

Mérnöksondázási adatok kiértékelése súlyozott faktoranalízis alkalmazásával

BALOGH G. P.

Miskolci Egyetem, Geofizikai Tanszék, 3515 Miskolc-Egyetemváros
E-mail: gfbg@uni-miskolc.hu

Jelen tanulmányban egy többváltozós statisztikus eljárást mutatunk be, amely hatékonyan alkalmazható a mérnök-geofizikai és geotechnikai vizsgálatokban. A mérnök-geofizikai szondázási adatok faktoranalízisével megbecsülhető a felszínközeli rétegek víz- és levegőteltettsége, valamint egyéb geotechnikai paraméterek is feltárhatók. A klasszikus faktoranalízis algoritmus a Steiner-súlyok bevezetésével iteratív úton javítható. Ezáltal a szondaadatokat terhelő zaj, valamint kiugró adatok mellett is pontosabb kiértékelés valósítható meg. A módszert egy nyugat-magyarországi adatrendszeren teszteltük, melynek keretében összehasonlítottuk a hagyományos és a súlyozott faktoranalízisen alapuló eljárásokat. A mechanikai, elektromos és nukleáris adatok faktoranalízisével megadható a faktor–víztelítettség helyi összefüggése, annak statisztikai minősítő jellemzői, valamint a klasszikus és a továbbfejlesztett módszerrel számított víztelítettség-szelvények egyezése. A módszer hatékonyan felhasználható felszínközeli szerkezetek közetfizikai és geotechnikai jellemzőinek számítására, mely több fúrás adatainak egyidejű feldolgozásával területi információt is szolgáltat.

Balogh, G. P.: Weighted factor analysis applied to the evaluation of engineering geophysical sounding data

In this study, a multivariate statistical method is presented, which can be effectively applied to engineering geophysical and geotechnical tests. The water and air saturation of near-surface layers are estimated by the factor analysis of engineering geophysical data, and other geotechnical parameters can be explored. The classical algorithm of factor analysis can be improved by the introduction of Steiner weights in an iterative way. Thereby a more accurate evaluation can be made with noisy data sets including outliers. The method was tested in a western Hungarian data set to make a comparison between the procedures of traditional and weighted factor analysis. The factor analysis of mechanical, electrical and nuclear data can be used to give the factor vs. water saturation relationship, its statistical quality features, as well as the agreement between the calculated water saturation logs estimated by the classical and further-developed statistical method. The method can be efficiently used for the calculation of petrophysical and geotechnical features of near-surface formations, which provides multidirectional information by the simultaneous processing of several boreholes.

Beérkezett: 2016. május 24.; *elfogadva:* 2016. június 1.

Bevezetés

A felszínközeli szerkezetek vizsgálatára hatékonyan alkalmazható a mérnök-geofizikai szondázás (MGSZ), mely a CPT (*Cone Penetration Test*) módszer továbbfejlesztett változata (Fejes, Jósa 1990). A kezdetben csak mechanikai paramétereket szolgáltatató, környezetvédelmi problémák megoldására alkalmazott technológiát továbbfejlesztették, ami új utakat nyitott meg a geofizikai mérések előtt. A mérés során a talajba hidraulikus berendezés és rudazat segítségével juttatott kúp csúcscellenállása és palástsúrlódása mérhető, amiből geotechnikai talajparaméterekre lehet következtetni. Az MGSZ módszer többek között alkalmas sekély víztároló-képződmények in situ vizsgálatára, melyet az esz-

közbe szerelt geofizikai (mechanikai, elektromos és nukleáris) szondák adatai alapján lehet elvégezni. E mérések kvantitatív információt szolgáltatnak az paleotalaj összetételéről, porozitásáról, agyagtartalmáról és víztelítettségéről. Az MGSZ adatok kiértékelése általában determinisztikus vagy mélységpontokénti inverziós módszerekkel történik (Drahos 2005), melyeknek eredményei a szondaadatokat statisztikus feldolgozásával tovább javítható. Erre a célra megfelelően alkalmazhatók a többváltozós statisztika feltáró módszerei (pl. faktoranalízis, főkomponens és klaszteranalízis).

A hagyományos faktoranalízis ugyanazt a jelenséget kevesebb (korrelálatlan) változóval írja le, miközben közvetlenül nem megfigyelhető információkat tár fel a statisztikai mintából (Lawley, Maxwell 1962). A faktoranalízisnek számos al-

kalmazása ismert a geofizikában, pl. Simone és szerzőtársai (1994) vagy Xu és szerzőtársai (2010). Mélyfúrású geofizikai adatok faktoranalízisével számítható a szénhidrogén- és víztárolók agyagtartalma (Szabó 2011, Szabó, Dobróka, 2013, Szabó et al. 2014), ugyanakkor mérnök-geofizika adatokra alkalmazva megbecsülhető a felszínközeli üledékek víztelítettsége (Szabó et al. 2012). A statisztikai faktorok a száraz sűrűségről is hordozhatnak információt (Szabó 2012). Jöreskog (2007) gyors, nem iteratív megoldást kínált a faktoranalízis megoldására abban az esetben, amikor az adatok Gauss-eloszlást követnek. Azonban ez a követelmény a mérnök-geofizikai adatrendszerre általában nem teljesül, ilyenkor az adateloszlás aszimmetriája és a kiugró adatok jelenléte torzíthatja a megoldást. A hagyományos faktoranalízis algoritmusát továbbfejlesztve egy robusztus módszert javasolunk, mely felhasználja a Steiner-féle (1991) leggyakoribb érték módszerét (*Most Frequent Value*, röviden MFV). Hasonlóan az automatikus súlyozást használó inverziós módszerekhez (Drahoš 2008, Gyulai et al. 2014) az MFV eljárás alapuló súlyozott faktoranalízis (MFV-FA) iteratív úton javítja az adatokból leszármaztatott faktorok számítását, ezáltal hatékonyan csökkenti a mért és a számított adatok távolságát. A módszert egy nyugat-magyarországi adatrendszeren teszteltük, és összehasonlítottuk azt a hagyományos faktoranalízis és a mérnökszondázási görbék mélységpontenkénti inverziójával kapott eredményekkel.

A leggyakoribb érték módszere és a Steiner-súlyok

A mért adatok alkotta statisztikai minta jellemző (átlagos) értékének becslésére többféle eljárás ismeretes, melyek kö-

$$\varepsilon_{j+1}^2 = \frac{3 \sum_{k=1}^N \frac{(x_k - M_j)^2}{\left[\varepsilon_j^2 + (x_k - M_j)^2 \right]^2}}{\sum_{k=1}^N \frac{1}{\left[\varepsilon_j^2 + (x_k - M_j)^2 \right]^2}} \Leftrightarrow M_{j+1} = \frac{\sum_{k=1}^N \frac{\varepsilon_{j+1}^2}{\varepsilon_{j+1}^2 + (x_k - M_j)^2} x_k}{\sum_{k=1}^N \frac{\varepsilon_{j+1}^2}{\varepsilon_{j+1}^2 + (x_k - M_j)^2}} \quad (4)$$

Példaként tekintünk egy RCPT (MPa) csúcscellenállás-szelvényre, mely adatainak leggyakoribb értékét becsüljük meg! Az 1. ábrán a leggyakoribb érték és a dihézió optimumhoz történő konvergenciája látható. A dihézió az iterációs lépésszám növekedésével fokozatosan csökken, így az optimum közelében egyre kisebb súlyt kapnak a kiugró adatok. Az RCPT szelvényadatok leggyakoribb értékét és számtani átlagát a 2. ábra mutatja. Látható, hogy a leggyakoribb érték számítása rezisztens, míg a számtani közép meglehetősen érzékeny a kiugró adatokra.

