A főkomponens analízis néhány mélyfúrási geofizikai alkalmazása (kutak közti rétegtani korreláció, karotázs szelvények rétegekre bontása)

*ELEK ISTVÁN**

A főkomponens analízis (Principal Component Analysis) egy, a mélyfúrási geofizikában még alig alkalmazott többváltozós statisztikai módszer. Más tudományágak — mint például a biometria — sikerrel alkalmazzák. A dolgozat célja, hogy bemutassa a főkomponens analízis matematikai alapjait, valamint két eljárást, amelyek nagymértékben támaszkodnak ezen statisztikai módszerre:

a) *A különböző kutakban mért karotázs szelvények alapján rétegtani korreláció adható meg, amely jelentős segítséget nyújthat a geológusoknak. Az ismertetésre kerülő eljárás egy interaktív, félautomatikus korrelációs módszer.*

b) *Néhány példán keresztül bemutatásra kerül egy eljárás, amely tetszőleges szelvényválaszték alapján a felhasználó igényeinek megfelelő nagyságú rétegekre bontást képes elvégezni.*

Анализ главных компонент в ГИС — это многомерный статистический метод, который еще довольно мало применяется. Другие науки, как, например, биометрия, в данное время успешно используются. Цель статьи — показать математическую основу анализа главных компонент, а также два способа, которые в значительной степени опираются на эти статистические методы:

а) *По каротажным диаграммам полученных в различных скважинах можно получить послойную корреляцию, которая может оказать существенную помощь геологам. Способ, с которам мы знакомим в данной статье, — это интерактивный, полуавтоматический метод.*

б) *На нескольких примерах показан способ, при помощи которого на основе произвольно выбранной диаграммы проводят вскрытие пласта такой мощности, которая в данный момент необходима.*

The Principal Component Analysis is a rearly employed multi-variable statistical log interpretation method. It is successfuly used in other sciences e.g. biometry. The aim of this paper is to demonstrate the mathematical base of principal components analysis and to introduce two methods based on these statistical analysis:

a) *Based on logs taken in various borehales identified layers can be correlated for use of geologists. The introduced process is an interactive semi-automatic correlation method.*

b) *On some examples a process is demonstrated to show the correlation width as required by the geologist.*

A főkomponens analízis rövid áttekintése

I. Mikor érdemes főkomponens analízist alkalmazni:

a) Gyakran előfordulhat, hogy valamely változócsoport mögött egy szakmailag jól értelmezhető, közös komponens ismerhető fel, amelyet közös okváltozónak vagy háttérváltozónak tekinthetünk. Ha nem ismerjük fel a háttérváltozót, de több vizsgálat alapján hasonló változócsoportot kapunk, indokolt a háttérváltozóra hipotézist felállítani és kísérletileg ellenőrizni a hipotézis helyességét.

* SZKFI, Budapest

b) A főkomponens analízis egyik nagy előnye, hogy általa csökkentjük a változóink számát, ugyanakkor kitűnik, hogy melyek a jelentéktelen változók. A változók számszerű csökkenése a változók összevonásában nyilvánul meg, így tehát – a probléma természetétől függően – több változó együttes vizsgálata kevesebb számú változó vizsgálatával helyettesíthető.

II. Matematikai alapok:

Az adatrendszerünk legyen a következő szerkezetű: legyen p számú megfigyelési egységünk, amelyek egységenként n számú adatot tartalmaznak (p számú megfigyelési vektorunk van)

$X^{(1)}$	$X^{(2)}$	.....	$X^{(p)}$
$x_1^{(1)}$	$x_1^{(2)}$	.....	$x_1^{(p)}$
$x_2^{(1)}$	$x_2^{(2)}$	.....	$x_2^{(p)}$
$\vdots$			
$x_n^{(1)}$	$x_n^{(2)}$	.....	$x_n^{(p)}$

Tekintsük az $\mathbf{x}^{(j)}$ vektorokat valószínűségi változóknak, a vektorok elemeit a valószínűségi változók egy realizációjának. Így tehát minden megfigyelési egység megfelel egy vektor valószínűségi változónak. Mivel a megfigyelési egységek különböző dimenziójú fizikai mennyiségek is lehetnek, ezért standardizáljuk a változóinkat:

$$\underset{\sim}{\mathbf{x}}_i^{(j)} = \frac{x_i^{(j)} - \bar{x}^{(j)}}{s^{(j)}}$$

ahol $\bar{x}^{(j)}$ az j-edik vektor elemeinek átlaga $s^{(j)}$ az empirikus szórása.

Így dimenziótlanná, 0 várható értékűvé, 1 szórásúvá tettük a vektorokat.

