

## A talajszerkezeti és spektrális tulajdonságok összefüggésének meghatározása

**Kalmár János**  
MTA CSFK GGI  
kalmar@ggki.hu

**Neményi Miklós**  
Széchenyi István Egyetem  
nemenyi.miklos@sze.hu

**Nyéki Anikó Éva**  
Széchenyi István Egyetem  
nyeki.aniko@sze.hu

**ÖSSZEFOGLALÓ.** A mezőgazdasági termőtalajok hasznosítása és művelése a talajösszetétel alapján történik. A talajparaméterek meghatározása helyszíni mintavétellel és a talajminták laboratóriumi elemzésével történik. Ez költséges és hosszadalmas eljárás, ezért megvizsgáltuk, helyettesíthető-e hiperspektrális (digitális) felvételek készítésével és kiértékelésével.

**ABSTRACT.** Utilization and cultivation of agricultural soils is done on the basis of the soil composition. Soil parameters are determined by on-site sampling and by laboratory analysis of soil samples. This is a costly and lengthy procedure, so we have examined whether it can be replaced by making hyperspectral (digital) images and evaluating them.

### 1. Bevezetés

A precíziós mezőgazdaság egy technológia, amelyik úgy segíti a mezőgazdasági termelés optimalizálását, hogy a művelési táblákat kisebb egységekre lebontva „sűrű” mérési adatok alapján támogatja a döntés-előkészítést.

Két parcellán folytattunk adatgyűjtést, összesen 21 mintavételi helyen. A talajmintákat laboratóriumban értékeltük ki 17 összetevőre (talajparaméterre) vonatkozólag, melyből a sótartalmat állandósága miatt elhagytuk. A mintavételi helyeken hiperspektrális felvételeket is készítettünk a (350, 2500) nm hullámhosszú tartományban, 1 nm felbontással [1], [2].

Hiperspektrális felvétel műszerhiány miatt csak 13 mintavételi helyen készült, viszont pontonként két példányban: az egyik felvétel a bolygatlan (normál) talajfelszínről, a másik pedig a simítottól készült. Műszerhiba miatt az 1835 nm és 1905 nm hullámhosszú reflexiók közül 34 darab simított és 39 darab normál talajról készült adatot töröltünk.

## 2. A hiperspektrális képek vizsgálata

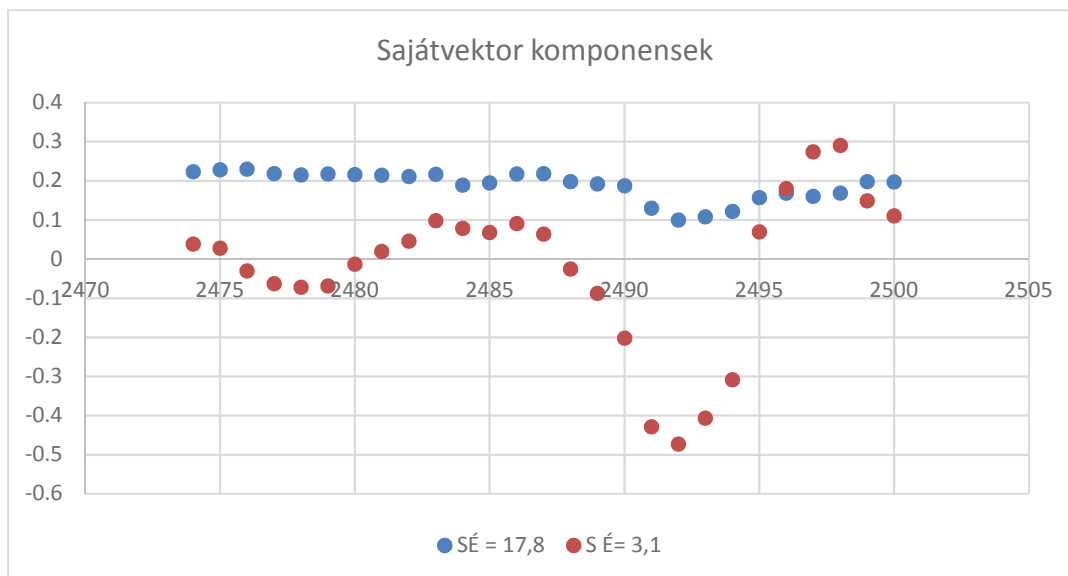
Először összehasonlítottuk a hiperspektrális reflexiók szomszédos hullámhosszait: várakozásainknak megfelelően az kaptuk, hogy a reflexiók korrelációja a szomszédos hullámhosszakon nagyon magas, csak az esetek 5 %-ban (106 mérési ponton) volt kisebb 0,99-nél. Elgondolkodtató eredményre vezetett a normál és simított reflexiók elkülönített összehasonlítása: a szomszédos adatsorok korrelációinak átlaga ugyan alig különbözik (0,9961 – 0,9963), viszont a korrelációk szórása (0,054 – 0,049), és különösen a minimuma már szignifikánsan eltér (0,304 – 0,594), ami a simított talajról felvett adatsorok előnyét (robosztusságát) mutatja.

