Gyanúgenerálás lehetőségei energia-fogyasztási idősoros adatok alapján

(Possibilities of suspicion generating based on data assets about energy uses)

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

Kivonat: Manapság a fogyasztók számára ellenőrző mérések kialakítása úgy technológiai, mint költségoldalról nem irracionális. Az adatok triviális felhasználásának számít a túlszámlázás elleni védekezés lehetősége. Ennél szofisztikáltabb feladatként jelentkezik az irracionális (nem várt) fogyasztások felismerése utólag helyszíni szemlék elrendelésének alapjaként és előrevetítve azonnali hibakeresést katalizálandó.

Kulcsszavak: homogenitás, riasztás, automatizálás, korai vészjelző rendszerek, ingatlanüzemeltetés

Abstract: Nowadays, the control measurements of energy uses are possible for the users both from technological and economic aspects. The most trivial using of data is the exploring of overbillings. More sophisticated approaches are: the derive norm values for energy uses. Based on these norms controlling actions on the spot can be prescribed in advance or belated.

Keywords: homogeneity, alerts, automation, early warning systems, facility management

# Bevezetés

Energiafogyasztási adatokat a kíváncsiságon túl rendszerszinten csak racionális okokból szokás mérni. A mért adatvagyon kialakításának költségei relatíve egyértelműsíthetők: mérőeszközök ára és ezek telepítésének költsége, adatkommunikációs rendszer költségei, adat-vizualizációs megoldások költségei, HR-költségek az adatok elemzése érdekében. A költségek fedezetét jelentő megtérülések forrása sokféle lehet: pl. a túlszámlázások feltárása, ill. túlfogyasztás vélelme esetén a helyszíni szemle keretében rá lehet jönni arra, hogy adott fogyasztó nem került lekapcsolásra okszerűen, ill. egyes készülékek elromlottak, és/vagy áramlopás esete áll fenn. Kevésbé direkt módon a felismert emberi/műszaki hibák elhárítása kihathat a szolgáltatás/üzemelés-biztonság színvonalára, így a partneri/vevői/munkavállalói elégedettségre is. A saját és esetleges publikus kísérleti fogyasztási adatok ismeretében energetikai beruházások előkészítése/tervezése is támogatható az adatvagyon létén keresztül…

# Az adatvagyon jellemzői

Jelen cikk/esettanulmány keretében csak a pl. 5 perces ritmusban pl. az áramfogyasztást mérő műszerek adatai álljanak idősorosan rendelkezésre pl. 1 évre és semmi egyéb adatforrás adatai ne. Ez nagyságrendileg 365 nap (á 24 óra, á 12 db 5 perces intervallum) esetén 105120 adatot jelentenek mérési pontonként. Az adatvagyon kapcsán előállhat adathiány, ami nem feltétlenül feleltethető meg a nulla értékű adatnak, mert lehet, hogy valóban nulla a valós fogyasztás az adott 5 perces időintervallumban. A helyes megoldás az adathiány és a fogyasztáshiány megkülönböztetésére szerencsés esetben a több párhuzamos mérés rendelkezésre állása lehet, ahol kis eséllyel fordulhat elő, hogy egyszerre 3-5 szenzor is egyszerre hibásodjon meg. Az adatkommunikáció azonban közös rendszeren fut, így formálisan a létező adat is elveszhet műszaki hibák következtében. A továbbiakban a nulla értékű adat NEM kerül adattisztítás keretében felülvezérlésre, vagyis első közelítésben nulla fogyasztásként kerül értelmezésre a nulla értékű adatcella.

# Elemzési lehetőségek

Az idősoros adatvagyon elemzése kapcsán alapvető elvárás, hogy minden lépése automatizálható legyen annak érdekében, hogy a szakértői bérköltség minimalizálható legyen és a szakértői tudás esetleges, ad hoc fluktuációjától minél védettebb legyen maga a rendszer.

