és annak az eredményeit ki is elemezni. Mind a két vizsgált adatsor esetében fontos, hogy lehetőleg tartalmazzák az összes reprezentatív esetet. A tanítási algoritmus során, ha hiányzik egy ilyen jellemző értékeket tartalmazó input adatsor, később a tesztelésnél ezekre az esetekre nagy eltérést generálhat, hiszen más összefüggéseket feltételez a tanítási eljárás során az adott input értéktartományra.

A továbbiakban vizsgált, eltérő felépítésű fuzzy rendszereknél, hogy a változatok könnyen megkülönböztethetőek legyenek, a következő jelölési rendszert alkalmaztam: **e-f-g-h**, ahol

**e:** Időpont adatsorhoz tartozó tagsági függvények száma

**f:** Munkarendhez tartozó tagsági függvények száma

**g:** Tmax, maximális napi hőmérséklethez tartozó tagsági függvények száma

**h:** csapadékmagassághoz tartozó tagsági függvények száma

Ahhoz, hogy a különböző változatokat ki tudjuk értékelni, nem feltétlenül elég egy a tanítástól független adatsorra tesztelni a betanított neurális hálót. A tesztelési adatsorok ideális esetben szintén tartalmaznak minden reprezentatív esetet, különösképp a szélsőértékeket. Jelen esetben

azonban a kevés rendelkezésre álló adat miatt ez nem fog teljesülni. A sok adathiány miatt 2007-től 2012 nyaráig megközelítőleg 400 napnyi adat, félórás idősor áll rendelkezésünkre. Külföldi példák alapján 5-6 éves szakadásmentes idősor lenne ideális <sup>[6][7]</sup>.

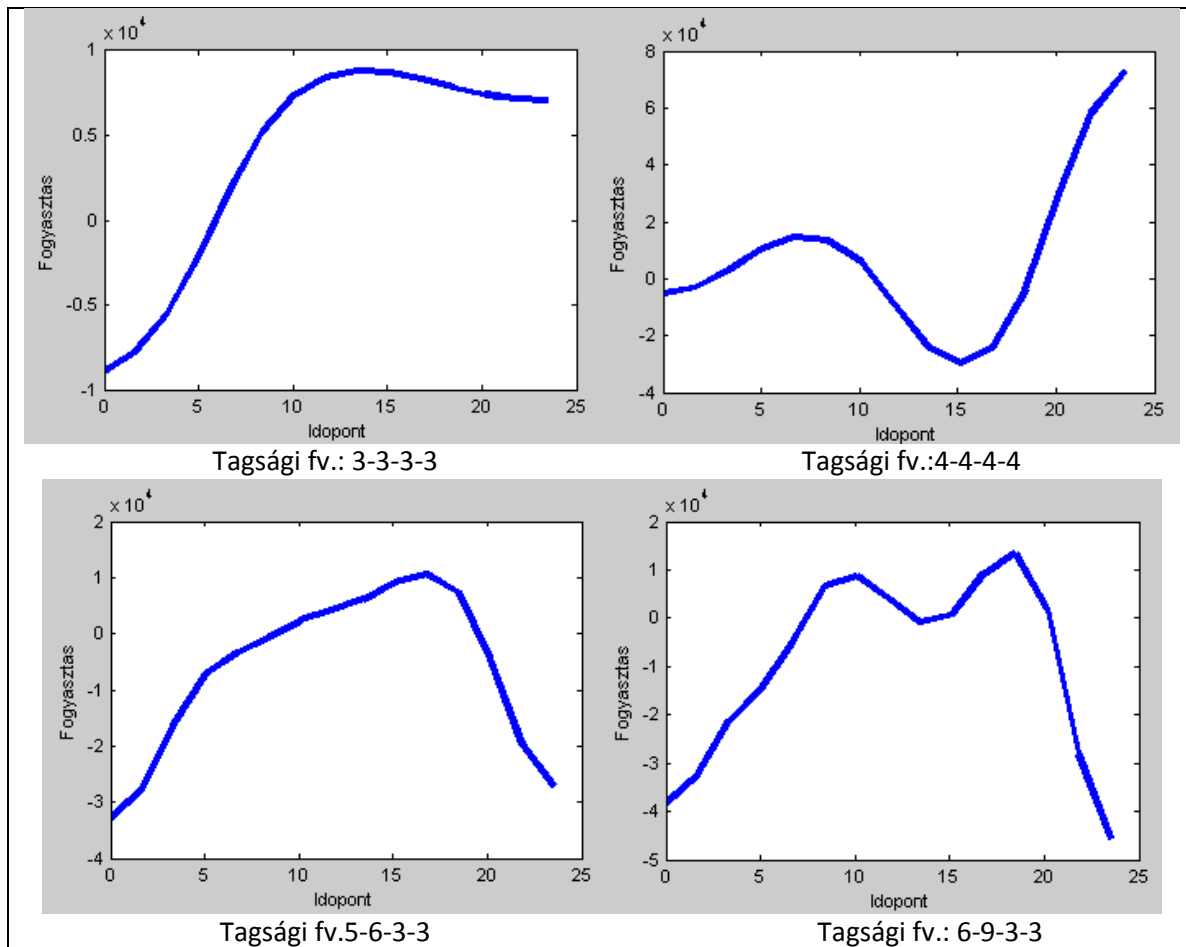
Hogy szemléletesebb legyen a szélsőértékek hiánya, táblázatba szedtem 1981 óta a 30mm-t meghaladó napi csapadékösszegeket, kiemelve azokra az időszakokra eső napokat, melyekhez tartozik archivált érték. Van két szélsőséges érték is, ezeket azért emeltem ki, mert Debrecen éves csapadékmagassága sokéves átlag alapján 540 milliméter <sup>[8]</sup>. Ehhez képest ezek az értékek irreálisan magasak. Ez a kézi adatbevitel során történt elütésnek tudható be. A komoly probléma viszont az,

Dátum	Csapadék [mm]	Dátum	Csapadék [mm]	Dátum	Csapadék [mm]
1981.06.06	40,0	1998.04.20	42,5	2003.07.30	45,7
1981.10.01	52,0	1998.06.30	33,0	2003.10.22	406,3
1982.07.27	39,0	1998.09.06	45,0	2004.07.27	33,0
1984.08.11	35,0	1998.10.18	31,0	2005.06.09	34,0
1985.05.02	33,0	1998.11.10	34,0	2007.07.04	48,8
1985.08.27	36,0	1999.02.10	46,0	2007.10.24	48,0
1988.09.15	42,0	1999.04.25	32,0	2008.07.14	39,0
1989.06.06	39,0	1999.06.13	35,0	2008.07.23	41,5
1989.08.28	41,0	1999.06.16	167,3	2010.08.16	39,5
1990.04.17	31,0	1999.07.12	45,0	2010.12.01	37,7
1995.07.16	74,0	1999.07.14	32,0	2011.07.11	50,0
1995.08.03	37,0	2001.06.04	32,5	2011.07.26	36,0
1995.09.03	40,0	2002.07.14	40,0	2011.07.29	32,0
1996.10.18	33,0	2002.08.08	45,0	2011.08.08	34,0
1997.08.09	70,0	2003.02.04	37,4		

hogy az üzemirányító rendszerben az archiválás 2007 szeptember 06-ai kezdete óta regisztrált 9 ilyen dátumból csupán kettőn van archivált értékünk a vízfogyasztásra vonatkozólag. Ez a pontos előrejelzés szempontjából kevés, arról nem is beszélve, hogy a fuzzy rendszerek sajátosságából adódóan egy adott intervallumra lehet előre jelezni értéket, extrapolálásra nem alkalmas. Ezért is nagyon fontos a tanítási adathalmazban a szélső értékek jelenléte.

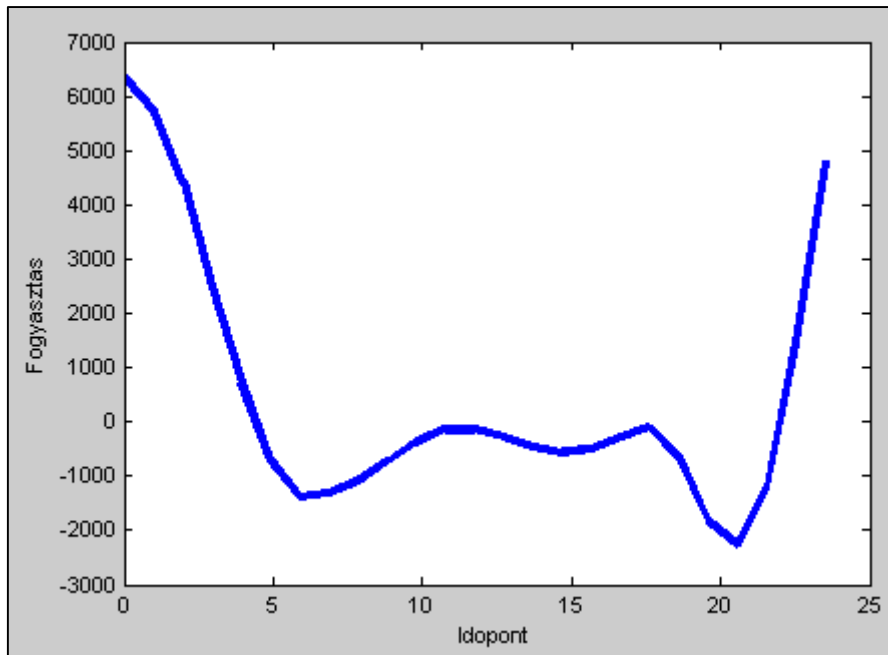
5.4.1-es táblázat: 30mm csapadékösszeget meghaladó napok 1980-tól