Az un. főkomponens analízis (továbbiakban PCA) képes az adatsorok számát variancia-vesztés nélkül csökkenteni, ezért főkomponens analízissel (PCA) vizsgáltuk meg a spektrális adatsorok 20 nm szélességű összefüggő (szomszédos) nyalábjait. A PCA az adatsorok keresztkorrelációs (vagy kovariancia) mátrixán alapul, mely mátrix sajátértékei mutatják a főkomponensek fontosságát, a főkomponenseket pedig az eredeti adatsorok olyan lineáris kombinációjaként állítja elő, ahol a súlyok a sajátvektor komponensei. Tudjuk, hogy a kovariancia mátrix pozitív szemidefinit, ezért minden sajátértéke pozitív, és a sajátértékek összege megegyezik a mátrix méretével.

- Vizsgálatunkban a kovariancia mátrix sajátértékeinek összege elméletileg 20, de a legfontosabb komponens sajátértéke általában közel húsznak (>19) adódott, ezért a többi főkomponens szerepe elhanyagolható.
- A legnagyobb sajátértékhez tartozó sajátvektor koordinátái közelítőleg egyformák voltak, ami a sajátérték dominanciájával együtt azt jelenti, hogy a szomszédos 20 spektrális adatsor információ-vesztés nélkül helyettesíthető az átlagukkal.
- A kivételek közé tartozott a (2474, 2500) nm hullámhosszú intervallum: az itt elvégzett PCA szerint a 27 nm széles spektrális adatsort legalább 2 főkomponens adatsorral (sajátértékeik 17,8 és 3,1) kell helyettesíteni könyökszabály értelmében (1. ábra):
  - A domináns sajátvektor komponensei most is közel egyenlőnek bizonyultak (átlag: 0.19, szórás: 0.037, 2. ábra)
  - A második sajátvektor komponensei már nem egyformák, és előjelben is igen változatosak, a sajátvektorok egyben bázisvektorok is, ezért skaláris szorzatuk 0.



1. ábra. A domináns sajátvektorokat a sajátértékekre alkalmazott könyökszabály jelöli ki