## Triviális feladatok

Bár az alapadatok között az energia-szolgáltató által kiszámlázott adatok nincsenek jelen, mégis a leginkább triviális adatfelhasználás a számlák rekonstruálása lenne, mely egyértelműen képes kell, hogy legyen a fogyasztóvédelmi vitákat megérő szintű túlszámlázások gyanúját feltárni…

Az idősorok vizuális értelmezése a rendelkezésre álló adatmennyiség kapcsán szinte lehetetlen, lévén a felbontás nem teszi értelmezhetővé a jelsűrűségnek megfelelő részletgazdagságot (vö. 1., 2. és 3. ábra).



1. ábra: Egyedi idősor 5 perces nézete (forrás: saját ábrázolás, ahol az X-tengely az idő, az Y tengely a fogyasztás)



1. ábra: Egyedi idősorok párhuzamos nézete (forrás: saját ábrázolás, ahol az X-tengely az idő, az Y-tengely a fogyasztás, s a színek az egyes telephelyeknek felelnek meg)



1. ábra: Napi aggregálású telephelyi idősorok (forrás: saját ábrázolás, ahol az X-tengely az idő, az Y-tengely a fogyasztás – az adathiányos/nullaértékű fogyasztások pontjai telephely-színkódonként a kép alján láthatók).

## Rendszerhomogenitás-vizsgálat

### Abszolút fogyasztások értelmezése

Amennyiben az 5 perces adatokat például napi átlaggá/összeggé sűrítjük (n=365 sor) és vesszük az azonos műszaki konstrukciónak vélt (fogyasztástípusba sorolt) objektumok (pl. 6 telephely - oszlop) adatait, s kimutatás-varázslás keretében minden egyes adatcella a saját telephely (oszlop) átlagától való eltérést tartalmazza %-ban kifejezve, akkor ezzel a telephelyek kitettsége, egyedisége okozta eltérések standardizálhatóvá válnak.

Ha ezek után vesszük az adott időpontra vonatkozó telephelyi átlagoktól mért egyedi telephelyi átlagok távolságának abszolút értékét, akkor ezzel eljutunk egy olyan nézethez, mely kapcsán feltehető a kérdés: Igaz-e, hogy minden egyes nap másként inhomogén a fogyasztások szerkezetét tekintve?

Egy antidiszkriminációs modellel (vö. <http://miau.gau.hu/myx-free/coco/beker_y0.php>) elemezve a kérdést eljuthatunk a 4. ábrához, mely megadja, hogyan alakult a fogyasztások napi átlaggal való összevetésének homogenitása telephelyenként.



1. ábra: Rendszerhomogenitás-vizsgálat eredményei (forrás: saját számítások, ahol az X-tengely az idő, az Y tengely a rendszer inhomogenitás-indexe, ahol piros szín jelzi a magas és zöld az alacsony inhomogenitást, a sárga szín pedig az átlagos inhomogenitást mutatja)

A 4. ábra bal szélén látható permanens piros hullámzás az egyik telephely hosszabb időszakot felölelő adathiányát jelzi, míg a kép közepe táján látható zöld „gödör” a teljes adatszolgáltatás hiányát, ahol minden telephelynek nulla volt a rendelkezésre álló adata. Emellett megfigyelhető, hogy a 3. ábra bal oldali fele színgazdagabb, pulzálóbb, mint a jobb oldali ábrafél, vagyis valahol az egy éves idősor felénél egy fajta „rendszerváltás” történt.

Ha az idősor kezdete egy adott év június 01-e, míg a vége értelemszerűen a következő év május vége, akkor a közel félidős időpont lehetne éppen egy évfordulóváltás. A szakaszhatár egzakt meghatározásának demo-állománya itt érhető el: <http://miau.gau.hu/miau/200/szakaszolas.doc>

Ha nem naptári jellegű a karakterisztika változásának oka, akkor ez lehet pl. a telephely-használati elvekre (inkl. személyzetcsere) és/vagy szenzorcserére visszavezethető is. Ilyen esetekben a vizsgált telephelyek üzemeltetőivel mindenképpen szakmai konzultációt érdemes folytatni, mert olyan jellegű rendszer-változások esetén, melyek kapcsán az inhomogenitás gyanúja visszaigazolható, nem feltétlenül célszerű a továbbiakban az adatsort egységes idősorként kezelni.