Ezek után számítsuk ki az adatrendszerünk korrelációs mátrixát

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & & & \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix}$$

ahol

$$r_{ij} = \mathrm{Korr}\,(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})\ i, j = 1, p$$

Határozzuk meg a korrelációs mátrix sajátértékeit és a hozzájuk tartozó sajátvektorokat. Legyen $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_p|$ a mátrix sajátértékei, és $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots$ $\ldots \mathbf{v}^{(p)}$ a hozzájuk tartozó sajátvektorok.

A j-edik főkomponens a következőképpen számítható ki:

$$C_i^{(j)} = x_i^{(1)} \cdot v_1^{(j)} + x_i^{(2)} \cdot v_2^{(j)} + \ldots + x_i^{(p)} \cdot v_p^{(j)} \qquad \begin{array}{l} i = 1, n \\ j = 1, p \end{array}$$

Amint az látható, maximum annyi főkomponenst tudunk számítani, mint ahány megfigyelési egység áll rendelkezésünkre.

Az így kapott főkomponensek néhány tulajdonságát érdemes megemlíteni:

A főkomponensekből képzett korrelációs mátrix diagonális, azaz a főkomponensek korrelálatlanok. (Átlagértékük nulla, szórásnégyzetük pedig a megfelelő sajátértékekkel egyenlő).

$$\mathbf{R}_C = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \cdot & \\ & & & \cdot \\ & & & \cdot \\ & 0 & & \lambda_p \end{bmatrix}$$

A sajátvektorok a következő tulajdonságokkal rendelkeznek:

normáltak, azaz $<\mathbf{v}^{(i)}, \mathbf{v}^{(j)}> = \delta_{ij}$ ($\langle\rangle$) jel skaláris szorzás, δ_{ij} Kronecker delta. A sajátvektorokból képzett mátrix transzponáltjából vett oszlopvektorokra is ugyanez igaz.

Vizsgáljuk meg az $\mathbf{R}_C$ mátrixot. Belátható, hogy igaz a következő:

$$\sum_{j=1}^{p} \lambda_j = p = \sum_{i=1}^{p} s_i^2 = \sum_{j=1}^{p} s_j^2 = p$$

azaz a standardizált változók és a főkomponensek szórásnégyzeteinek összege, azaz az összvarianciája azonos.

III. A főkomponens analízis szemléletes jelentése:

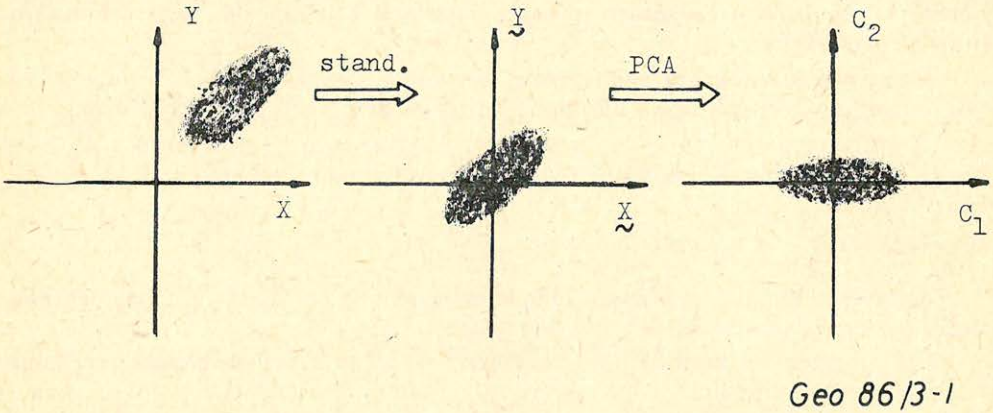
Amint az a fenti tulajdonságokból látszik, a főkomponens analízissel átrendeztük a varianciákat. Míg a standardizált változók mindegyikének variancia szempontból azonos a jelentősége, addig a főkomponensek közül az első magában foglalja az összes változó varianciájának legnagyobb közös részét. A második főkomponens a megmaradt varianciák legnagyobb közös részét tartalmazza, és így tovább. Az utolsó főkomponensre már alig marad varianciarész, ezért szokták ezeket jelentéktelennek tekinteni. Ebből látható az, hogy számos esetben ele-

gendő az első (vagy az első néhány) főkomponenst figyelembe venni egy adott probléma megoldásához. Az a kérdés viszont, hogy mikor mondható valamely főkomponens jelentéktelennek, a probléma jellegétől, a vizsgálat céljától stb. függ.

A főkomponens analízis geometriai jelentése

A szemléletesség kedvéért tekintsünk csak két változót (x, y).

Az *1. ábráról* leolvasható, hogy amit csináltunk az nem más, mint egy koordináta transzformáció (főtengelytranszformáció), azaz a standardizált adatrendszert olyan koordináta-rendszerbe transzformáltuk, ahol a transzformáltak korrelációs mátrixa diagonális.



