

AZ ACE ALGORITMUS ALKALMAZÁSA KARSZTVÍZ-SZINT ADATOK HIDROGEOLÓGIA VIZSGÁLATÁRA

Szűcs Péter¹, Darabos Enikő², Németh Ágnes²

¹tszv egyetemi tanár, ²PhD hallgató

^{1,2}Miskolci Egyetem, Környezetgazdálkodási Intézet, Hidrogeológiai-Mérnökgeológiai Intézeti Tanszék, 3515, Miskolc-Egyetemváros, hgszucs@uni-miskolc.hu

Összefoglalás

A hidrogeológia vizsgálatokban, illetve a modellezési feladatok előkészítésében rendkívül nagy a szerepe a regressziós vagy kiegyenlítő eljárásoknak. A különböző típusú vizsgálatok során világossá vált, hogy a regressziós eljárások esetében szükség van új módszerek kidolgozására is, ahol a vizsgált földtani és vízföldtani változók közötti természeti kapcsolatok kvantitatív leírása nem találgatáson, illetve szubjektív szakmai megérzésen alapul. Végrehajtottuk a Stanford University intézményben Breiman és Friedman (1985) által kidolgozott ACE nem paraméteres algoritmus adaptációját hidrogeológiai többváltozós regressziós vizsgálatokhoz, ahol a vizsgált változók közötti kapcsolatot a priori nem ismert.

Kulcsszavak: ACE algoritmus, regressziós vizsgálat, hidrogeológiai monitoring

Abstract

This paper introduces the alternating conditional expectation (ACE) algorithm for estimating the 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 correlation 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.

Keywords: ACE algorithm, regression analysis, hydrogeological monitoring

1. Regressziós módszerek hidrogeológiai vizsgálatokban

A Breiman and Friedman [1] által kidolgozott ACE („Alternating Conditional Expectation”) algoritmus adaptációja, módosítása után lehetőségünk nyílt különböző típusú hidrogeológiai és vízbányászati többváltozós regressziós problémák megoldására. Ez a nagy hatékonyságú és magas határfokú nem-paraméteres regressziós vagy kiegyenlítő eljárás könnyen alkalmazható a vizsgált változók elemzésére [2]. Az ACE algoritmus a különböző hidrogeológiai modellváltozók olyan optimális transzformációját hajtja végre, ahol maximális korrelációra számíthatunk a transzformált függő változó és a transzformált független változók összegeként

előálló becslés között. Ez a tulajdonság nagyon kedvező, hiszen a földtudományok területén igen gyakran alkalmazott tradicionális regressziós vizsgálataink mindig valamilyen változók közötti függvénykapcsolat feltételezésével indul [3]. A vizsgált paraméterek között fennálló komplex, és sokszor jósolhatatlan jellegű kapcsolatok miatt sokszor igen nehéz a megfelelő típusú függvénykapcsolatot megadni a függő és független változók esetében [4]. A hidrogeológiai paraméterek értéktartományának nagy változékonysága esetében például a rutinszerűen alkalmazott hagyományos többváltozós regressziós eljárások gyakran nem reális eredményeket produkálnak [5].

2. Az ACE algoritmus elméleti háttere

A vizsgált változók nem-lineáris transzformációja bevett gyakorlatnak tekinthető a különböző típusú regressziós problémák megoldása során [6]. Az ACE algoritmus olyan transzformációt alkalmaz az egyes vizsgált változók tekintetében, hogy a lehető legjobb kiegyenlítést érjük el az analízisbe bevont változók között. Az ACE algoritmus matematikai alapjait Breiman and Friedman [1] dolgozta ki a Stanford Egyetemen. Az eljárás elméleti háttere az alábbiakban megismerhető. További részletek az eljárással kapcsolatban megtalálhatók Breiman and Friedman [1] eredeti munkájában.

