Idősoros energia-fogyasztási adatok kiugró értékeinek automatizált feltárási lehetőségei

(Possibilities of automated outlier detection in time series in case of energy consumptions)

Pitlik László (SZIE MY-X team), Jacsó Ferenc (Rubin Zrt.)

Kivonat: A kiugró értékek keresése végtelen sok modell/elmélet mentén realizálható. Ezek egymással versengő alternatív megoldások. A felhasználó és a modellrendszerek, vagyis a gép-ember szimbiózis csak fokozatosan érlelhető ki.

Kulcsszavak: hasonlóságelemzés, modellezés, benchmarking, ok-okozatiság, szimuláció

Abstract: The outlier detection can be executed based on unlimited models/theories. These alternative solution can be seen as competitors. The users and the potential models, with other words, the level of the human-machine symbiosis can only be increased in a step-by-step approach.

Keywords: similarity analysis, modelling, benchmarking, causality, simulation

# Bevezetés

Minőségbiztosítási rendszerek (pl. ISO 50001) bevezetése kapcsán és/vagy racionális, big-data-jellegű menedzsment szemlélet mellett egyes fogyasztási helyek (a továbbiakban: telephelyek) 5-15 perces ritmus szerint mért (pl. elektromos) energia-fogyasztási (és esetlegesen meteorológiai: pl. hőmérsékleti) adatainak mérésére támaszkodva felmerül a kérdés: miként lehet a riasztást igénylő kiugró értékeket követő jelleggel automatikusan beazonosítani. Kiugró érték fogalma alatt elsődlegesen a kiugróan magas értékek értelmezendők – amennyiben a cél a megtakarítások generálása. A mindkét irányban extrém értékek pl. az üzemelésbiztonság kapcsán lehetnek érdekesek.

Az elemzések operatív szintre tereléséhez elsőként a kiugró érték fogalmát kell tisztázni elméleti síkon:

* Kiugró érték lehet egy fajta *abszolút* érték (extrém érték – vö. outlier detection): pl. éven belül legnagyobb napi fogyasztás, vagy napon belül legnagyobb óránkénti fogyasztás, ill. legmagasabb valaha is mért pl. 5 perces fogyasztási érték. Ezek kapcsán a riasztási szintek kézzel állítandók be: pl. az értékek felső 1-10-25-stb. %-ára.
* Egy fajta quasi racionális (*relativáló*) összehasonlítási küszöbérték lehet pl. az egy évvel korábbi időszakhoz mért eltérés, ill. ennek szélsőséges volta. Ez feltételezi, hogy az időjárás függvénye az energia-felhasználás, ami racionális, de azt is feltételezi, hogy az időjárás mintázatai pl. 365 napos ritmusban szigorúan léteznek – ami pedig egyszerűen nem igaz.
* Emellett létezik a *relatív* kiugró érték fogalma, mely kapcsán előbb be kell tudni vezetni a NORMA fogalmát, s a normától való eltérés maximumaira ugyanaz érvényes, mint a fentebb írt direkt (nyers) mérési adatokra. A norma az az erőtér, mely a keretfeltételek függvényében engedi levezetni az elvárt fogyasztást. A normától való eltérés tehát ab ovo riasztást vonhat maga után. Keretfeltétel lehet pl.
  + egy adott mérési pont múltja és/vagy
  + ok-okozati összefüggésrendszere (pl. hőmérséklet, ill. egyéb meteorológiai adatok), ill.
  + minden (fogyasztási és környezeti) mérés egymással összehasonlító elemzése (vö. benchmarking)…

A változásokat KÖVETŐ elemzések per definitionem utólagosak. Mégis létezik egy szimulációs jellegű átmeneti modellezés (vö. norma-becslés), mely alapján quasi előrejelzések (az aktuális keretfeltételek alapján a következő időszakban elvárható értékek) válnak normává, riasztási küszöbértékké.

A nem követő adatfeldolgozás olyan klasszikus vagy újszerű (öntanuló) előrejelzési modelleket jelent, melyek célja ab ovo a minél nagyobb találati arány realizálása (a változások irányát, ill. mértékét illetően) minél kisebb találati pontosság ingadozás mellett (azaz fenntarthatóan, kalkulálhatóan).