1. ábra
Рис. 1.
Fig. 1.

Kutak közti rétegtani korreláció

Egy adott terület különböző kútjai között a rétegek követése egyik kútból a másikba fáradságos, hosszadalmas munka. A gyakorlatban főként az *1 : 1000* ellenállás szelvényt használják, bár előfordul más szelvények figyelembe vétele is, főként komplikáltabb rétegződés esetén. A rutinszerű korreláció azonban szinte sohasem használja föl valamennyi karotázsszelvényt. Nézzünk meg egy olyan ideális esetet, amikor minden karotázsszelvényt föl akar használni a korrelációt végző geológus. Ehhez három – négy (de nem ritkán hat) szelvényt is egyszerre kell figyelnie. Ha a vizsgált kútszakasz *4 – 500* m hosszú, a korreláció rendkívül fáradságos, nehézkes és hosszadalmas, viszont felhasznál minden információt. Az ellenállás szelvény alapján végzett korreláció gyorsabb, kevésbé fáradságos, viszont nem vesz figyelembe minden információforrást ami rendelkezésre áll.

Ezért célszerű egy olyan kompromisszumos megoldást választani, amely az egyetlen szelvénnyel végzett korreláció előnyeit megtartja és minden lényeges információt felhasznál.

Erre nyújt megoldási lehetőséget a főkomponens analízis. Számítsuk ki az összes karotázs szelvény első főkomponensét. Ezzel kapunk egy olyan dimenziót-

lan „karotázs szelvényt", amely a felhasznált szelvények varianciáinak legnagyobb közös részét fogja tartalmazni. Ezáltal az információvesztés mértéke sokkal kisebb lesz, mint ha csak az ellenállás szelvényt használtuk volna föl, viszont változatlanul kihasználható szemünk kiváló alakfelismerő képessége, ami a jelenlegi korrelációs módszerek alapja.

Az eljárás gépi megvalósítása:

Az eljárás kihasználja az SZKFI Geofizikai Főosztályán rendelkezésre álló *HP 9845/B* típusú számítógép fejlett grafikai adottságait, de bármely más jó grafikával rendelkező gépen könnyen megvalósítható.

A grafikus display rajzmezője két részre van osztva, az alsó részen az egyik kút első főkomponense van felrajzolva, a felső részen a másik kút első főkomponense. A cursorral az összetartozó mélységpontok kijelölhetők, amelyeket tárolunk a háttértárban.

Egy rajzoló program segítségével kutanként megjeleníthetők az első főkomponens szelvények és a különböző kutak összetartozó mélységpontjai.

A grafikus display a *2. ábrán* látható. A *3. ábrán* a Battonya – É – 4 – 5 – 6 kutak korrelációjának eredménye látható és a bemenő szelvények.

Karotázs szelvények rétegekre bontása:

Ezen probléma megoldására már létezik néhány megoldás, amely megtalálható a szakirodalomban. ([4], [7], [8], [10])

Az ismertetésre kerülő eljárás egyetlen szelvény figyelembevételével képes réteghatárokat kijelölni. Több szelvény egyidejű használatát a főkomponens analízis teszi lehetővé.

Először ismertetjük az egy szelvényen alapuló módszert, majd kiterjesztjük tetszőleges szelvényválasztékra.

Döntsük el, hogy mekkora az a legkisebb réteg, amelynek kimutatását még elvárjuk az eljárástól. Hajtsunk végre mediánszűrést a szelvényünkön a minimális rétegvastagságnak megfelelő ablakhosszal. Ezzel az élmegörző szűrési eljárással kapunk egy helyenként szögletes görbét. Ennek szögleteit eltüntethetjük egy megfelelően választott felső határfrekvenciájú szűréssel. Ennek eredménye látható a *4. ábrán.*

