

# KUKORICASZILÁZSOK ÖSSZETÉTELÉNEK BECSLÉSE ILS MODELL ALAPJÁN

## ESTIMATING THE COMPOSITION OF CORN SILAGE BASED ON THE ILS METHOD

Végh Attila<sup>1\*</sup>

<sup>1</sup> Alaptudományi Tanszék, GAMF Műszaki és Informatikai Kar, Neumann János Egyetem, Magyarország

<https://orcid.org/0000-0003-3193-8600>

<https://doi.org/10.47833/2026.1.CSC.004>

---

### **Kulcsszavak:**

chemometria  
többváltozós kalibráció  
NIR spektroszkópia  
inverz legkisebb négyzetek módszer  
szere

### **Keywords:**

chemometrics  
multivariate calibration  
NIR spectroscopy  
inverse least-square method

### **Cikktörténet:**

Received 2025. november 5.

Revised 2026. február 11.

Accepted 2026. április 14.

---

### **Összefoglalás**

A közeli infravörös (NIR) spektroszkópia lehetővé teszi anyagok roncsolásmentes vizsgálatát, mégpedig a mért spektrális értékek alapján meghatározva, megbecsülve az analizálandó anyag összetevőinek a mennyiségét. Az ismert többváltozós kalibrációs eljárások közül a matematikai nehézségek miatt kevésbé használt inverz legkisebb négyzetek (ILS) módszert mutatom be. Az egymást követő vetítési algoritmus (SPA) és a mátrix invertálásának alternatív módja segít az ILS módszer hatékonyabbá tételében. MATLAB-ban írt kód segítségével kukoricaszilázs mintákon szemléltetem az ILS módszerrel kapott eredményeket. Egy kivételével az összes összetevő esetén 10% alatti, a legtöbb összetevő esetén 5% alatti relatív hibát kaptam.

### **Abstract**

Near-infrared (NIR) spectroscopy enables non-destructive testing of materials by determining and estimating the quantity of the components of the material to be analyzed based on the measured spectral values. Among the known multivariate calibration methods, I present the inverse least squares (ILS) method, which is less commonly used due to its mathematical complexity. The successive projection algorithm (SPA) and an alternative method of matrix inversion help to make the ILS method more efficient. Using code written in MATLAB, I illustrate the results obtained with the ILS method on corn silage samples. With one exception, I achieved a relative error of less than 10% for all components and less than 5% for most components.

---

## 1. Bevezetés

A közeli infravörös (NIR) spektroszkópia témakörében a többváltozós kalibrációt ismertető számos összefoglaló mű jelent meg, ezek közül Olivieri könyve[17] részletesen tárgyalja a témához kapcsolódó fogalmakat. Ebben a cikkben a számos módszer és szerteágazó téma közül az inverz legkisebb négyzetek (ILS) módszerét és ennek egy alkalmazását mutatom be.

---

\*Kapcsolattartó szerző.

E-mail cím: [vegh.attila@nje.hu](mailto:vegh.attila@nje.hu)

Az inverz módszerek előnye, hogy lehetővé teszik olyan többkomponensű minták vizsgálatát, amelyekben csak egy vagy néhány analizálandó összetevő érdekes és a minta többi összetevőjének koncentrációja, spektruma és kémiai tulajdonsága általában ismeretlen. Ily módon lehetővé teszik a klasszikus legkisebb négyzetek (CLS) módszer fő hátrányának leküzdését: a kalibrációs fázisban az összes lehetséges mintaösszetevő koncentrációjának ismeretét.

Az inverz modellek fejlesztése az 1960-as évtizedben kezdődött, mikor a NIR- spektroszkópiát anyagvizsgálatra alkalmazták, mely roncsolásmentes elemzésre, különösen a vetőmagok és más mezőgazdasági termékek minőségi paramétereinek kimutatására volt hasznos. A fő hátrányt az jelentette, hogy az egyes elnyelési spektrumokban a különböző összetevők fényelnyelése okán mért elnyelési sávok nagymértékben átfedik egymást. A klasszikus legkisebb négyzetek (CLS) módszer ezekben az esetekben nem használható jól, ezért a kutatók alternatív matematikai modelleket fejlesztettek ki. A probléma új megközelítését leíró első munkát Karl Norris végezte az Egyesült Államokban, aki a NIR-adatok két hullámhosszon történő mérését korrelálta a vetőmagok nedvességtartalmával Norris és Hart [16]. A NIR-spektroszkópia területén az egyik legidézettebb cikk Ben-Gera és Norris tanulmánya [3]. A technológia több évtizedes fejlődése során számos új módszert és eljárást fejlesztettek ki [9] [2] [4] [12] [13] [6], melyekről a továbbiakban érintőlegesen szó esik, de alapvetően az inverz legkisebb négyzetek (ILS) modell és alkalmazása kerül bemutatásra.