2. ábra. (2474 - 2500) nm hullámhosszokhoz számolt PCA két domináns sajátvektora

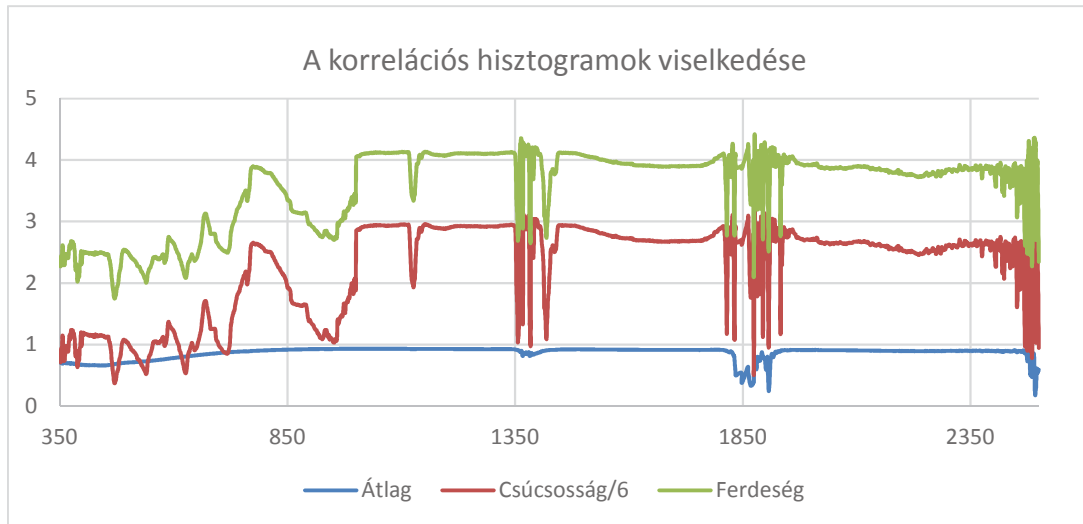
A tapasztaltak miatt tehát nem állítható az, hogy minden rövidebb spektrum-tartományban mért reflexiókat variancia-vesztés nélkül helyettesíthetünk a reflexiók átlagával, de a domináns sajátvektor komponensei közel egyenlőnek bizonyultak, vagyis a legfontosabb helyettesítő adatsor (lineáris kombináció) mindig az átlag.

Ezután meghatároztuk a 2120 db spektrális adatsor keresztkorrelációs mátrixát. A mátrix minden sorából (tehát hullámhosszanként) egy 20 oszlopos hisztogramot (gyakoriság függvényt) készítettünk. A 20 pontos gyakoriság függvényeket az átlag, ferdeség és csúcosság/6 statisztikákkal jellemeztük (3. – 4. ábrák).

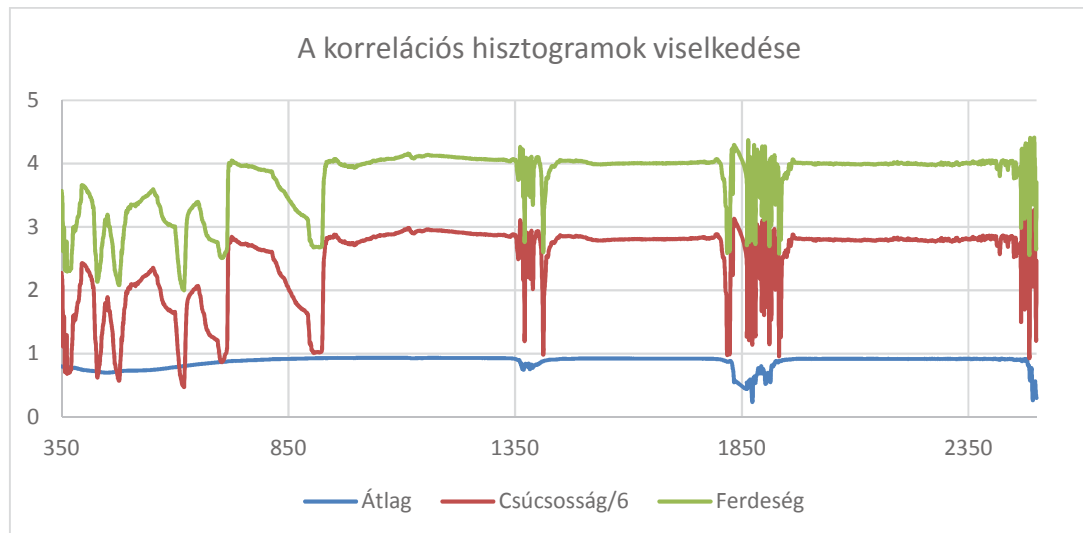
- A statisztikák diagramja rámutatott arra, hogy a görbék többnyire nyugodt, stabil szakaszokból állnak, a spektrum elején (kb. 1000 nanométerig) - az átlag kivételével - növekvő hullámzást mutatnak. Csak a simított képeken látunk intenzív kilengést az 1115-1150 nanométer hullámhossz tartományban, viszont 1350 és 1450 nanométer hullámhosszok között, 1800 és 1950 nanométer hullámhosszok között, és a spektrum végén (2400 nanométer fölött) mindegyik statisztika intenzív kilengést mutat.
- A hisztogramok a korrelációs mátrix sorai alapján készültek, ezért a kilengések arra utalnak, hogy a spektrum csak a jelzett hullámhossz tartományokban érzékeny a mérési pozícióra, tehát a talaj tulajdonságainak változására.

