KARSZTFEJLŐDÉS XVI.
Szombathely, 2011. pp. 21-29.

GEOMATEMATIKAI VIZSGÁLATOK A BKÉR ADATSORAIN ACE ALGORITMUS ALKALMAZÁSÁVAL

NÉMETH ÁGNES

Miskolci Egyetem, Hidrogeológiai - Mérnökgeológiai Intézeti Tanszék
3515, Miskolc, Egyetemváros, nemeth.agnes85@gmail.com

Abstract: This paper introduces the alternating conditional expectation (ACE) algorithm for estimatingthe transformations of a response and a set of predictor variables in multiple regression problems in hydrogeology. The proposed nonparametric approach can be applied easily for estimating the optimal transformations of different hydrogeological data to obtain maximum correlatio between observed variables. The approach does not require a priori assumptions of a functional form and the optimal transformations are derived solely based on the data set. The advantages and applicability of this new approach to solve different multiple regression problems in hydrogeology or in Earth sciences are illustrated.

Bevezetés

A föltudományok területén igen gyakran előfordul, hogy nyers mérési adatainkból szeretnénk a legtöbb hasznosítható információt kinyerni, és idősoraink között minél egyértelműbb, matematikailag is megfogalmazható összefüggéseket találni (*NÉMETH 2008, DARABOS 2008*). Az egyik ilyen alkalmazott megoldások közé tartoznak a regressziós vizsgálatok, melyek a változók közötti valamilyen függvénykapcsolat feltételezésén alapulnak.

A karsztos területek kutatása során egyre inkább előtérbe kerülnek a fizikai és hidrokémiai paraméterek segítségével történő vizsgálatok. Ezek a viszonylag könnyen mérhető paraméterek betekintést engednek a karsztos vízbázisokba. Jelen tanulmányunk a Bükki Karsztvízszint Észlelő Rendszer (BKÉR) vízszint-idősorait vizsgálja. A rendszer keretein belül az első műszerek telepítése 1992-ben történt. A méréseket a bükki karsztvizet kitermelő vízmű Rt.-k kezdeményezték, a miskolci és Miskolc környéki lakosok megfelelő mennyiségű ivóvizének biztosítása érdekében. Ennek egyik kiváltó oka az volt, hogy a korábbi években gyakran veszélybe került a terület vízellátása. Az elmúlt 47 évben tartósan is előfordult rendkívüli aszály és rendkívüli csapadékbőség. A rendszer alapjait a Böcker Tivadar által 1983-ban tervezett és kivitelezett kutak képezték. Ezen mennyiségi monitoring keretein belül jelenleg 34 helyen, kutakban, megfigyelő kutakban és forrásokban történik folyamatos vízszint, vízhőmérséklet, ill. részben elektromos vezetőképesség mérés (*LÉNÁRT* 2007). A rendszer adatsoraiból kiválasztot-

tuk a 3 legjellemzőbb megfigyelőkutat, és ezen vízszintadatsorok közti összefüggéseket próbáljuk megtalálni. Az eredményektől függően azt szeretnénk meghatározni, hogy az összefüggések alkalmasak-e a karsztos területek modellezésére, jellemző paramétereinek meghatározására, pontosítására.

Mindezek mellett tudni kell, hogy a Miskolci karsztvízbázis jelenleg az ország egyik legnagyobb területileg is összefüggő sérülékeny vízbázisa. Mint tudjuk a karsztosodott víztartókra kettős áramlási rendszerjellemző, a karszt érzékenységét legnagyobb mértékben a kőzet hidrogeológiai jellemzői határozzák meg, jelen esetben a felszín alatti vízrendszer a kőzettest törésrendszeréhez kapcsolódik, ezért például a terület modellezése esetén a kezdeti – biztosnak hitt – paraméterek egy felszín alatti nagyobb repedés miatt könnyen megdőlhetnek (*MADARÁSZ 2005*).