### 5.4.1. Napon belüli, időpont szerinti eloszlás



5.4.2-es ábra: Időponttól való függés eltérő változatokra

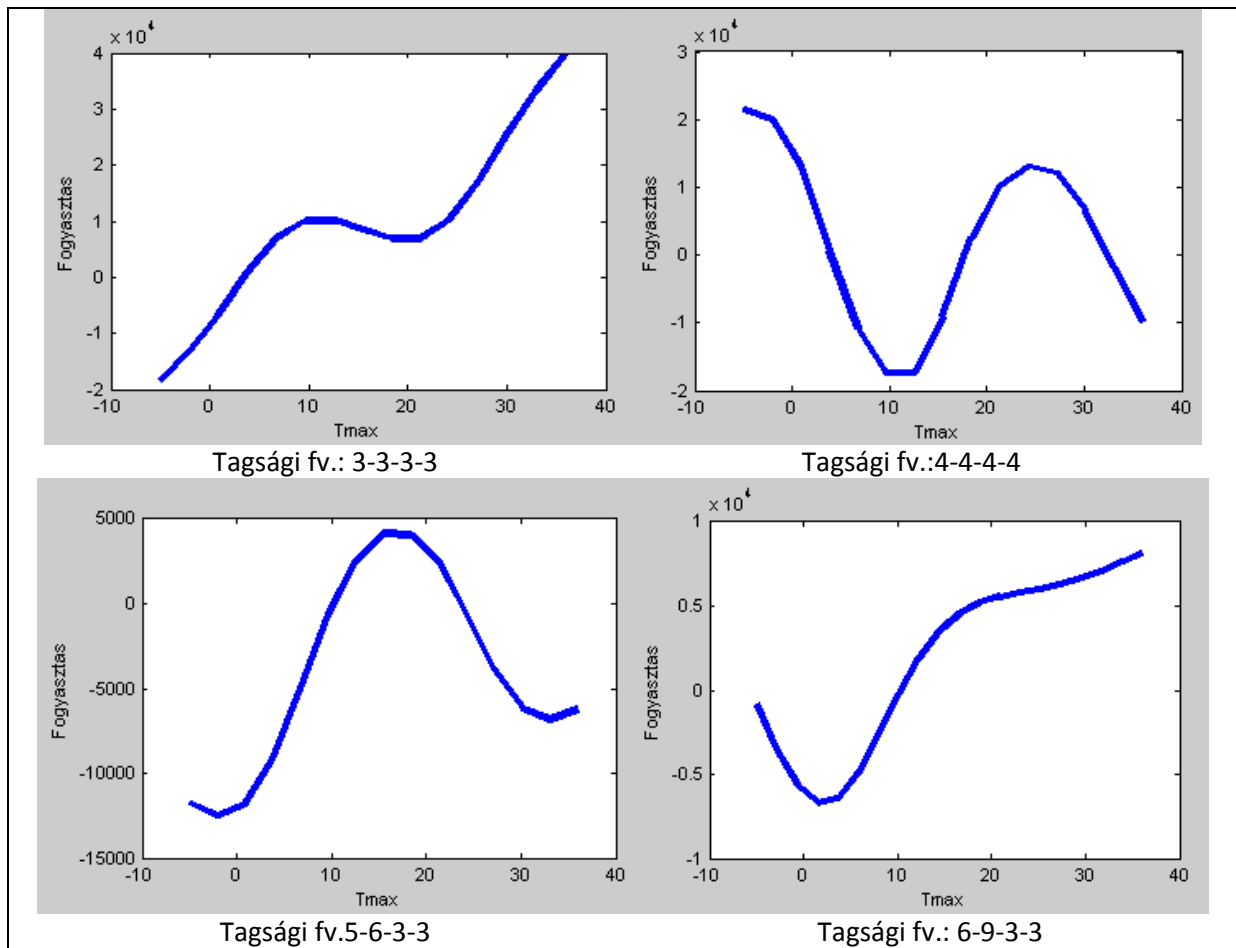
Ezeknél a grafikonoknál a kimeneti érték, mely jelen esetben a *fogyasztás* független a többi paramétertől. Az ábrákon jól látható, hogy a tagsági függvények számának növelésével hogyan közelíti a jellemző fogyasztási menetgörbét az időponttól való függés. Látható, hogy a reggeli és esti csúcs mértéke kisebb a vártnál, ez annak tudható be, hogy a hőmérséklet nagyobb hatással van a fogyasztási csúcsokra. A 9-10 óra környéki reggeli csúcs csupán a hétvégékre jellemző, így érdemes a tagsági függvények eltolásával próbálkozni, úgy, hogy a reggeli csúcsra hétköznap jellemző 7 órára essen egy tagsági függvény 1,0 értéke.



5.4.3-as ábra: Időponttól való függés a 6-9-3-3\_v4d változatra

A véglegesnek tekinthető változatban a várttal ellentétben nem az általános fogyasztási menetgörbének megfelelő függés látható, hanem gyakorlatilag annak a tükörképe. Ez a tanítási eljárás során bekövetkező hibából ered, melyre a modell értékelésénél térek ki bővebben. Ez a változat visszalépésként értelmezhető a sima 6-9-3-3-ashoz képest, azonban ott a meteorológiai adatokra egyértelműen hamis összefüggések adódtak. Az előrejelzés készítésénél az elsődleges a meteorológiai adatoktól való függés meghatározása volt, mivel ez nem olyan nyilvánvaló összefüggés, mint például a napon belüli eloszlása a fogyasztásnak, mely a lakosok napi rutinjától és életvitelétől függ. Az értékelés során az itt látható, téves összefüggések a fuzzy felületek (5.5.6-os ábra) vizsgálata során egyértelműen a szélsőséges meteorológiai értékek esetén jelentkező „félretanítás”-ból következnek. Véleményem szerint ezeket a téves összefüggéseket ellensúlyozza ilyen módon a modell.

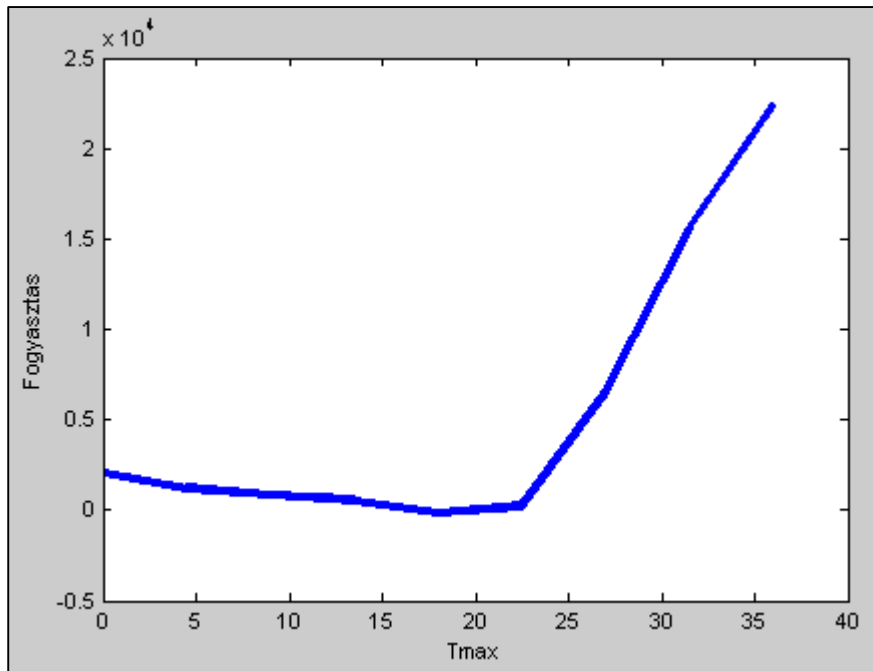
## 5.4.2. Meteorológiai tényezők Hőmérséklettől való függés



5.4.4-es ábra: Hőmérséklettől való függés eltérő változatokra

A hőmérséklettől való függéstől azt várjuk, hogy a hőmérséklet növekedésével a vízfogyasztás is nőni fog. A 6-9-3-3 tagsági függvényeknél a hőmérséklet csökkenésével elkezdi nőni a fogyasztás. Ez lehet a kevés – 3 – tagsági függvényből következő pontatlanság, vagy az elfagyásból következő veszteségnövekedés következménye is.

Azért, hogy egyértelműsítsem a fenti kérdést, létrehoztam egy olyan input adattáblát, ahol a hőmérsékleti oszlopban a negatív értékek helyett mindenhol 0-t veszek fel. Ezzel kiküszöbölhető a fuzzy halmazok miatti zárt intervallumok problémája a negatív hőmérsékleti csúcsok esetében.