### 3. A talaj-összetevők laboradatainak vizsgálata

- A mért 16 talajparaméterre is főkomponens analízist alkalmaztunk, hogy feltárjuk a közöttük található rejtett összefüggéseket.
- A könyökszabály szerint (5. ábra) a laboradatok már három főkomponenssel leírhatók.
- A 6. ábrán láthatók a legnagyobb sajátértékekhez tartozó sajátvektorok.
- A 7-8. ábrákon a talajösszetevők már fontosságuk (sajátértékkel szorzott sajátvektor komponensek összegeként meghatározott) sorrendjében láthatók.



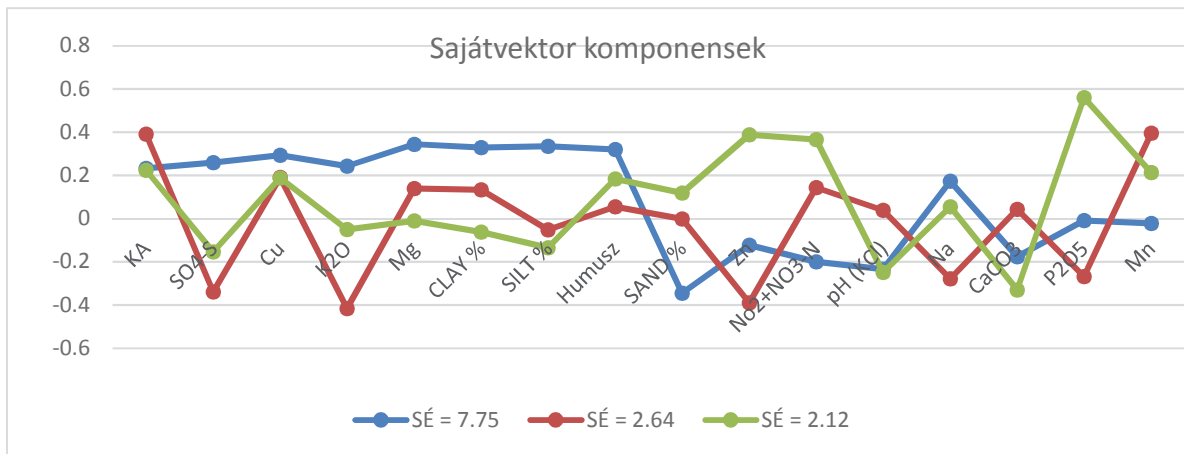
3. ábra. A simított talajú képek statisztikái



4. ábra. A normál talajú képek statisztikái



5. ábra. A domináns sajátvektorokat a sajátértékekre alkalmazott könyökszabály jelöli ki



6. ábra. A laboradatokból számolt PCA legfontosabb sajátvektorai

#### 4. A hiperspektrális és a laboradatok kapcsolata

A képi és laboradatok kapcsolatát úgy határoztuk meg, hogy korrelációt számítottunk minden hullámhossz-reflexió és minden labor adatsor (a komponensek mérési-hely függők) között. A korrelációk csak szűk hullámhossz tartományokban mutattak (abszolút értékben) 0,5-nél nagyobb értékeket, ezért a spektrummal kapcsolatos korábbi PCA-k alapján feltételeztük, hogy a spektrális adatok legfeljebb tíz hullámhossznyi összefüggő tartománya helyettesíthető a spektrális adatok átlagával. Minden labor-adatsorhoz megkerestük azon 1, 2, ..., 10 hosszú hullámhossz tartományt, ahol a labor-adatsor és a reflexiók átlagának korrelációja (abszolút értékben) maximális volt. A 10 maximum tartomány közül választottuk ki ezután a legjobb korrelációjút. Az eredményeket az 1. táblázatban foglaltuk össze, és a 7. – 8. – 9. ábrákon szemléltettük – természetesen a normál és a simított talajról készült felvételeket végig megkülönböztettük. Látható, hogy csak N és Mn talajparaméterek esetén korrelál lényegesen jobban ( $\approx 0.1$ ) a labor adat a normál talajról készült képpel, mint a simított talajjal. Ettől eltekintve a laboradatok és a simított talajról készült felvételek átlagosan 0.075-tel jobb korrelációt mutatnak, mint a normál talaj esetén.