\*\*\*

Egy-egy telephely kapcsán a hőmérsékleti adatok és az energiafogyasztási adatok birtokában kialakítható egy fajta arányossági tényező (pl. regressziós modell).

Az 1 évvel korábbi „önmagához való viszonyítás” az időjárás nem szabályos mintázatai kapcsán több téves riasztást ad, mint racionálisat…

A szimpla arányosság mellett komplexebb (dinamikusabb) modellek is kialakíthatók az időjárás mozgásának és a fogyasztás változásának vélt párhuzamai kapcsán, ami még mindig egy-egy telephelyet érint csak.

A telephelyek és hőmérsékleti adatok egymással való összevetése lenne az holisztikus modell, mely az univerzális normaszámítási elvárásokat fedi le.

Több éves hőmérsékleti adatokat a telephelyhez földrajzilag közeli pontokhoz elvileg fel lehet tárni és ezek alapján (egyszerre többféle – vö. páratartalom, csapadék, szél, légnyomás, stb. - meteorológiai adat) és az energia-fogyasztás között modelleket lehet építeni, ami egyszerre pontosítja a normaszámítást és egyszerre (indirekt módon) becsli a ténylegesen érzékelt hőmérsékletet (wind chill), ill. környezeti kitettséget…

# Operatív adatfeldolgozási rétegek

Az alábbiakban a bevezetés által felvázolt kiugró érték-feltárási lehetőségek kerülnek tételesen bemutatásra fiktív adatvagyonokon:

## Energia-felhasználási idősor önmagában való értelmezései

Az idősoros adatok kapcsán a feladat a kiugró érték-státusz definíciós lehetőségeinek és a riasztásra gyakorolt hatások feltárása:

### Nyers adatok

Legyen adott pl. egy 3 éves, 5 perces ritmusban mért (pl. elektromos) energia-fogyasztási idősor (vö. több, mint 300.000 egyedi mérési adat). Legyen ebben tetszőleges arányban hiányzó/nulla érték, mely a mérőműszer zavarai és/vagy az adattovábbítás esetleges anomáliai miatt vész el. Az adathiány a kiugró értékek feltárását nem érinti (közvetlenül), mert addig, amíg legalább 3 adat ismert, kettő állhat egymáshoz közelebb, így a harmadik a kiugró érték, vagyis az adathiány erre az értelmezésre nem hat. Ha három adat azonos vagy a két szélső adat a középsőtől arányosan tér el (bármilyen definíció is álljon is az arányosság hátterében: pl. a középsőt 100 %-nak tekintve a két másik azonos mértékű, de eltérő előjelű %-ban térjen el, vagy éppen azonos abszolút értékékkel, de eltérő előjellel – ami nem azonos számmisztikus helyzet[[1]](#footnote-1)), akkor nincs riasztás. Minden egyéb esetben az extrémebb eltérés már riasztást jelenthet.

Operatív feladat egy vállalati riasztási rendszer kialakítása során:

* Dönteni arról, vajon csak felfelé, csak lefelé, vagy mindkét irányban keresendők a kiugró értékek?
* Dönteni arról, hogy az abszolút maximum és/vagy abszolút minimum mellett milyen közelségben számít még riasztásra alkalmasnak az eltérés maga? Ezen belül:
  + Mi legyen az idősoros adat vagyon centruma (pl. átlag, medián)?
  + Mi legyen a centrumtól való távolság mérésének módja (pl. abszolút eltérés, relatív eltérés)?

Lehetséges vállalati megoldás a fenti paraméterek lebegtetése, vagyis egy univerzális paraméter-befogadó keretrendszer elvárása, ahol a fejlesztés során előkészített összes értelmezés szerint születik egy riasztási lista. Sőt, ezen listák egymással párhuzamosan is értelmezésre kerülnek: mely időpontok, időintervallumok a leginkább gyanúsak egy minden időpont másként egyforma ellenhipotézis-vizsgálat keretében (vö. anti-diszkriminatív hasonlóságelemzés), mely egy aggregált kockázat-indexet képezve sorrendbe állítja az ellenőrzési szükséglet szerint a riasztásokat.

A fenti paraméterek kezelése automatizálható (inkl. a párhuzamos riasztásokat egységes riasztási rendszerré összefogó anti-diszkriminatív hasonlóságelemző réteg is).

