

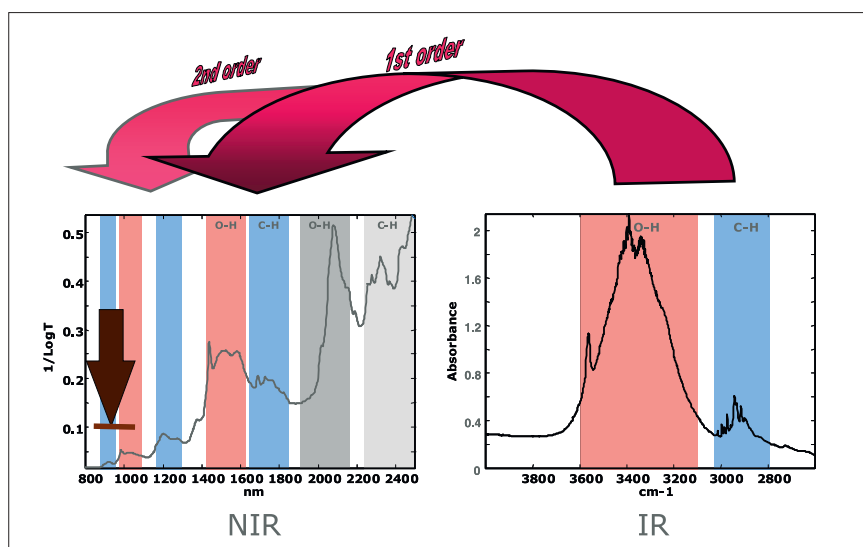
En spektroskopisk hurtigmetode: Protein i hvedekerner

Korn- og foderindustrien var blandt de første, der fik øjnene op for spektroskopiske hurtigmetoder og kemometri. Det handler om kvalitet og effektivitet i kvalitetskontrollen

Af Soren Balling Engelsen, Rasmus Bro & Lars Norgaard, Institut for Fødevidenskab, Det Biovidenskabelige Fakultet, Københavns Universitet

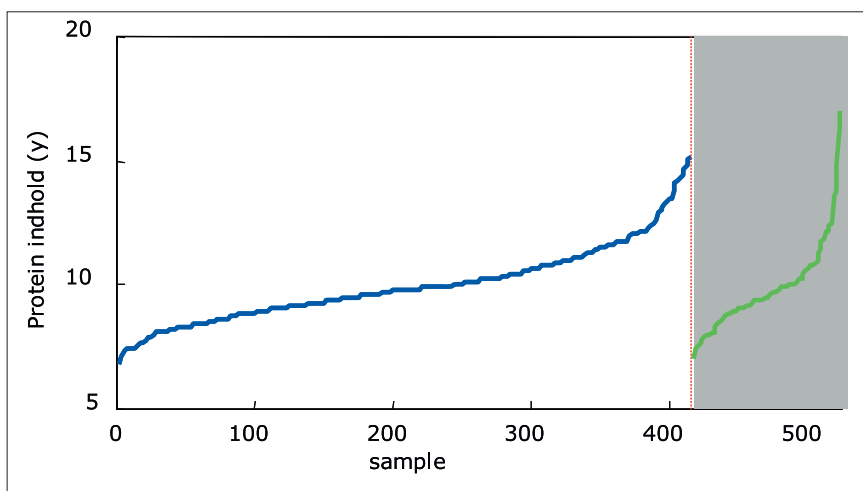
I de senere år har hurtige spektroskopiske målinger revolutioneret kvalitetskontrollen i den primære fødevarerproduktion. Nærinfrarød spektroskopi (NIR og NIT) er blevet implementeret til præcis og hurtig kvalitetsovervågning af millioner af korn-, mælke- og kødprøver. Vi har været vidne til et paradigmeskift i fødevarerindustrien, hvor ingeniører nu udforsker processer ved at anvende "modelfrie" empiriske metoder og spektroskopiske sensorer til at supplere de traditionelle fysiske/kemiske "hårde" modeller og univariate målinger. Nøglen til denne succes ligger i synergien mellem spektroskopi og kemometri. Vi vil her se på et eksempel fra kornindustrien, der demonstrerer de unikke fordele ved disse teknologier.