Módszer

Az idősorok elemzésére 3 különböző regressziós vizsgálatot végeztünk el. A regressziók, más néven kiegyenlítések során a mért adatainkat (y_i) egy számított adatsorral (y_i^{cal}) próbáljuk közelíteni. A számított adatsort pedig egy függvénykapcsolat segítségével állítjuk elő. A számítás során a függvénykapcsolat paramétereit határozzuk meg, melyből megkapjuk az egyes idősorok közötti korrelációs tényező értékét, illetve a mért és számított értékek közötti ún. RMSE („roor mean square error”) hibajellemző értékét.

A vizsgált megfigyelőhelyek a következők:

- Garadna-forrás
- Szinva-forrás
- Nv-17 megfigyelőkút

Az elemzés alapjául mindhárom monitoringhely esetében ugyanazon 2 éves időszak (2003.09.01-2005.08.31.) közötti intervallumot tekintettük, a mintavételezés mindhárom esetben napi gyakoriságú (Δt=1nap) volt, a minták száma pedig (N) háromszor 732 db, és ezekre alkalmaztuk a 3 különböző regressziós eljárást:

- legkisebb négyzetek elvére épülő többváltozós lineáris regresszió
- leggyakoribb érték (MFV) elvére épülő regresszió
- ACE („Alternating Conditional Expectation”) algoritmus

A legkisebb négyzetek elvére épülő többváltozós lineáris regresszió esetében a vizsgált függő változó értékét több független változó lineáris kombinációjának segítségével közelítjük. Ugyanezen metodikára épül az MFV módszer is, azonban előnye fentebb említett regresszióhoz képest, hogy nem érzékeny a kieső adatokra, és az adatok eloszlás típusának

változására sem. Az ACE algoritmus ezekkel szemben viszont egy olyan nem-paraméteres regressziós eljárás, mely olyan transzformációt alkalmaz az egyes vizsgált változók tekintetében, hogy a vizsgálatba bevont változók között a lehető legjobb kiegyenlítést érjük el, úgy, hogy nem szükséges semmilyen „priori" kapcsolat az egyes komponensek között, a létrehozott függvénytranszformációk csak a mérési adatainktól függenek (*SZŰCS et al. 2006, HORNE-SZŰCS 2007*).

Korrelációs számítással kiválasztottuk a 3 adatsor közül a legjobban korreláló megfigyelőhelyeket, így függő változóként a Garadna-forrást (y) állapítottuk meg, első független változóként a Szinva-forrást (x_1), második független változóként pedig az Nv-17 (x_2) megfigyelőhelyet. Az *1. ábra* ezen mérőhelyek egymáshoz képesti viszonyát mutatja be.

Eredmények

A három eljárás elvégzése után az alkalmazott regressziós vizsgálatok megbízhatóságának a jellemezésére bevezethetjük az alábbi gyakran alkalmazott kifejezést a mért és számított vízszintek különbségének a jellemzésére (*SZŰCS-HORNE 2009*). Az RMSE (root mean square error) az alábbi kifejezéssel adható meg:

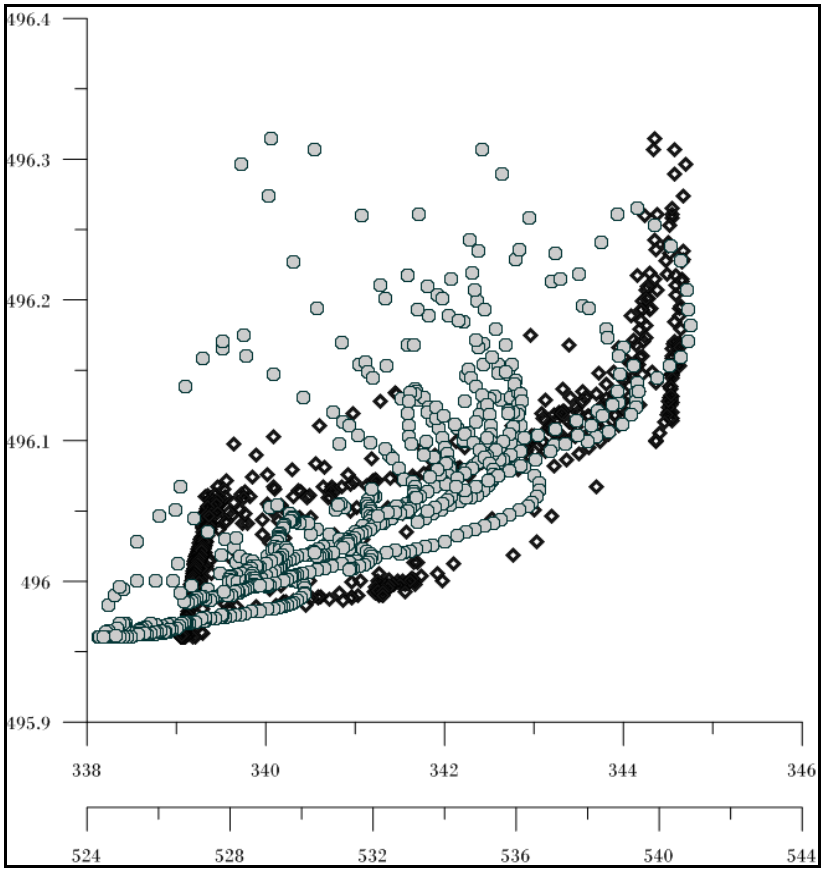
$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - y_j^{cal})^2}$$

Az *I. táblázat* a három vizsgálat által kapott eredményeket mutatja be. Ez alapján megállapítható, hogy a 3 számítás közül a legkisebb különbséget a mért és a számított értékek között az ACE algoritmus szolgáltatta, a továbbiakban ezen módszer alkalmazását mutatjuk be.

I. táblázat
Table I.

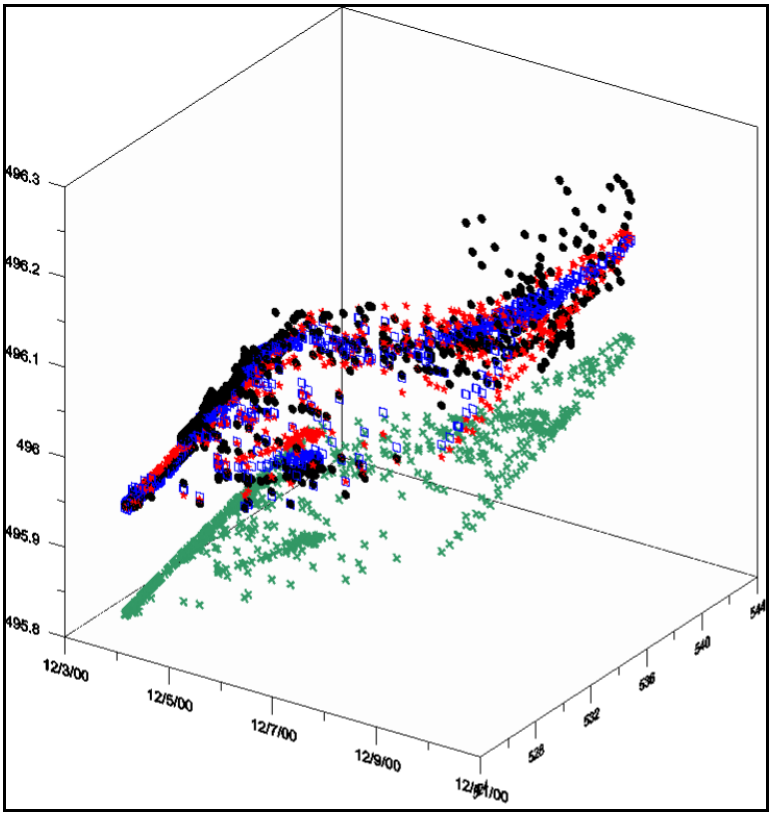
A mért és számított adatok közötti szórások különböző regregssziók alkalmazásával
The mesaured and the calculated data-series