A fenti nyers-adat-alapú riasztási alrendszer első alkalmazása egy fajta intuitív megrendelői tesztként értelmezhető, vagyis a megrendelő/felhasználó/alkalmazó szubjektív módon elégedetté válhat az elvileg már ismert jelenségek mögötti esetleges/vélt/kivizsgált okok és a riasztrások közötti nagy átfedések láttán.

Az éles üzem onnantól indul, hogy az addig rendelkezésre álló időszak minden adata feldolgozásra került. Az éles és a tesztidőszak fogalmainak bevezetésével egy új benchmarking lehetőség is megjelenik: a teszttől való eltérések értelmezésének lehetősége. Azaz pl. az is lehet riasztási szabály, ha a fogyasztások eddigi (teszt-)átlagát az éles adatok átlaga meghaladja. De ez már átvezet az aggregált értelmezések világába (vö. következő alfejezet). Megjegyzendő, hogy az eltérő logikák mentén születő riasztások végső soron átfedést mutathatnak egymással, mely átfedési mértéket a már felemlített anti-diszkriminatív modellezésből származó gyanúindex fogja össze egységes rendszerbe…

### Aggregált adatok

A nyers (pl. 5 perces) mérési adatokat a felhasználói igények (reakció idők) alapján lehet aggregálni pl. óránként, naponta, hetente, havonta – ill. tetszőleges egyéb bontásban (pl. dekádonként, kéthetente, stb.). Az aggregáció kapcsán lehet a hiányzó adatokat nullaként értelmezve összeadni, átlagolni és lehet a nulla értékek kiszűrése után átlagokat képezni. A minden adatot feldolgozó megközelítésben az adathiány átlagcsökkentő hatása révén a felfelé kiugró értékek keresésekor valódi riasztási eseményeket leplezhet le. A negatív extrém értékek üzemelés biztonsági értelmezésekor azonban már értelmezhető riasztások alapját adhatja. A nullás adatok nem mindenkor adathiányt jelölnek, így ezek értelmezése mindenkor egyedi megfontolásokat igényel. A nem nullás adatpozíciók aggregációs időintervallumra vonatkozó átlagai látszólag kizárják az adathiány hatásait. Ez alapvetően igaz, ha racionálisan kevés az adathiány aránya. Ha azonban egy 1-órás intervallumban a 12 db 5-perces adat helyett egyetlen egy érték létezik csak, akkor ennek extremitása nagyobb eséllyel állhat elő, mint ha egyetlen egy adat sem hiányzik, ill. csak egyetlen egy adat/mérés hiányzik.

Speciális értelmezéseknek ad teret, ha az aggregálandó adatok szórása veszi át a szerepet, s nem az átlagok eltéréseit, hanem a szórások eltéréseit vizsgálja az elemző. Ebben az esetben is érvényesek a már leírt döntési pontok.

Operatív feladatok:

* Dönteni arról: mi számít adathiánynak?
* Miként kell ezen pozíciókat bevonni/kizárni?
* Milyen aggregációk (milyen intervallumokra, milyen módokon – pl. összeg, átlag, szórás, stb.) szükségesek?
* Az aggregált adatsorra az előző (nyers adatokat) érintő döntések érvényesek-e? (vagyis mi a centrum és mi a távolságmérés?)
* Milyen adathiány esetén nem kell esetlegesen semmilyen aggregációt végrehajtani (vö. mennyi adat hiányozhat egy pl. ideálisan 12 elemű listából, vagy éppen minden esetben legyen aggregáció)?

Ebben az esetben is értelemszerűen kialakítható (és automatizálható) egy minden paraméter-variációt kezelni képes keretrendszer és a párhuzamos riasztási paraméterezéseket egységesen összefogó anti-diszkriminatív modellezés is.

Az aggregált adatok és a nyers adatok között nincs elsődlegesen érdemi értelmezési különbség: az aggregált adat olyan adat, amit akár nyers/direkt/közvetlen módon mérhettünk is volna. Így minden, ami érvényes lehet egy nyersadatra, a nyersadatok teljességére, az érvényesíthető az aggregált adatokra is.