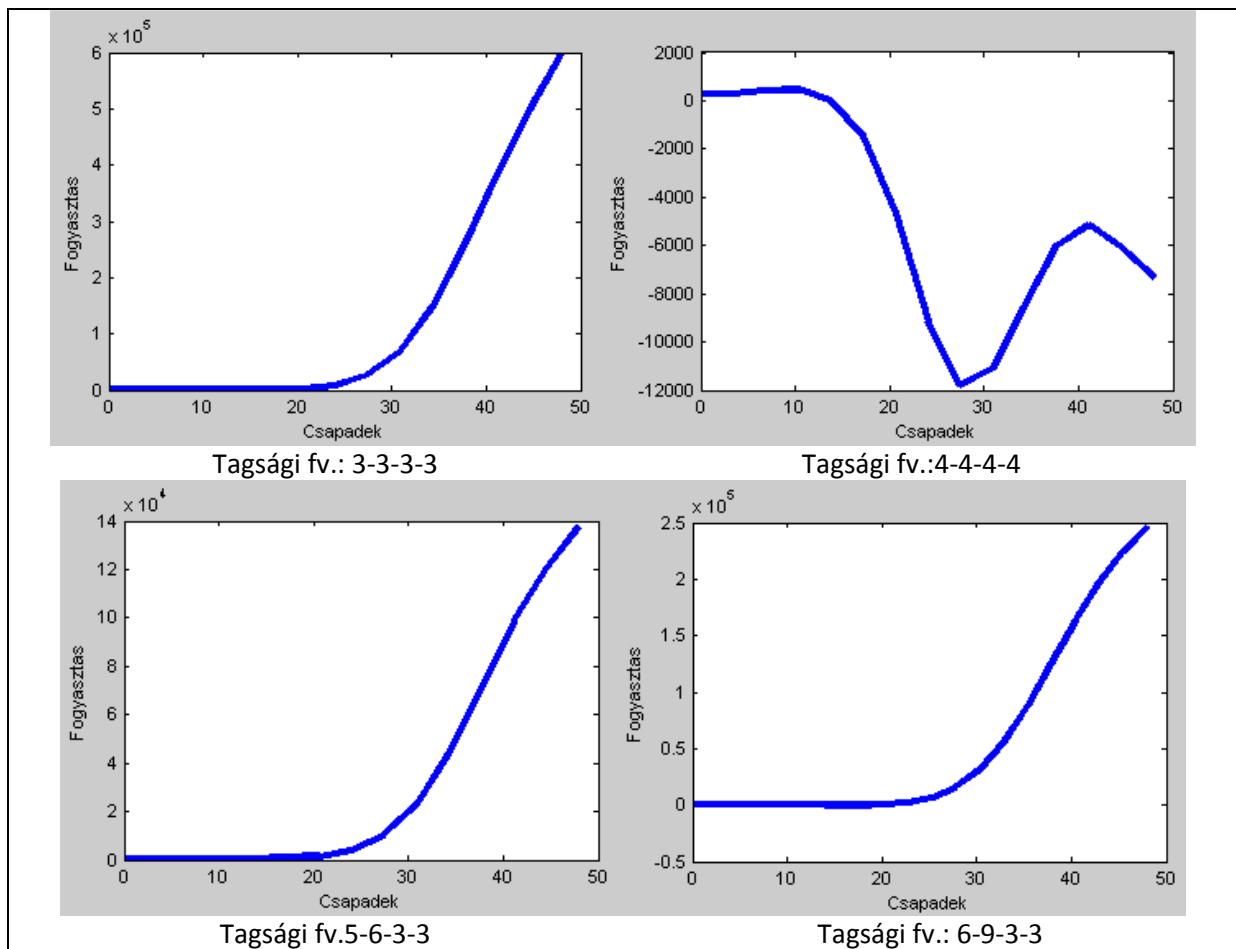


5.4.5-ös ábra: Hőmérséklettől való függés a 6-9-3-3\_v4d változatra

A 6-9-3-3\_v4d változat esetében nem az automatikusan generált Gauss-féle eloszlású tagsági függvényeket alkalmaztam, hanem a sejtéseknek megfelelő tagsági függvényeket, melyekkel jobban közelíthető (5.5.7-es ábra) a vélt összefüggés a napi maximális hőmérséklet és a fogyasztás között. Az 5.4.5-ös ábrán jól látszik, hogy a 22-23 °C felett befolyásolja a vízfogyasztást a napi maximális hőmérséklet. Ettől a határhőmérséklettől szinte lineárisan nő a fogyasztás a hőmérséklet függvényében. Ez igazolja a napi fogyasztási értékek vizsgálatánál azt a feltételezést, hogy csak adott hőmérséklet fölött befolyásolja a vízfogyasztást a napi maximális hőmérséklet (4. fejezet).



## Csapadéktól való függés

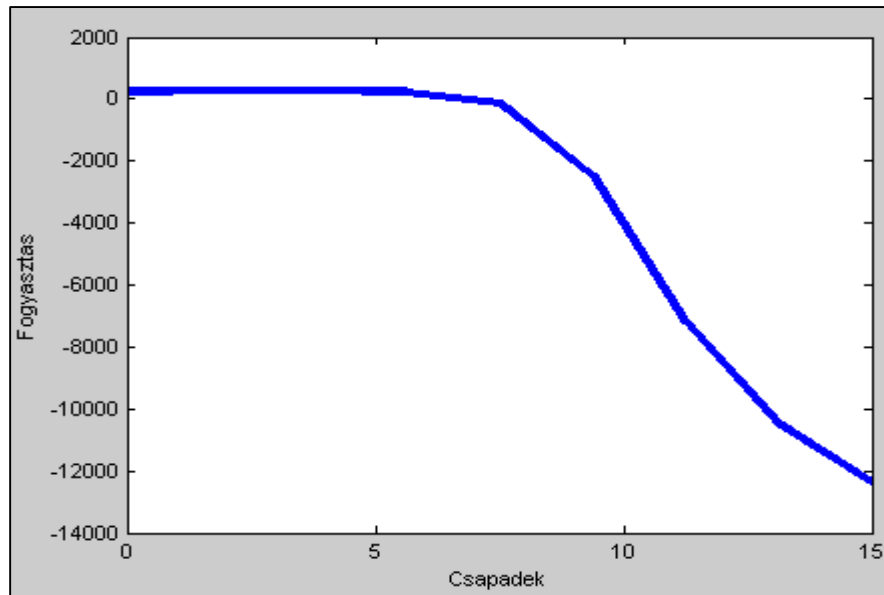


5.4.6-os ábra: Csapadéktól való függés eltérő változatokra

A lehullott csapadék mennyiségétől való függés esetében is hasonló probléma áll fent, mint a hőmérsékletnél. Van két szélsőséges érték a napi csapadékmagasságra, melyek félreviszik a modellt. Így a várttal ellentétben a lehullott csapadékkal arányosan a fogyasztás is nő, mely a feltételezéseinknek ellentmond.

A fenti hibát kiküszöbölendő a 15 mm-nél nagyobb csapadékmagassági értékeket felülírtam 15 mm-re. Ennek köszönhetően a szélsőséges értékek nem torzítják a modellt. Az előző állítás csak abban az esetben igaz, ha a valóságnak is megfelel az a feltételezés, hogy adott csapadékmagasság felett már nem csökken tovább a fogyasztás. Szemléletes példa erre, hogy attól függetlenül nem megy ki az ember locsolni a kertbe, hogy egész nap zuhogott, vagy csupán négy órán át esett aznap az eső.

Ha az előrejelzés során is felülírjuk a 15mm-nél nagyobb értékeket, akkor megoldódott a zárt intervallumok problémája a csapadékok esetén is.

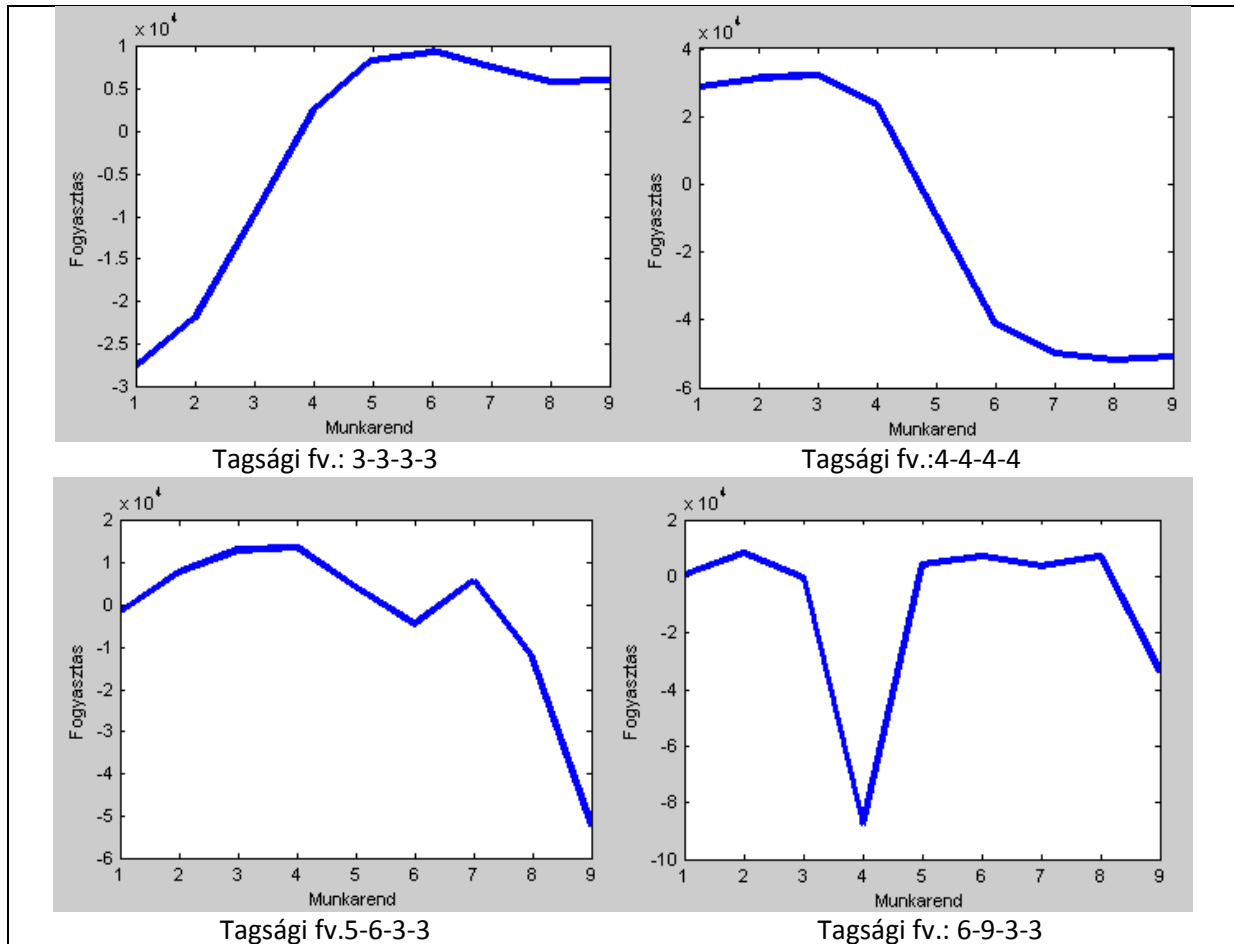


5.4.7-es ábra: Csapadékmagasságtól való függés a 6-9-3-3\_v4d változatra

A tagsági függvények megváltoztatásával (5.5.7-es ábra) sikerült a sejtéseinknek megfelelő összefüggést találni a csapadékmagasság esetében. Ezek a tagsági függvények nagyon hasonlóak a maximális hőmérsékletnél használtakhoz. Ennek az az oka, hogy mindkét esetben azzal a feltételezéssel éltem, hogy bizonyos hőmérséklet, illetve csapadékmagasság alatt nem befolyásolják az értékek a vízfogyasztás mennyiségét.