A súlyozott faktoranalízis elmélete

A hagyományos faktoranalízis (TFA) mint statisztikai eljárás a faktorok bevezetésével csökkenti statisztikai probléma mé-

zül egyesek igen zajérzékenyek, de léteznek a zajra rezisztens módszerek is. Ha az adatok (x_k) többségétől távoli pontokhoz kis, a legnagyobb adatsűrűségi helyen lévő adatokhoz pedig nagyobb súlyt (w_k) rendelünk, a számtani átlagnál jóval megbízhatóbb helyparaméter jellegű statisztikai jellemzőt kapunk, amely nem más, mint egy súlyozott átlagérték

$$M = \sum_{k=1}^N x_k w_k \left[\sum_{k=1}^N w_k \right]^{-1}, \quad (k = 1, 2, \dots, N). \quad (1)$$

A k -adik súlyt Steiner (1991) az alábbiak szerint választotta meg:

$$w_k = \varepsilon^2 / [\varepsilon^2 + (x_k - M)^2], \quad (2)$$

ahol N az adatok száma és ε egy skálaparaméter jellegű mennyiség, a dihézió. Ha az ε értéke nagy, akkor minden adathoz közel ugyanakkora súlyt rendelünk, és a kiugró adatok elrontják a jellemző érték becslését, viszont túl kicsi ε érték esetén vigyázni kell arra, nehogy a centrumhoz közeli adatok figyelmen kívül maradjanak. Az (1) egyenlet alapján definiált leggyakoribb értéknek (M) nevezett súlyozott átlagot ismerni kellene előre ahhoz, hogy a helyén maximális értékű és attól távolabb egyre kisebb súlyok kerüljenek kiosztásra. Ezért ehhez az eljáráshoz iteratív algoritmus szükséges, melynek során M -et és az ε -t együttesen határozzuk meg. Az első iterációs lépésben a dihézió a mintaterjedelemből az alábbi formula alapján becsülhető:

$$\varepsilon_1 = (\sqrt{3}/2)[\max(x_k) - \min(x_k)], \quad (3)$$

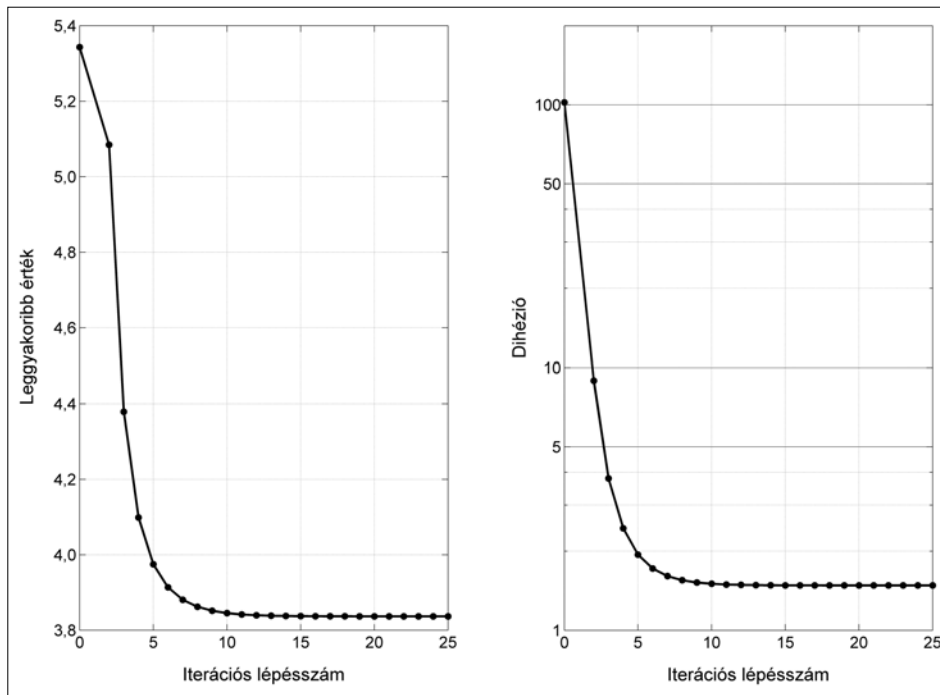
míg az M_1 kezdeti értéknek célszerű a mintaátlag vagy a medián értékét választani. Az ezt követő iterációs lépésekben M és ε egymásból származtatható a következő eljárás szerint:

retét. Jelen kutatás során a faktoranalízis bemenő mennyiségeit a standardizált mérnök-geofizikai szelvényadatok adják. A belőlük képzett \mathbf{D} adatmátrix sorai a mélységpontoknak (N), oszlopai pedig egy-egy szelvénytípusnak (K) felelnek meg. Az adatmátrix további két mátrixra bontható fel:

$$\mathbf{D} = \mathbf{F}\mathbf{L}^T + \mathbf{E}, \quad (5)$$

ahol az \mathbf{F} a faktorok $N \times Q$ méretű mátrixa, Q a faktorok száma ($Q < K$), \mathbf{L} a faktorsúlyok $K \times Q$ méretű mátrixa, \mathbf{E} a hibakomponens mátrix (\mathbf{T} a mátrix transzponáltját jelöli). Feltevétezzük, hogy az $\mathbf{F}\mathbf{L}^T$ és \mathbf{E} mátrixok korrelálatlanok és a faktorok lineárisan függetlenek ($\mathbf{F}^T\mathbf{F}/N = \mathbf{I}$ egységmátrix). Ekkor a mért szelvények $K \times K$ méretű korrelációs mátrixa kifejezhető a faktoregyütthatókkal:

$$\mathbf{R} = N^{-1}\mathbf{D}^T\mathbf{D} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}, \quad (6)$$



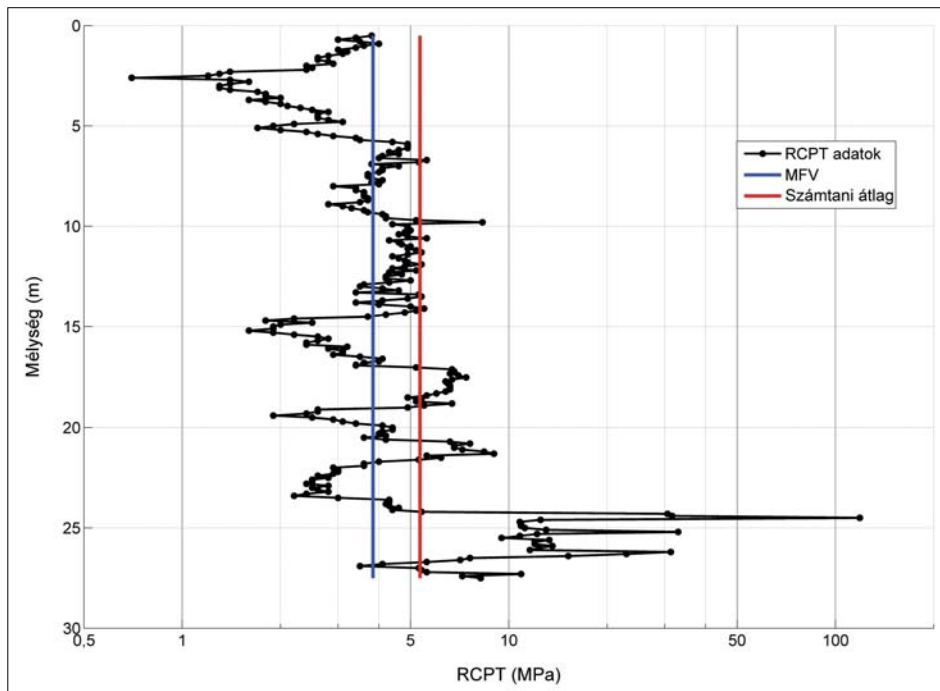
1. ábra | Az RCPT szelvény leggyakoribb értéke (bal oldal) és a dihéziója (jobb oldal) változása az iterációs eljárás során
Figure 1 | The convergence of the most frequent value (left panel) and dihesion (right panel) of RCPT data by the iteration procedure

ahol a $K \times K$ méretű Ψ diagonális mátrix a kiinduló változók szórásnégyzeteinek a közös faktorokkal nem értelmezhető részét képviseli. A faktorszelvények számítása érdekében először a faktorsúlyokat kell megbecsülni amellett, hogy a maradékfaktorok varianciáját tartalmazó mátrix is ismeretlen. A *Maximum Likelihood* módszer alkalmazása esetén közös eljárásban becsüljük meg a faktorsúlyokat és a hiba-

varianciákat (Móri 1999). Ennek keretében az L és Ψ mátrixot iteratív eljárásban számítjuk az alábbi célfüggvény optimalizálásával:

$$A(L, \Psi) = \text{tr}(R - LL^T + \Psi), \quad (7)$$

ahol tr a mátrix nyoma, azaz a főátlóbeli elemek összege. A A függvényt minden egyes faktorsúly és hibavariancia



2. ábra | Az RCPT szelvényadatok leggyakoribb értéke és számítási átlaga
Figure 2 | The most frequent value and mean of RCPT logging data

szerint egyidejűleg minimalizálni kell. A faktorokat a faktorsúlyok ismeretében kiszámíthatjuk. E probléma megoldása céljából leggyakrabban az alábbi log-likelihood függvényt maximalizálják

$$\lg P = -\frac{1}{2} \left[\lg |2\pi\Psi| + (\mathbf{D} - \mathbf{F}\mathbf{L}^T)\Psi^{-1}(\mathbf{D} - \mathbf{F}\mathbf{L}^T)^T \right] = \max, \quad (8)$$

melyből a faktorok Bartlett (1953) szerint torzítatlanul becsülhetők:

$$\mathbf{F}^T = (\mathbf{L}^T\Psi^{-1}\mathbf{L})^{-1}\mathbf{L}^T\Psi^{-1}\mathbf{D}^T, \quad (9)$$

melyben az \mathbf{F} mátrix oszlopai a faktorok különböző mélységpontonál becsült értékeit (a faktorszelvényeket) tartalmazzák. Jöreskog (2007) gyors, nem iteratív, közelítő megoldást kínált a hagyományos faktoranalízis elvégzésére, mely elkerüli a Ψ hibavariancia-mátrix becslését. A hagyományos faktoranalízis (TFA) eljárás alkalmazása az adatok Gauss-eloszlását feltételezi, mely mérnök-geofizikai adatrendszerekre általában nem érvényesül, így az adateloszlás aszimmetriája és a kiugró adatok jelenléte torzíthatja a megoldást.