Vi har i en tidligere klumme været inde på fordelene ved nærinfrarød spektroskopi, som har været hoveddrivkraften i udviklingen af den tidlige kemometri i 80'erne. Nogle af disse fordele er specielt indlysende for applikationer inden for korn- og foderstofområdet, og det er heller ikke noget tilfælde, at *referencebogen* inden for nærinfrarød spektroskopi: "Near-Infrared Technology in the Agricultural and Food Industries" er udgivet af den amerikanske forening for cerealie kemikere (AACC) i 1987 [1]. Nærinfrarød spektroskopi måler overtoner og kombinationstoner af de fundamentale molekulære vibrationer, som ligger i det infrarøde område. Figur 1 viser, hvorledes information om de fundamentale molekulære vibrationer gentages som overtoner i det nærinfrarøde spektrum. Det ses, at der kan opnås flere fordele ved at måle på overtonerne frem for de fundamentale svingninger. For det første absorberer første overtone kun ca. 1/10 så kraftigt og den anden overtone kun ca. 1/100 så kraftigt som den fundamentale vibration. Dette gør det muligt at se gennem større mængde prøve, hvilket alt andet lige vil øge samplingeffektiviteten. Desuden ses også, at den spektrale opløsning er større ved første overtone, og at man derfor kan forvente bedre/mere information. I denne klumme vil vi dog se på et andet område af det nærinfrarøde spektrum nemlig det snævre spektrale område fra 850 nm til 1050 nm (figur 1 – sort pil). Dette spektrale område indeholder anden overtone af O-H- og N-H-stræknings-

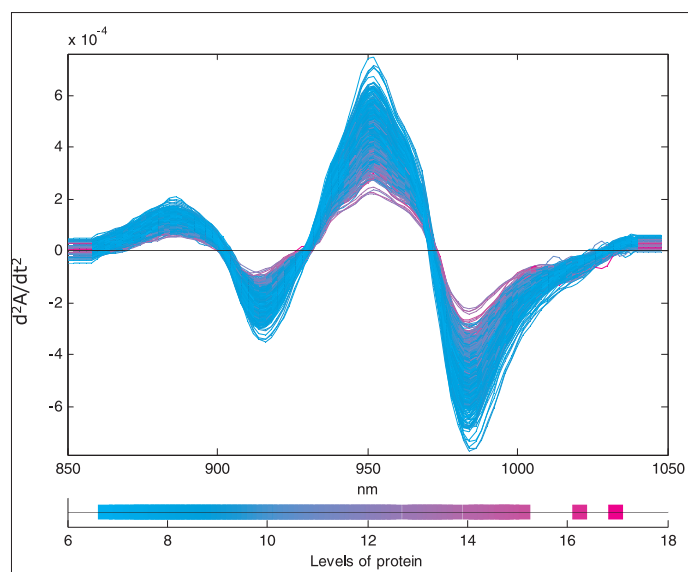
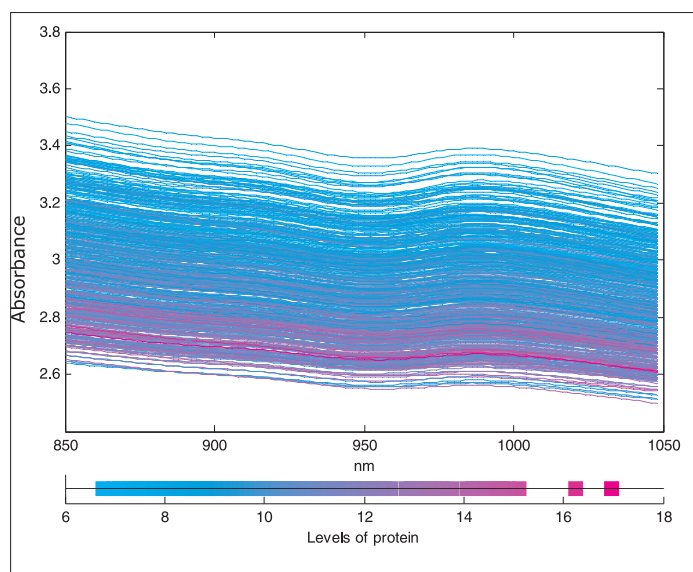


Figur 1. Illustration af hvorledes fundamentale molekulære vibrationer i det infrarøde spektrum giver anledning til overtoner i det nærinfrarøde område. Fundamentale C-H- og O-H-strækningsvibrationer for sukrose gentages som første overtone og anden overtone i det nærinfrarøde spektrum.

vibrationer og tredje overtone af C-H-strækningsvibrationer og dermed i princippet information fra alle organiske/biologiske



Figur 2. Hvedekernernes protein indholdssorteret efter proteinindhold og opdelt i henholdsvis kalibrerings sæt og testsæt.