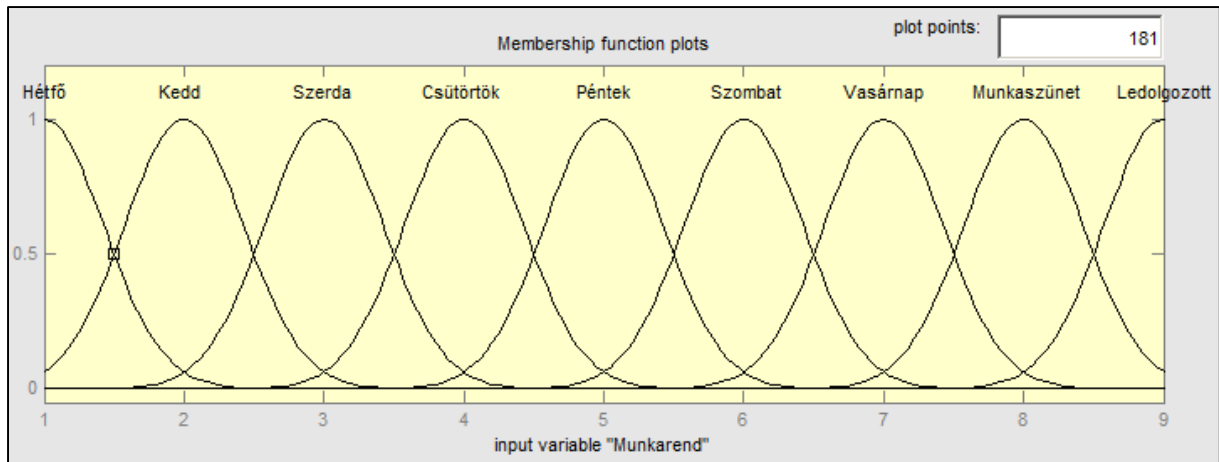
### 5.4.3. Munkarendtől való függés

A munkarendtől való függésnél a hét napjai szolgáltak kiindulási alapul. Az egyéb, általános munkarendtől eltérő napokat az 5.2-es fejezetben leírt módon kezeltem. Ehhez a [http://www.naptarak.com/naptarak\\_munka.html](http://www.naptarak.com/naptarak_munka.html) oldal forrásait használtam fel. Az ünnepek miatti munkaszüneti napokat un. *hardcodeolva* vittem be az adatfeldolgozó programba. Az egyházi ünnepeket 2022-ig tartalmazza, mivel ezek meghatározása bonyolult, és a keresztény egyház által meghatározott napok a mértékadók. Az egyéb, áthelyezett munkanapokat a vizsgált években a munkaidő naptárat végig nézve (2007-től) egyesével vittem be a programba.

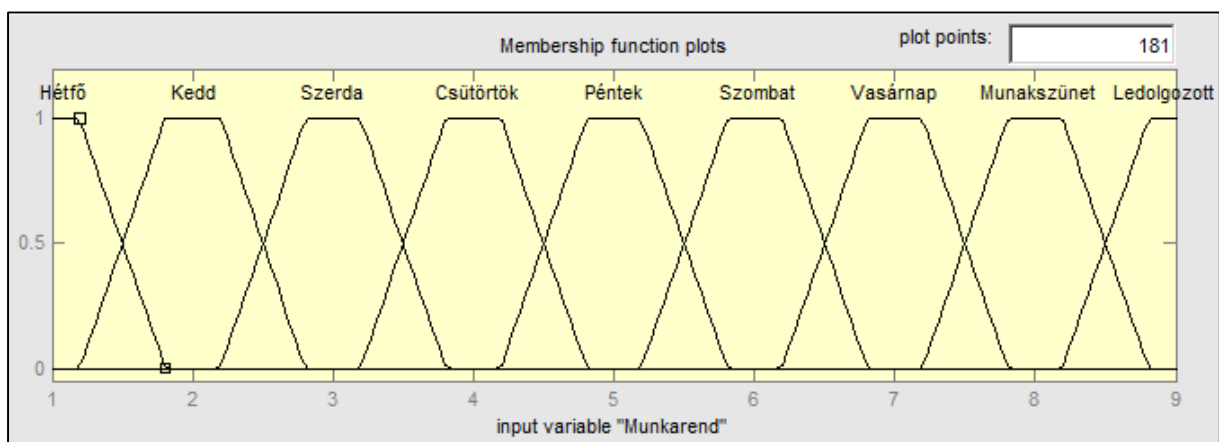


5.4.8-as ábra: Munkarendtől való függés eltérő változatokra

A munkarendtől való függés vizsgálatánál fontos megjegyezni, hogy a vizsgált események egymást kizárják. Nem lehet például egyszerre hétfő és csütörtök, ezek teljesen független események. Ezt azért fontos kihangsúlyozni, mert ha a hétfőhöz tartozó tagsági függvény nem zérus értéket ad vissza 2-es értéknél (ami a keddet jelöli), az torzítja a keddi tagsági függvény súlyszámát. Vagyis ha 2-nél a „Tf\_Hetfo” mondjuk, 0.2 értéket vesz fel, az azt jelenti, hogy a hétfői nap (1) súlyszámának egy ötöde hozzáadódik a keddi értékhez (5.4.9-es ábra). Ennek abban az esetben lenne realitása, ha az „ldopont” input helyett a két nap közti átmenetet  $1/48$ -ad értékekkel jelölnénk. Ebben az esetben „trimf”, avagy háromszög alakú tagsági függvényeket kéne alkalmazni, hiszen az idő egyenletesen telik, vagyis lineáris átmenet van két nap között. Így viszont nem tudnánk kezelni a pihenő-, vagy ledolgozott napokat. Azonban ha csak kerek értékeket adunk meg, ez a felépítés is megfelel a pihenőnapok vizsgálatára, hiszen kerek értékek használatánál a szomszédos tagsági függvények 0-t vesznek fel, függetlenül tőle.

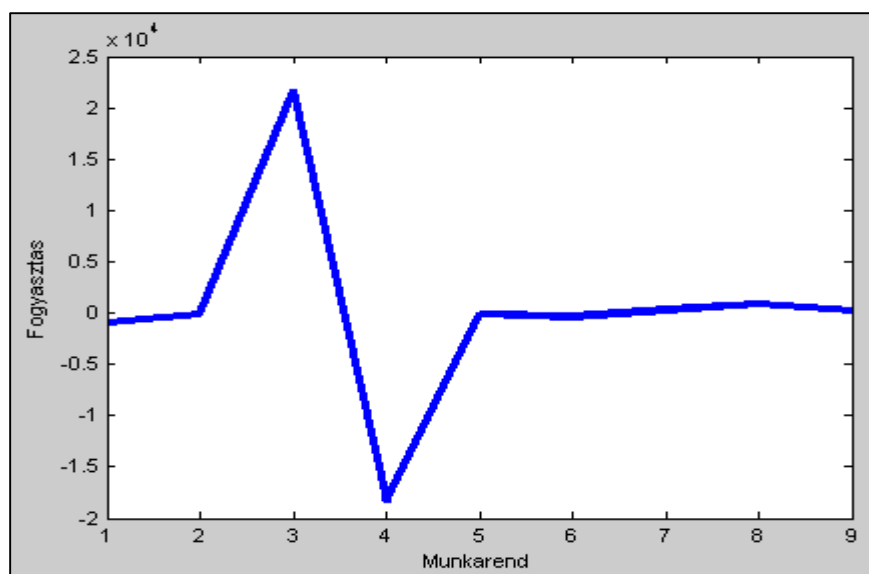


5.4.9-es ábra: Munkarendhez tartozó, „átfedéses” tagsági függvények



5.4.10-es ábra: Munkarendhez tartozó egymástól független tagsági függvények

A végleges megoldás végül a tagsági függvények teljes elkülönítése lett. Így az számítás során meghatározott súlysúlyszámok a valóságnak jobban megfelelnek. Ezekben az esetekben az egyszerűbb „trapmf” függvényeket alkalmaztam.



