

## Búza minták közeli infravörös spektrumainak minőségi vizsgálata

Kónya Éva<sup>1</sup> – Kovács Géza<sup>2</sup> – Györi Zoltán<sup>1</sup>

<sup>1</sup>Központi Élelmiszer-tudományi Kutatóintézet, Budapest

<sup>2</sup>Debreceni Egyetem Agrár- és Gazdálkodástudományok Centruma, Mezőgazdaság-, Élelmiszertudományi és Környezetgazdálkodási Kar, Minőségbiztosítási és Mikrobiológiai Intézet, Debrecen  
eva.konya@gmail.com

### ÖSSZEFOGLALÁS

*A közeli infravörös spektroszkópiát, mint gyors, roncsolásmentes analitikai eljárást több iparágban elterjedten alkalmazzák. A technika alkalmas mind azonosításra, mind minőségi, mind pedig mennyiségi elemzések elvégzésére. A technika közvetett voltából adódóan a mennyiségi kalibrációk készítéséhez megfelelő mintasereg van szükség. Ezen mintasereg spektrumainak felvétele, majd pedig minőségi vizsgálata előzi meg a mennyiségi elemzéseket, kalibrációs modellek felállítását. Munkánk célja volt megvizsgálni, hogy a búzaminták termesztési területének ismerete hatással van-e a közeli infravörös spektrumokra.*

**Kulcsszavak:** NIR, főkomponens analízis

### SUMMARY

*The near infrared spectroscopy is widely used in the different industries as a rapid, non-invasive analytical tool. It is suitable for identification, qualification and quantitative analysis as well. As this technique is indirect, to make accurate calibration equations we need a proper sample population. Before the quantitative analysis, developing calibration models we have to collect and examine the spectra. In our study we examined wheat samples with known origins to find if there is any effect of the growing area on the NIR spectra.*

**Keywords:** NIR, principal component analysis

### BEVEZETÉS

A közeli infravörös spektroszkópia egy olyan roncsolásmentes analitikai eljárás, mely előnyös tulajdonságai révén számos iparágban elterjedté vált. A technika mérési idejének rövidege lehetővé teszi, hogy gyorsan, akár néhány másodpercen belül megbízható eredményt adjon. Ezt a módszert nem csak mennyiségi meghatározásra, de emellett azonosításra és minőségi elemzésre is alkalmazzák. A mezőgazdaságban, élelmiszeriparban, de más iparágakban is egyre nagyobb jelentősége van olyan módszereknek, melyek rövid időn belül megfelelő pontossággal tudnak adatot szolgáltatni. Legyen az alapanyag azonosítás, vagy konkrét paraméter becslése, tájékoztató jellegű mérése, a közeli infravörös spektroszkópia jól alkalmazható ilyen területeken. A mérés közvetett volta megköveteli, hogy az azonosításhoz, minőségi és mennyiségi mérésekhez megfelelő mintasereg álljon rendelkezésre. Ennek a mintaseregnek kellően nagyszámúnak és az adott tulajdonságra (pl. származási hely, gyártó, évjárat), paraméterre (pl. nedvesség-, fehérjetartalom) reprezentatívnak kell lennie. Ezen mintasereg spektrumainak felvétele és azok minőségi elemzése szükséges a később erre alapozandó mennyiségi kalibrációs modellek

elkészítéséhez. A spektrumok elemzése többféle kemo-metriai módszerrel lehetséges, melyek közül a polár minősítő rendszer (Polar Qualification System) és főkomponens analízis (Principal Component Analysis) módszerek a legelterjedtebbek.

A spektrumok komplex információk hordozói, mind fizikai, mind pedig kémiai szempontból. A felvett spektrumok vizsgálata elkerülhetetlen. A kalibrációk készítéséhez a megfelelő mintasereg összeállítása fontos szempont. Ha laboratóriumi vizsgálatok azt is mutatják, hogy a mintasereg reprezentatív, akkor is a spektrumokban lehet olyan változékonyság, ami a későbbiekben problémát okozhat. Különböző matematikai előkezeléseket (pl. deriválás, multiplicative scatter correction, standard normal variate and detrend) alkalmaznak a mintasereg fizikai tulajdonságaiból (pl. különböző szemcseméret) adódó eltérések kiküszöbölésére, valamint például a deriválás megoldás a spektrumban lévő átlapoló csúcsok szétválasztására (Osborne, 2006).