A regresszió hibája a következőképpen írható fel:

$$e^2(\theta, \phi_1, \phi_2, \dots, \phi_p) = E \left(\left[\theta(Y) - \sum_{i=1}^p \phi_i(X_i) \right]^2 \right). \quad (1)$$

A $\phi_i(X_i), \dots, \phi_p(X_p)$ és $\theta(Y)$ transzformáltakra vonatkozó e^2 hiba minimalizációt egy speciális, egy függvényre vonatkozó minimalizációs sorozaton keresztül érhetjük el az alábbi két egyenlet alkalmazásával.

$$\phi_i(X_i) = E \left[\theta(Y) - \sum_{j \neq i}^p \phi_j(X_j) \mid X_i \right] \quad (2)$$

$$\theta(Y) = E \left[\sum_{i=1}^p \phi_i(X_i) \mid Y \right] / \left\| E \left[\sum_{i=1}^p \phi_i(X_i) \mid Y \right] \right\|. \quad (3)$$

A (2) és (3) egyenletekben ún. feltételes elvárásokat megvalósító matematikai operátorok is szerepelnek az iterációs minimalizálási procedúra során. A minimalizációs iterációs eljárás végeredményeként kapott végső $\phi_1(X_1), \phi_2(X_2), \dots, \phi_p(X_p)$ és $\theta(Y)$ függvénytranszformáltak becslései az optimális,

legjobb regressziót biztosító $\phi_1^*(X_1)$, $\phi_2^*(X_2)$, ..., $\phi_p^*(X_p)$ és $\theta^*(Y)$ transzformáltaknak. Vagyis a transzformált paraméterek terében a függő és független változók közötti kapcsolat a következő egyszerű alakot veszi fel:

$$\theta^*(Y) = \sum_{i=1}^p \phi_i^*(X_i) + e^* \quad (4)$$

ahol e^* az ACE regressziós közelítés (zérus helyparaméterű eloszlással jellemezhető) hibáját fejezi ki. Az ACE eljárással elérhető minimális regressziós hiba tehát e^* , míg a többváltozós korrelációs koefficiens ρ^* , és értéke a regresszió hibájával a következő kapcsolatban áll: $e^{*2} = 1 - \rho^{*2}$.

$$Y^{pre} = \theta^{*-1} \left[\sum_{i=1}^p \phi_i^*(X_i) \right] \quad (5)$$

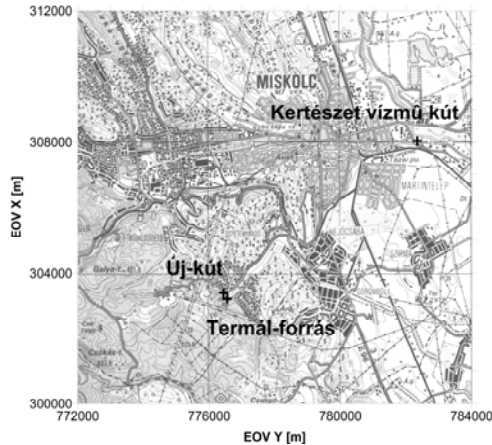
3. Esettanulmány: bükki karsztvíz monitoring adatok regressziós vizsgálata

Az ország egyik legnagyobb vidéki városa, Miskolc mintegy 170000 lakossal a Bükk keleti lábánál fekszik. A város, illetve a környék vízellátásában igen fontos szerepet játszik a bükki karsztvíz. A hideg karsztvíz mellett a hegység peremén megtalálhatjuk a meleg karsztvíz feltöréseket is. Miskolc-Tapolcán például a híres Barlangfürdő és városi vízmű egyik üzeme esetében 100 méteres távolságon belül található a nagy hozamú hideg és meleg vizes karsztforrások. A hévíz regionális áramlási pályák mentén egészen mélyről tör fel a Bükk déli és keleti peremein. A Barlangfürdő meleg vízzel történő ellátását a Termál-forrásra épült kút biztosítja. A Barlangfürdő szomszédságában található vízműtelepen működik ún. Új-kút, amely normál üzemben kb. 30000 m³/nap ivóvíz minőségű hideg karsztvizet termel. A Termál-kút látja el a Barlangfürdőt kb. 2700 m³/nap vízhozammal. Nyilvánvaló, hogy Miskolc-Tapolcán a Bükk lábánál egymás közelébe kerültek a hideg és meleg vizes áramlási rendszerek, ahol egyensúly szempontjából egy igen érzékeny hidraulikai rendszer jött létre [7].