5.4.11-es ábra: Munkarendtől való függés a 6-9-3-3\_v4d változatra

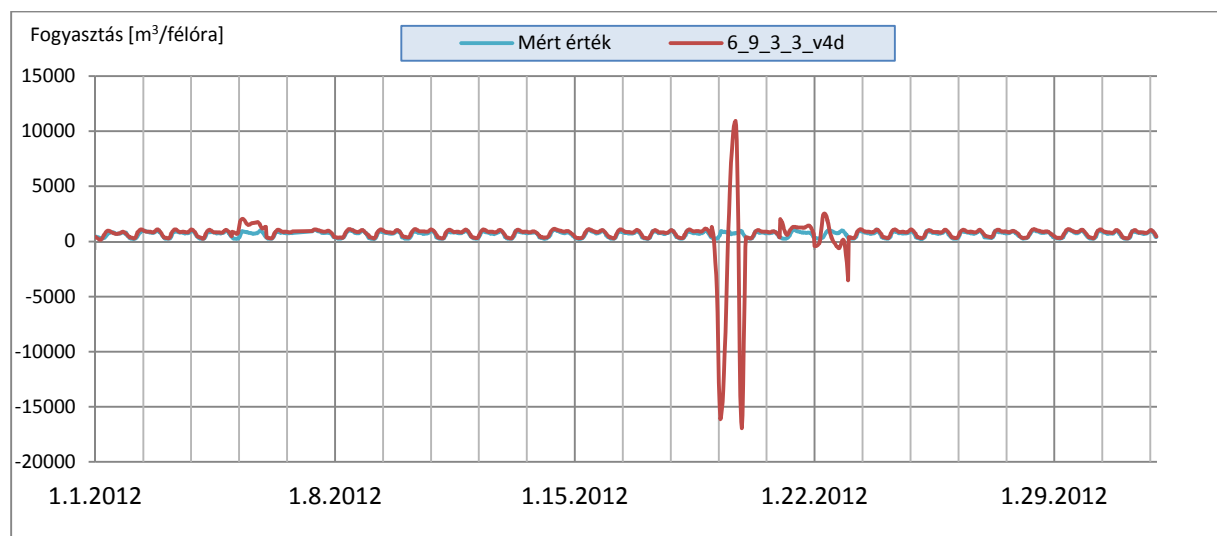
A TDK munka során véglegesnek tekinthető változatban szélsőséges súlyszámok adódtak a 3, illetve 4-es értékhez. Ez a már az időponttól való függés vizsgálatánál is említett félretanítás eredménye. A tanítás során talált hamis összefüggések kompenzálja így, melynek köszönhetően a tanítási adatsorra jó értékeket ad vissza.

### 5.5. Modell értékelése

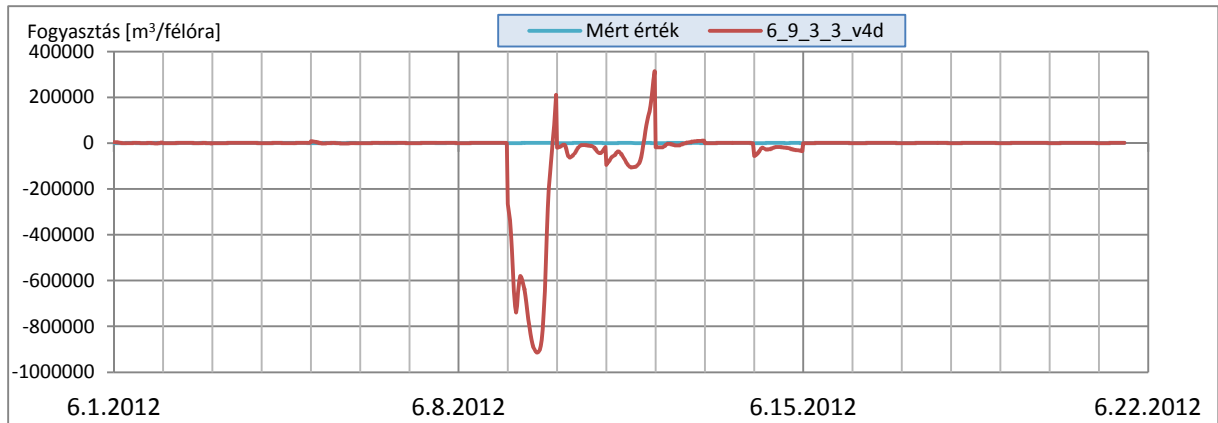
Kijelenthető, hogy az ANFIS modell alkalmas a rövid távú vízfogyasztás előrejelzésére. A téli időszakban, amikor a hőmérséklettől és a csapadéktól jóval kevésbé függ a fogyasztás, igen jó pontosságot sikerült elérni. Ilyenkor gyakorlatilag csak a korábbi fogyasztási adatokra támaszkodik az előrejelzés. A meteorológiai változók súlyszáma zérushoz közeli (5.4.5-ös és 5.4.7-es ábra).

A modell gyenge pontja a szélsőséges időjárási viszonyok esetén történő előrejelzés. Főleg a csapadéktól való függés, melyet rosszul követ. Ez a kis elemszámú tanítási adatsor következménye. Sajnos ezen a téren nem lehet rövidtávon eredményt elérni. Az adatok folyamatos és kitartó archiválásával egyre pontosabb lesz szélsőséges értékek esetén is az előrejelzés.

Ameddig nem áll rendelkezésre kellő számú adat, át lehet hidalni a problémát közelítő eljárásokkal. Ilyen megoldás lehet a modell korlátozása olyan tartományra, melyre kellő számú adat áll rendelkezésre, és a teljes intervallumon kielégítő az előrejelzés pontossága. Amennyiben a tartományon kívülre esne, a szélső értékek alapján extrapolálással kisebb eltérés érhető el. A tesztelési adatsoroknál jól látszik a szélső értékek esetében a tízes nagyságrendű eltérés is (5.5.1-5.5.2-es ábrák), aminél akkor is közelebb kerülnénk a valósághoz, ha korlátozott tartomány szélső értékét adnánk meg.