További érdekes megfigyelés, hogy a normál talajról készült képeknél az optimális korrelációhoz 10 talaj-összetevőnél csak egy hullámhosszat szabad figyelembe venni, és 5 esetben két szomszédost. A simított talajnál a spektrális környezet jobban korrelál a talaj-összetevővel, 7 esetben kell egy hullámhosszat figyelembe venni, 3 esetben pedig legalább kettőt, de 7 hh is előfordul (KA).

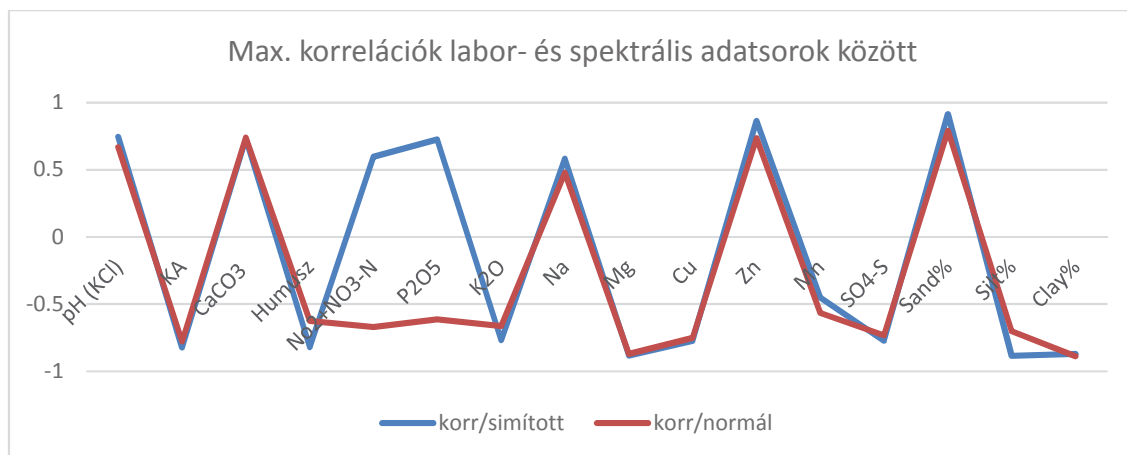
Vegyük észre, hogy K<sub>2</sub>O és SO<sub>4</sub>-S talaj-összetevőket ugyanazon az 1849-es hullámhosszon a legbiztosabb észlelni, sőt, ugyanazon korrelációval. Ugyanezt figyelhetjük meg az Mg és Clay összetevőknél is – vajon nincs közöttük valami direkt kapcsolat? A korrelációs számítás szerint igen, mert a korrelációjuk 0.93-nál nagyobb. Hasonló megállapítást tehetünk, ha megvizsgáljuk Sand és Silt összetevők kapcsolatát. Sand talajparaméter hullámhossz tartománya tartalmazza Silt-ét, és korrelációjuk összege közelítőleg 0. Ebben az esetben az adatsorok közötti korreláció ( $-0.97$ ) még erősebb is, mint az előző példákban.