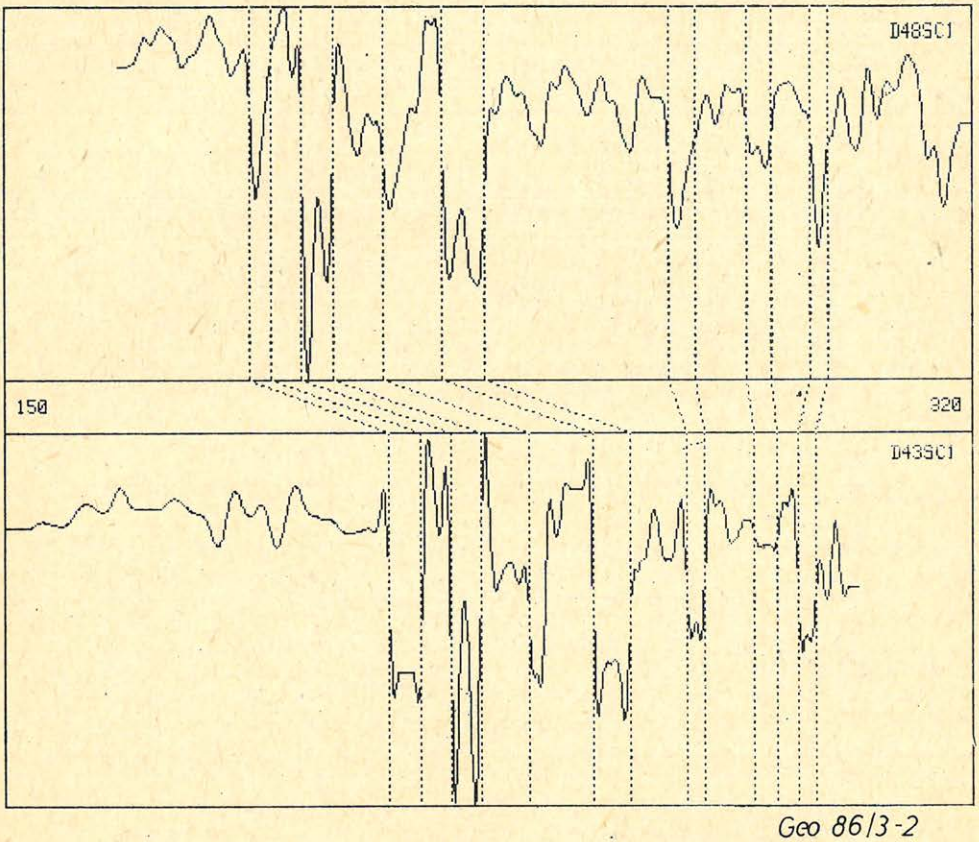
Ezek után a simított szelvényen jelöljük meg az összes csúcs és völgy között a legmeredekebb pontokat, mint a keresett réteghatárokat.

Tekintsük egy rétegen belül a szelvények értékeit állandónak, és nevezzük ezt az értéket a rétegre jellemző értéknek. Ezzel a valóságot egy olyan modellel helyettesítjük, amely feltételezi, hogy egy rétegen belül a fizikai tulajdonságok homogének, a határokon ugrásszerűen változnak, és a mérési eredmények nem „szögletes" voltát a mérőműszerek véges felbontóképessége okozza.

A jellemző érték megállapítása a következő igen egyszerű módon történik:

a) ha a rétegen belül a maximum a szélső pontoktól eltérő helyre esik, akkor a rétegjellemző értéke a maximum.

b) ha a minimum a szélső pontoktól eltérő helyre esik, akkor a jellemző érték a minimum.



2. ábra. A Demjén – K – 443 és 448 kutak első főkomponensei láthatók az alsó, ill. felső rajzterületen, valamint az összetartozó pontok, amelyeket a szaggatott vonalak jelölnek. Az ábra a grafikus display-ről készített hard-copy.

Рис. 2. Первые главные компоненты скважин Демьен – К – 443 и 448 изображенные в верхней и нижней части рисунка, прерывистые линии изображены пунктиром.