A spektrumok minőségi elemzéséhez, a mintaseregben lévő különbözőségek kiszűrésére alkalmas módszer a polár minősítő rendszer (PQS). A polár minősítő rendszer (polar qualification system=PQS) egy olyan módszer, melynek segítségével a Descartes koordináta-rendszerben lévő spektrum átalakítható polár koordinátarendszerbelivé. A sugár a spektrumérték, míg a szög a hullámhossz függvénye. Adott hullámhosszhoz tartozó abszorpciós értékek mindkét rendszerben megegyeznek. A minta spektrumát a kétdimenziós „minőségsíkon” ábrázoljuk. Az így keletkező spektrum középpontja a „minőségpont”, mely jellemző az adott mintára. A különböző minták minőségpontjai közötti távolság jellemzésére a „polár távolság” használható (Kaffka és Gyarmati, 1998.) A főkomponens analízis matematikailag egy más módszer, de ezt alkalmazzák elterjedtebben a spektrumokban lévő változékonyság vizsgálatára, a változók csökkentésére.

Munkánk célja volt olyan búzaminták vizsgálata, melyek különböző termő területekről származtak. Ezen minták közeli infravörös spektrumainak minőségi vizsgálatát végeztük főkomponens analízis segítségével arra keresve a választ, hogy a földrajzilag eltérő területekről származó minták spektrumai mutatnak-e szeparálódást a mintaseregben belül, s így ez a mintasereg felhasználható-e későbbi mennyiségi kalibrációk készítéséhez vagy egyes részeit külön kell-e kezelni.

### ANYAG ÉS MÓDSZER

A vizsgált minták 2009-es évből, 11 különböző termő területről (Magyarországon belül) származó búzaminták voltak. A vizsgálatokat a Debreceni Egyetem

Agrár- és Gazdálkodástudományi Centruma, Mezőgazdaság-, Élelmiszertudományi és Környezetgazdálkodási Kar, Élelmiszertudományi, Minőségbiztosítási és Mikrobiológiai Intézet laboratóriumában végeztük.

### Közeli infravörös spektroszkópiai mérések

Méréseinket Infratec 1241 Grain Analyzer (Foss Analytical AB, Svédország) transzmissziós módban működő készülékkel végeztük. A spektrumok felvétele búzaszemekből a 850–1048 nm-es hullámhossztartományban 2 nm-ként történt. A spektrumok feldolgozására és azok elemzésére WinISI II v1.50 (Infrasoft International LLC., USA) szoftvert használtuk.

### Főkomponens analízis (PCA)

Adatredukció céljából a spektrumok tanulmányozásához első lépésként főkomponens analízis (principal component analysis=PCA) végezhető. A PCA a spektroszkópiai adatmátrix meghatározott számú korreált változója (oszlopa) közötti összefüggések vizsgálatának megkönnyítésére az eredeti változókat egy olyan transzformációnak veti alá, mely új, korrelálatlan változókat eredményez. Ezek az új változók a főkomponensek, amelyek az eredeti változók lineáris kombinációi. Úgy vannak sorba rendezve, hogy elől állnak azok, amelyek az eredeti változók együttes varianciájának legnagyobb részéért felelősek.

A PCA alkalmas módszer arra, hogy az adatokat úgy adjuk vissza kevesebb dimenzióban, hogy új korrelálatlan változók keletkeznek. Azért fontos, hogy az adatokat kevesebb dimenzióban képezzük le, mert a háromnál nagyobb dimenziós adatkészletet az emberi agy nem képes felfogni. Azonban 1, 2, vagy 3 dimenzióban az emberi agy mintázatfelismerő képessége nagyon jó; a kapcsolatokat, hasonlóságokat, különbségeket könnyen észreveszi. Ezek alapján a főkomponensek egy új (derékszögű) koordináta-rendszer tengelyeinek, értékei pedig az eredeti adatmátrixban levő oszlopvektor elemeinek ezekre a tengelyekre történő vetítéseinek tekinthetők.