Mérnök-geofizikai adatrendszerek hatékonyabb feldolgozása céljából új, iteratív újraszűzőszen alapuló faktoranalízist javasolunk. Ennek első lépésében a Jöreskog-algoritmussal (2007) megbecsüljük a faktorsúlyokat és faktorszelvényeket. Ezután a kezdeti becslés eredményeit az iteratív faktoranalízis eljárásának segítségével tovább finomítjuk. Ebben a fázisban a hagyományos faktoranalízis mátrixos elrendezése helyett oszlopvektorba rendezzük a bemenő adatokat, így az (5) egyenlet az alábbi alakot ölti:

$$\mathbf{d} = \tilde{\mathbf{L}}\mathbf{f} + \mathbf{e}, \quad (10)$$

ahol \mathbf{d} a KN elemű mért adatvektor, $\tilde{\mathbf{L}}$ a $KN \times QN$ méretű faktorsúly-mátrix, \mathbf{f} a QN elemű faktorok értékeit tartalmazó vektor és \mathbf{e} a hibát képviselő KN elemű vektor. A faktorokat az \mathbf{e} eltérésvektor súlyozott normájának minimalizálásával határozzuk meg. A q -adik iterációban a faktorsúlyokat az előző lépésben meghatározott faktorokból a csillapított legkisebb négyzetek módszerével, majd azok ismeretében a faktorokat az iteratív újraszűzőszen legkisebb négyzetek módszerével (Scales 1985) becsüljük:

$$\mathbf{L}^{T(q)} = (\mathbf{F}^{T(q-1)}\mathbf{F}^{(q-1)} + \alpha^2\mathbf{I})^{-1}\mathbf{F}^{T(q-1)}\mathbf{D}, \quad (11)$$

$$\mathbf{f}^{(q)} = (\tilde{\mathbf{L}}^{T(q-1)}\mathbf{W}\tilde{\mathbf{L}}^{(q-1)})^{-1}\tilde{\mathbf{L}}^{T(q-1)}\mathbf{W}\mathbf{d}, \quad (12)$$

ahol a \mathbf{W} diagonális adattérbeli súlymátrix, α a numerikus stabilitást biztosító csillapítási tényező. A \mathbf{W} diagonális súlymátrix az adatok automatikus súlyozását végzi, mely elemeinek a Steiner-súlyokat választjuk ($k = 1, \dots, KN$):

$$W_{kk} = \varepsilon^2 / [\varepsilon^2 + e_k^2]. \quad (13)$$

A fenti iterációs eljárás minden lépésében a faktorokkal elvi szelvényadatokat ($\mathbf{d}^e = \tilde{\mathbf{L}}\mathbf{f}$) számítunk. Abban az esetben, amikor a (10) egyenletből kifejezhető \mathbf{e} eltérésvektor értéke nagy, azaz az elméleti adatok távol állnak a mérési adatoktól, akkor (13) súlyok kicsik lesznek. Ellenkező eset-

ben, jó adattérbeli illeszkedésnél nagyobb súlyokat kapunk. A súlyozás eredményeként a jobb minőségű adatok nagyobb mértékben befolyásolják a megoldást, míg a kiugró értékek hatását csaknem teljesen elnyomjuk. A fenti statisztikai eljárást MFV-FA eljárásnak nevezzük.

Az MFV-FA módszer egydimenziós (1D) algoritmusa könnyen továbbfejleszhető többdimenziós (2D vagy 3D) algoritmussá. A szomszédos fúrásokból szerzett adatok egyidejű feldolgozásával két- vagy háromdimenziós faktor változók képezhetők. Egyesítsük az összes fúrás valamennyi szelvényadatát egy adatvektorban! Ekkor a mért és a számított adatok eltérése $\mathbf{e}^* = \mathbf{d}^* - \tilde{\mathbf{L}}^*\mathbf{f}^*$, mely az alábbi egyenletrendszerre vezet:

$$\begin{pmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_h \\ \vdots \\ \mathbf{d}_H \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{L}}_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \tilde{\mathbf{L}}_h & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \tilde{\mathbf{L}}_H \end{pmatrix} \begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_h \\ \vdots \\ \mathbf{f}_H \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_h \\ \vdots \\ \mathbf{e}_H \end{pmatrix}, \quad (14)$$

ahol $\tilde{\mathbf{L}}_h$ a faktorsúlyok mátrixa és \mathbf{f}_h a faktorok oszlopvektora a h -adik fúrólukban. Az N_h a mélységpontok számát jelöli a h -adik fúrásban ($h = 1, 2, \dots, H$), a feldolgozott mélységpontok teljes száma $N^* = N_1 + N_2 + \dots + N_H$. Ha minden fúrásban ugyanazokat a szelvényeket regisztráljuk, akkor a \mathbf{W} diagonális súlymátrix mérete $KN^* \times KN^*$. A két-dimenziós MFV-FA eljárás során az \mathbf{L}^* faktorsúlyok mátrixa és az \mathbf{f}^* faktor értékeinek vektora a Jöreskog-módszerrel meghatározható és a súlyozott faktoranalízissel tovább javítható.

Mélységpontonkénti inverzió eljárása

A geofizikai mérésekkel kapcsolatba hozható közetfizikai paramétereket az inverz feladat megoldásával is meghatározhatjuk, melynek során a mért adatokat egy előzetes modellen számított adatokkal hasonlítjuk össze (Dobróka, Szabó 2012; Dobróka et al. 2012, Dobróka et al. 2016). Modelünkben a felszínközeli rétegek közetmátrixa durva- és finomszemcsés komponensekből állnak, míg a térfogati pórusteret édesvíz és levegő tölti ki. Az inverz probléma modellvektora a következőképpen definiálható:

$$\mathbf{m} = [V_{cl}, V_s, V_w]^T, \quad (15)$$

ahol V_{cl} az agyag, V_s a homok, V_w a víz fajlagos térfogata. Mivel a gáz térfogatát az alábbi formulával számoljuk $V_g = 1 - V_w - V_{cl} - V_s$, az inverz probléma rendszerint túlhatározott. A víztelítettség a fenti petrofizikai paraméterekből a következő szerint számolható:

$$S_w = V_w / (V_w + V_g). \quad (16)$$

Általában a mérnök-geofizikai szondákkal az alábbi szelvényadatokat mérjük: természetesgamma-intenzitás (GR),

sűrűség (*DEN*), neutronporozitás (*NPHI*) és fajlagos ellenállás (*RES*). Az adatok és petrofizikai paraméterek kapcsolatát a következő válasz egyenletek írják le, amelyek alapján elvi szelvényadatokat számíthatunk:

$$GR = V_{cl}GR_{cl} + V_sGR_s, \tag{17}$$

$$DEN = V_w\rho_w + V_{cl}\rho_{cl} + V_s\rho_s, \tag{18}$$

$$NPHI = V_w\Phi_{N,w} + V_{cl}\Phi_{N,cl} + V_s\Phi_{N,s}, \tag{19}$$

$$RES = a \left(V_w + V_g + V_{cl} \right)^{-m} \times \left(\frac{V_{cl}/(V_w + V_{cl})}{R_{cl}} + \frac{1 - [V_{cl}/(V_w + V_{cl})]}{R_w} \right)^{-1} \times \left(\frac{V_w + V_{cl}}{V_w + V_g + V_{cl}} \right)^{-n}, \tag{20}$$

ahol a kőzetösszetevőket, a pórufolyadékot és azok fizikai állandóit az alábbiak szerint jelöljük: *cl* az agyag, *s* a homok, *w* a víz, *g* a gáz, ρ a tömegsűrűség, Φ_N a neutronporozitás, *m* a cementációs kitevő, *a* a tortuozitási tényező, *n* a telítettségi kitevő. Az elvi adatokat egy vektorba rendezzük ($\mathbf{d}^{(sz)}$). Hasonlóan a mért adatok vektora egy adott mélységben a következő:

$$\mathbf{d}^{(m)} = [GR, DEN, NPHI, RES]^T. \tag{21}$$

A lokális inverz problémát ugyanebben a mélységben a mért és számított adatok különbsége súlyozott euklideszi normájának négyzete minimalizálásával oldjuk meg

$$\Omega = (\mathbf{e}, \mathbf{W}\mathbf{e}) = \mathbf{e}^T \mathbf{W}\mathbf{e} = (\mathbf{d}^{(m)} - \mathbf{d}^{(sz)})^T \mathbf{W} (\mathbf{d}^{(m)} - \mathbf{d}^{(sz)}) = \min. \tag{22}$$