Figur 3. (A) Rå nærinfrarød transmittans (NIT)-spektre (850-1050 nm) af 415 enkelthvedekerner. (B) Anden afledte af NIT-spektre. De enkelte hvedespektre er farvede i forhold til deres proteinindhold.

forbindelser. Desuden har dette område meget lav absorptions, hvorfor det er muligt at måle igennem selv stærkt absorberende materialer som korn.

Hvededata

Indtil for nyligt er næsten alle kornkvalitetsanalyser foretaget på prøver bestående af hundredvis af kerner og meget ofte i formålet form. Derved tabes informationen vedrørende de enkelte kerner. Hvis man vil vide noget om homogeniteten af et lot korn er der ingen vej uden om enkeltkerneanalyser. Enkeltkerneanalyser bidrager til en øget forståelse for variationen i et lot korn og kan benyttes til at optimere valget af kornsort og anvendelse. Hurtig ikke-destruktiv kvalitetsanalyse er derfor et værdifuldt værktøj i kornforædling og kvalitetssortering.

I denne klumme vil vi evaluere et sæt hvedekerner (415) fra 43 forskellige hvedesorter dyrket to forskellige steder i Danmark – vi bruger det som et kalibreringssæt. Som testsæt, til at verificere at vores model virker, benytter vi et sæt hvedekerner (108), der repræsenterer 11 forskellige hvedesorter målt to måneder senere og repræsenterende et lidt større range [2].

Vi vil fokusere på hvedekernerens proteinindhold, da det i store træk bestemmer anvendelsen af de forskellige hvedesorter. Det er da også proteinindholdet, der ofte belønnes for ved afregning til landmændene. For at udvikle en spektroskopisk hurtigmetode er det normalt nødvendigt at foretage referencemålinger, i dette tilfælde proteinindholdet som responsvariabel. Når man ønsker at bestemme proteinindholdet i korn, er det stadig den gode gamle Kjeldahl-metode (år 1883), der benyttes til at bestemme nitrogenindhold, som via en faktor omregnes til proteinindhold. Metoden er baseret på forudsætningen, at protein indeholder 16% nitrogen og at ikke-protein nitrogen kan negligeres. Ikke særligt elegant, men effektivt. Figur 2 viser proteinvariationen i de to prøvesæt. Som det fremgår af figuren varierer proteinindholdet i hvedekernerne fra ca. 6% til ca. 17%.