Az határozza meg a főkomponensek sorrendjét, hogy minden soron következő főkomponens annak a varianciának a legnagyobb részét írja le, amelyet nem magyaráznak az előző főkomponensek. Ennek értelmében az első főkomponens hordozza az adatokban lévő variancia legnagyobb részét. A másodikban több információ van, mint a harmadikban, és így tovább. Annyi főkomponenst kell kiszámítani, amennyi ahhoz szükséges, hogy magyarázza a teljes variancia előre meghatározott százalékát. Az összes főkomponenst használva, a teljes variancia 100%-át meg tudjuk magyarázni. Általában azonban csak annyi főkomponenst veszünk figyelembe, amennyivel a variancia bizonyos százalékát tudjuk leírni.

A főkomponens elemzés eredményeinek láthatóvá tételét rendszerint a főkomponens együttható (loading) és főkomponens (score) ábrákkal oldhatjuk meg. A főkomponens együttható ábrából az eredeti tulajdonság változók hasonlóságaira, korrelációira lehet következtetni. A főkomponens ábrák hasonlóak a főkomponens együttható ábrákhoz, de skálájuk eltér: a főkomponens együttható értéke -1 és 1 közé esik, míg a főkomponensek nem szoríthatók határok közé (Borosy et al., 2001).

### EREDMÉNYEK ÉS ÉRTÉKELÉS

Az 1. ábrán a minták felvett spektrumai láthatóak.

A spektrumokat főkomponens analízisnek vetettük alá, hogy megvizsgáljuk egyrészt, hogy a mintaseregünket a későbbiekben tervezett kalibrációs modellek készítéséhez tudjuk-e egy populációként, egyben kezelni. Másfelől pedig a termőhely ismeretének függvényében, az egy termőhelyről származó minták spektrumai egy csoportot alkotnak-e, és ezek a csoportok elkülönülnek-e egymástól. Ha az elemzés azt mutatja, hogy élesen elszeparálódnak ezen csoportok, akkor a mintapopulációt nem lehet egyként kezelni, az egyes szétváló csoportokra kellene a kalibrációs modelleket elkészíteni. Ez utóbbi eset lokálisan specifikálódott kalibrációt eredményezne.

1. ábra: Felvett spektrumok

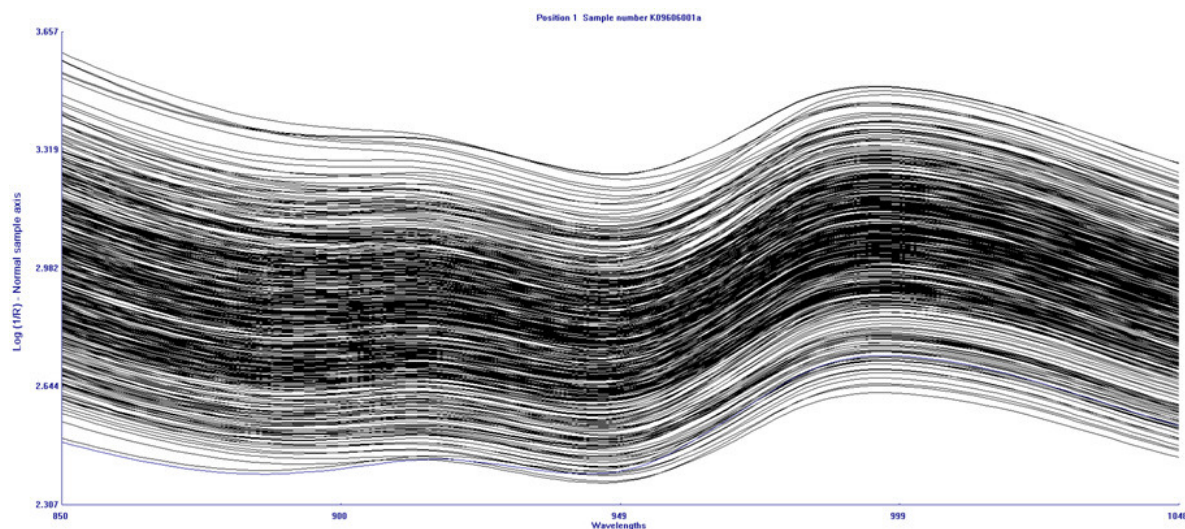


Figure 1: The collected spectra

A főkomponens analízis elvégzése után kapott eredményeket az 1. táblázat és a 2. ábra mutatja.