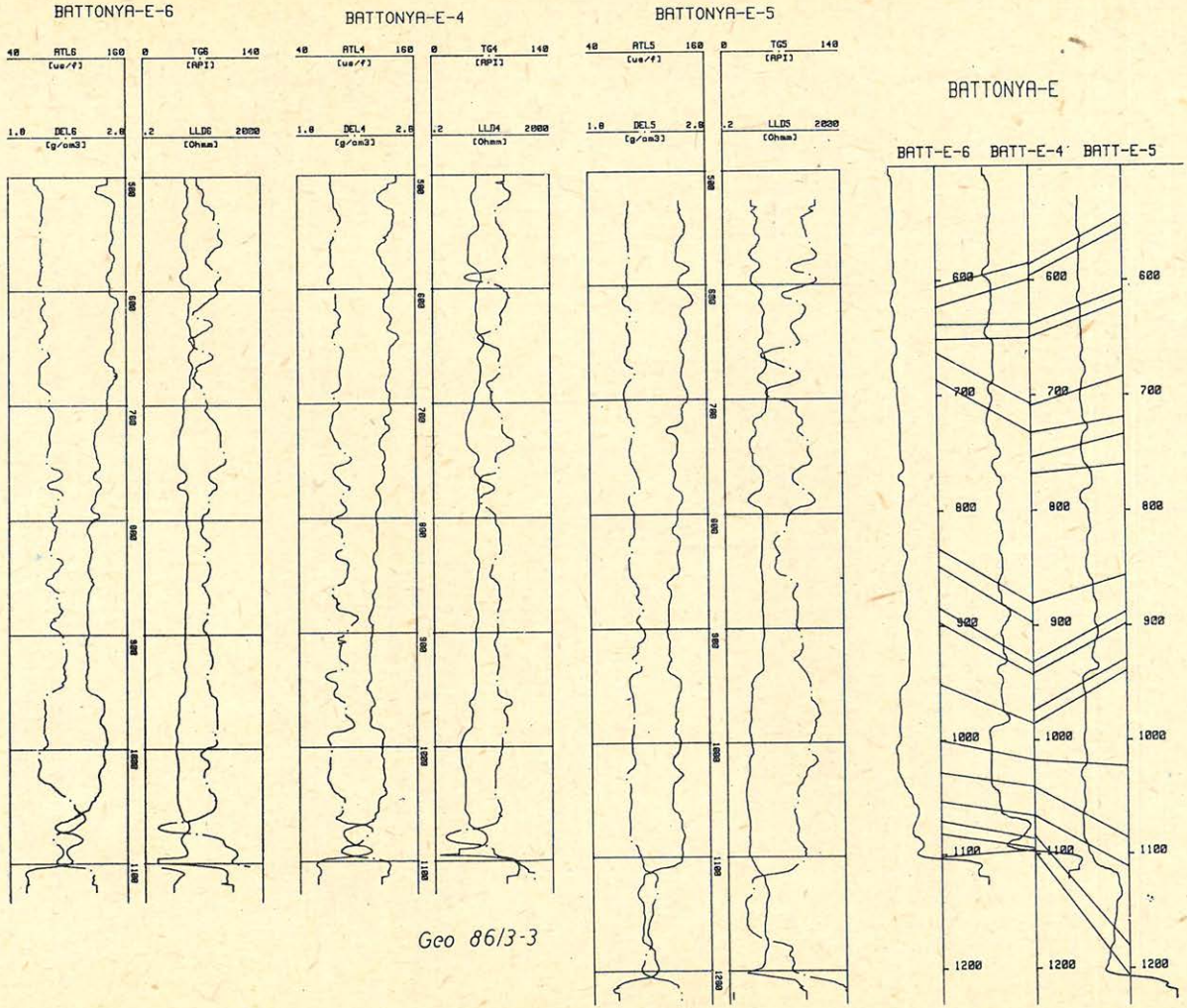
Fig. 2. Principal components of wells Demjén – K 443 and 448 seen on the upper and lower part. Dashes show the corresponding points. The hard copy was prepared on the garphic display.

(Az eljárás természetéből adódóan olyan eset nem fordulhat elő, hogy a szélső pontokra essen akár a maximum akár a minimum.)

A jellemző értékképzésre mutat példát az *5. ábra.*

Az így kapott rétegek elég finom felbontásúak, ezért fölmerülhet az igény nagyobb egységek kijelölésére, amit a meglevő réteghatárok szelektálásával oldhatunk meg.

Erre a problémára egy lehetséges megoldási mód a következő: adjunk meg előre egy kritikus értéket, amely a rétegzettség finomságára lesz jellemző (tapasztalatilag kell megállapítani).

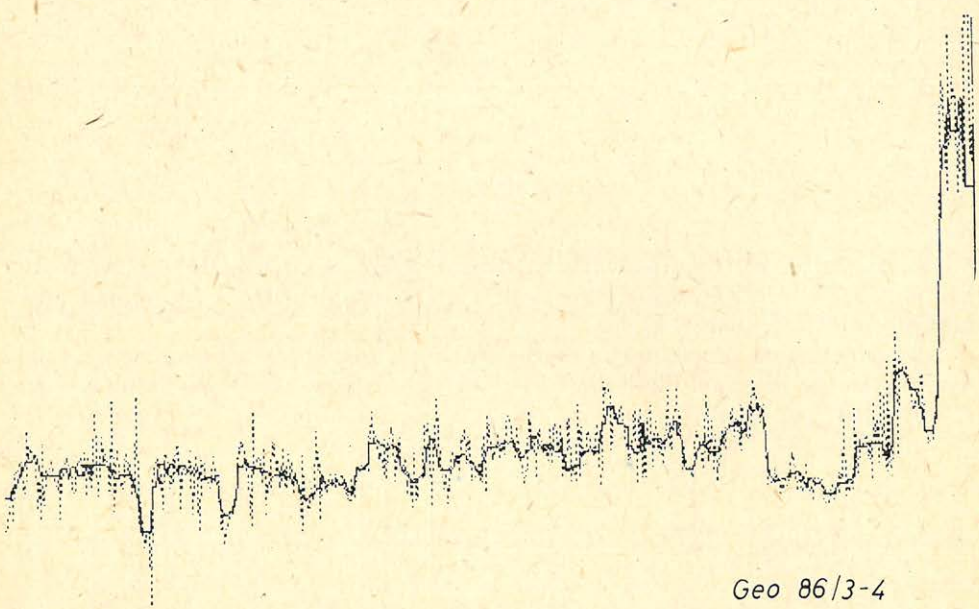

BATTONYA-E-6
BATTONYA-E-4
BATTONYA-E-5
BATTONYA-E
BATT-E-6
BATT-E-4
BATT-E-5
ATL6
TG6
DEL6
LLD6
ATL4
TG4
DEL4
LLD4
ATL5
TG5
DEL5
LLD5
[us/f]
[API]
[g/cm3]
[Ohmm]
600
700
800
900
1000
1100
1200