Az adattérbeli súlymátrix elemei a szelvények hibájával (σ) fordítottan arányosak

$$\mathbf{W} = \text{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \dots, \sigma_N^{-2}), \tag{23}$$

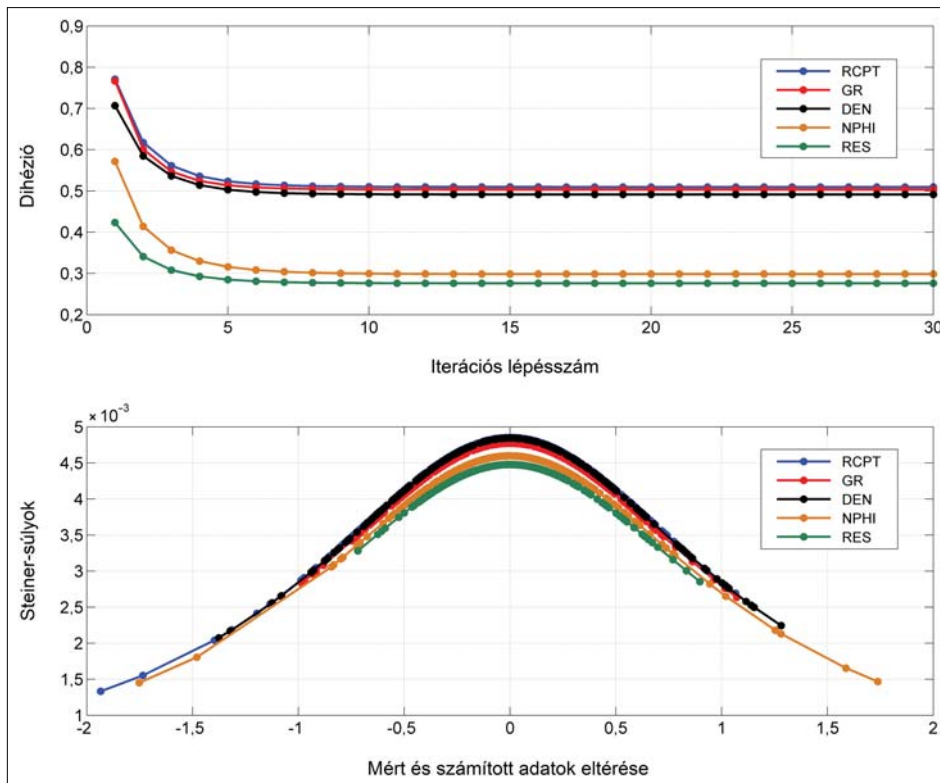
ahol *N* az alkalmazott szondák száma. A fenti inverz probléma megoldására a súlyozott legkisebb négyzetek módszere alkalmazható (Drahos 2005), mely stabil megoldást ad a modellparaméterek vektorára

$$\mathbf{m} = (\mathbf{G}^T \mathbf{W}\mathbf{G})^{-1} \mathbf{G}^T \mathbf{W}\mathbf{d}^{(m)}, \tag{24}$$

ahol **G** a Jakobi-mátrixot jelöli. Az eljárás magában foglalja a becsült térfogatjellemző mennyiségek hibájának számítását is (Menke 1984).

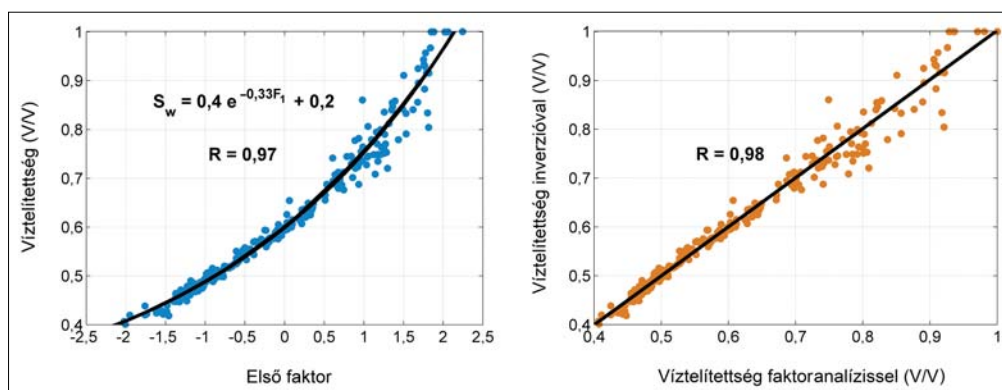
Terepi vizsgálatok

A hagyományos faktoranalízis (TFA) és a Steiner-súlyokon alapuló faktoranalízis (MFV-FA) eljárásainak összehasonlítását a bátaapáti nukleáris hulladék-lerakó felett a gránit ala-



3. ábra Az MFV-FA eljárás során a dihézió alakulása (felső ábra), és a mérnökshondázási adatok Steiner-súlyainak optimális értékei az adattávolság függvényében (alsó ábra)

Figure 3 Development of convergence of dihesion during the MFV-FA procedure (upper panel), optimal values of Steiner's weights applied to EGS data in the function of data misfit (lower panel)



4. ábra Az első faktor és a víztelítettség regressziós függvénye (bal oldal), az MFV-FA eljárással és a pontonkénti inverzióval becsült víztelítettség kapcsolata (jobb oldal)

Figure 4 Regression relation between the first factor and water saturation (left panel), connection between water saturations estimated separately by the MFV-FA procedure and local inverse modeling (right panel)

pon települt laza szerkezetű löszös üledékben gyűjtött MGSZ-adatok segítségével végeztük. A szondázási program az alábbi mért paraméterekre terjedt ki: *RCPT* (MPa) csúcshellállás, *GR* (cpm) természetesgamma-intenzitás, *DEN* (g/cm³) sűrűség, *NPHI* (V/V) neutronporozitás és *RES* (ohmm) fajlagos ellenállás. Míg az *RCPT*-szelvény a talaj szilárdságára, addig a *GR* főként az agyagtartalomra és a litológiára érzékeny, a *DEN*-szelvény a porozitásról és a tömegsűrűségről hordoz információt, az *NPHI*- és a *RES*-szelvényeket elsősorban a víztartalom befolyásolja. A TFA és MFV-FA eljárások ellenőrzését független mélységpontonkénti inverzióval végeztük el.

Az MFV-FA statisztikai eljárás a Jöreskog-algoritmusból indul ki, melyet az (11)–(12) egyenletek segítségével tovább finomítjuk. A mért és számított adatok relatív különbsége az iterációs lépések előrehaladtával folyamatosan csökkenthető. A faktorsúlyokat és faktorokat 15 iterációs lépésben állítjuk elő, míg minden egyes lépésben egy további iteratív (belső) ciklusban 30-szor újraszámoljuk a Steiner-súlyokat. Ennek eredményeképpen, a (13) egyenletben szereplő dihéziót automatikusan csökkentjük, szelvényenként különböző mértékben (3. ábra). A dihézió ugyanazon értéke mellett nagyobb mért és számított adatok eltérése esetén kisebb súlyt kapnak az adatok, így a nagyobb eltérések (kiugró adatok) kevésbé járulnak hozzá a megoldáshoz. A 3. ábrán a Steiner-súlyok optimális értékei láthatóak az

adattérbeli eltérés (eltérésvektor elemei) függvényében. Példánkban az MFV-FA eljárással két független faktort számítunk. Az első faktor magyarázza a mérnökszondázási adatok teljes varianciájának meghatározó részét (76,2%-át). Az egyes szelvények faktorsúlyai: $-0,41$ (*RCPT*), $0,31$ (*GR*), $0,88$ (*DEN*), $0,93$ (*NPHI*), $-0,98$ (*RES*). A faktorsúlyokból kiderül, hogy a neutronporozitás- és fajlagosellenállás-adatok nagy súllyal vesznek részt az első faktor meghatározásánál, azaz az első faktor a víztelítettségéről hordoz információt. Ennek bizonyítására regresszióanalízist végeztünk. Az első faktor (F_1) és a víztelítettség (S_w) közötti összefüggést a 4. ábra mutatja, melynek regressziós függvénye