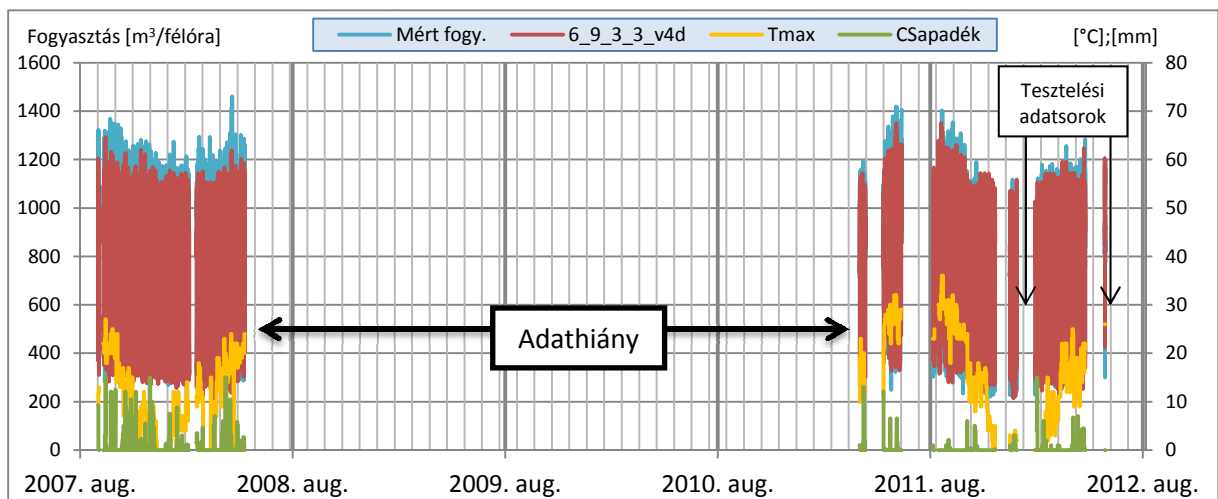


5.5.1-es ábra: Téli független időszak ellenőrzése

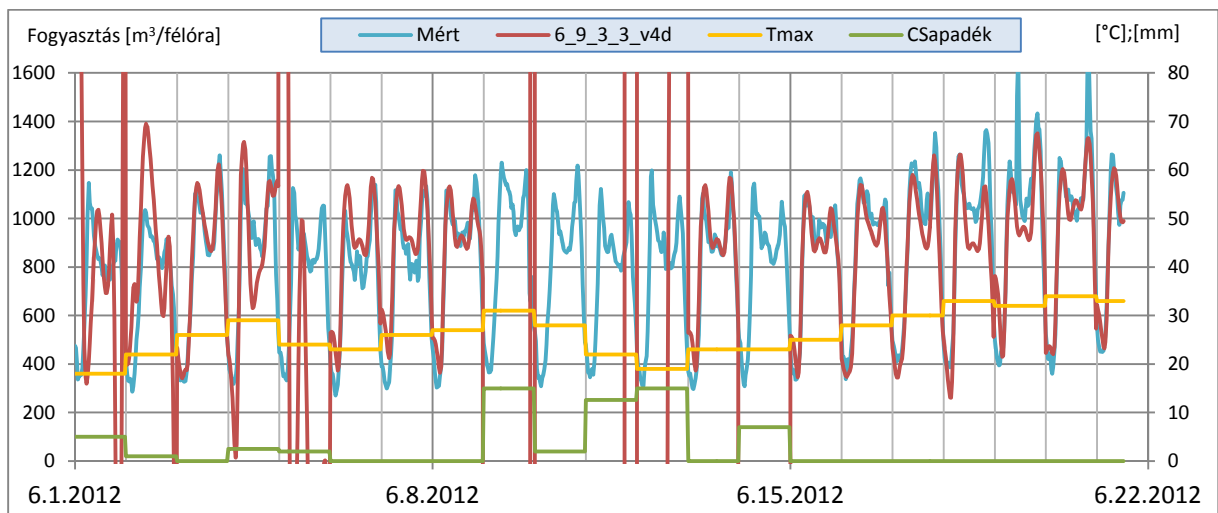


5.5.2-es ábra: Nyári független időszak ellenőrzése

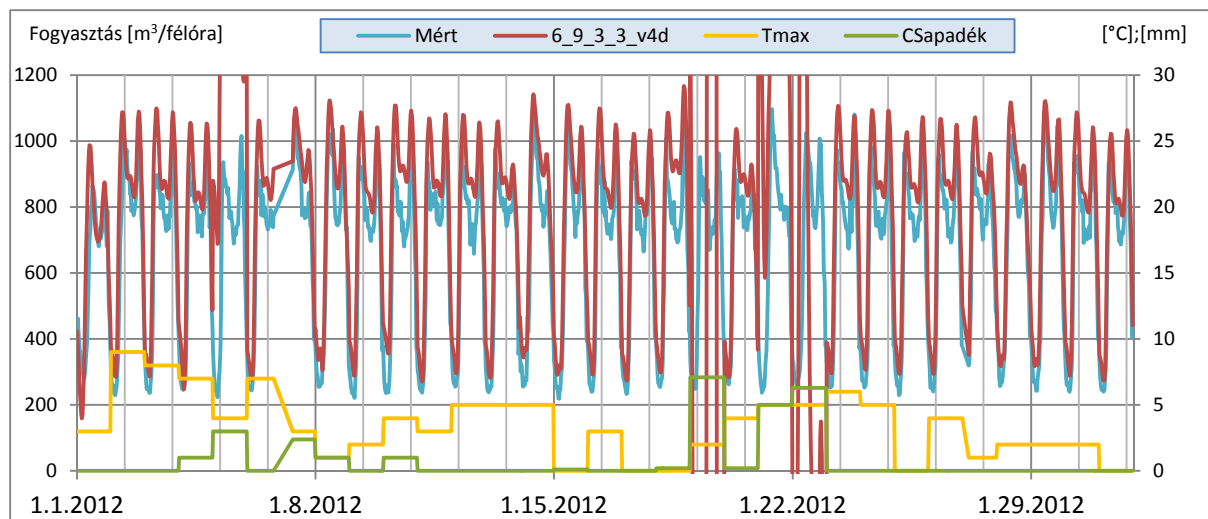
Jól látható, hogy a nyári tesztelési időszakban sokkal nagyobb hibát generált a modell. A negatív fogyasztási értékek azonban még nagyobb problémát jelentenek, hiszen ez fizikailag lehetetlen, normális üzemállapot esetén. Ilyen mértékű hibát az előrejelzési modell rossz kalibrálása, vagy a reprezentatív események hiánya jelent. A torzulások okának kiderítését az input értékek vizsgálatával kezdjük.



5.5.3-as ábra: Napon belüli értékek tesztelése a tanítási időszorra



5.5.4-es ábra: Napon belüli értékek tesztelése a független nyári időszorra



5.5.5-ös ábra: Napon belüli értékek tesztelése a független téli időszorra

Jól látható mind az 5.5.4-es és az 5.5.5-ös ábrán, hogy a magasabb csapadékmagassági értékek esetében „száll el” szélsőséges értékekbe a becsült fogyasztás. A nyári időszak vizsgálatából kiderül, hogy a becsült érték jól követi a maximális napi hőmérséklet változását. Ha napon belül vizsgáljuk az idősorokat, a gyenge pontnak a két fogyasztási csúcs közötti időszak tekinthető, melyet következetesen fölé becsül a modell. Utóbbi probléma megoldása a történeti függés vizsgálata, történeti értékek felvétele a bemeneti értékek közé lehet. Ennek a vizsgálatát jelenleg a kis mintaszám miatt elvettem. Azonban a későbbiekben, ha az előrejelzést fel szeretném használni egyéb célokra, illetve pontosítani, ez elengedhetetlen lesz.

A fenti előrejelzési anomáliák okait az összefüggések vizsgálatával próbáltam kideríteni. A fuzzy rendszereknek utólagos kiértékelésének jó eszköze az úgynevezett „fuzzy felületek” vizsgálata (5.5.6-os ábra). A zárt intervallumokon belül, bármely két input érték függvényében legenerálható egy ilyen, folytonos felület. A tanítási eljárás során talált összefüggések és azok valóságtartalma így ellenőrizhetők. Például, ha a hőmérséklet hatására csökken a vízfogyasztás a modell szerint, az minden bizonnyal félretanítás eredménye. Hamis összefüggéseket a modell több okból találhat:

- **Szélsőséges input értékek jelenléte:** Ilyen esetre példa az 5.4.1-es táblázat kiugró csapadékmagassági értékei. Fontos, hogy amennyiben elegendő számú minta lenne ebben a csapadékmagassági tartományban, ez a probléma kevésbé lenne jelentős. Ennek a megoldására, csak több éves tanítási adatsor lenne a megoldás, mely tartalmaz szélsőségesen száraz, és csapadékos éve(ke)t is.
- **Tagsági függvények rossz megválasztása:** Ha van feltételezésünk az összefüggésekre, kevesebb tagsági függvénnyel is elérhető a modell betanítása. Ilyenkor különböző alakú tagsági függvényeket veszünk fel az általános háromszöges, trapéz, vagy gauss féle eloszlású változók helyett. Jelen esetben a maximális hőmérsékletre, és a csapadékmagasságra alkalmaztam ilyen tagsági függvényeket. Ezzel javulást értem el összefüggések terén, bár a szélsőséges értékek továbbra is jelentkeztek.
- **Hiányzó input változók:** Amennyiben az előrejelzés szempontjából fontos változók hiányoznak, az erre vonatkozó összefüggések más változónál jelentkezhetnek, tévesen megváltoztatva azok tagsági függvényeinek a súlyszámát.

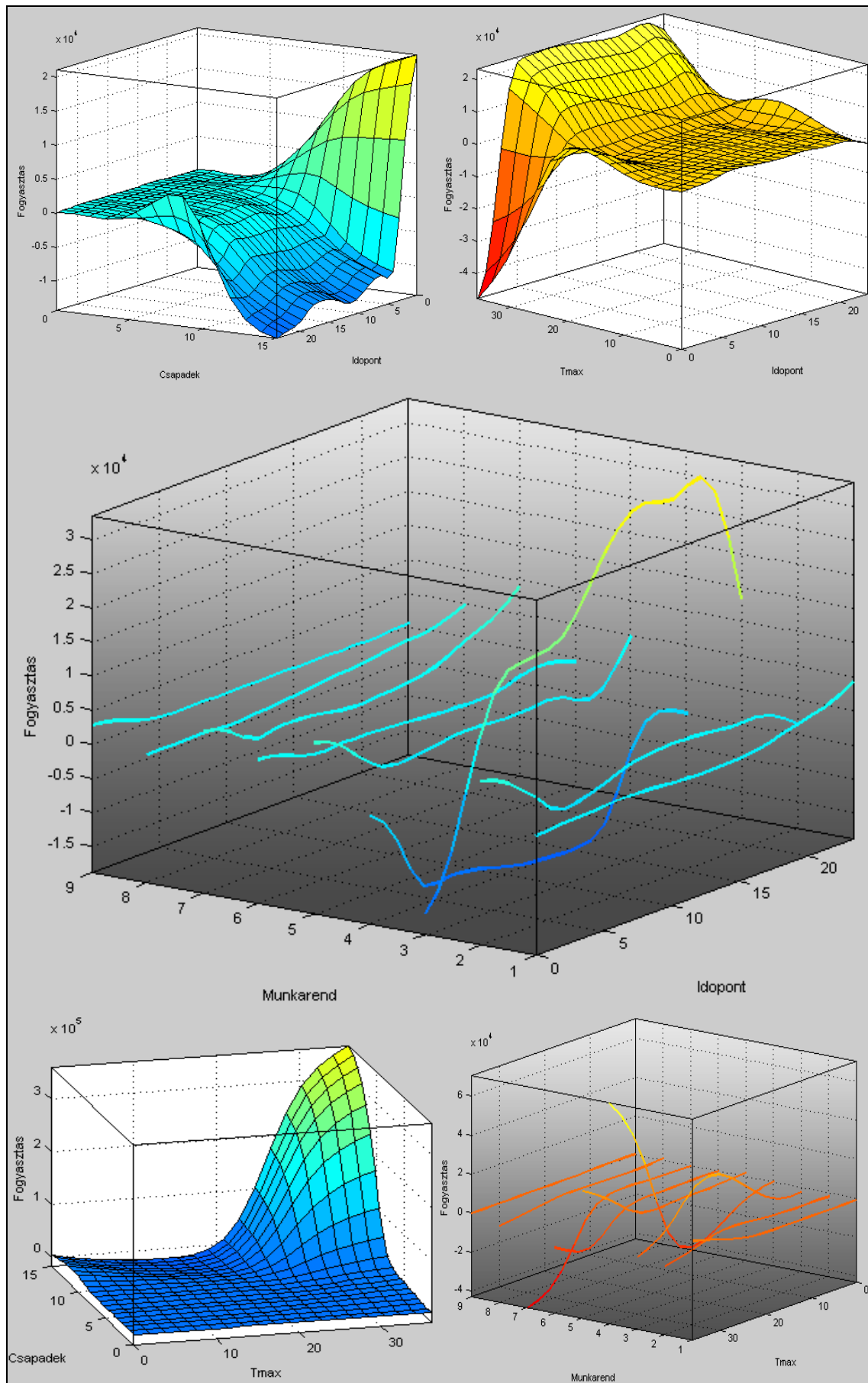
A fuzzy felületek vizsgálata során jól látható, hogy az első félórai időpont kritikus az értékek becslése során. Mind a *Tmax-Idopont*, mind a *Csapadék-Idopont* felületen a szélsőséges értékek esetén jelentkeznek az anomáliák. Ezt a vizsgált maximális csapadékmagasság további csökkentésével nem sikerült kiküszöbölni.