Végeredményben a szerzők azt javasolják, hogy a spektrális mérések csak simított talajon történjenek, mert:

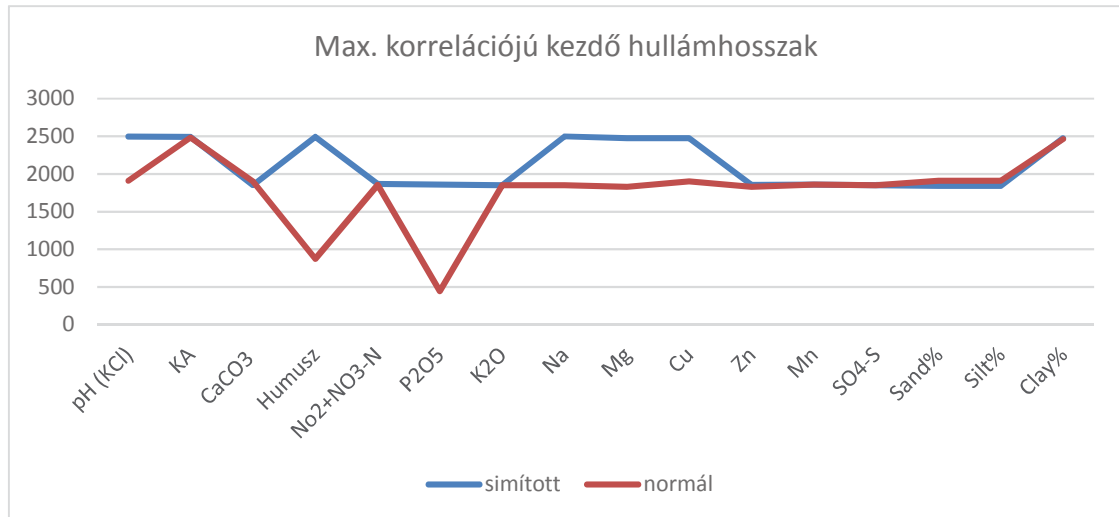
- a mérések (illetve átlaguk) jobban korrelálnak a laboradatokkal,
- a szélesebb optimális hullámhossz tartományok átlagolása nagyobb hibatűrést (robosztusságot) eredményez,
- csak az (1841, 1870) és (2474, 2500) hullámhossz tartományokat érdemes mérni, viszont normál talajon a (1831, 1910) és (2463, 2482) tartományokon kívül szükség van a 446 és 873 nanométeres hullámhosszakra is.

Labor adat	korr/simított	kezdő hh	utolsó hh	korr/normál	kezdő hh	utolsó hh
pH (KCl)	0,74404744	2494	2494	0,66945753	1909	1909
KA	-0,82240143	2493	2499	-0,77818156	2482	2482
CaCO <sub>3</sub>	0,75805876	1851	1853	0,73946585	1901	1901
Humusz	-0,82147033	2492	2497	-0,62540187	873	873
No <sub>2</sub> +NO <sub>3</sub> -N	0,57043033	1869	1870	-0,6714987	1855	1858
P <sub>2</sub> O <sub>5</sub>	0,71732301	1858	1858	-0,61360923	446	446
K <sub>2</sub> O	-0,76824831	1849	1849	-0,66255002	1849	1849
Na	0,58352248	2498	2500	0,47748388	1851	1852
Mg	-0,88173063	2475	2475	-0,87102881	1831	1831
Cu	-0,77196251	2474	2475	-0,75079347	1900	1901
Zn	0,77043110	1857	1858	0,73627430	1831	1832
Mn	-0,46224779	1859	1859	-0,56489117	1858	1859
SO <sub>4</sub> -S	-0,77323049	1849	1849	-0,73280416	1849	1849
Sand %	0,91297538	1841	1845	0,79055039	1909	1910
Silt %	-0,88239173	1841	1844	-0,70036498	1910	1910
Clay %	-0,87035868	2475	2475	-0,88816105	2463	2463