A táblázatban látható, hogy az első három főkomponens összesen a spektrumokban lévő variancia 99,97%-át leírja, vagyis a spektrumokban lévő változékonyságot közel teljesen magyarázza. A 2. ábra a főkomponens analízissel kapott háromdimenziós koordináta-rendszer egyes vetületeit tartalmazza, az egyes tengelyeken a főkomponenseket feltüntetve. Ahogy a 2. ábrán is látszik, a minták nem különülnek el nagyobb csoportokra, közöttük nincsen szeparálódás.

1. táblázat

Az első három főkomponenshez tartozó egyedi és összesített variancia

		Egyedi(1)	Összesített(2)
PC1	PC1	98,41	98,41
PC2	PC1+PC2	1,51	99,92
PC3	PC1+PC2+PC3	0,05	99,97

Table 1: Individual and cumulative variance to the first three principal components

Individual(1), Cumulative(2)

2. ábra: PCA eredmények ábrázolása különböző vetületekben az eredeti mintahalmaz esetén

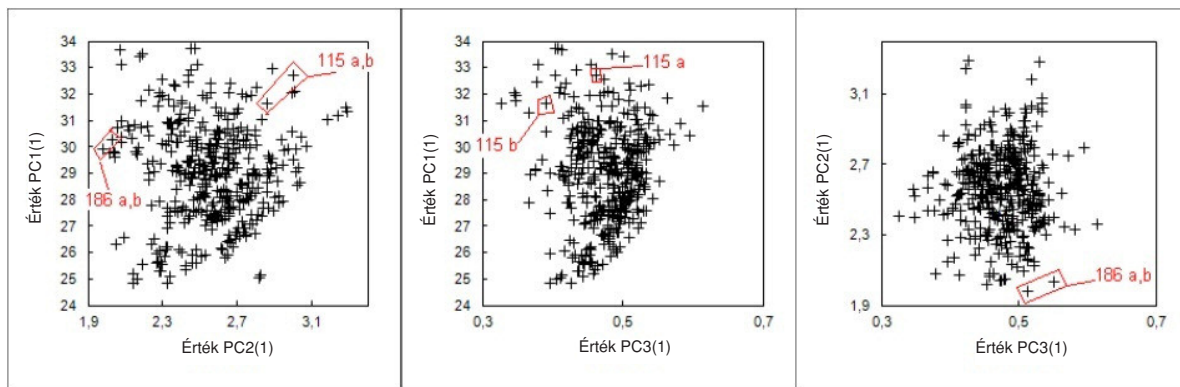


Figure 2: Projections of the PCA results for the original sample population Value(1)

Az egyes termőhelyekről származó mintákat külön színnel jelöltük (3. ábra). A 3. ábra mutatja a színezés után kapott, megjelölt mintákat tartalmazó vetületeket. Az alábbi színek kódokat alkalmaztuk: piros – Gesztely,

sötétkék – Látókép, zöld – KSZ, rózsaszín – Tápió, sárga – Tiszavasvári, halványvörös – Jánoshalma, szürke – Somogyszil, sötétnarancs – Csorvás, lila – Körösszeg-apáti, vízkék – Harta és olívdzöld – Komádi.

3. ábra: Termesztési helyek szerinti színezéssel kapott eredmények az eredeti mintahalmaz esetén

