

# PCP: NYTT OG OPTIMALISERT PLOTT AV SKÅRER OG LADNINGER

Øyvind Langsrud  $\square$  og Tormod Næs  $\square$

$\square$  MATFORSK, Osloveien 1, 1430 Ås. • Avdeling for statistikk, Matematisk institutt, Universitet i Oslo

## INTRODUKSJON

Plott basert på PLS foretrekkes ofte fremfor PCR/PCA-plott fordi PLS-plottene inneholder mer  $Y$ -informasjon. Et interessant spørsmål er hvorvidt det går an å lage et plott som er enda mer optimalt med tanke på  $Y$ -informasjon.

Svaret er ja og den nye metoden, som er beskrevet i Langsrud and Næs (2001, 2003), heter *principal components of predictions* (PCP). PCP baseres på en allerede valgt regresjonsløsning fra for eksempel PLS.

## PCP

- Utfør PCA (SVD) på de predikerte  $Y$ -verdiene (fra for eksempel PLS):

$$\hat{Y} = USVT$$

- PCP-skårene kan skrives som lineære funksjoner av  $X$ :

$$U = XBVS^{-1}$$

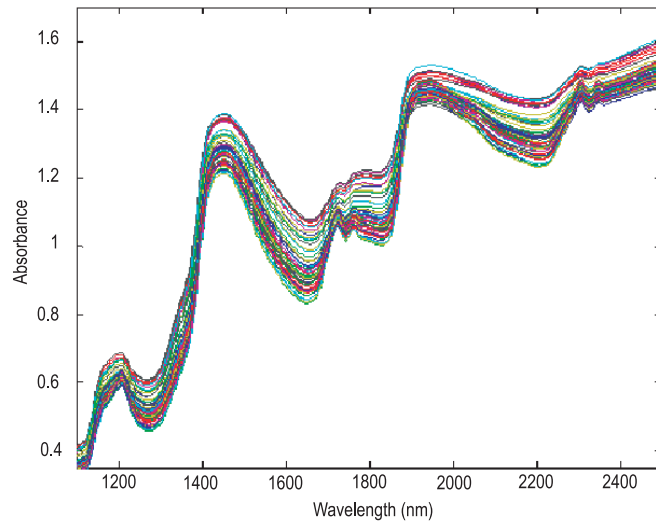
- $X$ - og  $Y$ -ladningene for PCP kan skrives henholdsvis som  $X^T U$  og  $V S$ .

PLS skårer og ladninger plottes som for PLS og PCR.

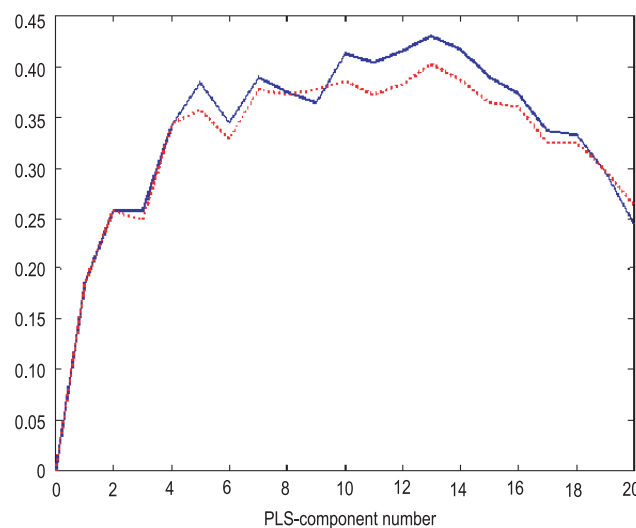
Med kun én responsvariabel finnes bare en komponent (proporsjonal med predikerte verdier). Den andre komponenten velges ved å maksimere forkortet  $X$ -varians ortogonalt på den første.

## EKSEMPEL

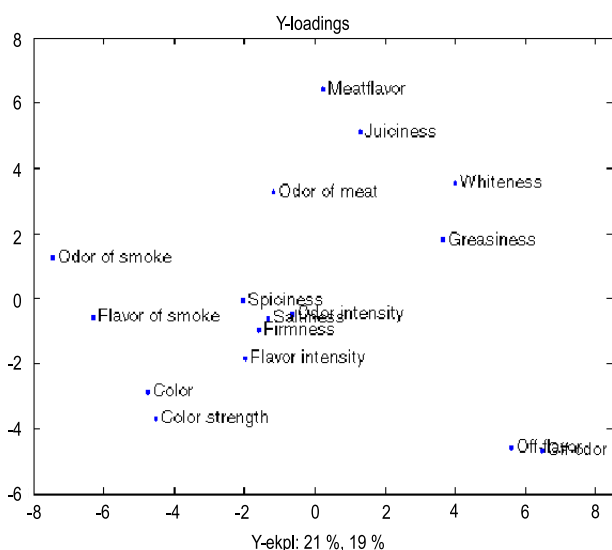
57 ulike pølsetyper ble målt med både NIR-spektroskopi (Figur 1) og ved sensorisk analyse (Ellekjær et al., 1994). To-komponent PCP gir nesten like bra resultat som PLS basert på et relativt stort antall komponenter. PLS med kun to komponenter forklarer en relativt bekjeden del av  $Y$ -variansen (Figur 2). PCP  $Y$ -landinger vises i Figur 3. I Figur 4 illustreres forskjellen mellom et skår-plott basert på PCP og et basert på PLS. Figur 5 viser et PCP skår-plott der *Meatflavor* er valgt som eneste responsvariabel.