Módszer	Szigma
Regresszió	0,031000
MFV	0,032247
ACE	**0,019636**



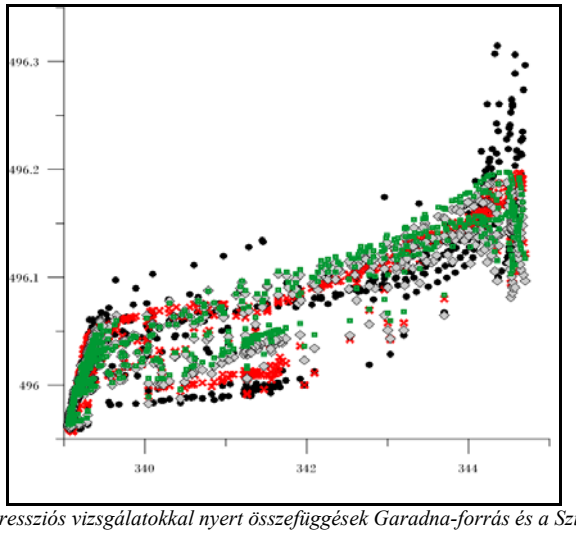
1. ábra: A vizsgált adatsorok Garadna-forrással való kapcsolata (kör: Nv-17, négyszög: Szinva-forrás, függőleges tengely: Garadna-forrás karsztvízszintjei (mBf), vízszintes tengely: Szinva forrás, Nv-17 karsztvízszintjei (mBf))
Fig.1.: Relationship between the Garadna-gauge and the examined data-series (round: Nv-17, square:: Szinva-gauge, x axle: Szinva-gauge, and Nv-17 waterlevel (mBf), y axle: Garadna-gauge waterlevel (mBf))

A *2. ábra* a három eljárással kapott eredményeket szemlélteti a számítások elvégzése után a mért értékekhez képest. A legkisebb egyezést a hagyományos regresszió által szolgáltatott adatok adják, a második legjobb egyezést a leggyakoribb érték elvén működő regresszió szolgáltatja, a legjobb korrelációt pedig az ACE algoritmussal nyertük.

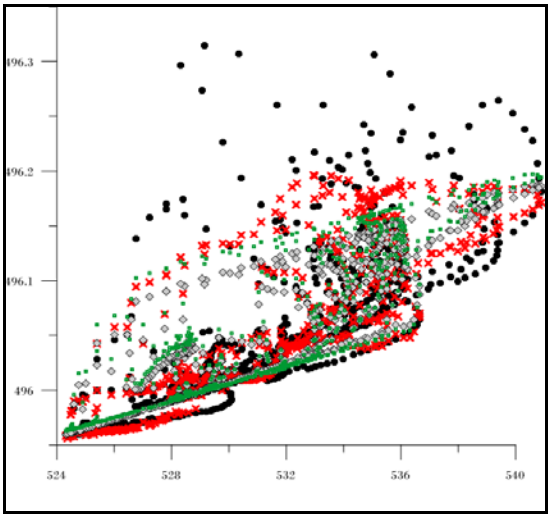


2. ábra: Különböző regresszióval elvégzett számítások a vizsgált adatsorokon
Jelmagyarázat: kereszt:hagyományos regresszió, csillag: MFV módszer, kör: mért adatsor, négyszög: ACE algoritmus, x tengely: Szinva-forrás vízszintjei (mBf), y tengely: Nv-17 vízszintadatai (mBf), z tengely: Garadna-forrás vízszintjei (mBf))
Fig.2.: Analysis of the examined data-serieswith using diffferent regression methods
Legend: cross: conventional regression method, star:MFV method, round: measured data-series, square: ACE algorithm, x axle: Szinva-gauge waterlevel (mBf), y axle: Nv-17 waterlevel (mB), z axle: Garadna-gauge waterlevel (mBf))