## 2. Kalibráció

A többváltozós kalibrációs eljárások közül az inverz legkisebb négyzetek regresszióját (ILS) vizsgálom, amely többszörös lineáris regresszió (MLR) néven is ismert. A jól ismert klasszikus legkisebb négyzetek (CLS) és a részleges legkisebb négyzetek (PLS)-modellek esetében a kalibráció az alábbiak szerint történik [15]:

$$Jel = \text{Érzékenység} \times \text{Koncentráció} + \text{Hibák}$$

Az inverz kalibrálás során viszont a lineáris törvényt fordítottan írjuk fel [9] [17]:

$$\text{Koncentráció} = Jel \times \text{Regressziós együttható} + \text{Hibák}$$

ahol a kalibrált összetevők koncentrációja és a megfelelő műszeres válaszok között arányosságot feltételezünk a becsülendő regressziós együtthatókon keresztül. Vegyük észre, hogy a két egyenletben a modellhibákat gyűjtő kifejezés eltérő: az előbbi esetben spektrális hibáknak, az utóbbiban koncentrációs hibáknak felel meg. Az inverz modellekkal kapcsolatos irodalom bőséges; Massart [14] klasszikus könyvét, valamint Haaland és Thomas [9] kiváló cikkét ajánlom. Nézzük most a fenti modell matematikai leírását. A Lambert-Beer-törvény inverzén alapuló általános modell a következőképpen fejezhető ki:

$$\mathbf{Y} = \mathbf{X}^T \mathbf{B} + \mathbf{E}$$

ahol az  $\mathbf{X}$  mátrix (mérete  $J \times I$ )  $I$  kalibrációs minta  $J$  hullámhosszon mért műszeres jeleit tartalmazza. Másrészt az  $\mathbf{Y}$  mátrix tartalmazza az egyes  $I$  mintákban az egyes  $N$  mintaösszetevő kalibrációs koncentrációit, és mérete  $I \times N$  ( $N$  az összetevők teljes száma).

A fenti egyenletben  $\mathbf{B}$  egy  $J \times N$  méretű mátrix, amely a koncentrációkat inverz módon kapcsolja össze a válaszokkal, és amelyet regressziós együtthatók mátrixának nevezünk. Végül az  $\mathbf{E}$  a modellhibák mátrixa, amelynek mérete megegyezik az  $\mathbf{Y}$  mátrixéval. Ha a probléma túldeterminált, több egyenletet tartalmaz, mint ismeretlent, a paraméterek becslése a legkisebb négyzetek illesztésével történik, minimalizálva az  $\mathbf{E}$  négyzetelemeinek összegét.

Tegyük fel, hogy nem a minták összes kémiai összetevőjére vagyunk kíváncsiak, hanem csak néhányra, esetleg egyetlenegyre. Ebben az esetben nem kell ismernünk a  $\mathbf{B}$  mátrix összes elemét, hanem csak az érdeklődésünkre számot tartó analizálandó összetevőnek megfelelő oszlopot, amelyet az  $n$  index azonosít. Így lehetséges egy olyan egyszerűsített modellt felírni, amelyben csak az

$n$ -edik analizálandó összetevő koncentrációi ismertek, elkülönítve a fenti egyenletből az utóbbi analizálandó összetevőnek megfelelő részt:

$$\mathbf{y}_n = \mathbf{X}^T \mathbf{b}_n + \mathbf{e},$$

ahol  $\mathbf{y}_n$  az analizálandó összetevő koncentrációjának  $n$  vektora,  $\mathbf{b}_n$  az adott analizálandó összetevőnek megfelelő  $\mathbf{B}$  mátrix oszlopai, és  $\mathbf{e}$  a koncentrációs maradékok vektora erre a redukált ILS modellre.

Formailag a  $\mathbf{b}_n$ -t úgy kaphatjuk meg, hogy mindkét oldalt balról szorozzuk  $\mathbf{X}$  -szel, majd szorozzuk az inverz  $(\mathbf{X} \mathbf{X}^T)^{-1}$  mátrixszal:

$$\mathbf{b}_n = (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{y}_n.$$

Ha  $(\mathbf{X} \mathbf{X}^T)^{-1}$ -et  $\mathbf{X}^+$ -szal jelöljük és  $\mathbf{X}$  általános inverzének nevezzük, akkor a fenti egyenletet az alábbi tömör formában írhatjuk fel:

$$\mathbf{b}_n = \mathbf{X}^+ \mathbf{y}_n.$$

A fenti egyenlet jelenti az ILS előnyét, mivel csak az analizálandó összetevők koncentrációit kell ismerni és nem kell tudni a kalibrációs mintában szereplő összes összetevő összes koncentrációját. Vegyük észre, hogy a fenti egyenlet nem jelenti azt, hogy az ILS modell csak egyetlen analizálandó összetevő vizsgálatát teszi lehetővé. Ha több analizálandó összetevő érdekel, elvileg mindegyikre külön modellt írhatunk, az egyenletünkhöz hasonlóan. Az ILS modell lehetővé teszi a minta globális tulajdonságainak vizsgálatát, mivel az egyenletben az  $\mathbf{y}_n$  vektor nem a specifikus analizálandó összetevő koncentrációkat, hanem a minta globális tulajdonságait tartalmazza.

### 3. Matematikai nehézségek

Az  $\mathbf{y}_n = \mathbf{X}^T \mathbf{b}_n$  egyenlet több ismeretlent tartalmazó szimultán egyenletrendszer mátrix-formája. Az egyik ilyen egyenlet

$$y_1 = x_{11}b_1 + x_{21}b_2 + \dots$$

ahol az első mintában az analizálandó összetevő ( $y_1$ ) koncentrációját az adott minta minden egyes hullámhosszon mért ( $x_{11}, x_{21}, \dots$ ) értékeinek és a megfelelő regressziós együtthatók ( $b_1, b_2, \dots$ ) szorzatának összegeként fejezzük ki [17].