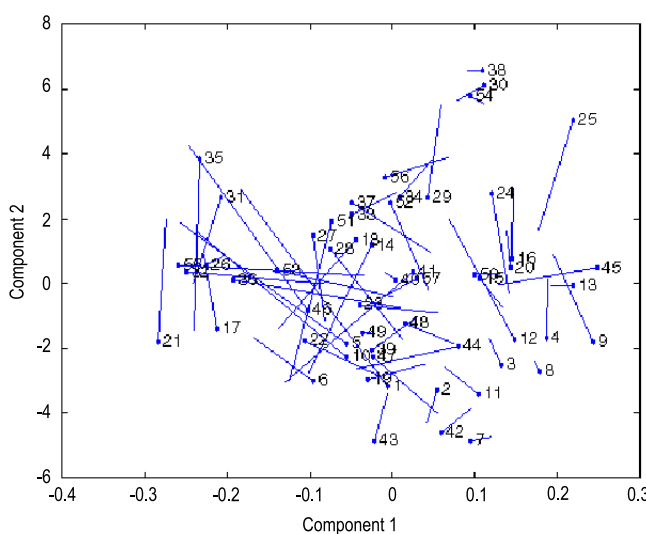
Figur 1: De 57 NIR spekterene.



Figur 2: Kryssvalidert forklart  $Y$ -varians for PLS (—) og for PLS etterfulgt av to-komponent PCP (---).



Figur 3: PCP  $Y$ -ladninger basert på den optimale PLS-løsningen (13 komponenter). Prosentverdiene for de to komponentene refererer til full kryssvalidering.



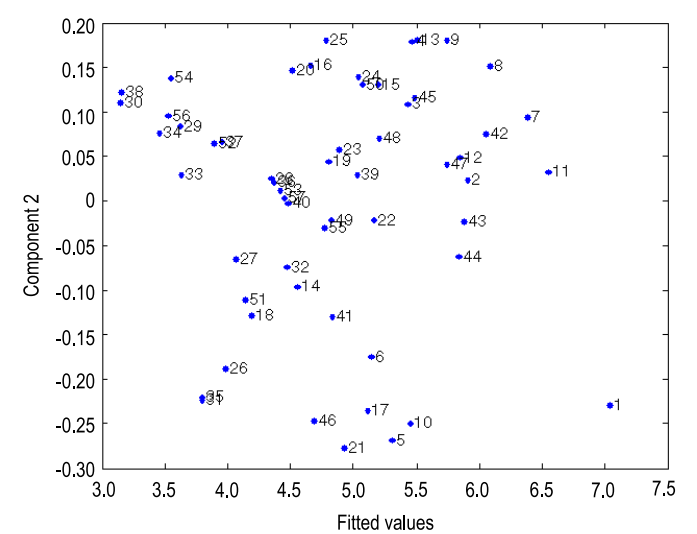
Figur 4: PLS skårer og roterte PCP skårer (den nummererte enden av linjene). Forskjeller som kun skyldes rotasjoner er eliminert ved bruk av Procrustes-rotasjoner.

## KONKLUSJON

- PCP gir nye plott som er optimale med tanke på  $Y$ -informasjon.
- PCP er spesielt viktig i situasjoner der man trenger mange PLS komponenter for å få gode prediksjoner.
- PCP erstatter ikke PLS og PCR, men kan ledsage disse metodene.
- PCP kan gi informasjon om dimensjonen på den delen av  $Y$  som kan predikeres fra  $X$ .
- Enhver regresjonsmetode kan være basis for PCP.
- PCP er en generalisering av *reduced rank regression*.
- Ved én responsvariabel kan predikerte verdier leses fra første akse.

## REFERANSER

- Ellekjær, M.R., Isaksson, T., Solheim, R. (1994), Assessment of Sensory Quality of Meat Sausages Using Near Infrared Spectroscopy, *Journal of Food Science*, **59**, 456-464.
- Langsrud, Ø. and Næs, T. (2001), Optimised score plot by principal components of predictions, *Proceedings of the 2nd International Symposium of PLS and Related Methods*, (Eds. Vinzi, Lauro, Morineau, Tenenhaus) Cisia-Ceresta Montreuil, France.
- Langsrud, Ø. and Næs, T. (2003), Optimised score plot by principal components of predictions, *Chemometrics & Intelligent Laboratory Systems*. Submitted for publication in the special PLS'01 issue.



Figur 5: PCP skår-plott når *Meatflavor* er eneste responsvariabel. Regresjonsløsningen er basert på fem PLS-komponenter. De to komponentene forklarer 52% av  $Y$  og 96% av  $X$ . Til sammenligning forklarer de to første PLS-komponentene 11% av  $Y$  og 99% av  $X$ .