Próbáltam eltolni az *Idopont* értékeket 12 órával (hogy ne a felület sarkára essen a szélsőséges érték), és vele együtt a tagsági függvényeket. Így ami jelen esetben 12.00 az a módosított változatban 0:00-ra esik, a 3:00, pedig 15:00-re. Az anomália azonban továbbra is fenn állt. Ez alapján megállapítható, hogy nem a tanítási eljárás okolható az anomáliáért.

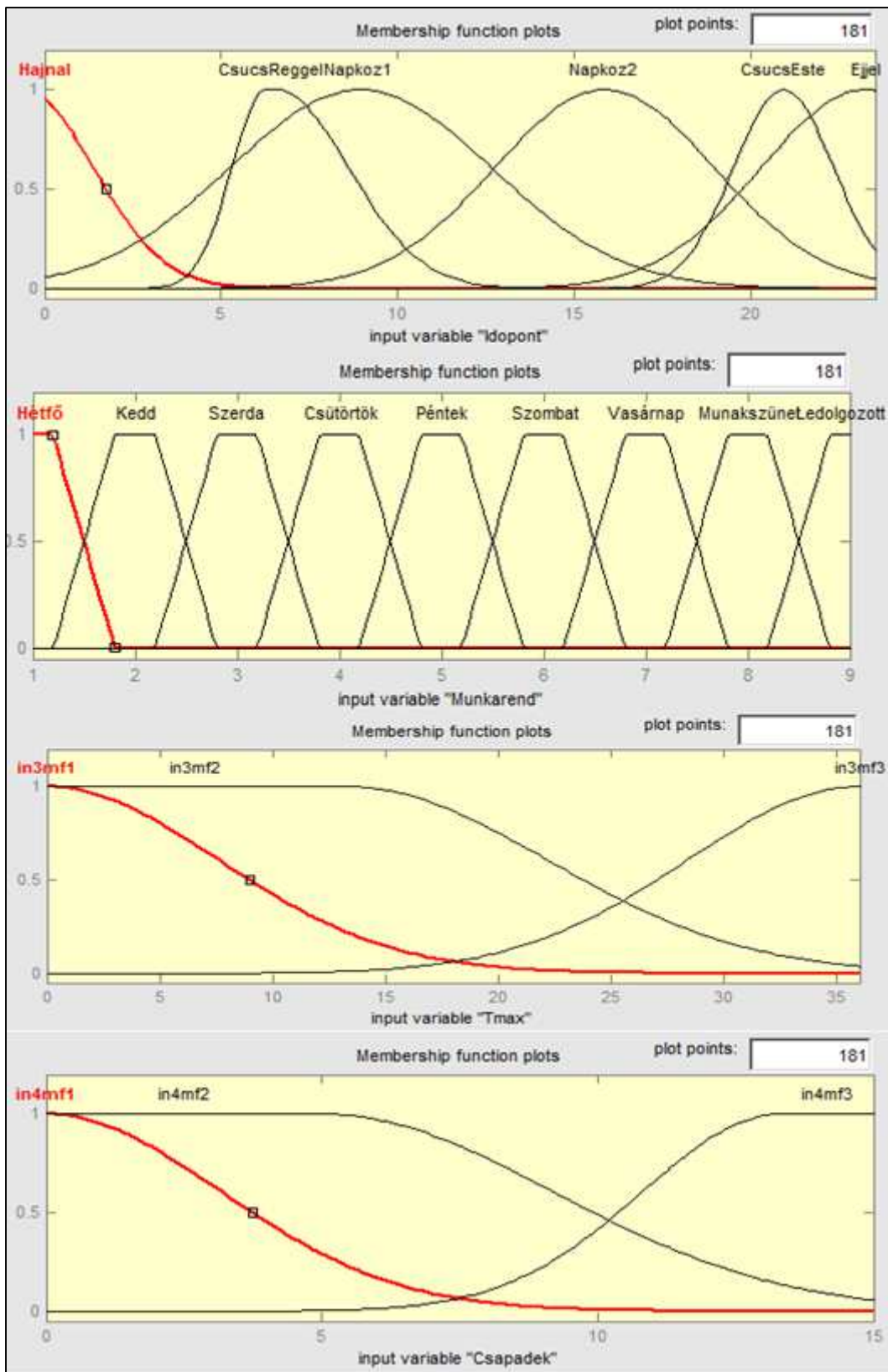
Azon adatként, melyek a félretanítást okozzák, szerdán és csütörtökön jelentkeznek a szélsőséges értékek alapján (5.4.11-es és 5.5.6-os ábra). Ezen kívül az 5.5.6-os ábra jobb alsó felületén, mely a „*Munkarend*” és a „*Tmax*” értékek függvényében ábrázolja a fogyasztást, található egy anomália. Ebben az esetben vasárnapra esik egy szélsőségesen negatív érték.

A dolgozat megírásáig nem sikerült ezt az anomáliát eltüntetni a betanított fuzzy rendszerből. Maga az eljárás ígéretesnek tűnik, ezért a TDK dolgozat befejeztével folytatni fogom a finomítását. A következő fontos lépés a történeti értékek felvétele az input adatsorok közé lesz. Ha az új bemeneti értékek felvételével sem zökken ki a betanított modell ezekből a szélső értékekből, a meteorológiai értékek mélyrehatóbb vizsgálatára lesz szükség. Az idő múlásával a további archivált adatok feldolgozása, majd a tanítási adatsor bővítése szintén fontos feladat a rendszer finomításához.





5.5.6-os ábra: A 6-9-3-3\_v4d változat fontosabb fuzzy felületei



5.5.7-es ábra: A végleges, 6-9-3-3\_v4d változat tagsági függvényei

## 6. Összefoglalás

A TDK munka célja olyan előrejelzési modell keresése volt, mely igazolhatóan alkalmas félórás fogyasztási idősorok előrejelzésére. Megvizsgáltam a napi fogyasztási értékeket a meteorológiai adatok és a munkarend függvényében. A vizsgált időintervallumot leszűkítettem a 2003.01.01 utáni értékekre az összefüggések könnyebb vizsgálatának céljából. Hőmérsékleti adatok közül a maximális napi hőmérséklet befolyásolta legnagyobb mértékben a vízfogyasztást. A napi csapadékmennyiség vizsgálata során bebizonyosodott, ami sejthető is volt, hogy a vizsgált több napi csapadékösszeg majdnem annyira befolyásolja a vízfogyasztást, mint az aznapi csapadékmennyiség. Igen sok változatot megvizsgálva kijelenthető, hogy ilyen módon regressziós eljárással nem érhető el a további felhasználáshoz szükséges pontosság. A legjobb esetben 91,05% pontosságot sikerült elérni napi szinten (ez több mint 5%-kal kevesebb az ANFIS modell esetében mérttől, melyet még így is terhel a félretanításból eredő hiba).

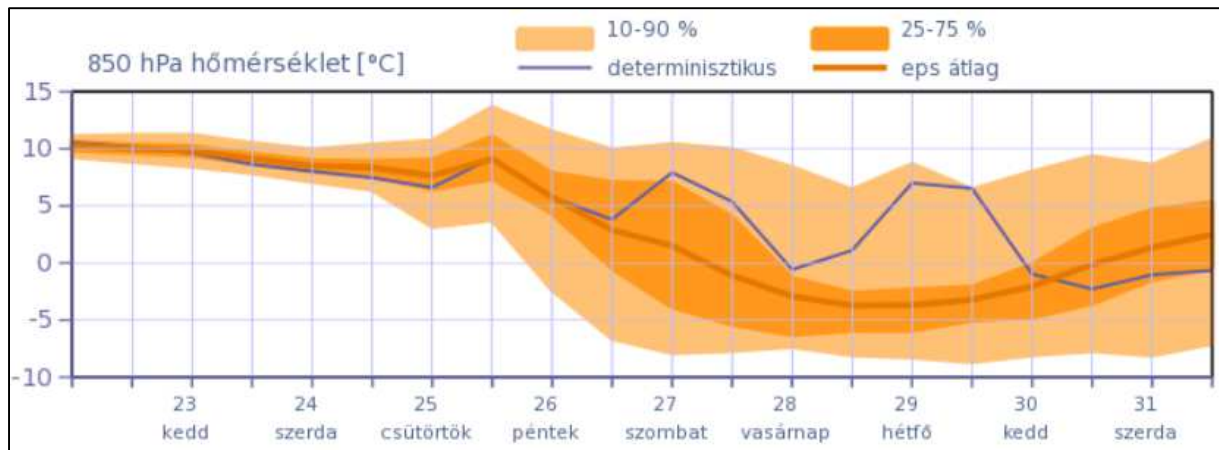
Napon belüli fogyasztási értékekből jóval rövidebb időintervallum állt rendelkezésre, mint napi nagyságrendűből. E miatt a korábbi értékektől való függőség vizsgálatát elvettem, mivel így még tovább szűkült volna a tanításra felhasznált adatsor. Az előrejelzési modell elkészítéséhez a MatLab ANFIS Toolboxát használtam. A fuzzy rendszer tagsági függvényeinek a kalibrálása során több változatot is kipróbáltam. Ezek közül a 6-9-3-3\_v4d változat szolgáltatta a legjobb eredményeket és összefüggéseket. Annak ellenére, hogy csak a munkarendet, illetve az aznapi meteorológiai értékektől való függést vizsgáltuk, ígéretesnek mondható a végeredmény. Téli időszakban félórai értékeket tekintve független, téli adatsorra 84,13%-os, míg a tanításra 85,46% pontosságot ért el a fent említett változat. Napi szinten az eltérés kiegyenlítődik, a független téli adatsorra 89,80%, míg a tanítási adatsorra 96,74% a pontosság. A nyári független adatsor vizsgálata rosszabb eredményeket generált: 82,87% és 87,25%-os pontosságot ért el napi szinten. Sajnos a kevés szélsőséges meteorológiai értékhez tartozó adatsor miatt ilyen esetekben a modell nagyságrendekkel mérhető hibával dolgozott, mely elfogadhatatlan. Ameddig nem áll rendelkezésünkre kellő számú szélsőséges hőmérsékleti, illetve csapadékmagassági adat, a legjobb megoldás ezeknek az értékeknek a felülírása annak a tartománynak a szélső értékére, mely még pontos előrejelzési adatokat szolgáltat.