Vizsgáljuk meg az egyenletek és ismeretlenek számát. Összesen  $I$  egyenlet van. (Az  $\mathbf{y}_n$  elemek száma, a kalibrációs minták száma). Másrészt  $J$  ismeretlent kell becsülni, azaz a regressziós együtthatók  $\mathbf{b}_n$  vektorának elemeit, amely szám egybeesik a spektrális hullámhosszok számával. Itt szembesülünk az első problémával, amikor egy ILS-modellt próbálunk kalibrálni: több kalibrációs mintára van szükség, mint hullámhosszra ( $I > J$  feltétel).

Az egyik lehetőség a kalibrációs minták nagyszámú, a hullámhosszok számánál nagyvonalúan nagyobb számú kalibrációs minta elkészítése. Ez költség- és időigényes lehet, és több ezer mintát igényelhet. Egy másik alternatíva a teljes spektrumtartományból a hullámhosszok egy csökkentett csoportjának kiválasztása, hogy a minták száma meghaladja a hullámhosszokét. Ez egy jó algoritmust igényel a hullámhosszok kiválasztására, és az érzékenység jelentős csökkenéséhez vezethet, mivel az elemzéshez potenciálisan hasznos hullámhosszokat hagyhatunk el.

Másik probléma az  $\mathbf{X} \mathbf{X}^T$  mátrix invertálása. Az  $\mathbf{X}$  oszlopai a kiválasztott hullámhosszon nem mutathatnak jelentős mértékű korrelációt vagy átfedést. Ellenkező esetben az  $\mathbf{X} \mathbf{X}^T$  mátrix determinánsa nulla lesz, tehát a mátrix nem invertálható, vagy ha közel nulla, akkor az inverzió instabil lesz.

## 4. Előrejelzés

Az előrejelzés során a kalibrációs fázishoz hasonló inverz egyenletet írunk fel, amely az ismeretlen mintára alkalmazott Lambert-Beer-féle inverz törvényen alapul [17]:

$$y_n = \mathbf{x}^T \mathbf{b}_n,$$

ahol  $\mathbf{x}$  az ismeretlen minta spektruma. A fenti egyenletből következik, hogy  $\mathbf{b}_n$  úgy viselkedik, mint az  $n$  alkotóelem regressziós együtthatója. Ha egynél több érdekes analizálandó összetevő fordul elő, a fenti egyenletet annyiszor kell alkalmazni, ahányszor szükséges, minden alkalommal az egyes analizálandó összetevőkhöz tartozó  $\mathbf{b}_n$  vektor felhasználásával.

## 5. Validáció

Az ILS-modell kalibrálása után fontos annak validálása. Ebből a célból egy független validálási mintakészletet készítünk. A validációs készletre becsült analizálandó összetevő koncentrációk összehasonlítása megfelelő statisztikai mutatók segítségével történik. Ezek egyike az előrejelzés átlagos négyzetes hibájának négyzetgyöke (Root Mean Square Error, RMSE), amelyet az alábbiak szerint számítunk ki [15]:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

ahol  $n$  a validálási minták száma,  $y_i$  az analizálandó összetevő névleges koncentrációja a validálási mintákban, és  $\hat{y}_i$  a ILS által előre jelzett koncentráció ugyanezen mintákban.

Egy másik gyakori mutató az előrejelzés relatív hibája (REP), amelyet %-ban adunk meg, és a következőképpen határozzuk meg:

$$\text{REP} = 100 \frac{\text{RMSE}}{\bar{y}}$$

ahol  $\bar{y}$  az analizálandó összetevő kalibrációs koncentrációinak átlagértéke. Az előre jelzett vs. névleges analizálandó összetevő koncentráció, valamint az előrejelzési hibák vs. előre jelzett értékek ábrázolása, vizuális elemzése is hasznos lehet.

Két egymással szoros kapcsolatban lévő statisztikai mutatószámot ismertetünk még. A Pearson-féle lineáris korrelációs együttható ( $r$ ), mely a lineáris kapcsolat erősségét méri jelen esetben a névleges analizálandó összetevő koncentráció és az előrejelzett értékek között. Ez az alábbiak szerint kapható meg:

$$r(y, \hat{y}) = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}}$$

Melyből pár soros levezetéssel az alábbi összefüggést kapjuk

$$\sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

ahol a gyökjel alatt lévő számláló a magyarázott eltérésnégyzetösszeg ( $SSR$ ), a nevező pedig a teljes eltérésnégyzetösszeg ( $SST$ ), tehát a fenti képlet definíció szerint a determinációs együttható ( $R^2$ ) négyzetgyöke. Ha felhasználjuk, hogy  $SST = SSR + SSE$ , ahol  $SSE$  a nem magyarázott eltérésnégyzetösszeg, akkor az  $R^2$  determinációs együtthatóra az alábbi összefüggést kapjuk:

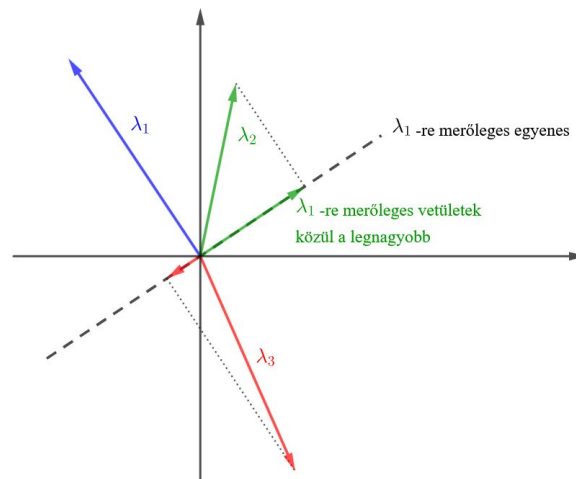
$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

## 6. Egymást követő vetítési algoritmus(SPA)

Ennek az algoritmusnak a célja, hogy kis számú, alacsony fokú kollinearitással rendelkező hullámhosszokat válasszon ki, hogy az ILS-modell biztonságosan alkalmazható legyen [1] [19] [18]. Ha 33 kalibrációs mintánk van és 125 hullámhosszon mérjük a spektrumokat egy sikeres ILS-modell felépítéséhez a hullámhosszok számának kisebbnek kell lennie, mint a minták száma, így ebből a 125 hullámhosszból összesen  $10^{30}$  nagyságrendű  $1,2,3, \dots, 32$  hullámhosszú részhalmaz építhető fel. Az egymást követő vetítési algoritmus (SPA) összesen csak  $125 \cdot 32 = 4000$  részhalmazt fog létrehozni. Általános esetben a részhalmazok száma  $J \times (I - 1)$  lesz.

Az első 125 egy hullámhosszot tartalmazó részhalmazt egyszerű meghatározni: mindegyik tartalmaz a 125 eredeti változó közül egyet-egyét. A következő elemek a 125 hullámhossznak a két hullámhosszot tartalmazó részhalmazai lesznek. Ahelyett, hogy a 125 változó mindegyikének a fennmaradó 124 változóval való összes lehetséges kombinációját kiválasztanánk, az SPA az elsőt követő második hullámhosszként azt a változót választja ki, amelyik leginkább merőleges az elsőre.

Az SPA logikája a következő: minden hullámhosszot egy vektorral vagy  $I$  számok halmazával lehet reprezentálni: az  $X$  kalibrációs mátrix sorának elemeivel. Ezek a vektorok egy többdimenziós mintatérben egy bizonyos irányba mutatnak, és a vektor csúcsa jelzi a hullámhossz helyzetét.



1. ábra. Ortogonális vetítés [17]

A maximális ortogonális vetítés fogalmát az 1. ábra szemlélteti. A lehető legegyszerűbb, mindössze két mintával rendelkező mintatérben három változót tartalmazó példát mutatunk be. Ha egy adott két hullámhosszot tartalmazó részhalmaz első változója a kék ( $\lambda_1$ ), akkor a második változó megtalálásához a fennmaradó két változó ( $\lambda_2$  és  $\lambda_3$ ) vetületét képezzük a  $\lambda_1$ -re merőleges irányra. A legnagyobb a zöld lesz, így a  $\lambda_2$ -t választjuk a két hullámhosszot tartalmazó részhalmazban a  $\lambda_1$  melletti második változónak.

A 125 változóból és 33 mintából álló példánkban a három hullámhosszot tartalmazó részhalmazokkal folytatva hasonló kritériumot alkalmazunk: a 125 két hullámhosszot tartalmazó részhalmaz mindegyikéhez tartozó harmadik kiválasztott hullámhossz az lesz, amelyik a fennmaradó 123 hullámhossz közül a legnagyobb ortogonális vetületet mutatja az első két változó által meghatározott síkra. És így tovább, amíg el nem érjük az egyenként 32 változóból álló részhalmazt.

Miután a 4000 részhalmaz létrejött, mindegyiküket egy ILS-modell segítségével megvizsgáljuk, hogy melyik adja a legjobb előrejelzési eredményeket a mérendő analizálandó összetevő-koncentráció vagy mintatulajdonság tekintetében. A legjobb részhalmazok legjobb változóinak hatékony szűrésére további eljárásokat dolgoztak ki, ami az optimális analitikai eredményekkel rendelkező hullámhosszok még szűkebb halmazához vezet [5]. Továbbá megemlítjük, hogy ridge regresszió alkalmazásával az ILS modell tovább javítható [10], [11].

## 7. Kukoricaszilázs elemzése

Ebben a szakaszban kukoricaszilázs mintákat elemzem az ILS modell segítségével. A kukoricaszilázs összetétele meghatározó szerepet játszik tejelő tehenek esetén a tej minőségére [7].

Az adatok a következők: 33 különböző, laborban is bevizsgált minta állt rendelkezésünkre. A labor a 33 minta mindegyikét 10 analizálandó összetevő ( N.víz: nedvességtartalom, Fehérje: nyersfehérje, Zsír: nyerszsír, Rost: nyersrost, Hamu: nyershamu, Kemény: keményítő, ADF: savdetergens rost, NDF: neutrális detergens rost, ADL: savdetergens lignin, Cukor: összes cukor.) esetében adta meg. Az alábbi táblázatokban csak a rövidítéseket használom. A minták mindegyikéről egyenként 30-30 NIR spektroszkópiás felvétel is készült 908,1 – 1676,2nm hullámhossz közötti tartományban 6,2nm lépésközzel, így összesen 125 hullámhosszon mért értékkel rendelkezünk 990 esetben. A táblázat az alábbi formát ölti, ahol (1+990) sor és (1+10+125) oszlop van.