Nærinfrarøde spektre

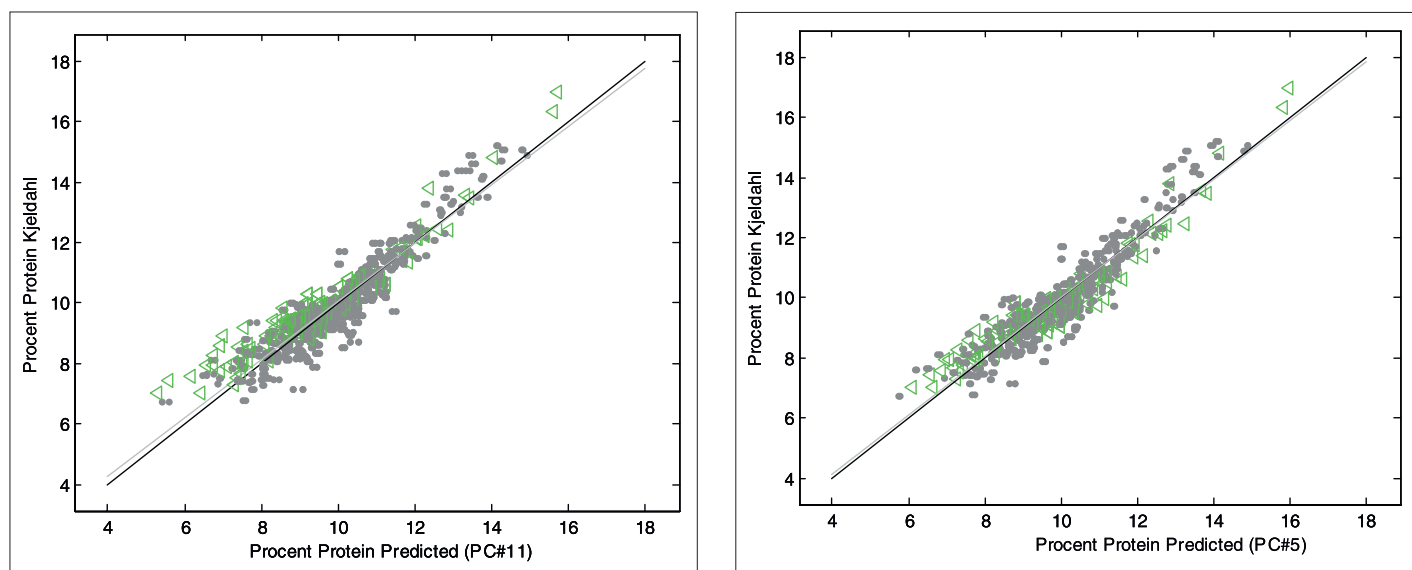
Enkeltkerne nærinfrarøde transmissions (NIT)-spektre blev målt på en Infratec 1255 Food and Feed Analyzer (FOSS). Hver kerne blev placeret i en karrusel med plads til 23 enkeltkerner og transmissionsspektre blev målt fra 850-1050 nm (disse data

kan downloades på www.models.life.ku.dk). Det tager ca. 90 sekunder at måle de 23 kerner i en karrusel. Figur 3A viser de næstest featureløse NIT-spektre af de 523 hvedekerner, og det bemærkes, at der kun er en svag tendens til at kerner med højt proteinindhold har lave absorptioner. Det bemærkes også, at NIT-spektrene trods den lave absorption ved den anden over-toner har betragtelige absorptionsniveauer. Ved klassisk kvantitativ IR-spektroskopi skal man holde sig under 0,5 absorptionsenheder. I disse NIT-spektre ses absorptionsniveauer på over 3, hvilket svarer til at under 0,1% af lyset slipper igennem prøven til detektoren. Der er af forskellige årsager tradition for, at man tager de anden afledte af nærinfrarøde spektre, da de anden afledte fremkalder variationen i spektrene og fjerner simpelt offset og lineære trends, som skyldes forskellige lysspredningsfænomener (vi vil vende tilbage til dette i en senere klumme). Figur 3B viser de anden afledte af de 523 hvedekernespektre, og først efter denne transformation træder proteinvariationen tydeligt frem.

Kalibrering af hurtigmetode

Partial Least Squares Regression (PLS) er gennemgået i den forrige klumme og bruges rutinemæssigt til at sammenholde spektroskopiske data (hurtige målinger) med kemiske/fysiske data (langsomme målinger). Vi demonstrerer nu opbygningen af en prædiktiv PLS-model, som er i stand til at prædiktere proteinindholdet i en hvedekerne ud fra et målt NIT-spektrum på selvsamme kerne. Til dette formål udvikles en PLS-kalibreringsmodel ud fra de 415 NIT-spektre (X) og 415 referencebestemmelser af proteinindhold (y) for hvedekernerne i kalibreringssættet. For at gøre PLS-modelleringen så simpel som mulig har vi centreret alle data og benyttet fuld krydsvalidering (denne teknik beskrives i en senere klumme) til bestemmelse af antallet af signifikante PLS-komponenter.