### 6.1. Eredmények

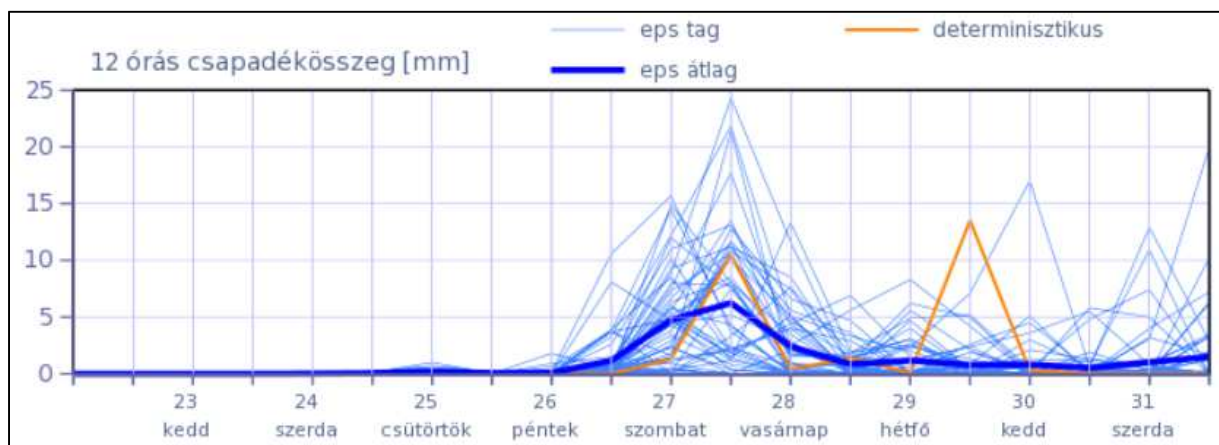
A napi szintű előrejelzéssel nem értem el a kívánt pontosságot. A félórás adatsorok létrehozása során alkalmazandó fogyasztási menetgörbék pedig további hibát vittek volna az előrejelzésbe. Ezt a meteorológiai előrejelzés bizonytalansága még tovább terhelné, ami így teljesen alkalmatlan lenne az eredeti cél eléréséhez, az energiafelhasználás optimalizálásához.

A másik megközelítés, mely a félórás adatsorokból indult ki, ígéretesebbnek bizonyult. Habár a kis elemszámból következő gyermekbetegségeit még nem sikerült kiküszöbölni a neuron hálózatos (ANFIS) rendszernek, további folyamatos archiválással –amennyiben bekövetkeznek szélsőséges időjárási események - a célnak megfelelő pontosság érhető el a független idősorokra is. A továbbiakban érdemes ezen a vonalon folytatni a kutatást. Természetesen a meteorológiai előrejelzés pontossága ugyanúgy terheli az így kapott értékeket is. E miatt érdemes az előrejelzés során a meteorológiai előrejelzés várható értékeihez tartozó fogyasztáson kívül adott valószínűségű felső és alsó határhoz tartozó fogyasztási értékeket is megadni. A 6.1.1-es képen a sötétnarancssárga sáv felső és alsó határa jó példa erre. Szemmel látható, hogy a négy nap után kezd el látványosan nőni a különbség a felső és az alsó határérték között. Ez a későbbi felhasználás szempontjából az

olyan területek esetében lényeges, ahol a heti periodicitás a jellemző a vízfogyasztásra (pl.: üdülőterületek).



6.1.1-es ábra: Hőmérséklet előrejelzés pontossága (forrás: [www.met.hu](http://www.met.hu))



6.1.2-es ábra: 12 órás csapadékösszeg előrejelzése (forrás: [www.met.hu](http://www.met.hu))

Lényeges, hogy a fenti grafikonokhoz tartozó valószínűségi előrejelzési adatok az OMSZ által ingyenesen hozzáférhetőek oktatási és kutatási célra:

**„Az információs díjat nem az OMSZ, hanem az ECMWF határozza meg.(...) Kutatási és oktatási célra nincs információs díj, csak kezelési és távközlési költséget lehet felszámolni. A fenti szabályok csak a „nyers” adatátadásra vonatkoznak, értéknövelt szolgáltatások esetén tárgyalásos ár érvényesül.”<sup>[9]</sup>**

## 6.2. Alkalmazási lehetőségek

A neurális hálózatok és fuzzy rendszerek előnyeit egyesítő ANFIS modell kellő pontossággal tudja kiszámítani a várható vízfogyasztási értékeket (jelenleg még csak a tanítási időszorra tesztelve). Ennek köszönhetően lehetőség nyílt a debreceni ivóvízhálózat energetikai optimalizálására. A magastározók térfogatát a lehető legjobban kihasználva a csúcsidőben felhasznált energia mennyiségét minimalizálni lehet. Ennek köszönhetően az üzemeltetés is jóval költségkímélőbb lenne.

A vízfogyasztás a 90-es évek elején érte el a tetőpontját Magyarországon. Azóta a fogyasztás főként az árak emelkedésének köszönhetően 50-60%-kal esett. A vízellátó hálózatok tározókapacitása erre, vagy némely esetben még ennél is nagyobb fogyasztásra lettek méretezve. A

vízigény csökkenésével a tározókapacitás túlméretezetté vált. Ez egyrészt növeli az üzembiztonságot, másrészt lehetőséget nyújt a tározókapacitás kihasználásával az energiaköltségek mérséklésére.

Az előrejelzésen alapuló szivattyúmenetrend optimalizálásból eredő költségcsökkentés mértékét és hatékonyságát természetesen vizsgálni kell. Jelenleg a víztorony 3000 m<sup>3</sup>-es kapacitásából ténylegesen csak 1300m<sup>3</sup> van üzemben a szigetelési problémákból adódóan. Kérdés, hogy a szigetelési hibák javítása, és az ebből fakadó költségek mennyi idő alatt térülnének meg, figyelembe véve, hogy a nagyobb tárolt térfogat hatékonyabb szivattyúmenetrend optimalizálást tesz lehetővé.

Mind az archivált adatokat feldolgozó programom, mind az ANFIS modell tanítását végző toolbox integrálható egy üzemirányítást optimalizáló programba – a MatLabben rendelkezésre állnak a toolboxok C nyelvben írt forráskódjai.

A debreceni ivóvízhálózat egyszerűségéből fakadóan – egy nyomásháló, egy magastározóval – ideális az előrejelzés és optimalizálás szempontjából. Összetettebb rendszerek esetében is megoldható ugyanez a feladat (3. fejezet), csupán a zónakapcsolati adatbázist kell elkészíteni a megfelelő archivált adatsorok adatkódjaival.

## 7. Irodalomjegyzék

[1] **Short-Term Water Demand Forecast Modelling at IIT Kanpur Using Artificial Neural Networks** ASHU JAIN\*, ASHISH KUMAR VARSHNEY and UMESH CHANDRA JOSHI Department of Civil Engineering, Indian Institute of Technology, Kanpur, India 27 August 2001

[2] **LIN.ILL Excel – Office.com**

<http://office.microsoft.com/hu-hu/excel-help/lin-ill-HP005209155.aspx>

[3] **Rövidtávú on-line terhelésbecslés az ÉMÁSZ KDSZ-ben**, Dr. Kádár Péter, Dr. Mergl K. Attila, Vashegyi György, Hrivnyák Gyula, Orlay Imre (1997)

[4] **Hybrid intelligent Systems for Time Series Prediction Using Neural Networks**, Fuzzy Logic, and Fractal Theory, Castillo, O., Melin P. (2002)

[5] Kardos László, Budapest, Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Kar (2009)

[6] **Forecasting operational demand for an urban water supply zone**, [S.L.Zhou<sup>a</sup>](#), [T.A.McMahon<sup>a</sup>](#), [A.Walton<sup>b</sup>](#), [J.Lewis<sup>b</sup>](#) Journal of Hydrology, Volume 259, Issues 1–4, 1 March 2002, Pages 189–202

[7] **Short-Term Forecasting for Urban Water Consumption**, Aly A. and Wanakule, N. (2004). J. Water Resour. Plann. Manage., 130(5), 405–410. doi: 10.1061/(ASCE)0733-9496(2004)130:5(405)

[8] **Debrecen éghajlati jellemzői**

[http://www.met.hu/eghajlat/magyarorszag\\_eghajlata/varosok\\_jellemzoi/Debrecen/](http://www.met.hu/eghajlat/magyarorszag_eghajlata/varosok_jellemzoi/Debrecen/)

[9] **Az Országos Meteorológiai Szolgálat szolgáltatásainak díjai - Érvényes: 2012. június 5-től visszavonásig**

<http://www.met.hu/doc/tevekenyseg/OMSZdijkatalogus.pdf>