$$S_w = a \exp\{-bF_1\} + c, \quad (25)$$

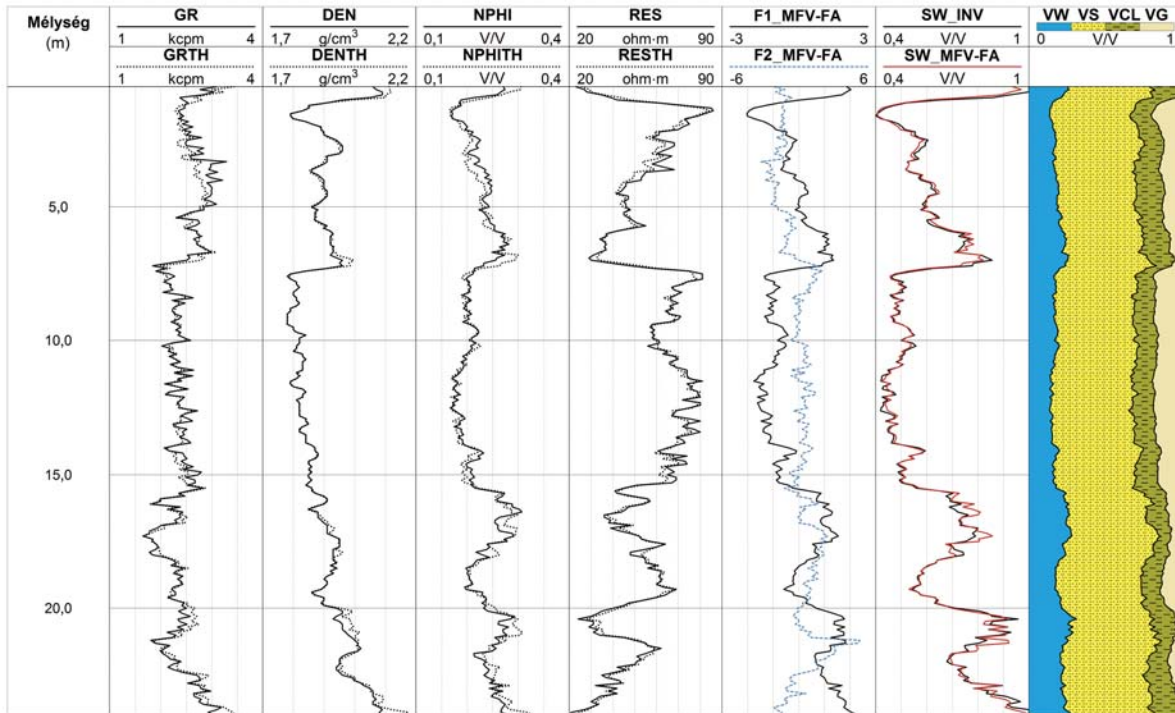
ahol a regressziós együtthatókra 95%-os konfidencia-intervallumon belül $a = 0,40 \pm 0,04$, $b = -0,33 \pm 0,03$, $c = 0,20 \pm 0,04$ értékek adódtak. A két változó kapcsolata erős, ezt a $R = 0,97$ értékű Pearson-féle korrelációs tényező mutatja.

A mélységpontonkénti inverzió során választott zónaparaméterek értékeit az 1. táblázatban tüntetjük fel, ahol a levegő fizikai jellemzőit gyakorlatilag nullának vehetjük. A direkt feladatot (17)–(20) egyenletek felhasználásával oldottuk meg. Az inverz feladat megoldását a (24) egyenlet adja. A 4. ábra a faktoranalízissel és a pontonkénti inverzióval

1. táblázat Mélységpontonkénti inverzióval alkalmazott zónaparaméterek értékei

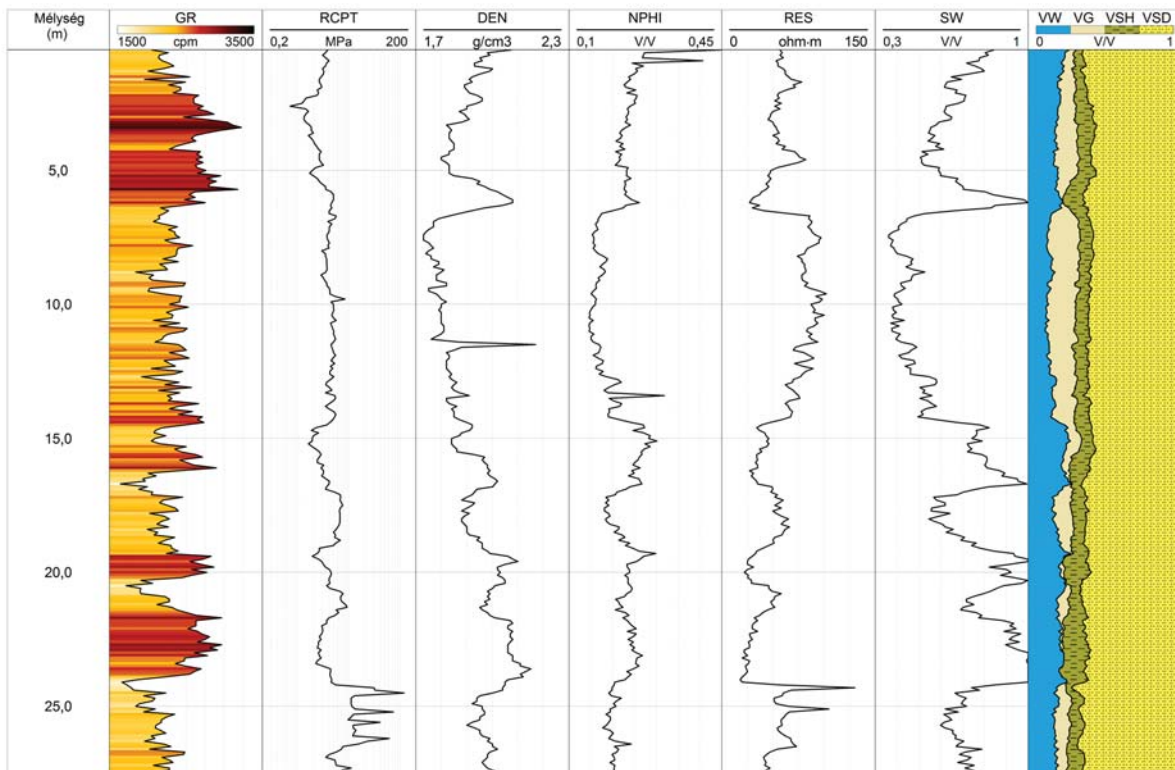
Table 1 Zone parameters applied in depth-by-depth inversion

Zónaparaméterek	Textúra	Agyag	Homok	Víz	Egység
<i>GR</i>	–	11.6	1.45	0	cpm
<i>DEN</i>	–	2.10	2.60	1.0	g/cm ³
<i>NPHI</i>	–	0.23	0	1.0	V/V
<i>RES</i>	–	6.50	–	9.0	ohmm
<i>m</i>	1.68	–	–	–	–
<i>a</i>	1.0	–	–	–	–
<i>n</i>	2.0	–	–	–	–



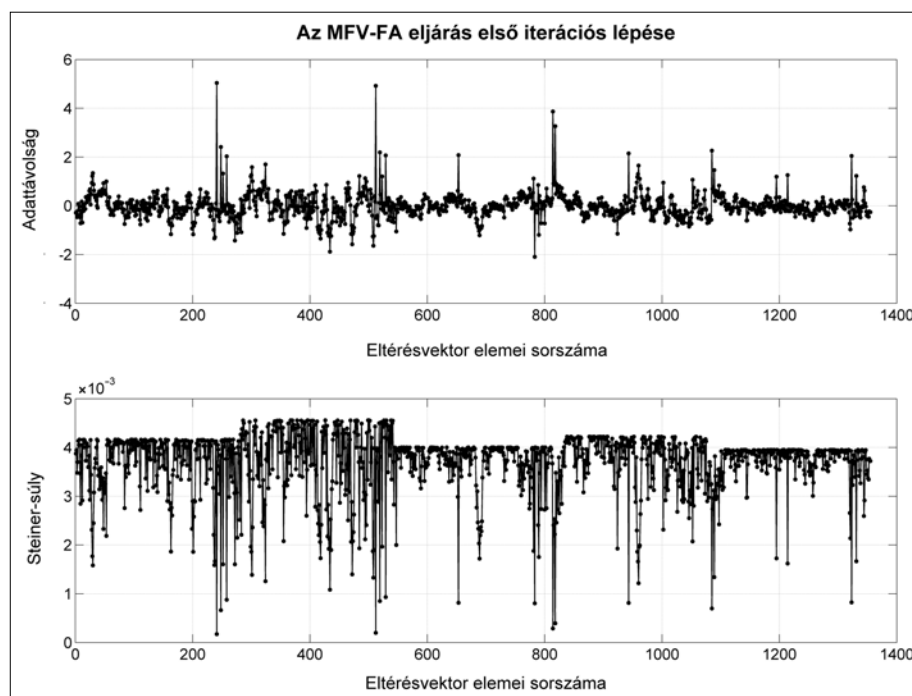
5. ábra A mért és pontonkénti inverzióval számított mérnöksondázási adatok, a faktorok, a faktoranalízissel (SW_MFV-FA) és inverzióval (SW_INV) becsült víztelítettség-, a víz-, a gáz-, a homok- és az agyagtartalom-szelvény