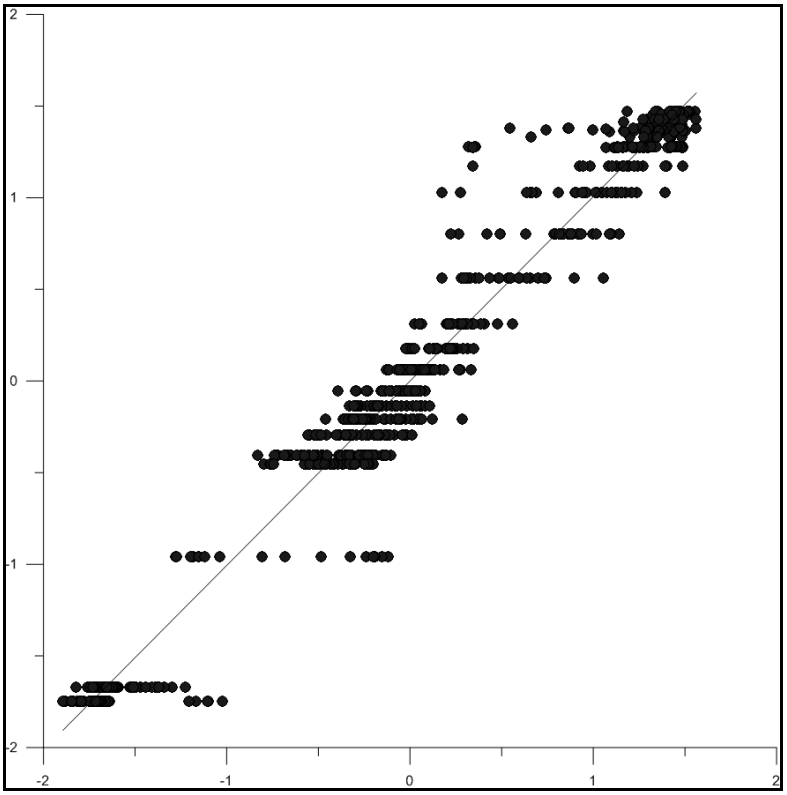
A *3. és 4. ábrákon* szintén ugyanezek az eredmények láthatók, csak 2 dimenzióban ábrázolva, hogy jobban kivehetők legyenek az egyes mérőhelyek közötti összefüggések. A *3. ábra* a Garadna-forrás és a Szinva forrás összefüggéseit mutatja be a különböző regressziós vizsgálatok elvégzése után. A 4. ábra pedig a Garadna-forrás Nv-17-tel való viszonyait szemlélteti.



3. ábra: Különböző regressziós vizsgálatokkal nyert összefüggések Garadna-forrás és a Szinva-forrás között
Jelmagyarázat: kereszt:hagyományos regresszió, csillag: MFV módszer, kör: mért adatsor, négyszög: ACE algoritmus, x tengely: Garadna-forrásvízszintjei (mBf), y tengely: Szinva-forrás vízszintjei (mBf)
Fig.3.: Analysis of the examined data-series with using diffferent regression methods between the Garadna and the Szinva-gauge
Legend: cross: conventional regression method, star:MFV method, round: measured data-series, square: ACE algorithm, x axle: Garadna-gauge waterlevel (mBf), y axel: Szinva-gaugewaterlevel (mBf)



4. ábra: Különböző regressziós vizsgálatokkal nyert összefüggések Garadna-forrás és a Szinva-forrás között
Jelmagyarázat: kereszt:hagyományos regresszió, csillag: MFV módszer, kör: mért adatsor, négyszög: ACE algoritmus, x tengely: Garadna-forrásvízszintjei (mBf), y tengely: Nv-17 vízszintjei (mBf)
Fig.4.: Analysis of the examined data-series with using diffferent regression methods between the Garadna and the Szinva-gauge
Legend: cross: conventional regression method, star:MFV method, round: measured data-series, square: ACE algorithm, x access: Garadna-gauge watrelevel (mBf), y tengely: Szinva-gauge (mBf), x axle: Garadna-gauge waterlevel (mBf), y axel: Nv-17 waterlevel (mBf)))



5. ábra:A transzformált függő változó értékei a 2 független változó transzformáltjainak összege függvényében
Fig. 5.: Relationship between the transformed dependent and the sum transformed independent variable

Az *5. ábrán* a transzformált függő változó értékeit tekinthetjük meg (függőleges tengely) a 2 független változó transzformáltjainak összege (vízszintes tengely) függvényében. Az ACE transzformált térben a lineáris regresszió a következő eredményre vezetett:

$$\theta\ (y_j) = 1.00622\left[\phi_1\ (x_{1j}) + \phi_2\,(x_{2j})\right]$$

elég jó korrelációs értékkel (R^2= 0.9611), mely sokkal magasabb, mint amelyeket a hagyományos többváltozós eljárások során kaptunk.