### Fogyasztásváltozások értelmezése

Amennyiben az abszolút fogyasztásváltozások adatait két szomszédos időintervallum (pl. nap) kapcsán egymásból kivonjuk (értelemszerűen a későbbiből a korábbit) és az így kapott adatokat az egyes telephelyek maximális és minimális napi fogyasztásváltozásainak intervallumában relativáljuk és ezen relatív adatok naponkénti átlagától való abszolút távolságot vesszük minden telephelyre vonatkozóan a dinamikus inhomogenitás alapjaként, ismét fel lehet tenni a kérdést: Vajon minden nap (n=365) másként azonos az inhomogenitások telephelyenkénti megoszlását tekintve?

Az inhomogenitás-indexek napi értéke ismét csak egy anti-diszkriminatív modellel állítható elő (vö. 5. ábra):



1. ábra: Dinamikus inhomogenitás alakulása (forrás: saját ábrázolás, ahol az X-tengely az idő, az Y-tengely az inhomogenitás mértéke).

A 5. ábra kapcsán az vélelmezhető, hogy a telephelyek nap-nap változásainak összhangja a középső harmadban karakterisztikusan eltérőbb, mint a kezdeti és a záró intervallumban.

Ha feltételezzük, hogy az idősor kezdete egy adott év június első napja, akkor az impulzívabb szakasz a szeptember-március időszak, vagyis a fűtési/téli/sötétebb szezon. A telephelyek közötti harmónia tehát az intenzívebbnek vélt energia-igény kapcsán kevésbé adott, vagyis a telephelyek egymáshoz való hasonlóságát alapvető karakterisztikák különbségei felerősítik. A valódi okok kiderítése ismét csak az üzemeltetőkkel folytatott, helyszíni bejárást/szemlét feltételező akciók keretében végezhető el.

A rendszer inhomogenitások feltárása egyben az elemzések kockázat-menedzsmentjét is jelenti: minél inhomogénebb egy rendszer-időszak, az ott felmerülő riasztások annál kockázatosabbak, hiszen a helyszíni szemle keretében a riasztás okának megállapítása racionális ok-okozatokat mutathat a helyszíni szemle esetleges költségeivel tovább terhelve az adatvagyon-gazdálkodás profit-centerét.

Természetesen a kockázatok fordítva is értelmezhetők: az alacsonyabb inhomogenitású szakaszokra jutó riasztás (l. következő fejezet) valódi problémaként, azaz megtakarítási esélyként való realizálása nagyobb mértékben várható el.

### Telephely-inhomogenitás

Amennyiben a telephelyek kapcsán pl. a havi átlagos fogyasztások átlagtól való eltéréseit telephelyenként értelmezzük és keressük, lehet-e minden telephely másként egyforma, akkor a telephelycsoport homogenitását kapjuk meg (esetlegesen több lépcsőben) eredményként



1. táblázat: Inhomogenitás-index telephelyenként (forrás: saját ábrázolás, ahol a sárga a csoportátlaghoz legközelebbi objektumokat jelzi a becslésben.)

## Riasztás-generálás

Az adatvagyon-gazdálkodás, elemzés-készítés egyik tipikus outputja a riasztás, vagyis már bekövetkezett anomáliákra való rámutatás:

### Együttmozgások elemzése

Ha a tényleges fogyasztások időegységről időegységre történő változásainak előjeleit minden telephely esetén kiszámítjuk és keressük azon együttállásokat, ahol csak egyetlen egy telephely változik a többivel ellentétes irányba, akkor egyrészt a vizsgált időszakra kumulálva ezen gyanús eltérések telephelyenkénti arányai feltárhatók (vö. 6. ábra)



1. ábra: Adott telephely aszinkronitásának alakulása időben (forrás: saját ábrázolás, ahol az X-tengely az idő, az Y-tengely az aszinkronitás léte = 1)

Ezen arányok (közel) nagyságrendi eltérést is felmutathatnak (vö. 5% vs. 41% - 2. táblázat). Az iránytartás homogenitásának egy telephelyet érintő megsértése riasztási eseményként is értelmezhető. Helyszíni vizsgálat keretében lehet feltárni, vajon mi az oka annak, hogy adott objektum, mely formálisan a vizsgált telephely-csoport legitim tagja, adott pillanatban vagy éppen tartósan a többiekre jellemző nap-nap-változásokkal szembe megy. Itt értelemszerűen a minél tartósabb, annál kockázatosabb elv érvényesül.