Figure 5 Observed and calculated EGS data, factors, water saturation estimated by factor analysis (SW_MFV-FA) and inversion (SW_INV), the logs of water content, gas content, sand and shale volumes



6. ábra A kiugró adatokkal terhelt (bemenő) mérnöksondázási adatok, súlyozott inverziós eljárással meghatározott víztelítettség, víz-, gáz-, agyag- és homoktartalom-szelvény

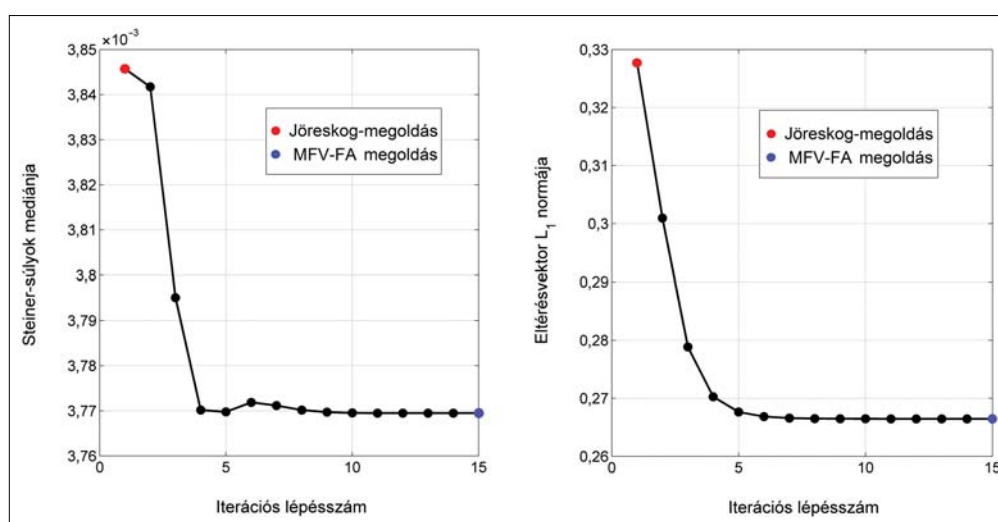
Figure 6 EGS data contaminated by outlying data, water saturation estimated by weighted inversion, the logs of water content, gas content, sand and shale volumes



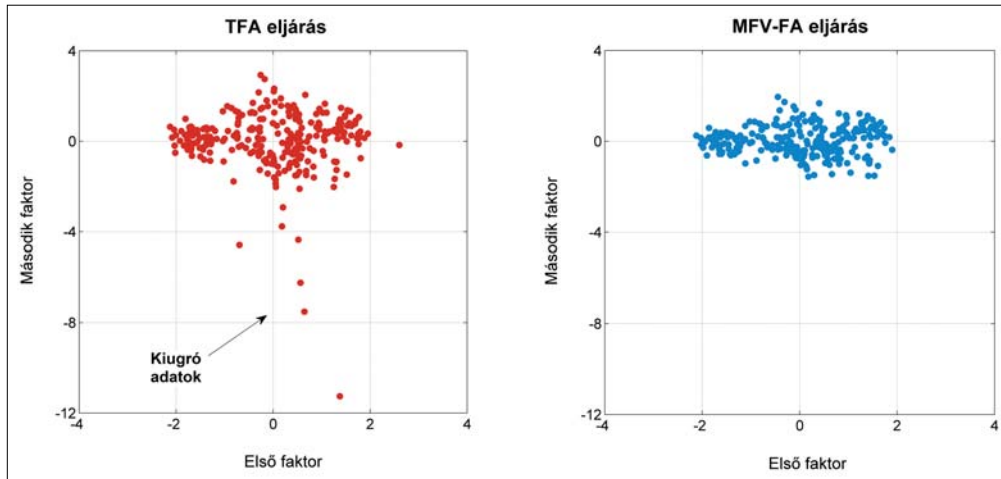
7. ábra Az egyesített adatvektorhoz számolt adattérbeli távolságok (felső ábra) és a Steiner-súlyok (alsó ábra) az MFV-FA eljárás első iterációs lépésében
Figure 7 Data distances calculated for the integrated data vector (upper panel), Steiner's weights in the first iteration of the MFV-FA procedure (lower panel)

óval számolt víztelítettség jó egyezését mutatja. Itt a korreláció szintén erős ($R = 0,98$). Az 5. ábrán a mért és számított szelvények, az inverzióval becsült kőzetalkotók térfogatarányai, a két faktor ($F1_{MFV-FA}$, $F2_{MFV-FA}$) és a víztelítettség szelvények láthatók. A mért és számított adatok közötti átlagos négyzetes eltérés $RMS = 3,9\%$. A robusztus faktoranalízissel becsült (SW_{MFV-FA}) és az inverziós eljárással számolt (SW_{INV}) víztelítettség szelvények közötti távolság $RMS = 2,4\%$.

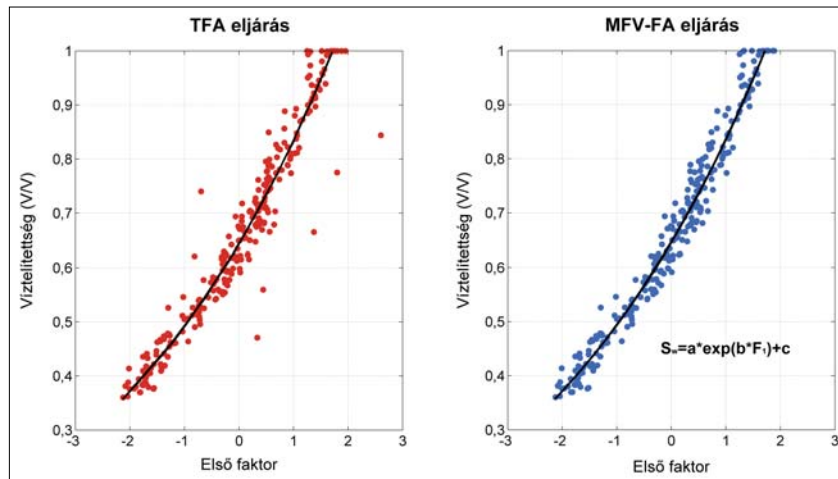
A robusztus eljárás leghatékonyabban kiugró adatokat tartalmazó adatrendszer feldolgozására használható. A hagyományos (TFA) és az iteratív újrásúlyozáson alapuló faktoranalízis (MFV-FA) összehasonlításához szükséges bemenő paramétereket a bátaapáti kutatási terület egy másik fúrásának mért szelvényei szolgáltatták. A 6. ábrán a mért szelvényadatok mellett megtalálható a súlyozott inverziós eljárással meghatározott SW víztelítettség, VW víz-, VG gáz-, VSH agyag- és VSD homoktartalom-



8. ábra A Steiner-súlyok mediánjának (bal oldal) és az eltérésvektor L_1 normájának (jobb oldal) konvergenciája az MFV-FA eljárás során
Figure 8 Median of Steiner's weights (left panel) and the convergence of the L_1 norm of the prediction error in the MFV-FA procedure (right panel)