Következtetések

Jelen tanulmány az ACE nem paraméteres algoritmus adaptációját mutatja be hidrogeológiai többváltozós regressziós vizsgálatokhoz, ahol a vizsgált változók közötti kapcsolat a priori nem ismert. Az ACE algoritmus egyik

nagy előnye, hogy a legkülönbözőbb típusú és nagyságú adatok együtt kezelhetőek. Az ACE által szolgáltatott transzformáltak részletes vizsgálata új következtetésekre vezethet a vizsgált függő és független változók közötti kapcsolatok feltárásában (pl. ekvivalencia hatás, érzékenység vizsgálat, stb.). Természetesen az ACE algoritmusnak is megvannak a maga korlátai számtalan előnyei mellett. Bizonyos esetekben az eljárás különböző eredményekre vezethet, ha megcseréljük a független változó sorrendjét. Másrészt az ACE algoritmus extrém kieső adatokra nagyon érzékenyen reagál. Termesztésen itt is ki kell hangsúlyozni, hogy az ACE algoritmus csak akkor lehet tényleg hatékony, modern statisztikai eljárás, ha a vizsgált változók között létezik tényleges fizikai, vagy egyéb természettudományos kapcsolat, és a mérési adataink minőségellenőrzöttek.

Köszönetnyilvánítás

A cikkben ismertetett kutatómunka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként - az Új-Magyarország Fejlesztési Terv keretében – az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

IRODALOM

DARABOS E. (2008): A Bükki Karsztvízszint Észlelő Rendszer által szolgáltatott adatok kapcsolatainak vizsgálata - OTDK dolgozat, Szombathely, 2009., p. 34-36.

LÉNÁRT L. (2007): A bükkben keletkezett kitermelhető karsztvízkészlet folyamatos meghatározásának módszere, XVI - Az 1992.10.10. – 2007.01.01 közötti mérések értékelése, 2007. február, Miskolc, p. 35-43

NÉMETH Á. (2008): A Bükki Karsztvízszint Észlelő Rendszer hiányzó mérési adatainak pótlása, a pontosított adatsorok alapján előrejelzések kidolgozása - diplomamunka, Miskolci Egyetem, Környezetgazdálkodási Intézet, Hidrogeológiai - Mérnökgeológiai Intézeti Tanszék , p. 71-78.

SZUCS, P.–CIVAN, F.–VIRAG, M. (2006): Applicability of the most frequent value method in groundwater modeling. - Hydrogeology Journal (2006), 14: p. 31-43. Springer-Verlag, DOI 10.1007/s10040-004-0426-1

HORNE, R. N.-SZUCS, P. (2007): Inferring Well-toWell Connectivity Using Nonparametric Regression on Well Histories. - PROCEEDINGS, Thirty-Second Workshop on Geothermal Engineering, Stanford University, Stanford, California, January 22-24, 2007, SPG-TR-183, p. 1-8.

SZUCS P. - HORNE R. N. (2009): Applicability of the ACE Algorithm for Multiple Regression in Hydrogeology. - DOI: 10.1007/s10596-008-9112-z

COMPUTATIONAL GEOSCIENCES : (13) p. 123-134 (2009). Springer
WOLFBAUER J.-STIBITZ M.-MADARÁSZ T.-SZABO I. (2005): Quality assurance in field remediation - Project Num.: Hungarian-Austrian S&T Cooperation Proj. A-9/2002 Period 2003-2004 (WTZ Ungarn/ÖAD); p. 1 – 25, 2005; Leoben;