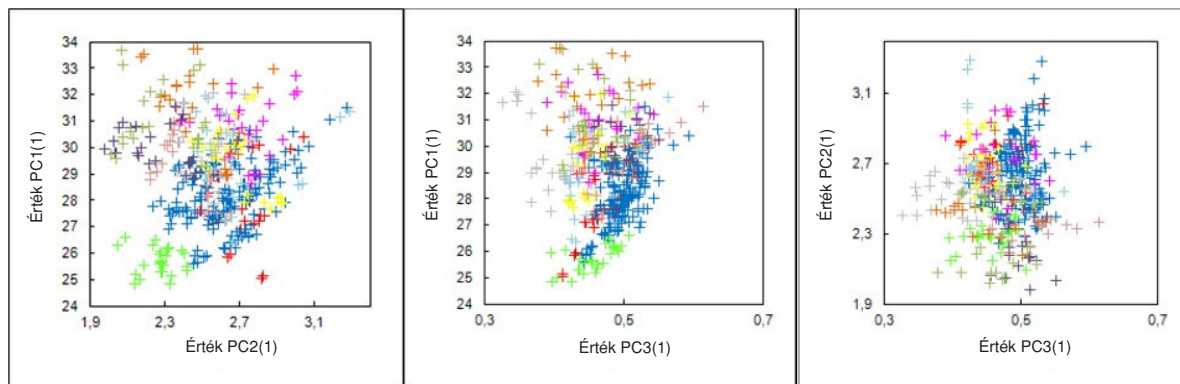


Figure 3: Crop site based coloured presentation of samples for the original sample population Value(1)

A 3. ábrán látható, hogy egy termesztési területhez tartozó minták egy-egy csoportot alkotnak, de az egyes csoportok nem különülnek el élesen, köztük átfedések vannak. Ebből arra lehet következtetni, hogy a termesztési hely a rendelkezésünkre álló információk alapján nincs jelentős hatással a spektrumokra. A termőhely hatásvizsgálatát olyan információkkal lehetne pontosítani, mint például a fajta ismerete egy-egy termőhelyen.

A főkomponens analízis két mintát kiugrónak, „outliernek” tekintett, melyeket eltávolítva ismét elvégeztük az elemzést. A főkomponensek által leírt varianciák nem változtak (2. táblázat), viszont ezen minták ismerete azért is fontos, mert a későbbiekben, kalibrációs modellek felállításánál ezen mintákra figyelni kell, sorsukat nyomon kell követni.

2. táblázat

Az első három főkomponenshez tartozó egyedi és összesített variancia a módosított mintahalmaz esetén

		Egyedi(1)	Összesített(2)
PC1	PC1	98,44	98,44
PC2	PC1+PC2	1,48	99,92
PC3	PC1+PC2+PC3	0,05	99,97

Table 2: Individual and cumulative variance to the first three principal components a módosított mintahalmaz esetén  
Individual(1), Cumulative(2)

Az elvégzett főkomponens analízis alapján elmondható, hogy ezen a mintaseregből felvett spektrumok halmaz nem tartalmaz szeparálódást, így a további vizsgálatokhoz nem szükséges szétválasztani, kisebb csoportokban kezelni a populációt, hanem alkalmas kalibrációs modellek készítésére, ha a laboratóriumi vizsgálatok eredményei alapján az adott paraméterre vonatkozóan a mérési tartomány egészében reprezentatív.

## KÖVETKEZTETÉSEK

A közeli infravörös spektroszkópia egyszerűségének, a mérési idő rövidegének, roncsolásmentes voltának köszönheti elterjedését. A technika közvetett voltából adódik, hogy egyrészt fontos jól megválasztani a kalibráló mintasereget, hogy nagyszámú, a kívánt paraméterre megfelelően reprezentatív legyen. Ezen feltételek teljesülése esetén is adódhatnak a spektrumokban olyan eltérések, melyek nem teszik lehetővé a populáció egy csoportként való kezelését.

A vizsgálataink arra irányultak, hogy az adott mintapopulációk esetében a spektrumokra van-e hatása a különböző termőhelyeknek, külön kell-e kezelnünk az egyes területekről származó mintákat. A főkomponens analízis eredményei alapján elmondható, hogy a vizsgált populáció nem szeparálódik külön csoportokra, a csoportok között átfedések vannak. A későbbi vizsgálatokban így nem szükséges termőhelyenként külön választani a mintákat.

## IRODALOM

- Borosy A.P.–Héberger K.–Horvai Gy.–Kolossváry I.–Lengyel A.–Paksy L.–Rajkó R.–Szepesváry P. (2001): Sokváltozós adatelemzés (kemometria) Nemzeti Tankönyvkiadó Rt. Budapest.
- Kaffka, K.J.–Gyarmati, L.S. (1998): Investigating the polar qualification system. *Journal of Near Infrared Spectroscopy*. 6: A191–A200.
- Osborne, B.G. (2006): Near-Infrared Spectroscopy in Food Analysis. *Encyclopedia of Analytical Chemistry*.