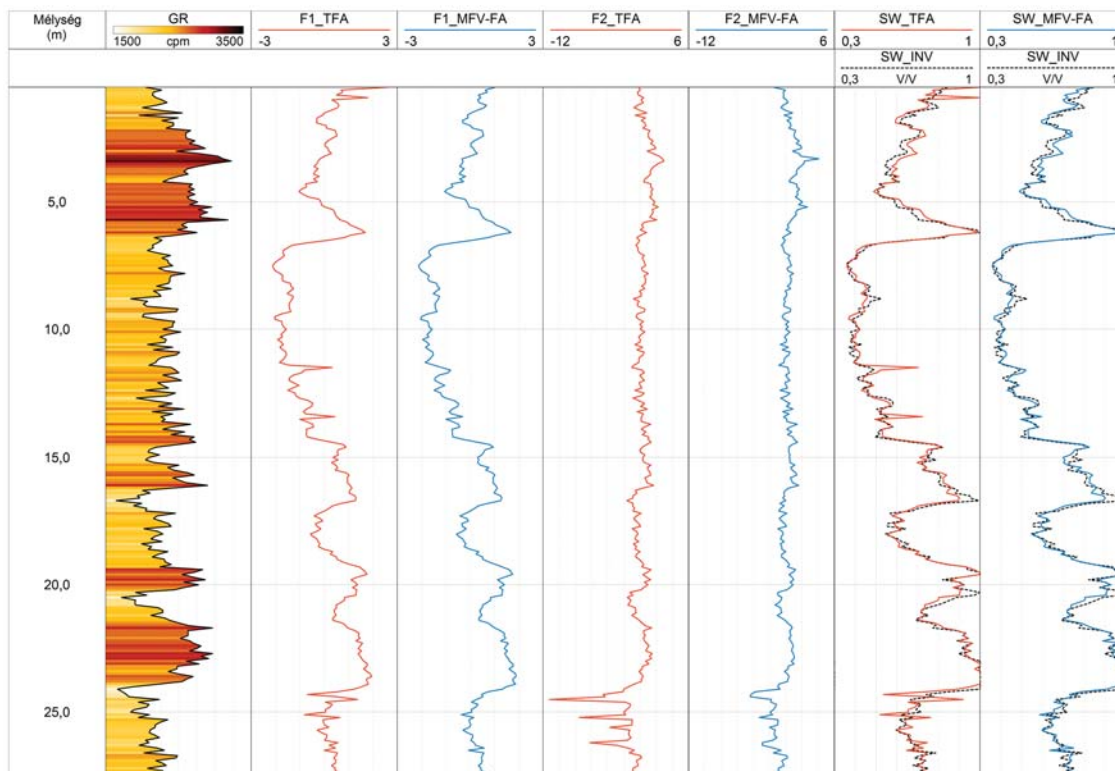
9. ábra | A TFA és MFV-FA eljárásokkal meghatározott első és második faktor keresztábrái
Figure 9 | Crossplots of two statistical factors estimated by the TFA and MFV-FA procedures



10. ábra | A kiugró adatokat tartalmazó adatsorra számított első faktor és a víztelítettség regressziós kapcsolata a TFA (bal oldal) és MFV-FA (jobb oldal) eljárások esetén
Figure 10 | Regression relation between the first factor and water saturation for the case of outlying data given by TFA procedure (left panel) and local inverse modeling (right panel)

szelvény. Előzetes normalitás vizsgálatából kiderült, hogy a *GR* természetesgamma-intenzitástól eltekintve a szelvény adatok nem Gauss-eloszlást követnek, így a TFA eljárást a kiugró adatok jelenléte torzítja a megoldást. A szelvénytípusok eltérő nagyságrendjét figyelembe véve, minden szelvénytípust a súlyozáson túl normáltuk a súlyok összegével (7. ábra). Az ábrán látható, hogy a kiugró adatokhoz (nagyobb adattávolságokhoz) kisebb súlyt rendelt az MFV-FA eljárás. A 8. ábrán megfigyelhető, hogy az iterációs lépések számának növekedésével mind a (standardizált mérési és számított adatokból képzett) eltérésvektor normája, mind a Steiner-súlyok közepes értéke a kezdeti relatíve meredek csökkenésből már néhány lépés után konvergál az optimum felé, ami bizonyítja az MFV-FA eljárás stabilitását. Az MFV-FA módszerrel számolt faktorsúlyok: $-0,01$ (*RCPT*), $0,17$ (*GR*), $0,86$ (*DEN*), $0,73$ (*NPHI*), $-0,88$ (*RES*). A korábbi eredményekhez hasonlóan az első faktor (F_1) és a víztelítettség (*SW*) között mindkét statisztikai eljárás estében erős korreláció figyelhető meg

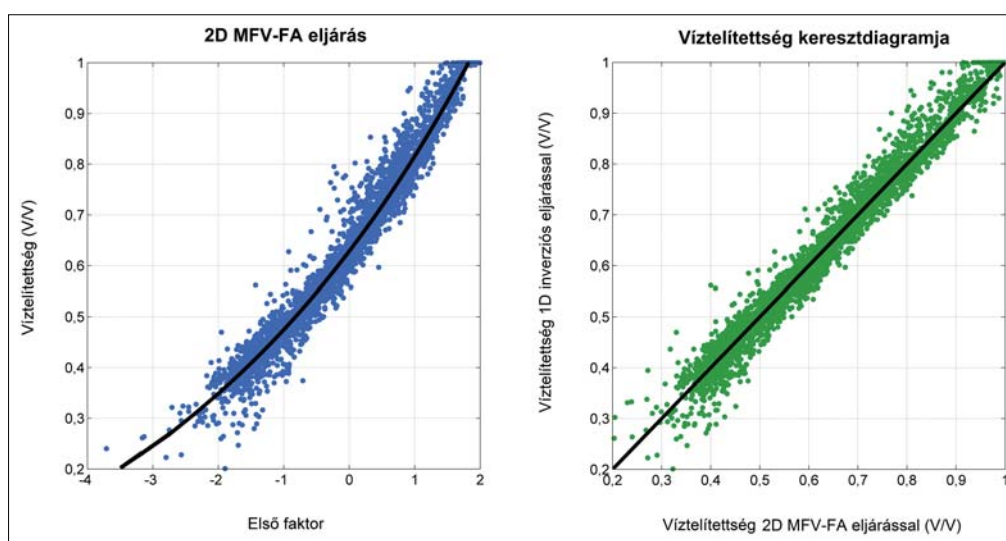
(TFA: $R = 0,96$, MFV-FA: $R = 0,98$). Az exponenciális kapcsolatot jellemző regressziós együtthatók az MFV-FA módszer esetén: $a = 0,77 \pm 0,14$, $b = 0,22 \pm 0,04$, $c = -0,12 \pm 0,11$. A 9. és 10. ábrán jól látható, hogy az MFV-FA eljárás a kiugró adatok hatását hatékonyan elnyomja. A 11. ábra a hagyományos (*FI_TFA*) és az általunk kifejlesztett új robusztus eljárással számított faktorokat (*FI_MFV-FA*) és a víztelítettség-szelvényeket mutatja. Ez utóbbiakat összehasonlíthatjuk a független pontonkénti inverzió eredményével. A hagyományos faktoranalízissel és az inverzióval számított víztelítettség-szelvények eltérése $RMS = 7,4\%$, míg az iteratíván újrásúlyozott statisztikai eljárással az $RMS = 4,6\%$. Észrevehető, hogy a kiugró adatok mélységeiben a zaj kevésbé befolyásolja az MFV-FA eljárás eredményeit. Például a 10–15 m között vagy a 25 m mélység környezetében az első faktor és az abból származtatott víztelítettség-szelvény (*SW_MFV-FA*) nem tartalmaz kiugrásokat, ami a statisztikai eljárás rezisztenciáját és megbízhatóságát bizonyítja.



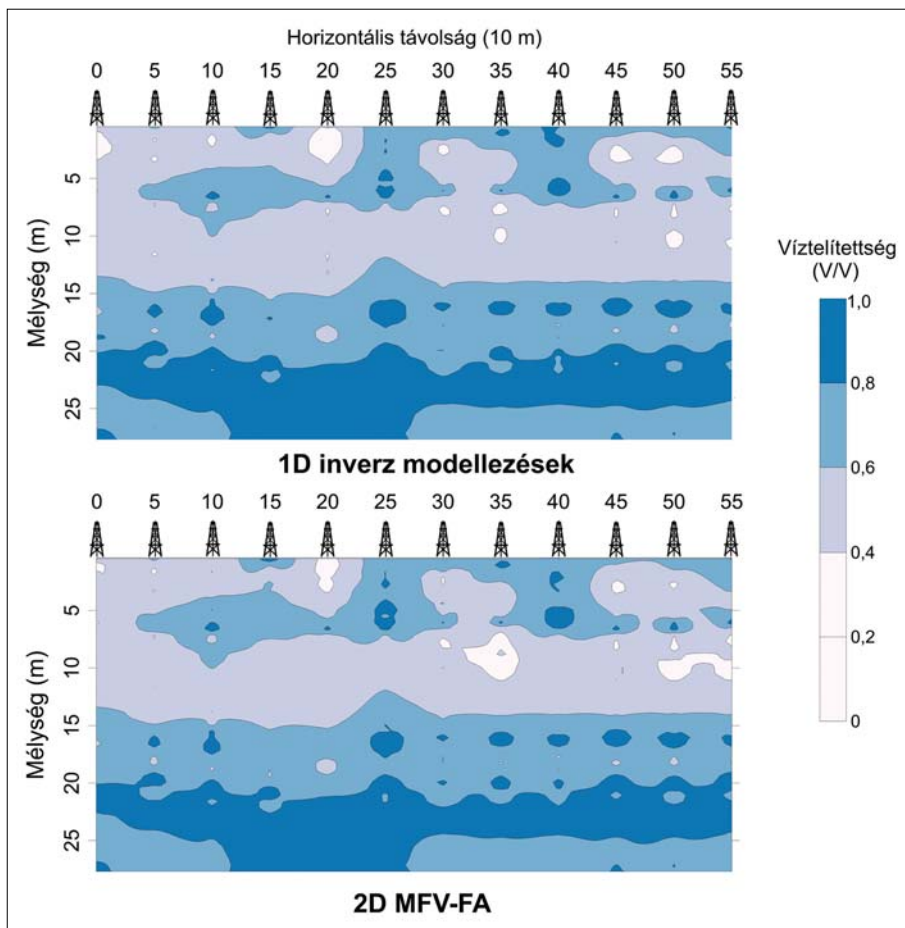
11. ábra A TFA és MFV-FA eljárással meghatározott faktorok és a belőlük számított, illetve a pontonkénti inverzióval becsült víztelítettségi szelvények
Figure 11 Factor logs estimated by the TFA and MFV-FA procedures and water saturation logs estimated by factor analysis and depth-by-depth inversion