1. táblázat. Bemeneti adatok

No.	Y N..víz	...	Y Cukor	X908.1	X914.294	...	X1676.2
1850	70,2	...	0,1	0,433608	0,422146	...	0,7009554

A spektrumokat előkezelés nélkül alkalmaztam az elemzés során, nem használtam simítás, normalítás (SNV), szóródáskorrekció (MSC), vagy deriválás eljárásokat [9] [2] [8]. A későbbiekben látni fogjuk, hogy a nyers adatok segítségével felépített ILS modell kielégítő eredményeket adott. További vizsgálatok pedig azt mutatták, hogy más módszerek esetén se kaptam lényegesen jobb eredményeket. Ezt az Excel táblázatot ("2022nyaroszksz20.xlsx") olvastam be a MATLAB-ba (2. sor).

```

1 clear
2 B=readmatrix("2022nyaroszksz20.xlsx");
3 size(B)
4 M=[];
5 sz=0;
6 T= B(1,2:136);
7 for k=1:989
8   if B(k,1)== B((k+1),1)
9     T=T+B((k+1),2:136);
10    sz=sz+1;
11 else
12   T=T/(sz+1);
13   M=[M;T];
14   sz=0;
15   T= B((k+1),2:136);
16 end
17 end
18 M=[M;T/(sz+1)];
19 size(M)

```

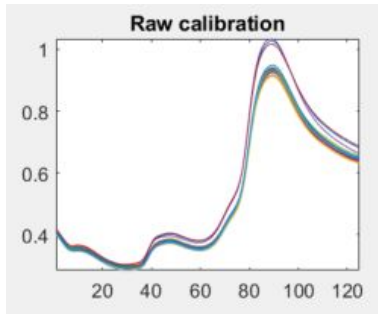
A következő lépésben mindegyik minta esetén kiszámoltam a 30 mért spektrális adat átlagát (4-18. sor). Így mind a 33 mintához pontosan egy spektrális adat tartozik, így az  $M$  mérete már csak  $33 \times 136$ .

```

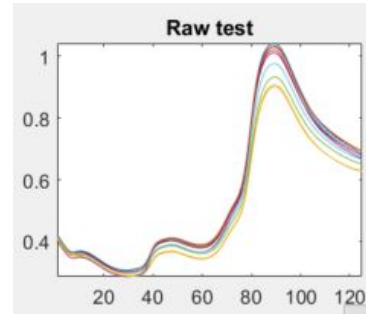
20 Bemenet=M;
21 X=Bemenet(11:33,11:135)';
22 Y=Bemenet(11:33,1:10);
23 Xteszt=Bemenet(1:10,11:135)';
24 Yteszt=Bemenet(1:10,1:10);

```

Ezután az  $M$  mátrixot kétszer két részre osztottam. Egyrészt a kalibrációs (2. ábra) és a validációs (3. ábra) készletre, másrészt az analizálandó összetevők és hullámhosszok szerint. (21-24. sor).



2. ábra. Kalibrációs minták



3. ábra. Teszt minták

Mint az ILS módszer esetén fentebb is írtam, mindegyik analizálandó összetevőt külön kell vizsgálni és egyes összetevőkre vonatkozóan különböző számú hullámhossz kerül kiválasztásra, így a kalibrációs együtthatók vektorának hossza is eltérő az egyes összetevők esetében. A 27. sor ezt egy ciklusba foglalja, mind a 10 analizálandó összetevőre egyenként végrehajtva az alábbi eljárást. Tehát most csak egy analizálandó összetevőt tekintek.

```

25 statmut=[];
26 fontpar=zeros(10);
27 for n=1:1:10
28 hbrmse=10;
29 A=X';
30 for cser=1:1:125
31 hvek=cser;
32 T=A(:,cser);
33 TM=[];
34 for elso1=2:1:23
35     vet=0;hvekt2=[];
36 for oszl=1:1:125
37 Tuj=A(:,oszl);
38 [Q,R]=qr([T,Tuj]);
39     if vet < Q(:,elso1)'*Tuj
40         vet=Q(:,elso1)'*Tuj;
41         hvekt2=oszl;
42         TM=Tuj;
43     end
44 end
45 hvek=[hvek,hvekt2]; T=[T,TM]; Xu=T';

```

A 28-tól a 45. sorig az SPA eljárás részletei találhatóak. Jelen példánkban egy (a kalibrációs minták számával megegyező) 23 dimenziós térben (a hullámhosszok számával megegyező) 125 vektort képzelünk el. Így egy vektornak 23 koordinátája van, mely a 23 minta spektrális értékeit tartalmazza egy adott hullámhosszon. A 30. sorban induló ciklus az első kiválasztandó vektort adja meg a 125 vektor közül (32. sor). Talán a könnyebb érthetőség kedvéért a 36. sorral folytatom. Az is egy ciklus kezdete, mely a második vektort keresi végignézve mind a 125 vektort és azt választja, melynek az első vektorra merőleges vektorra eső vetülete a legnagyobb (37. sor). Elsőre furcsa, hogy mind a 125 vektort végignézzük, de technikailag egyszerűbb és nem okoz problémát, hiszen, ha két vektor megegyezik, a vetület 0, így nyilván nem ezt fogjuk választani. Az eljárásban a mátrixok QR felbontását (38. sor) használtuk, mely a szokásos Gram-Schmidt-féle ortogonalizációs eljárás mátrix alakja. A QR felbontás a Gauss-Jordan eliminációt használja, mely csak sorokkal végzett alapműveleteken