Geo 86/3-3

3. ábra. A Battonya – É – 4 – 6 – 5 kutak szelvényei és a korreláció eredménye. A kutak rajzon látható távolságai nem arányosak a valódi távolságokkal. A három fúrás közel egy egyenesbe esik. Az ábrán látható szelvények az áttekinthetőség kedvéért 10 m-es ablakhosszal mediánszűrt és simított szelvények. A korrelációt ábrázoló oszlopban a kutakhoz tartozó első főkomponensek vannak felrajzolva. Az összetartozó mélységpontok az egyes fúrások tengelyére mint mélységtengelyre vannak vonatkoztatva.

Рис. 3. Диаграммы скважин Баттоня – Е – 4 – 6 – 5 и результаты корреляции. Расстояние между скважинами, показанными на рисунке, условно. Три буровые приблизительно попадают на одну прямую. Изображенные на рисунке диаграммы в целях наглядности подвергнуты медианной фильтрации и сглаживанию при длине окна 10м. В графе, где изображена корреляция, отмечены первые главные компоненты относящиеся к скважинам. Соответствующие глубинные точки относятся к осям единичных скважин, как к осям глубин.

Fig. 3. Profiles of wells Battonya – É 4 – 6 – 5 and results of correlation. Well distances are not real. All three wells are lined up approximetaly along a straight line. Profiles shown were filtered with a 10 Meters window length and smoothed. Correlation coloums contain the first principal well components. Common depth points are corrected to the various well axis.



4. ábra. A szaggatott görbe az eredeti, szűretlen görbe, a folytonos görbe a mediánszűrt majd simított görbe.

Рис. 4. Прерывная кривая, кривая без фильтрации, непрерывная кривая после медианной фильтрации и сглаживания.

Fig. 4. Dashed line is the original unfiltered curve. Continuous line is median filtered smoothed curve.

Tegyük fel, hogy már benne vagyunk egy „nagy" rétegben az i-edik adatig (az i a réteg kezdetétől számítandó).

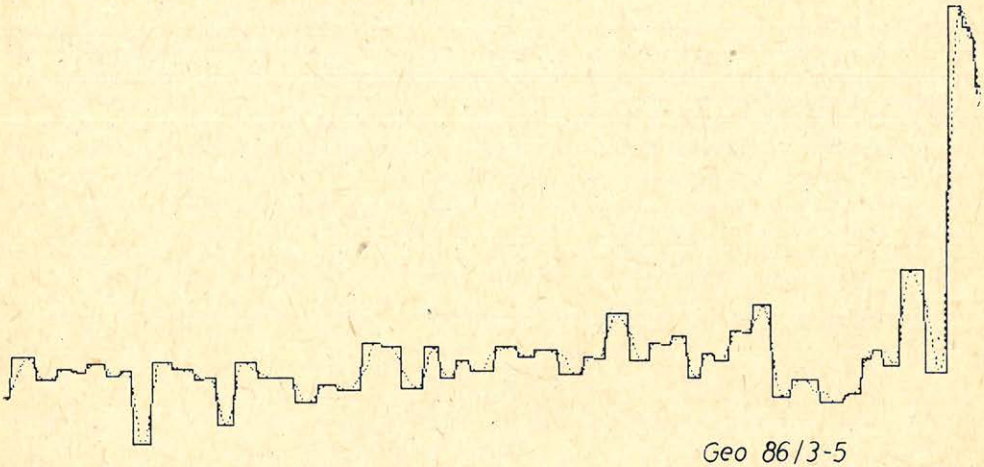
Az i-edik adatig a szelvény átlagértéke m.

$m = \frac{1}{i} \sum_{j=1}^{i} a_j$ ahol a_j a szelvény j-edik pontban felvett értéke.