A 2D karotázs inverziós módszer (Dobróka et al. 2009) analógiájára új faktoranalízisen alapuló eljárást fejlesztünk. Az MFV-FA eljárás kétdimenziós kiterjesztésére a bátaapáti terület 12 fúrását használtuk fel. Ennek keretében valamennyi fúrás összes adatát a (14) egyenletnek megfelelően egy oszlopvektorba rendeztük. Az adatokból 2D fak-

torszelvényeket képeztünk, majd az első faktort a víztelítettséggel kapcsolatba hoztuk. Az egyedi fúrásokhoz képest a két nagyságrenddel nagyobb statisztikai minta (15500 adat) megbízhatóbb megoldást jelent. A 12. ábra jól mutatja az első faktor és a víztelítettség erős exponenciális, valamint az inverzióval és a faktoranalízis által kapott eredmények szo-



12. ábra Az első faktor és a víztelítettség regressziós függvénykapcsolata (bal oldal), a 2D MFV-FA eljárással és az 1D mélységpontonkénti inverzióval becsült víztelítettség egyezése (jobb oldal)
Figure 12 Regression relation between the first factor and water saturation (left panel), crossplot of water saturations estimated by 1D depth-by-depth inversion and 2D MFA-FA procedures (right panel)



13. ábra 1D pontonkénti inverziós eljárások sorozatából előálló (felső ábra) és a 2D MFV-FA eljárással (alsó ábra) becsült víztelítettség-szelvények

Figure 13 Water saturation sections estimated by a set of 1D local inversion runs (upper panel) and 2D MFV-FA procedure (lower panel)

ros kapcsolatát. A 13. ábrán megjelenő kontúrtérkép segítségével a víztelítettség mindkét irányú, horizontális és vertikális eloszlása is megvizsgálható. Szembetűnő a hagyományos 1D inverziós eredmények alapján szerkesztett és a 2D faktoranalízissel kapott víztelítettség-szelvények egyezése.

Összefoglalás

Mérnök-geofizikai szondázási adatrendszerek kiértékelésére új robusztus statisztikai eljárást javasoltunk. A szelvényadatok együttes feldolgozásával statisztikai faktorokat számítottunk, ami a vizsgált terület valamennyi fúrásában jól korrelált a felszínközeli laza rétegek víztartalmával. A robusztus faktoranalízis algoritmusában tér el a klasszikus módszertől, hogy nemcsak a hibavariációkkal súlyoz, hanem figyelembe veszi az egyes adatok megbízhatóságát is és ahhoz optimális súlyt rendel. A faktor–víztelítettség regressziós kapcsolat alapján független víztelítettség-becslés valósítható meg, ami növelheti a szelvényértelmezés megbízhatóságát. A relatív javulás mértéke az MFV-FA módszer javára a bemutatott példákban legalább 38%-os

volt, ami jelentősnek mondható. A faktoranalízis szolgáltatott eredményeket megerősítettük (független) mélységpontonkénti inverzióval is. A víztelítettség-becslési módszer becslési pontossága a statisztikai minta adatszámának növelésével érhető el. A kétdimenziós faktoranalízis megvalósítását is bemutattuk. A javasolt módszer elősegíti a heterogén konszolidálatlan képződmények közzefizikai modellezését és jól alkalmazható mérnöki és környezeti problémák megoldására.

Köszönetnyilvánítás

A szerző mint a PD109408 sz. Ifjúsági OTKA részt vevő kutatója köszönetet mond az Országos Tudományos Kutatási Alap támogatásáért, emellett dr. Szabó Norbert Péter témavezetőnek illetve dr. Dobróka Mihály professzor úrnak, akik elősegítették a munkáját és megosztották szakmai tapasztalatukat. Hálás köszönet illeti Stickel János igazgató urat és dr. Drahos Dezső tanár urat a szakmai együttműködésükért.

A tanulmány szerzője

Balogh Gergely Pál

Hivatkozások

- Bartlett M. S. 1953: Factor analysis in psychology as a statistician sees it. *Nordisk Psykologi's Monograph Series 3*, Almqvist and Wiksell, Uppsala, pp. 23–34
- Dobróka M., Szabó P. N., Cardarelli E., Vass P. 2009: 2D inversion of borehole logging data for simultaneous determination of rock interfaces and petrophysical parameters. *Acta Geodaetica et Geophysica Hungarica* 44/4, 459–479
- Dobróka M., Szabó N. P. 2012: Interval inversion of well-logging data for automatic determination of formation boundaries by using a float-encoded genetic algorithm. *Journal of Petroleum Science and Engineering* 86–87, 144–152
- Dobróka M., Szabó N. P., Turai E. 2012: Interval inversion of borehole data for petrophysical characterization of complex reservoirs. *Acta Geodaetica et Geophysica Hungarica* 47/2, 172–184
- Dobróka M., Szabó N. P., Tóth J., Vass P. 2016: Interval inversion approach for an improved interpretation of well logs. *Geophysics* 81/2, D163–D175
- Drahos D. 2005: Inversion of engineering geophysical penetration sounding logs measured along a profile. *Acta Geodetica et Geophysica Hungarica* 40, 193–202
- Drahos D. 2008: Determining the objective function for geophysical joint inversion. *Geophysical Transactions* 45/3, 105–121
- Fejes I., Jósa E. 1990: The engineering geophysical sounding method. Principles, instrumentation, and computerised interpretation. In: *Geotechnical and environmental geophysics, 2*. Ward S. H. (ed.), *Environmental and groundwater: SEG*, 321–331
- Gyulai Á., Baracza M. K., Szabó N. P. 2014: On the application of combined geoelectric weighted inversion in environmental exploration. *Environmental Earth Sciences* 71/1, 383–392
- Jöreskog K. G. 2007: Factor analysis and its extensions. In: *Factor analysis at 100, Historical developments and future directions*, Cudeck R., MacCallum R. C. (eds.). Lawrence Erlbaum Associates, New Jersey
- Lawley D. N., Maxwell A. E. 1962: Factor analysis as a statistical method. *The Statistician* 12, 209–229
- Menke W. 1984: *Geophysical Data Analysis: Discrete Inverse Theory*. Elsevier
- Móri T. 1999: Főkomponens- és faktoranalízis. ELTE Valószínűségelméleti és Statisztika Tanszék, egyetemi jegyzet, pp. 1–13
- Scales L. E. 1985: *Introduction to non-linear optimization*. Macmillan Computer Science Series. Springer-Verlag
- Fraiha S. G. C., Silva J. B. C. 1994: Factor analysis of ambiguity in geophysics. *Geophysics*, 59/7, 1083–1091
- Steiner F. 1991: *The most frequent value. Introduction to a modern conception of statistics*. Academic Press, Budapest
- Szabó N. P., 2011: Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophysica* 59, 935–953
- Szabó N. P., Dobróka M., Drahos D. 2012: Factor analysis of engineering geophysical sounding data for water saturation estimation in shallow formations. *Geophysics* 77, WA35–WA44
- Szabó N. P., 2012: Dry density derived by factor analysis of engineering geophysical sounding measurements. *Acta Geodaetica et Geophysica Hungarica* 47/2, 161–171
- Szabó N. P., Dobróka M. 2013: Extending the application of a shale volume estimation formula derived from factor analysis of wireline logging data. *Mathematical Geosciences* 45/7, 837–850
- Szabó N. P., Dobróka M., Turai E., Szűcs P. 2014: Factor analysis of borehole logs for evaluating formation shaliness: a hydrogeophysical application for groundwater studies. *Hydrogeology Journal* 22/3, 511–526
- Xu C., Dai F., Yao X., Zhao Z., Xiao J. 2010: GIS platform and certainty factor analysis method based Wenchuan earthquake-induced landslide susceptibility evaluation. *Journal of Engineering Geology* 2010-01