Amennyiben több párhuzamos nyers adatsor egyidejű feldolgozása a cél, akkor a nagyobb felbontású adatsorokat a legkisebb felbontásúhoz érdemes igazítani, ami ismét csak erősíti a fentebb vélelmezett elveket. S ez át is vezet a következő fejezethez, ahol a nyers energia-fogyasztási adatok és a párhuzamosan létező nyers meteorológiai adatok együttértelmezése történik meg:

## Két párhuzamos idősor együttes értelmezése

Ebben a fejezetben az energia-felhasználási idősor mellett egy hőmérsékleti idősor rendelkezésre állásából indulunk ki és továbbra is keressük a kiugró értékek értelmezésének lehetőségét. Itt és most még nem lesz szó arról, ha két telephely párhuzamos adatai állnak rendelkezésre, akkor milyen értelmezési többlettel kell elméleti szinten megbirkózni.

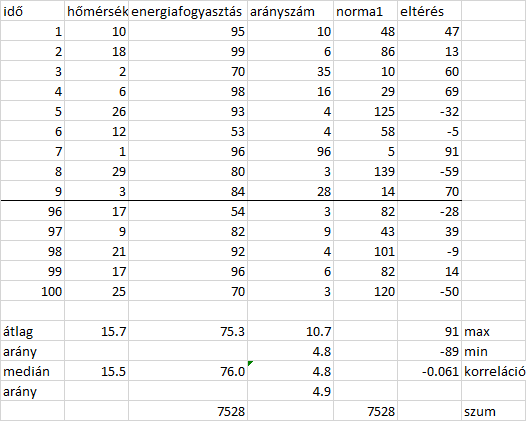
A hőmérsékleti és a fogyasztási adatok nem feltétlenül kerülnek azonos ritmusban mérésre. Elsődleges elvként deklarálandó, hogy a ritkábban mért jelenséghez kell aggregálni a sűrűbben rendelkezésre álló adatokat a fentebb már jelzett aggregáció paraméterekről születő döntések mentén.

A következőben feltételezzük, hogy a szükséges aggregáció megtörténtek, vagy a nyers adatok azonos mérési ritmus szerint állnak ab ovo rendelkezésre.

### Arányossági tényezők levezetése

A hőmérséklet és a fogyasztás között a legegyszerűbb megközelítésben egy arányszám kereshető, mely megadja, vajon egy egység hőmérsékletváltozás mennyi egység fogyasztásváltozással társítandó. Az alábbi (véletlen számokkal kialakított, 100 időegységes) példa arra hívja fel a figyelmet, hogy (vö. 1. táblázat):

* az átlag és a medián, mint centrum eltérhet egymástól
* az átlagok aránya és a mediánok aránya is eltérhet egymástól
* az egyedi arányok átlaga és mediánja is eltérhet egymástól és a többi aránytól (akár jelentősen is: vö. 10.7 vs. 4.8)
* a norma a hőmérséklet és a kiválasztott arányszám (pl. 4.8) alapján keletkezik (hőmérséklet\*arányszám alapon)
* a norma és a tényleges fogyasztás közötti korrelációnak azonban nem illik tetszőlegesen alacsonynak és tetszőleges előjelűnek lennie (a jelenlegi negatív, alacsony érték pl. a véletlenszerű adatok leleplezésére alkalmas mutatószám)
* a norma és a tényleges fogyasztás közötti eltérések távolsága (max, min) az abszolút riasztás alapja, vagyis a maximális/minimális normától való eltérések mindenképpen gyanúsak
* a maximum/minimum mellett további riasztási intervallum tetszőlegesen definiálható, akár egyesével is lehet léptetni lefelé az eltérésmértékeket… (ezek abszolút értékét)…
* annyiféle norma számítható, ahány arányszám
* ezek közül a pozitív és minél magasabb korrelációt mutató megközelítés önmagát legimitálóbb, mint bármely más
* a normák összege és a tényleges fogyasztások összege nem illik, hogy eltérjenek egymástól
* az alternatív megoldásokból az egyes időszakokhoz levezetett riasztások ténye/mértéke (ahol a mérték maga a távolság) egy antidiszkriminatív modellben egységes gyanúindex-értékké konvertálható