Az említett kutak Miskolc EOV koordinátákkal rendelkező térképlapján elhelyezve a *1. ábrán* láthatóak. A térképen az említett két kút mellett látható még a Kertészeti vízmű kút is, amely szintén szerepelt a későbbiekben ismertetett komplex regressziós vizsgálatban. Az utóbbi időben sajnos egyre gyakrabban fordul elő a Termál-forrás vízének hőmérséklet ingadozása és csökkenése, amely a Barlangfürdő hosszú távú üzemelését komolyan befolyásolhatja [8]. A vízszintek mellett

például a hozam és hőmérséklet viszonyok is mérésre kerülnek. A Barlang-fürdőben várható vízszint és hőmérséklet viszonyok előrejelezhetősége érdekében komplex hidrogeológiai többváltozós regressziós vizsgálatot hajtottunk végre a Termál-forrás, az Új-kút és a Kertészeti-kút monitoring adatainak felhasználásával.

A Miskolc keleti szélén található Kertészeti-kút kb. 8 km távolságban található észak-kelet irányban a tapolcai kutaktól. A mintegy 460 méter mélységben szűrőzött Kertészeti-kút a város azon a részén már több száz méter üledék összlettel fedett karszt rendszerből termel hévizet. A korábbi vizsgálatok már bizonyították, hogy a Kertészeti-kút is a Bükk felől kapja az utánpótlódását regionális nagymélységű áramvonalpályák mentén.



1. ábra. A komplex hidrogeológiai vizsgálatba bevont monitoring kutak elhelyezkedése Miskolcon.

Először a Termál-forrás vízszint adatait hasonlítottuk össze a közelben lévő Új-kút, és a jóval távolabb lévő (bár a hidraulikai kapcsolat meglétét szintén feltételező) Kertészeti-kút vízszint adataival.

Az egyes kutakban mért adatok közötti kapcsolat tisztán felismerhető, de az is világos, hogy az alkalmazott egyszerű, két kút vízszint adataira épülő lineáris kiegyenlítés nem vezet kielégítő eredményre, amelyet a kiegyenlítések eredményeül adódó regressziós együtthatók is megerősítenek [9].

A továbbiakban a vízszint adatok mellett egyéb típusú adatokat (pl. vízhőmérsékleti és hozam adatok a Termál-forrásból) is felhasználtunk annak érdekében, hogy a Termál-forrás vízszint adatait pontosabban tudjuk közelíteni egy többváltozós lineáris regressziós összefüggés segítségével.

Ezért a következő lépésként az ACE algoritmust alkalmaztuk ugyanannak az adatrendszernek a vizsgálatára, hogy növelni tudjuk a becslések pontosságát és megbízhatóságát. Ebben az esetben is a Termál-forrás tengerszint feletti magasságban mért vízszint adatait tekintettük az y függőváltozónak. Az ACE algoritmus alkalmazásával ki tudtuk választani azt a négy független változót, amelyek segítsé-

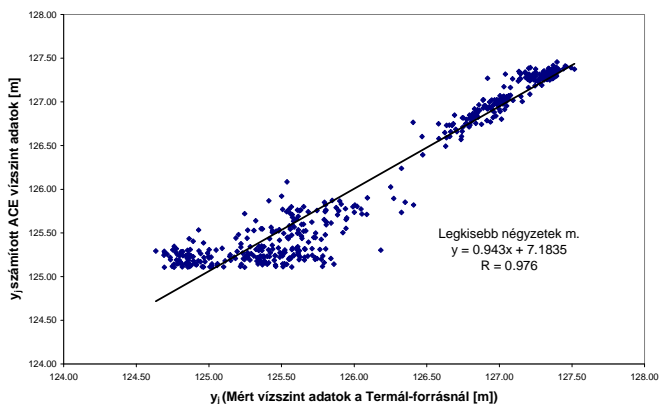
gével a legjobban közelíthető volt az y függő változó értéke: x_1 az Új-kút vízszint adatai [m], x_2 a Termál-forrás vízhőmérsékleti adatai [Celsius], x_3 a Termál-forrás napi termelési hozam adatai [m^3/nap] és x_4 a Kertészeti-kút vízszint adatai [m]. Az összehasonlítás kedvéért ugyanezekkel a változókkal (y és x_1, x_2, x_3 , valamint x_4) végrehajtottuk a hagyományos többváltozós lineáris kiegyenlítést is mind a legkisebb (L_2 -norma), mind pedig a leggyakoribb érték (P-norma) módszerével. A kapott eredményeket az alábbiakban tárgyaljuk részletesen.