alapuló stabil eljárás. Valójában csak a Q ortonormált mátrixot használjuk, jelen lépésben a második vektorát, mely az eredeti vektorra merőleges és az első és második által meghatározott síkba esik. A következő lépésben számítjuk ki a merőleges vetületet (39-40. sor) és ha ez nagyobb mint az eddigi legnagyobb érték volt, ezt a sorszámot (41. sor), illetve vektort (42. sor) eltároljuk. Végignézve mind a 125 vektort a végén azt hagyjuk meg, melynek a legnagyobb a vetülete és ezt hozzávesszük az elsőhöz (45. sor). Így a T mátrix már két vektort tartalmaz.

```

46  if abs(det(Xu*Xu'))>10^(-110)
47      S=((Xu*Xu')\Xu)*Y(:,n); % n. analizálandó összetevőt vizsgáljuk
48      y3=Xteszt(hvek,:)*S;
49      if hbrmse>rmse(y3,Yteszt(:,n))
50          hbrmse=rmse(y3,Yteszt(:,n));
51          hre=100*hbrmse./mean(Y(:,n));
52          rreg=corrcoef(y3,Yteszt(:,n));
53          reg1=rreg(1,2);
54          Bn=S;XT=Xu;
55          hvekt=hvek;
56      end
57  end
58 end
59 end
60 fontpar(n,1:length(hvekt))=hvekt;
61 statmut=[statmut;hbrmse,hre,reg1];

```

Itt nem ugrunk vissza hanem egyből kiszámoljuk a kalibrációs együtthatókat (47. sor), illetve ez alapján az előrejelzett analizálandó összetevő arányát is minden egyes tesztminta esetén (48. sor). Gyakorlatilag ez a két sor az ILS eljárás lényege. Ezzel kapcsolatban megjegyeznénk a MATLAB egy invertálással kapcsolatos előnyét. Mégpedig azt, hogy az `inv(Xu*Xu')*Xu` helyett lehet használni a `(Xu*Xu')\Xu` parancsot, mely Gauss-Jordan eliminációval vagy elemi bázistranszformációval oldja meg az egyenletrendszert és nem a Cramer-szabállyal, mely esetén a 0 közeli determináns miatt nagyfokú számolási pontatlanság lép fel. Érdeemes kipróbálni az `inv` parancsot, hogy lássuk, teljesen rossz eredményeket kapunk.

Az átlagos négyzetes hiba gyöke (RMSE) alapján eldöntjük(49-50. sor), hogy megtartjuk-e az eddig vizsgált vektorrendszert, a hozzá tartozó hullámhossz értékeket és kalibrációs együtthatókat (58-60. sor). Az 51-53. sorok további statisztikai mutatókat adnak meg. Ezeket nem használjuk a fenti döntésnél, csak a későbbi elemzés miatt számítjuk ki.

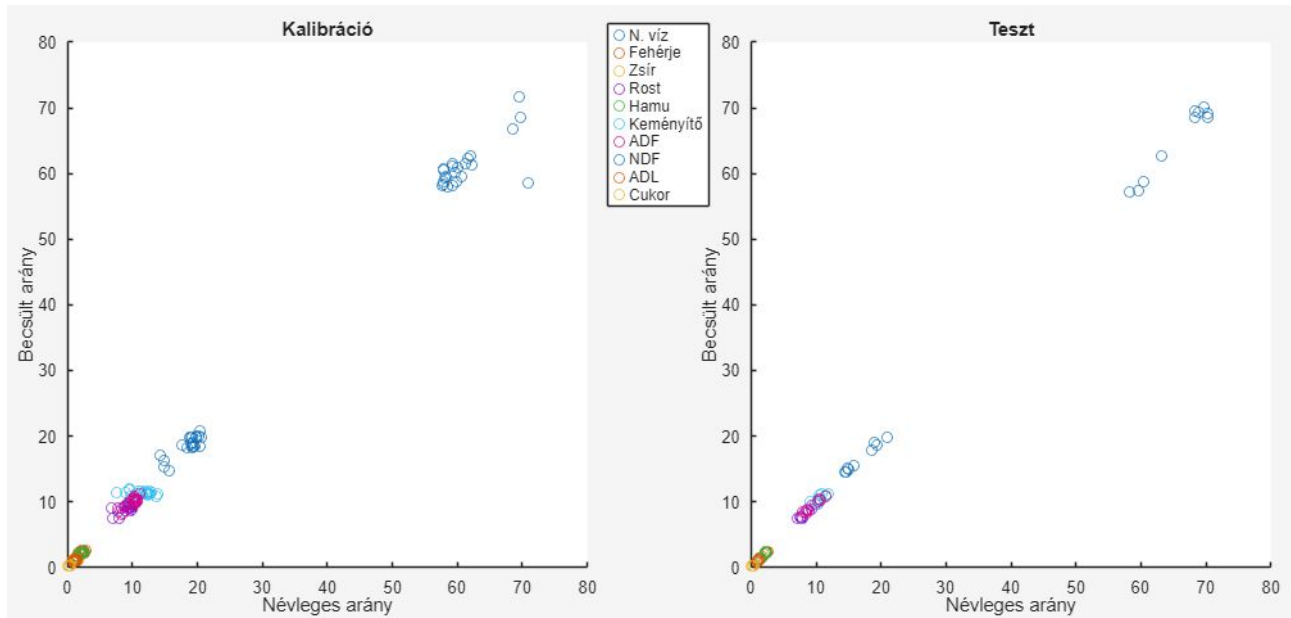
Majd most ugrunk vissza a 34. sorra, ahol a 3. vektort keressük. A 36. sortól végignézzük újra az összes vektort, hogy melyiknek a legnagyobb a vetülete. Majd ezt újra hozzávesszük a T mátrixhoz és újra megnézzük, hogy ez jobb eredményt ad-e. Majd a 4. vektort keressük egészen a 23. vektorig (34. sor). Tovább nem tudunk menni, mert a terünk 23 dimenzós.

```

62 S=Bn;
63 y3=XT'*S;
64 subplot(1,2,1)
65 hold on
66 plot(Y(:,n),[y3'],'o')
67 y2=Xteszt(hvekt,:)*S;
68 subplot(1,2,2)
69 hold on
70 plot(Yteszt(:,n),[y2'],'o')
71 end
72 fontpar'
73 statmut'