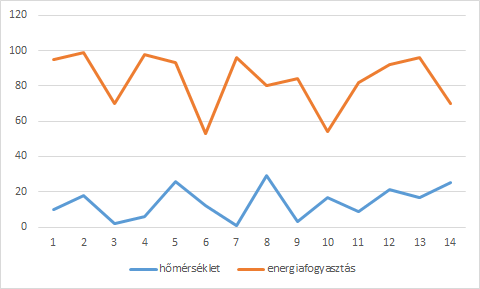
1. táblázat: arányszám-becslések variánsai (forrás: saját ábrázolás)

Operatív feladatok:

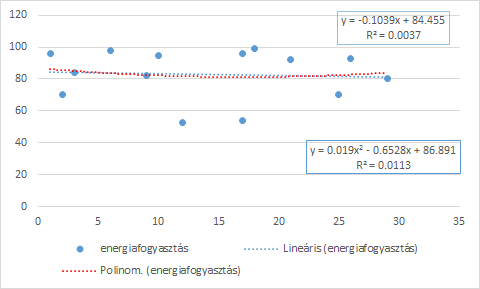
* eldöntendő, milyen típusú arányosság mentén kell norma-értéket és ebből következően normasértést számolni?
* eldöntendő, milyen elvek alapján kell a normától való eltérés extrém értékein túl riasztást értelmezni?

### Regresszió-alapú összefüggés

A hőmérséklet (X) és a fogyasztás (Y) közötti összefüggés lineáris alakzatának meredeksége -0.1. Az regressziós egyenlet Y=-0.1X+84.455 egy fajta újabb normaképzés alapja. A korreláció négyzeteként előálló R2=0.0037. A másodfokú polinom jobb (R2=0.01). A két görbe időben való együttmozgás (a véletlen számoknak megfelelő) esetleges. A fogyasztás trendje csökkenő, a hőmérsékleté növekvő…



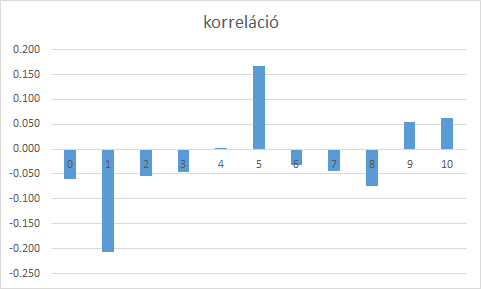
1/a ábra: Az egyedi mérések lefutása (forrás: saját ábrázolás, ahol az X tengely az idő, az Y tengely egyrészt hőmérsékletek, másrészt fogyasztások abszolút értékeit tartalmazza)



1/b ábra: A fogyasztás és a hőmérséklet pontpárjai (forrás: saját ábrázolása, ahol az X-tengely a hőmérséklet, az Y-tengely a fogyasztás)

Operatív feladatok:

* eldöntendő, milyen típusú regresszió alapján kell normaértékét számítani?
* eldöntendő, kell-e késleltetéssel számolni?

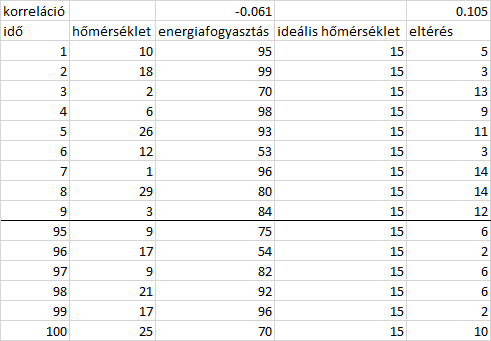


2. ábra: A késleltetés hatása a korrelációs értékekre (forrás: saját ábrázolás, ahol az X-tengely a késleltetés időegysége, az Y-tengely a hőmérséklet és a fogyasztás együttmozgását leíró korrelációs érték)

A késleltetés kapcsán az alapfeltételezés, miszerint a hőmérséklet csökkenése növeli az energiafelhasználást (vö. fűtés) 5 időegység késleltetéssel hat a leginkább, míg ennek ellentettje (vö. klimatizálás) 1 időegység késleltetést valószínűsít (a véletlen szám generátor által adott 100-100 érték esetén).