A Termál-forrás esetében a 2. ábra mutatja be a mért és az ACE algoritmus alapján számított vízszint adatok összehasonlítását. Az összetartozó pontpárok elhelyezkedése és a korrelációs tényező értéke alapján megállapíthatjuk, hogy az ACE algoritmus segítségével hatékony regressziós kapcsolat állítható fel a vizsgált változók között.

Az alkalmazott regresszió vizsgálatok megbízhatóságának a jellemzésére bevezethetjük az alábbi gyakran alkalmazott kifejezést a mért és a számított vízszintek különbségének a jellemzésére. Az RMSE („root mean square error”) hiba-jellemző a következő kifejezéssel adható meg.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - y_j^{cal})^2} . \quad (6)$$

Az ACE algoritmus alkalmazása esetében a Termál-forrás mért és számított vízszintjei esetében az RMSE értéke 0.2311 méter volt. A vizsgált változók (y és x_1, x_2, x_3 és x_4) regressziós analízisét természetesen elvégeztük a hagyományos többváltozós regresszió alkalmazásával is mind a legkisebb négyzetek módszere, mind pedig a leggyakoribb értékek módszere felhasználásával. A legkisebb négyzetes többváltozós lineáris regresszió esetében az RMSE értéke 0.4052 méter volt, míg az MFV regresszió esetében az RMSE nagyságára 0.3826 méter adódott.



2. ábra A mért és az ACE algoritmus alapján számított vízszint adatok összehasonlítása a Termál-forrás esetében.

A további vizsgálatokban a Bükki Karsztvízszint Észlelő Rendszer (BKÉR) egyéb kútjainak vízszint-idősorait is vizsgáltuk. A rendszer keretein belül az első műszerek telepítése 1992-ben történt. A rendszer adatsoraiból kiválasztottuk a 3 legjellemzőbb megfigyelőkutat, és ezen vízszintadatsorok közti összefüggéseket próbáljuk megtalálni.

Az idősorok elemzésére 3 különböző regressziós vizsgálatot végeztünk el [10]. 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 („root 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 ($\Delta t=1$ nap) volt, a minták száma pedig (N) háromszor 731 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.

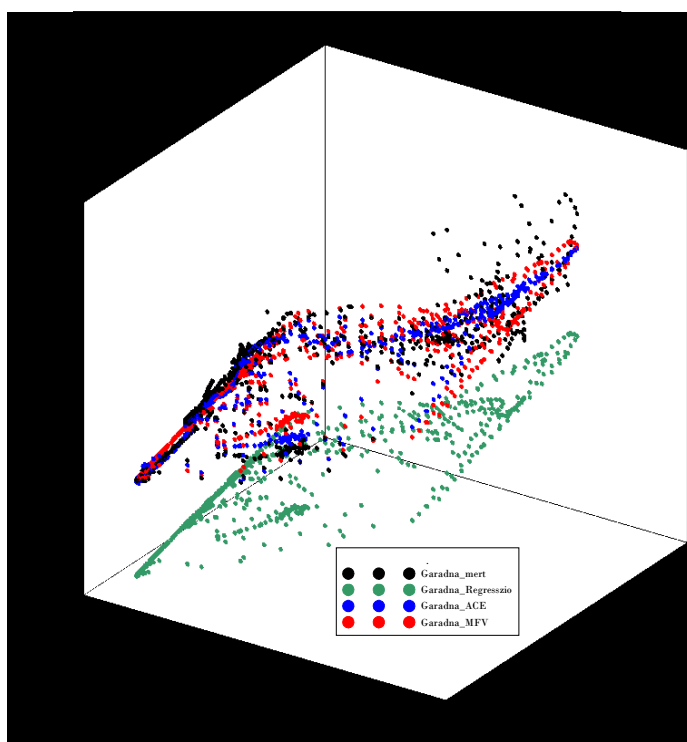
Korrelációs számításal kiválasztottuk a 3 adatsor közül a legjobban korreláló megfigyelőhelyeket, így függő változóként a Garadna-forrást (y) állapítottuk meg, első független változóként a Szinva-forrást (x_1), második független változóként pedig az Nv-17 (x_2) megfigyelőhelyet. A számítások alapján megállapítható, hogy a 3 számítás közül a legkisebb különbséget a mért és a számított értékek között az ACE algoritmus szolgáltatta (lásd 4. ábra).