```

Most egészen a 30. sorig megyünk vissza, és megismételjük az egész eljárást úgy, hogy a kezdővektor a 2. vektor, majd a 3. vektor, végül egészen a 125. vektorig. Így végignéztük az összes olyan vektorkombinációt, melyet úgy választottunk, hogy a legnagyobb vetületeket választottuk a merőleges térre. A kapott hullámhosszakat (60. sor) és statisztikai mutatókat (61. sor) letároltuk. Végül 62-70. sorban a kalibrációs és a tesztkészletre vonatkozóan is kirajzoltattuk az eredeti és az előrejelzett értékpárokat. Mind a kalibrációs, mind a teszt adatok viszonylag jól illeszkednek az  $y = x$  egyenesre. Az  $x$  tengelyen az eredeti adatok, az  $y$  tengelyen a számolt értékek láthatók (4. ábra).



4. ábra. Az összes vizsgált összetevő becsült és névleges arányának kapcsolata a kalibrációs és a teszt minták esetén

Ezután a 27. sortól kezdődően a 2., majd a 3., stb analizálandó összetevőre is elvégezzük a fenti számításokat. A kapott eredményeket két táblázatba foglaltuk össze: a 2. táblázat a kiválasztott hullámhosszakot tartalmazza az egyes analizálandó összetevőkre vonatkozóan, a 3. táblázat a RMSE, REP és az  $r$  fent tárgyalt statisztikai mutatókat.

2. táblázat. A vizsgált összetevők esetén az SPA algoritmus által kiválasztott hullámhosszok

	N..víz	Fehérje	Zsír	Rost	Hamu	Kem	ADF	NDF	ADL	Cukor
1. hhossz(nm)	1541	1355	1237	1348	1404	1262	958	1330	1392	1367
2. hhossz(nm)	1398	1404	1417	1404	1324	1417	1429	1398	1491	1404
3. hhossz(nm)	1522	1392	1392	1392	1386	1392	1386	1386		1392
4. hhossz(nm)			939		1398		1404	1392		
5. hhossz(nm)							1125	933		
6. hhossz(nm)							908	1100		
7. hhossz(nm)							1621	1249		
8. hhossz(nm)								908		
9. hhossz(nm)								1609		
10. hhossz(nm)								1454		

A validációra általánosan elfogadott 2-5-10 szabály: kevesebb mint 2% relatív hiba, kiváló; kevesebb mint 5%, elég jó; kevesebb mint 10%, jó; több mint 10%, rossz. A cukor kivételével a relatív hiba 10% alá, a vizsgált összetevők nagy részénél 5% alá csökkent, míg egy esetben kiváló értéket kaptunk. Ha az  $r$  korrelációs együttható értékeit nézzük a legtöbb analizálandó összetevő esetén az is megfelel az irodalomban általánosan elfogadott 0,8-nál nagyobb értéknek.

3. táblázat. A vizsgált összetevők esetén a tesztmintákra vonatkozó statisztikai mutatók

	N..víz	Fehérje	Zsír	Rost	Hamu	Kem	ADF	NDF	ADL	Cukor
RMSE	1.2087	0.1055	0.0405	0.2628	0.0997	0.4434	0.3270	0.4786	0.0804	0.0697
REP	1.9751	4.5977	3.5401	2.8971	4.8562	3.9341	3.2839	2.5771	7.5789	25.0507
r	0.9845	0.9832	0.9118	0.9545	0.9711	0.7710	0.9861	0.9925	0.8806	0.7709

## 8. Összegzés

Kukoricaszilázs minták NIR spektrumait elemeztem inverz legkisebb négyzetek módszert (ILS) használva. A módszer a matematikai nehézségek miatt kevésbé használatos. Összességében kijelenthetjük, hogy a hullámhossz kiválasztásra az egymás utáni vetítési algoritmust használva illetve a számolás stabilitását fokozó matematikai algoritmusokat alkalmazva a kapott eredmények a legtöbb vizsgált összetevő esetén a jó és kiváló kategóriába estek. További vizsgálatok tárgyát képezheti ezen eredmények más módszerekkel való összehasonlítása.