### Nem monoton hatásmechanizmusra alapozó norma-számítás

Amennyiben az energiafelhasználásnak optimuma van, vagyis adott hőmérséklet esetén még nem kell intenzív klíma-használatot, ill. fűtési igényt feltételezni, akkor a klasszikus monoton együttmozgások nem fogadhatók el.



1. táblázat: Az optimum-hatás leképezése (forrás: saját számítások)

Mint látható (2. táblázat), a korreláció értéke az optimumtól való eltérés és a fogyasztás között nagyobb értékét mutat, mint a monoton hatásmechanizmus esetén.

Operatív feladatok:

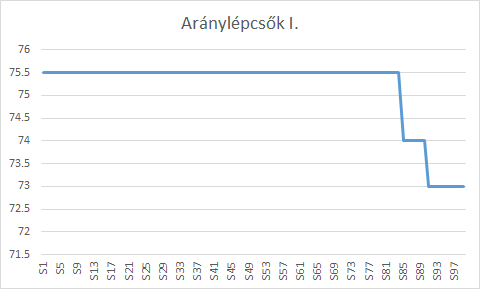
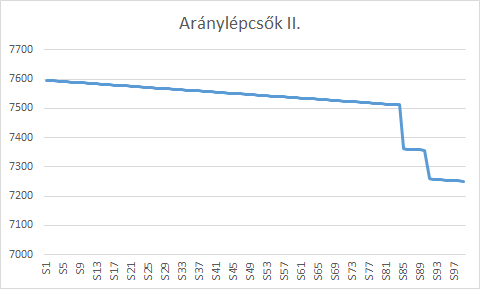
* dönteni arról, milyen jellegű az összefüggés a hőmérséklet és a fogyasztás között?
* kapcsolatot keresni a késleltetés és az optimum-hatás között?

Az összefüggések deklarálása feltételezi, hogy a szakirodalom maga egységes a monoton vs. optimum-jellegű hatásmechanizmusok közötti döntés kapcsán. Amennyiben ez nem adott, akkor exploratív modellezéssel kell a rendelkezésre álló adatvagyon kapcsán döntéstámogató modellszámításokat végezni.

A nem monoton összefüggésre alapuló norma számításakor az ideális külső hőmérséklettől való eltérés abszolút értékét tekintjük alapadatnak, majd az arányosság valamilyen formáját ezen adatsor kapcsán alakítjuk ki. Az arányosság (regressziós modell vs. arányszám) mibenlétének eldöntése után a normaszámítás és a riasztások hasonlóképpen történnek, mint a már korábban jelzett nyersadat-orientáció esetén.

### Nem monoton arányszámokra alapozó norma-számítás

A regresszió esetén minden egyes hőmérsékleti értékváltozás hatása azonos, ami nem triviális, sem műszakilag, sem logikailag. A regressziós együttható azonban lehet lépcsős függvény által becsült érték is, mint ahogy a nem lineáris esetekben sem monoton a két hőmérsékleti értékhez rendelt becslés aránya, de ezek a függvényalakok nem ismerik a lépcsőzetesség lehetőségét:

1. ábra: Aránylépcsők (forrás: saját számítások, ahol az X-tengely a bemeneti jelek szintjei, az Y-tengely az arányszám maga szintenként – a példa az optimális hőmérséklettől való eltérés 14 létező eltérés típusára vonatkozó klasszikus és módosított lépcsős függvénybecslést mutatja)

Operatív feladatok:

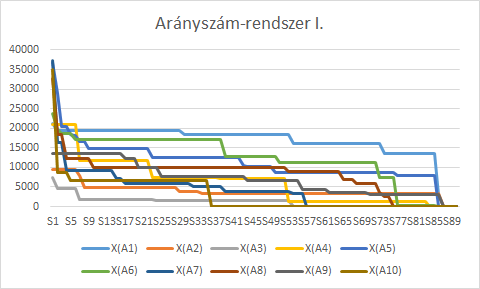
* dönteni arról, vajon monoton-e az arányszám alakulása?
* dönteni arról, milyen típusú lépcsős függvények hitelesek az adott műszaki valóságban?