Indledningsvis forsøger vi at lave en PLS-model uden at benytte anden afledte af spektrene. Hvis vi benytter NIT-råspektrene (figur 3A) opnås en *Root Mean Square Error of Cross Validation* (RMSECV), dvs. en gennemsnitlig fejl, på 0,55% protein ved brug af 11 PLS-komponenter. Denne kalibreringsfejl er i samme størrelsesorden som standardafvigelsen ved referencebestemmelsen. Det betyder, at fejlen i PLS-modellens prædiktioner primært kan henføres til referencebestemmelsen ▶



Figur 4. Predicted versus measured plot for proteinindhold i hvedekerner. (A) PLS-model (11 komponenter) baseret på de rå NIT-spektre og (B) PLS-model (5 komponenter) baseret på de anden afledte NIT-spektre. Kalibreringssættet er farvet gråt og testsættet er farvet grønt.

som den holdes op imod; eller sagt med andre ord, at PLS-modellen er tæt på at have en betydningsløs fejl. Korrelationskoefficienten for kalibreringsmodellen er 0,93 (tabel 1). Derefter præsikeres de 108 hvedekernespektre i testsættet, og vi ser at *Root*

Mean Square Error of Cross Prediction (RMSEP) er større end kalibreringsfejlen. Dette er normalt et tegn på en lettere overfittet model, hvilket betyder at der er anvendt for mange PLS-komponenter i modellen. Da scatterplottet for de prædikterede værdier mod de aktuelle værdier (figur 4A) ser lineært ud, kan modellen accepteres, men det er bekymrende, at fejlen i testsættet er så meget større end fejlen i kalibreringssættet.

Hvis vi i stedet benytter de anden afledte af NIT-spektrene som input til PLS-modellen opnås en langt mere simpel model med kun fem PLS-komponenter. Denne kalibreringsmodel giver ligesom for de rå NIT-spektre en RMSECV på 0,56% protein, men ved brug af kun fem PLS-komponenter. Derefter præsikeres de 108 hvedekernespektre i testsættet. I dette tilfælde observeres det, at RMSEP (0,52%) er ganske lig med RMSECV. Denne model er altså at foretrække frem for den tilsvarende model bygget på de rå NIT-spektre, først og fremmest pga. den mere simple PLS-model, men også pga. den bedre prædiktionsfejl.

PLS-modellen kan nu bruges som en spektroskopisk hurtigmetode til bestemmelse af proteinindhold ud fra et NIT-spektrum med samme eller mindre standardafvigelse som den langsommere kemiske Kjeldahl-metode. Derudover kan de optagne NIT-spektre samtidig bruges til bestemmelse af andre kvalitetsparametre for kernerne som f.eks. vandindhold eller måske bare til detektion af "syge" kerner.

Outro

Det er ikke svært at forestille sig, at denne hurtigmetode til bestemmelse af protein i hvedekerner kan benyttes industrielt

NIT data	# of PLS Komponenter	Korrelation		Prædiktionsfejl (% protein)	
		Kal. set (CV)	Test set	Kal. Set RMSECV	Test set RMSEP
Rå	11	0,93	0,96	0,55	0,70
2nd afledte	5	0,93	0,96	0,56	0,52

Tabel 1. Performance af PLS-kalibreringsmodel for bestemmelse af proteinindhold i enkelte hvedekerner.

til at sortere hvede efter forskellige kvalitetsparametre, og der er allerede nu på markedet forsøgsanlæg, som uden problemer kan sortere et ton enkelthvedekerner i timen. En af fordelene ved at benytte kemometri er, at det er nødvendigt at sortere efter kun én kemisk kvalitetsparameter. Det kan lige så vel være en kompleks kvalitetsparameter som f.eks. bagekvalitet.

I den næste klumme vil vi komme ind på, hvordan man kan korrigere sine nærinfrarøde spektre for uønskede lysspredningsfænomener, således at spektrene bliver bedre egnede til kemometrisk modellering; herunder hvorledes den gennemgængede NIT-hurtigmetode for protein i hvedekerner kan forbedres ved at bruge en dedikeret scatterkorrektion.

E-mail-adresser

Søren Balling Engelsen: se@life.ku.dk

Rasmus Bro: rb@life.ku.dk

Lars Nørsgaard: lan@life.ku.dk

Referencer

1. Near-Infrared Technology in the Agricultural and Food Industries (Eds. Williams and Norris), The American Association of Cereal Chemists, Inc., St. Paul, Minnesota, US, 1996 (ISBN: 0-913250-49-X).
2. D.K. Pedersen, H. Martens, J.P. Nielsen and S.B. Engelsen, Near-Infrared Absorption and Scattering Separated by Extended Inverted Signal Correction (EISC): Analysis of Near-Infrared Transmittance Spectra of Single Wheat Seeds, *Applied Spectroscopy* (2002), **56**, 1206-1214.