A statisztikai jellemzők alapján megállapíthatjuk, hogy a BKÉR 3 megfigyelőhelyének adatsorain elvégzett 3 regressziós vizsgálat nagyon jól közelítette a mért értékeket. A számított RMSE értékekből egyértelműen kitűnik, hogy az ACE algoritmus meggyőző eredményeket szolgáltatott egy komplex hidrogeológiai karsztrendszer adatainak vizsgálatánál [14].

4. Összefoglalás

Végrehajtottuk a Stanford University intézményben Breiman és Friedman [1] által kidolgozott ACE nem paraméteres algoritmus adaptációját hidrogeológiai többváltozós regressziós vizsgálatokhoz, ahol a vizsgált változók közötti kapcsolat a priori nem ismert [11]. Az ACE algoritmus egyik nagy előnye, hogy a legkülönbözőbb típusú és nagyságú adatok együtt kezelhetőek [12]. 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 előnye mellett. Bizonyos esetekben az eljárás különböző eredményekre vezethet, ha megcseréljük a független változó sorrendjét. Másrészt az ACE algoritmus extrém kieső adatokra nagyon érzékenyen reagál. Természtésen 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 minőség-ellenőrzöttek [13].



4. ábra. A mért és az ACE algoritmus alapján számított vízszint adatok összehasonlítása a három vízszint megfigyelőhely esetében.

5. 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.

6. Felhasznált irodalom

- [1] Breiman L. and Friedman J.H., 1985: *Estimating optimal transformations for multiple regression and correlation (with discussion)*. Journal of American Statistical Association, 80, 580-619 (September).
- [2] Steiner, F., (ed.) 1997: *Optimum methods in statistics*. Akademia Kiado, Budapest. 370 pp.
- [3] Kovács, B., 2004: *Hidrodinamikai és transzport modellezés I.* (Processing MODFLOW környezetben.) Miskolci Egyetem, Műszaki Földtudományi Kar, Szegedi Tudomány Egyetem, Ásványtani, Geokémiai és Kőzettani Tanszék, GÁMA-GEO Kft., pp. 1-159.
- [4] Kovács, B., Szanyi J., 2005: *Hidrodinamikai és transzport modellezés II.* (Processing MODFLOW és Surfer for Windows környezetben.) Miskolci Egyetem, Műszaki Földtudományi Kar, Szegedi Tudomány Egyetem, Ásványtani, Geokémiai és Kőzettani Tanszék, GÁMA-GEO Kft., pp. 1-213.
- [5] Kitanidis P.K., 1997: *Introduction to geostatistics: Applications to hydrogeology*. Cambridge University Press, p. 249.
- [6] Lee, T- C., 1999: *Applied Mathematics in Hydrogeology*. Lewis Publishers and CRC Press LLC, ISBN 1- 56670- 375- 1.
- [7] Lénárt, L. (2006): *A Bükk-térség karsztvízpotenciálja – a hosszú távú hasznosíthatóságának környezetvédelmi feladatai*. Észak-magyarországi Stratégiai Füzetek. III. évf. 2. sz. pp. 17-28. Miskolc.
- [8] Juhász J., 2002: *Hidrogeológia*. Akadémiai Kiadó, Budapest, pp. 1-1176.
- [9] 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 , pp.: 71-78.
- [10] Darabos E. 2008: *A Bükki Karsztvízszint Észlelő Rendszer által szolgáltatott adatok kapcsolatainak vizsgálata*, OTDK dolgozat, Szombathely, 2009., pp.: 34-36.
- [11] Wang D. and Murphy M., 2004: *Estimating Optimal Transformations for Multiple Regression Using the ACE Algorithm*. Journal of Data Science 2(2004), 329-346.
- [12] 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) pp. 123-134 (2009). Springer.
- [13] Székely F., 2006: *A háromdimenziós kúthidraulikai modellezési módszer és gyakorlati alkalmazása*. VITUKI 79, Budapest, pp. 1- 113.
- [14] Lénárt L. 2010: *The Interaction of Cold and Warm Karst Systems in the Bükk Region*. Proceedings of the 1th Knowbridge Conference on Renewables, pp. 111-118., Miskolc.