### Komplex arányszámok

A komplex arányszám-becslés lényege, hogy a rendelkezésre álló adatokból pszeudo-okozatiság-modelleket készítünk:

#### Fogyasztás-orientált bechmarking

Lehetséges a korábbi fogyasztások alapján a következő fogyasztás várható értékének becslése:



1. ábra: A fogyasztások önmagyarázó rendszere (forrás: saját számítások, ahol az X-tengely a rendelkezésre álló 90 adatsorhoz tartózó lépcsőszint, az Y tengely az arányszám, ill. a színes vonalak az 1-10 időegységgel a vizsgált fogyasztást megelőző fogyasztások idősorai – melyek alapján következik a 100 nyers adatsorból a 90 modell-adatsor)

A modell korrelációs értéke = 0.57. Vagyis még véletlen számokra alapozva is racionális normaszámítás kínálható.

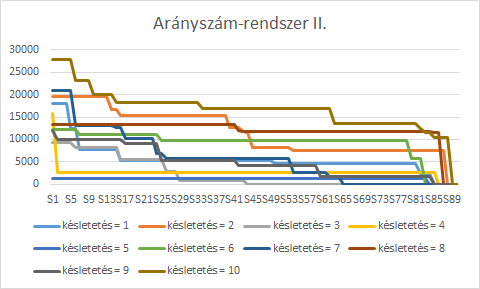
#### Időjárás-orientált benchmarking

Ha csak az időjárás korábbi 10 adatát használja fel valaki a 11. időszak fogyasztásának becslésére, akkor az alábbi arányszám-rendszerhez juthat, ahol a korreláció = 0.38 a becslés és a tényleges fogyasztások között (vö. 5. ábra).

A modell képes természetesen az egyidejű adatokat i s feldolgozni: ebben az esetben nem előrejelzésről, hanem szimulációs jellegű normabecslésről beszélünk.

Amennyiben a fogyasztások önmagyarázó ereje nagyobb, mint a hőmérsékleti adatok magyarázó ereje, úgy nem a hőmérséklet az érdemi ható tényező.

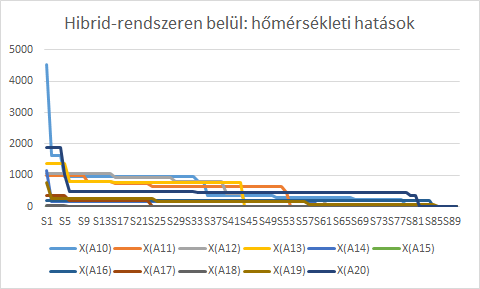
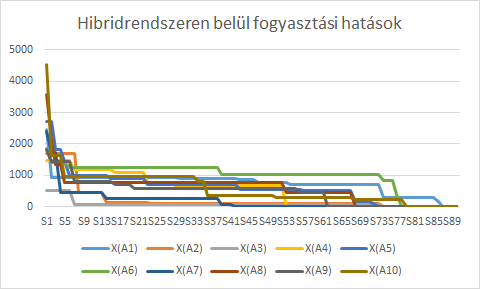
A két hatásmechanizmus hibridizálása további magyarázóerő-növekedéssel kecsegtet…



1. ábra: A fogyasztások hőmérséklet-alapú magyarázó rendszere (forrás: saját számítások, ahol az X-tengely a rendelkezésre álló 90 adatsorhoz tartózó lépcsőszint, az Y tengely az arányszám, ill. a színes vonalak az 1-10 időegységgel a vizsgált fogyasztást megelőző hőmérsékletzek idősorai – melyek alapján következik a 100 nyers adatsorból a 90 modell-adatsor)

#### Hibrid benchmarking

A fogyasztási és hőmérsékleti adatok egyidejű bevonása alapján a komplex normabecslő modell korrelációja nő: 0.75!



1. ábra: a Hibrid modell fogyasztási (balra) és hőmérsékleti (jobbra) rétegei (forrás: saját számítások, ahol az X-tengely a 90 lépcsőszint, az Y-tengely maga a hatásmérték, vagyis az arányszámok magyarázó tényezőnként a szimulátor tetszőleges inputkombinációt becslő képessége érdekében)

A hibrid-modell lehet nem csak lépcsősfüggvényre alapozó, hanem regressziós és bármilyen más jellegű is. A magas korrelációs érték (véletlen számok esetén is) azt engedi vélelmezni, hogy a riasztások, vagyis a normától való eltérés magas szinten megalapozott, míg az alacsony korrelációs értékű norma-becslő modellek esetén nagyobb arányban várhatók önkényes (számmisztikus) riasztások.