## Köszönetnyilvánítás

Köszönettel tartozunk a kutatás támogatásáért, amely a GINOP PLUSZ-2.1.1-21-2022-00144 „Takarmányok telepi mérésére alk. közeli-infravörös spektroszkópiás műszer-, módszer- és modell-fejlesztés” pályázat keretében valósult meg.

Külön köszönetemet fejezem ki a bírálónak a hasznos és segítő észrevételeiért.

## Hivatkozások

- [1] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvao, T. Yoneyama, H. C. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemom. Intell. Lab. Syst.*, vol. 57, pp. 65–73, 2001. [https://doi.org/10.1016/S0169-7439\(01\)00119-8](https://doi.org/10.1016/S0169-7439(01)00119-8)
- [2] G. Bázár, "Application possibilities of near infrared spectroscopy in the qualification of pork, meat products and lard," PhD thesis, Kaposvár University, 2011.
- [3] I. Ben-Gera and K. Norris, "Direct spectrophotometric determination of fat and moisture in meat products," *J. Food Sci.*, vol. 33, pp. 64–67, 1968. <https://doi.org/10.1111/j.1365-2621.1968.tb00885.x>
- [4] D. Cozzolino, M. J. Kwiatkowski, R. G. Damberg, W. U. Cynkar, L. J. Janik, G. Skouroumounis, and A. Gishen, "Analysis of elements in wine using near infrared spectroscopy and partial least squares regression," *Talanta*, vol. 74, pp. 711–716, 2008. <https://doi.org/10.1016/j.talanta.2007.06.045>
- [5] R. K. H. Galvao, M. C. U. Araújo, W. D. Fragoso, E. C. Silva, G. E. José, S. F. C. Soares, and H. M. Paiva, "A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm," *Chemom. Intell. Lab. Syst.*, vol. 92, pp. 83–91, 2008.
- [6] J. Farkas, „Selected achievements of the Central Food Research Institute, Budapest, from its establishment until the turn of the 21st century,” *Acta Alimentaria*, vol. 38, suppl., pp. 3–20, 2009.
- [7] Y. H. Gado, A. Omeralfaroug, A. Szabó, T. Tóth, and G. Bazar, „Feeding mixed silages of winter cereals and Italian ryegrass can modify the fatty acid and odor profile of bovine milk,” *Agriculture*, vol. 13, no. 2, art. 381, pp. 1–15, 2023.

- [8] M. Görbe, "An overview of sample and spectrum preparation methods for near-infrared spectroscopy", *GRADUS*, vol. 12, No.1, pp. 1–6, 2025. <https://doi.org/10.47833/2025.1.ENG.009>
- [9] D. M. Haaland and E. V. Thomas, "Partial least-squares methods for spectral analysis. 1. Relation to other quantitative calibration methods and the extraction of qualitative information," *Anal. Chem.*, vol. 60, pp. 1193–1202, 1988.
- [10] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–70, 1970.
- [11] J. H. Kalivas, "Basis sets for multivariate regression," *Anal. Chim. Acta*, vol. 428, pp. 31–40, 2001. [https://doi.org/10.1016/S00003-2670\(00\)01225-3](https://doi.org/10.1016/S00003-2670(00)01225-3)
- [12] K. J. Kaffka, K. H. Norris, J. Rerédi, and A. Balogh, „Attempts to determine oil, protein, water and fiber content in sunflower seeds by the NIR technique,” *Acta Alimentaria*, vol. 11, pp. 253–269, 1982.
- [13] K. J. Kaffka and Zs. Seregély, „PQS (Polar Qualification System): The new data reduction and product qualification method,” *Acta Alimentaria*, vol. 31, pp. 3–20, 2002.
- [14] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Amsterdam, The Netherlands: Elsevier, 1997, chs. 17 and 36.
- [15] T. Næs, T. Isaksson, T. Fearn, and T. Davies, *Multivariate Calibration and Classification*. Chichester, UK: NIR Publications, 2002.
- [16] K. H. Norris and J. R. Hart, "Direct spectrophotometric determination of moisture content of grain and seeds," in *Proc. 1963 Int. Symp. Humidity and Moisture*, vol. 4, pp. 19–25, New York, NY, USA: Reinhold Publishing Co., 1965.
- [17] A.C. Olivieri, *Introduction to Multivariate Calibration*, Springer, 2024. <https://doi.org/10.1007/978-3-031-64144-2>
- [18] H. M. Paiva, S. F. C. Soares, R. K. H. Galvao, and M. C. U. Araújo, "A graphical user interface for variable selection employing the successive projections algorithm," *Chemom. Intell. Lab. Syst.*, vol. 118, pp. 260–266, 2012.
- [19] S. F. C. Soares, A. A. Gomes, M. C. U. Araújo, A. R. Galvao Filho, and R. K. H. Galvao, "The successive projections algorithm," *Trends Anal. Chem.*, vol. 42, pp. 84–98, 2013. <https://doi.org/10.1016/j.trac.2012.09.006>