Vizsgáljuk meg az a_{i+1}-dik adat és m viszonyát:

ha $|a_{i+1}-m| \geq$ kritikus érték$/(i+1)^{1/2}$, akkor van réteghatár az $i+1$-edik, helyen,

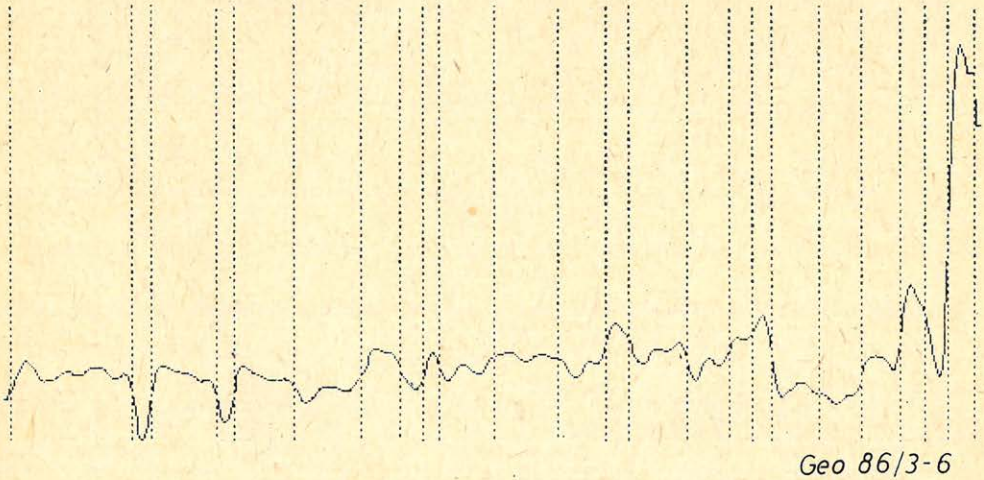
ha $|a_{i+1}-m| <$ kritikus érték$/(i+1)^{1/2}$ akkor nincs réteghatár.



5. ábra
Рис. 5.
Fig. 5.

A kritikus érték változtatásával a felbontott rétegek *markánssága* is változik, azaz a kritikus érték növekedtével csak a markánsabb változások fognak megmaradni mint réteghatárok.

Az ismertetett procedúrát alkalmazhatjuk a már kijelölt finom rétegek jellemző érték szelvényére. Ezáltal a finom rétegek határai közül elhagyjuk a jelentékteleneket. *(6. ábra)*



6. ábra
Рис. 6.
Fig. 6.

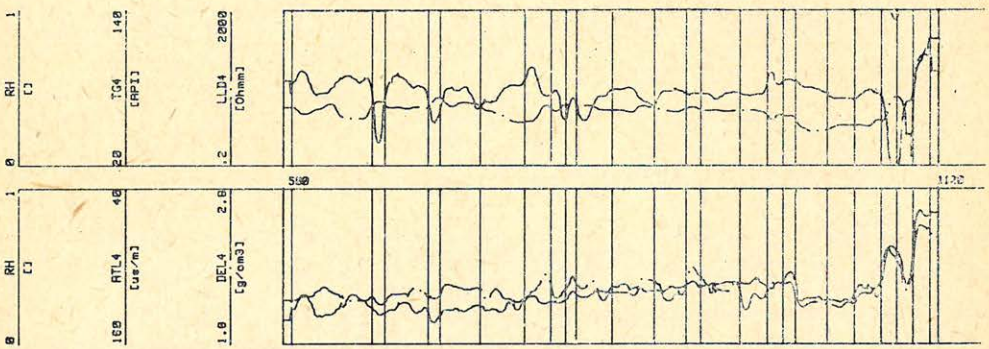
Az eddig ismertetett eljárás csak egyetlen szelvényt használ föl a réteghatár-kijelölésre. Azonban semmi okunk nincs arra, hogy kitüntessünk bármely karotázs szelvényt, és azt állítsuk, hogy valamelyik szelvény legyen a réteghatár kijelölés alapja. Célszerű volna minden nem technikai szelvényt fölhasználni. A szelvényenként történő réteghatár-kijelölés sajnos nem teszi lehetővé, hogy egy egységes, minden szelvényre igaz réteghatárképet hozzunk létre. Ennek oka, hogy a magyarországi szelvényezési gyakorlat nem néhány szondaszerelvény felhúzásával végzi a méréseket, hanem az egyes szondákat külön-külön húzza föl. Emiatt a mérési anyagon mélységeltérések vannak, amelynek oka a kábelnyúlás, az eszköz megszorulása stb. Ez tehát azt jelenti, hogy a valóságban azonos mélységbe eső pontok a regisztrátumokon nem esnek azonos mélységbe.

A fent ismertetett eljárás szelvényekénti alkalmazása tehát a különböző szelvényeken más-más helyre fogja rakni a réteghatárokat és ezek utólagos egyeztetése reménytelennek látszik, különösen nagyobb rétegek esetén. A szelvényenként kijelölt réteghatárok még a leggondosabb mélységegyeztetés után sem közösíthetők.

Az egységes réteghatárkép kialakítására nyújt lehetőséget a főkomponens analízis (I/b. pont).