A vizsgált időszakban az anonim esettanulmány szerint 16%-nyi a riasztások száma, ill. ezen riasztások 41%-a egyetlen egy telephelyhez kötődik. A riasztások súlyosságát a homogén jellegű szakaszok hossza jelzi, ami a 6. ábrán két olyan szakaszt is kiemelni enged, ahol a riasztás tényei az összes riasztás által érintett napok keretében anélkül ismétlődnek, hogy máshol riasztani kellett volna… (Ideális riasztási képlet szerint ezen napok valóban szomszédos napok illik, hogy legyenek – vö. piros kerettel jelölt 2 szakasz.)



1. táblázat: Iránytartási kockázatok arányszámai (forrás: saját ábrázolás, ahol a vizsgált időszak egy év, azaz 365 nap)

### Hazudj-ha-tudsz (lie2me) elemzések

Amennyiben a rendelkezésre álló idősorok valamilyen aggregációját (jelen esetben havi összevonását) úm. eltörjük pl. a 12 elemű idősorok kapcsán a 7. elemek után, akkor a 7+1 elemű ok-okozat-szimulációs mintázat minden telephelyre 5-5 ismétlést eredményez. Vagyis 30 objektum keletkeztethető, melyek a mindenkori utolsó 7-7 hónap (t-6, …, t) alapján igyekeznek megmagyarázni, levezetni a következő (t+1) hónapra várható fogyasztás (jelen esetben az átlagtól való eltérés) irányát és mértékét. Vagyis itt már egy klasszikus termelési függvény-becslésről van szó, ahol hét bemeneti jel alapján kell a nyolcadik értéket minél közelebbi becsléssel értelmezni.

A készített modell egy klasszikus lépcsős függvény, ahol az összes Xi esetére igaz, annál nagyobb lesz a fogyasztás átlagtól való eltérése a jövőben, minél nagyobb volt a múltban. A modell tény-becslés közötti korrelációs értéke: R = 0.996. A 7. ábra szerint egyetlen egy olyan hónap volt a 2. telephelyen, ahol 3%-ot meghaladó mértékű volt a becslés és a tény eltérése nem kívánatos irányba, azaz min. 3%-kal illett volna kevesebbet fogyasztani ott és akkor, mint az ténylegesen megtörtént. Ez már ok lehet utólagos riasztás kiadására és helyszíni szemle elrendelésére. A 7. ábra X-tengely alatti jelei azt mutatják, hogy értelemszerűen ahhoz, hogy túlfogyasztásról tudjunk egy kiegyensúlyozott modell esetén beszélni, alulfogyasztást is vélelmezni kell tudni. Ezen alulfogyasztások értelmezése is érdekes lehet, amennyiben olyan best-practice-jellegű gyakorlatokra mutatnak rá, melyek rendszerszintűvé tehetők.



1. ábra: Utólagos telephelyszintű riasztás (forrás: saját ábrázolás, ahol az X-tengely az telephelyek és időszakok által kiadott objektumok véletlenszerű sorozata, ezen objektumok ID-ja, az Y-tengely a riasztás mértéke, vagyis a tény és a becslés közötti eltérés %-os értéke a mindenkori tényértékre vetítve)

Az itt leírtakhoz hasonló számítások nem csak havi, hanem akár heti, napi, órás, vagy negyedórás ritmusban is kialakíthatók. Minél rövidebb az aggregációs szint, annál operatívabb lehet a gyanú kezelésének módja, ami végső soron átcsap az előrejelzésszerű norma-értelmezésbe:

## Normaszámítás/előrejelzés

Az utólagos tény-becslés értelmezések egyben normaszámításként is felfogható és az erre képes modellek egyben szakértői rendszerként, szimulátorként is felfoghatók. Az előző fejezet nagyon magas korrelációjú modellje keletkezése pillanatában minden telephelyre képes azonnal egy ténnyel még le nem fedett becslést adni, vagyis egy fajta előrejelzésként deklarálni, hol mekkora fogyasztás várható az utolsó 7 hónap már ismert adatai alapján. Ennek a norma-becslésnek az értéke napról napra (időről-időre) elsődlegesen arányosan összevethető az idő múlása során keletkező adatokkal telephelyenként, s a legnagyobb arányos eltérést produkáló telephely kapcsán az okok feltárása operatív feladatként is értelmezhető riasztási szintet jelenthet a kockázatmenedzsment kapcsán…

# Következtetések

Mint az a fentiek alapján világosan látható: a riasztások matematikai levezetése számos módon lehetséges. A riasztások közgazdasági/üzleti értékét csak a ténylegesen érintett objektumok szakértői értelmezése alapján lehet megítélni, hiszen akár az is fennállhat, hogy valódi mérések helyett szinte véletlenszerű számokat szállít az adatgyűjtés, mely riasztásai semmilyen valós műszaki eseményhez nem köthetők szisztematikusan.

Amennyiben az adatvagyon reális, akkor a benchmarking jellegű hasonlóságelemzéseknek illik az egymással összevetett objektumok alapján racionális okokkal magyarázható riasztási eseményeket feltárni tudni.

Amennyiben egy ilyen rendszer automatizálásra kerül, úgy az elemzői munkaidő formálisan kiesik, a riasztások valósidejűvé válnak, már „csak” a riasztási szintek és a kockázatkezelés vállalat-specifikus arányosságát kell ad hoc döntésekkel beállítani, vagy további modellekkel optimalizálni:

# Jövőkép

Az idősoros fogyasztási adatok analógiájára bárminemű rendszer naturális és pénzügyi adatai mellé felsorakoztatható a következményadatok tömege, vagyis a potenciális partnerek (piaci szereplők, fogyasztók, cégen belüli érintettek, stb.) reakcióinak (pl. ügyfélvesztés, piaci részarány-vesztés, munkaerő-fluktuáció, hatósági büntetések, stb.) ismét csak idősoros sokasága.

Ahhoz, hogy arról lehessen beszélni, milyen műszaki/működtetési beavatkozások érhetik meg a vállalkozás számára a beavatkozás költségeit olyan szimulátorok felállítása szükséges, melyek a potenciális partneri reakciók és a rendszerparaméterek között összefüggéseket képesek feltárni. Ezen szimulátorok alapján vélelmezhető, milyen rendszerparaméter-változás milyen piaci/elégedettségi hatásokat katalizálhat, s mennyi ezek pénzben is kifejezhető értéke. Így már a műszaki beavatkozások várható megtérülésének kérdése a mérleg két serpenyőjébe került becslés összevetéséből levezethető.

A szimulátorok felállításának demonstrálására az mellékelt táblázatkalkulációs állomány szolgál: <http://miau.gau.hu/miau/214/operation_marketing_mixed_model.xlsx>

\*\*\*

Tartalomjegyzék

[Bevezetés 1](#_Toc454022297)

[Az adatvagyon jellemzői 1](#_Toc454022298)

[Elemzési lehetőségek 2](#_Toc454022299)

[Triviális feladatok 2](#_Toc454022300)

[Rendszerhomogenitás-vizsgálat 3](#_Toc454022301)

[Abszolút fogyasztások értelmezése 3](#_Toc454022302)

[Fogyasztásváltozások értelmezése 4](#_Toc454022303)

[Telephely-inhomogenitás 4](#_Toc454022304)

[Riasztás-generálás 5](#_Toc454022305)

[Együttmozgások elemzése 5](#_Toc454022306)

[Hazudj-ha-tudsz (lie2me) elemzések 6](#_Toc454022307)

[Normaszámítás/előrejelzés 6](#_Toc454022308)

[Következtetések 7](#_Toc454022309)

[Jövőkép 7](#_Toc454022310)