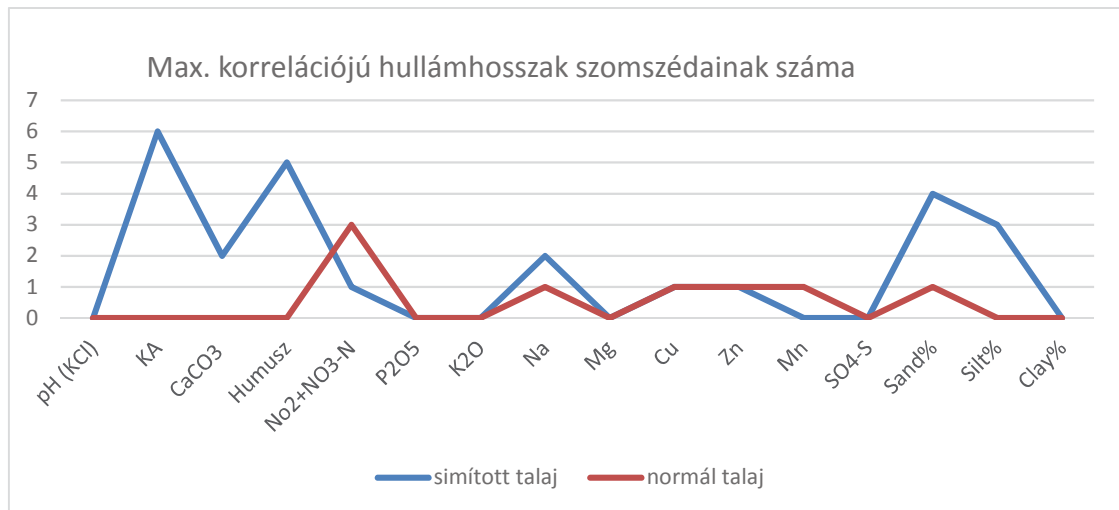
1. táblázat. A legjobb korrelációk a labor és a spektrális adatok között



7. ábra. A labor- és spektrális adatsorok legerősebb kapcsolatai



8. ábra. A labor adatra leginkább érzékeny hullámhosszak simított/normál talaj esetén



9. ábra. A max. korrelációnál figyelembe vett szomszédos hullámhosszak száma

## 5. A laboradatok becslése a hiperspektrális adatokból

Az előző fejezetben láttuk, hogy bizonyos spektrális adatsorok (illetve azok átlaga) jól korrelálnak a labor-adatokkal. A jó korreláció viszont közelítőleg lineáris kapcsolatot feltételez az adatsorok között, ami regressziós egyenes illesztéssel meghatározható. A kutatás zárásaként meghatároztuk a simított talaj és a talajparaméterek legjobb korrelációjánál érvényes egyenes-illesztés (lineáris regresszió) paramétereit és hibáját, amit a 2. táblázatban foglaltunk össze:

Talaj attribútum	kezdő hh	utolsó hh	korreláció	m* intenzitás	+ konstans	átlagos hiba
pH (KCl)	2494	2494	0,744047445	0,945893825	7,17114112	0,086870166
KA	2493	2499	-0,822401438	-41,38217159	55,002649	2,589002186
CaCO <sub>3</sub>	1851	1853	0,758058768	14,27983081	11,3285781	2,048742253
Humusz	2492	2497	-0,821470332	-3,611737425	3,33340928	0,199029439
No <sub>2</sub> +NO <sub>3</sub> -N	1870	1870	0,570430331	9,2108647	3,36817544	4,757651887
P <sub>2</sub> O <sub>5</sub>	1872	1872	0,717323014	147,3655419	134,48017	33,58685803
K <sub>2</sub> O	1849	1849	-0,768248319	-277,4079628	384,851677	67,88898619
Na	2498	2500	0,583522484	91,4145948	29,1253924	17,5190072
Mg	2475	2475	-0,881730636	-1371,854066	651,107612	37,22532888
Cu	2474	2475	-0,771962518	-23,91006587	12,307372	1,002542246
Zn	1908	1912	0,770431108	4,54802179	1,09366512	0,350396624
Mn	1868	1868	-0,46224779	-19,02886049	32,5268205	11,83677712
SO <sub>4</sub> -S	1849	1849	-0,773230494	-17,0688913	31,3539888	4,111215183
Sand%	1841	1846	0,912975388	53,92367293	4,75521429	5,204997516
Silt%	1842	1845	-0,88239173	-33,17047645	69,0639769	4,36215317
Clay%	2475	2475	-0,870358686	-73,32830072	40,716835	2,104065309

2. táblázat. A lineáris regresszió paramétereit és hibáját

Már utaltunk rá, ha szélesebb összefüggő tartományban detektálunk, akkor a spektrális adatok átlagolása javítja a hibátűrést. A talajparaméterek többségénél a legjobb korrelációt egyedi, nem szomszédos hullámhosszakon kaptuk, ezzel ez a hibajavító opció kiesett. Megvizsgáltuk ezért, ha megelégszünk az optimálisnál legfeljebb 0,1-el kisebb korrelációval, tudjuk-e növelni a detektáló tartomány hosszát. A válasz természetesen igen, és az eredményeket a 3. táblázat tartalmazza.