### Exploratív arányszámok

Az lineáris regresszió statikus, a nem lineáris szigorúan monoton arányszámait a lépcsős függvények arányszámai ún. mocsár-erőterekkel (változatlan hatású arányszám-szakaszokkal) írják le. Emellett azonban léteznek a szabály-generálás elvét, vagyis az exploratív modellezést használó arányszám-becslések is. Itt az elemző eldönti, mennyi objektum (időszak) egyidejű hasonló viselkedését tekinti már mértékadónak (jelen esetben ez 10%-nyi objektum lesz). Az így csoportokba kényszerített objektum egyedi arányszámot kapnak: vagy csak egyirányú vagy dupla mechanizmusú (direkt és inverz irányú bemeneti) jelekre támaszkodva.

Ezek az elemzések most a komplex arányszámbecslés 3 esetére készültek el szimpla és dupla kivitelezésben. Ennek eredményei - a bemeneti véletlen számokra támaszkodva – az alábbi korrelációhoz vezettek:

* a csak fogyasztási adatok alapján kialakított modell korrelációja 0.57-ről 0.87-re nőtt
* a csak hőmérsékleti adatok alapján kialakított modell korrelációja 0.38-ról 0.82-re nőtt
* a hibrid inputra támaszkodó modell korrelációja 0.75-ről 1.00-ra nőtt.

Ez utóbbi már felveti a riasztások elmaradását, hiszen minden úgy történt, ahogy történnie kellett. Ezen quasi túltanulás gyanús összefüggésrendszer ellen az egy csoportba sorolt elemek számának növelésével lehet védekezni, ill. az idősorok hosszának növelése is a csökkenti a túltanulás esélyeit.

A szimpla és a dupla modellek korrelációja azonos, mert a szabályrendszere egymás tükrözött alakja...

# Következtetések

A fentebb felsorolt lehetőségi kör további tetszőleges modellezési/elemzési paraméterekkel lenne bővíthető (pl. modellek hiba-definíciója: négyzetes hiba vs. abszolút hibaösszeg). Vagyis riasztási modell-logikából formálisan végtelen sok létezik, s ezek üzleti értelemben egymás versenytársai a megtérülésért folytatott versenyben. A legjobb modell automatikus keresés jelenleg meghaladja jelen dokumentum felvállalni kívánt kereteit, de ez is lehetséges.

A felhasználók számára azonban a szerves fejlődést azon automatizációs fok jelenti, ahol mindenről ők döntenek és mindent ők értelmeznek mindaddig, amíg a háttér emberi tudás nem segíti az elemzések és a szakértői intuíciók szinkronizálását.

Tartalomjegyzék

[Bevezetés 1](#_Toc455294070)

[Operatív adatfeldolgozási rétegek 2](#_Toc455294071)

[Energia-felhasználási idősor önmagában való értelmezései 2](#_Toc455294072)

[Nyers adatok 2](#_Toc455294073)

[Aggregált adatok 3](#_Toc455294074)

[Két párhuzamos idősor együttes értelmezése 4](#_Toc455294075)

[Arányossági tényezők levezetése 4](#_Toc455294076)

[Regresszió-alapú összefüggés 6](#_Toc455294077)

[Nem monoton hatásmechanizmusra alapozó norma-számítás 7](#_Toc455294078)

[Nem monoton arányszámokra alapozó norma-számítás 8](#_Toc455294079)

[Komplex arányszámok 9](#_Toc455294080)

[Fogyasztás-orientált bechmarking 9](#_Toc455294081)

[Időjárás-orientált benchmarking 9](#_Toc455294082)

[Hibrid benchmarking 10](#_Toc455294083)

[Exploratív arányszámok 11](#_Toc455294084)

[Következtetések 11](#_Toc455294085)

Melléklet: <http://miau.gau.hu/miau/214/idosoros_energiafogyasztas_kiugro_ertekei.xlsx>

1. További (fiktív centrumpontos) arányossági megközelítés lehet: pl. a rendelkezésre álló adatok átlagától való eltérések vizsgálata, stb. [↑](#footnote-ref-1)