Hajtsunk végre főkomponens analízist a standardizált szelvényeinkre. Ekkor annyi darab főkomponens szelvényt kapunk, mint ahány bemenő szelvényünk volt. E szelvények varianciájának legnagyobb közös részét az első főkomponens fogja magába sűríteni. Így célszerűnek látszik csak az első főkomponensre alkalmazni az ismertetett réteghatárkijelölő-eljárást, az összes szelvény helyett. Ezáltal létre tudunk hozni egy minden szelvényre igaz réteghatárképet. A bemenő szelvények kisebb mélységeltérései nem okoznak durva hibát, az így kapott réteghatárok többnyire minden szelvényre megfelelők.

Egy főkomponens analízissel kombinált réteghatár kijelölés eredménye látható a *7. ábrán.*



Geo 86/3-7

7. ábra
Рис. 7.
Fig. 7.

Összefoglalás

Az ismertetett két alkalmazás csak ízelítőt kívánt adni a főkomponens analízis használhatóságáról. Az ismertetett eljárások nem tekinthetők befejezettnek, noha a kitűzött feladatokat egy bizonyos szinten megoldják. A kutak közti korreláció megoldásában jelentős javulást hozhat a korrelációs procedúra alakfelismerésen alapuló automatizálása. Az elkészült programok már jelenlegi állapotukban is hathatós segítséget nyújthatnak a geológus kollégáknak, különösen olyan esetekben, amikor a vizsgált területen a rétegződés bonyolult, a rétegek követése egyik kútból a másikba nehezen valósítható meg. A kutak közti korreláció a szeizmikus interpretátorok munkáját is segítheti.

IRODALOM

[1] *Serra, O.* and *H. T. Abbot:* The contribution of logging data to sedimentology and stratigraphy 1982. SPE 9270

[2] *Carlos A. Pomalaza-Raez* and *Clare D. McGillen:* An adaptiv edge-preserving filter IEEE Transaction on Acoustics, speech and signal processing vol. ASSP – 32, No. 3, june 1984

[3] *Jan Biemond* and *Jan J. Gerbrands:* An edge-preserving recursive noise-smoothing algorhitm for image data IEEE Transaction systems man and cybernetics vol. smc. 9 No. 10. october 1979.

[4] *Elek István* és *Kovács György:* A Walsh-transzformáció alkalmazása a mélyfúrási geofizikában. Alkotó Ifjúság Pályázat, SZKFI, 1984.

[5] *Dr. Sváb János:* Többváltozós módszerek a biometriában, Mezőgazdasági Kiadó, 1984.

[6] *Donald F. Morrisson:* Multivariate Statistical Methods Mc Graw-Hill New York, 1979.

[7] *Vermes Mátyás:* Rétegsor-meghatározás karotázsszelvények számítógépes feldolgozásával, Magyar Geofizika 1984. 5 – 6.

[8] *Vermes Mátyás:* Szeizmikus rétegmodell számítása akusztikus karotázs mérések alapján, Magyar Geofizika 1983. 4. sz.

[9] Proc. Schlumberger Well Evaluation Conference, Algiers (dec. 1979.)

[10] *Lannig E. N.* and *Johnson D. M:* Automated identification of rock boundaries: an application of the Walshtransform to geophysical well-log analysis, GEOPHYSICS, v. 48, february, 1983.

СОДЕРЖАНИЕ

Турай Э: Математическое моделирование электромагнитного поля плоской волны с использованием интегрального уравнения 1

Тотх Й: Роль ГИС и каротажной интерпретации в поисках углеводородных месторождений с плохими коллекторскими свойствами 17

Элек И: Применение анализа главных компонент в ГИС 26

CONTENTS

Turai E: Mathematical modeling of planewave electromagnetic spaces by integral equations 1

Tóth J: Role of up-to-date well logging and log interpretation in the exploration of a hydrocarbon field having bad reservoir properties 17

Elek I: Some borehole geophysical application of the principal component analysis (layer correlation betwen boreholes, layer interpretation of logs) 26

MAGYAR GEOFIZIKA

A szerkesztésért felelős: Zelei András

A szerkesztőség címe: 1368 Budapest VI., Anker köz 1. Telefon: 429-754

Kiadja: a Delta Szaklapkiadó és Műszaki Szolgáltató Leányvállalat 1442. Budapest VII., Garay u. 5.

Telefon: 415-583, 215-440

Felelős kiadó: Faklen Pál igazgató

86.134. Állami Nyomda, Budapest – Felelős vezető: Mihalek Sándor igazgató

Terjeszti a MAGYAR GEOFIZIKUSOK EGYESÜLETE

Egyesületi tagoknak tagdíj ellenében

Megjelenik évente hatszor

Index: 26 507