A regisztrálható spektrális tartományok - természetesen most is simított talajnál - (1826, 1916) és (2476, 2500)-ra nőttek. Az alacsonyabb korreláció miatt a lineáris becslés átlaghibája ugyan nőtt, de az eljárás robusztusabb lett; az átlagszámítás miatt kevésbé érzékeny az egyedi mérési hibákra.

Talaj attribútum	kezdő hh	utolsó hh	korreláció	m * intenzitás	+ konstans	átlagos hiba
pH (KCl)	2492	2496	0,7141035	0,965679365	7,15539451	0,091019558
KA	2491	2500	-0,746042	-37,98773274	53,7732071	3,030392961
CaCO <sub>3</sub>	1826	1835	0,5611566	25,33492736	7,8283256	2,478766134
Humusz	2490	2499	-0,8414135	-3,482238028	3,31550214	0,188607857
No <sub>2</sub> +NO <sub>3</sub> -N	1869	1872	0,6546498	10,4461466	3,10281759	3,176786172
P <sub>2</sub> O <sub>5</sub>	2479	2481	0,6254813	520,5185891	27,350564	37,61195169
K <sub>2</sub> O	1848	1849	-0,7385674	-322,6904595	406,782762	71,49766462
Na	2497	2500	0,585624	97,00546181	28,3045939	17,48634292
Mg	2473	2476	-0,8411375	-1285,975252	620,6395504	42,67538593
Cu	2490	2499	-0,7376959	-13,79582055	8,4239765	1,064784557
Zn	1907	1916	0,6555008	4,634827774	1,10977595	0,415071272
Mn	1868	1869	0,5099243	28,00535249	14,5632884	11,72269405
SO <sub>4</sub> -S	1848	1849	-0,7438687	-19,86878481	32,7100108	4,332948023
Sand%	1841	1850	0,822889	43,71973523	8,96588654	7,24849115
Silt%	1841	1850	-0,8030664	-31,00772199	69,1223839	5,524497619
CLAY %	2473	2476	-0,8115942	-67,19020033	38,55142629	2,496333399

3. táblázat. A bővített tartományú lineáris regresszió és hibája



## 6. Összefoglaló

A talajmintákat vettünk és hiperspektrális felvételeket készítettünk mezőgazdasági művelésű parcellák ugyanazon mintavételi helyein. Korrelációs számításokkal határoztuk meg, hogy az egyes talaj-összetevők mely hullámhossz tartományokon mutathatók ki a legbiztosabban, illetve regresszió-analízissel következtettünk a hiperspektrális képekből a talajparaméterek mennyiségére. Ajánlást adtunk a felvételek készítésének körülményeire is. Megállapítottuk, hogy a felvételek pontosságvesztés nélkül készíthetők olcsóbb multi-spektrális kamerákkal is.

## Köszönetnyilvánítás.

Kutatásainkat a VKSZ\_12-1-2013-0034 Klíma2 pályázat támogatta.

## Irodalomjegyzék

- [1] **E Ben-Dor, Y Benyamini, N Goldschleger, M Agassi:** The Spectral Reflectance Properties of Soil Structural Crusts in the 1.2- to 2.5- $\mu\text{m}$  Spectral Region, in Soil Science Society of America Journal 67(1), January 2003. doi:10.2136/sssaj2003.0289
- [2] **G Grandjean, I Cousin, M Seger, J Thiesson, S Lambot, B Van Wesemael, A Stevens, K Samyn, A Bitri, S Bernardie,** 2009. From geophysical parameters to soil characteristics, Report N°BRGM/FP7-DIGISOIL-D2.1, 52 pages. [https://esdac.jrc.ec.europa.eu/public\\_path/Digisoil-D2.1.pdf](https://esdac.jrc.ec.europa.eu/public_path/Digisoil-D2.1.pdf)