

# Tucker3 kan håndtere alle typer af multivejsdata

PARAFAC støder indimellem på problemer med at modellere multivejsdata. Tucker3-modellen har ikke sådanne problemer, da den ligesom PCA altid kan modellere data.

Af Rasmus Bro, Søren Balling Engelsen, Institut for Fødevidenskabelig, Københavns Universitet og Lars Nørgaard, FOSS

PARAFAC har indimellem problemer med at modellere multivejsdata. Problemet er af matematisk art og er dermed ikke relateret til data. For sensoriske data (prøver  $\times$  deskriptorer  $\times$  dommere) er det f.eks. naturligt at ønske at anvende en PARAFAC-model. Også selvom der ikke ligger en decideret "hård" model, som Beers lov, bag sådanne data, som der gør ved modellering af fluorescensdata. Det synes blot rimeligt at visualisere sensoriske data vha. underliggende latente fænomener såsom "afsmag", "modenhed" eller lignende. Hvis en PARAFAC-model af sensoriske data virker, har man således basis for at fortolke og forstå data.

Men som vist i figur 1, så lykkes det desværre ikke altid. I figuren er vist plot for en PARAFAC-model af et sensorisk datasæt, hvor forskellige flødeoste blev vurderet af otte dommere mht. 23 forskellige attributter/deskriptorer. Data er også kort beskrevet i klummen i Dansk Kemi 5, 2012. I figuren ses et eksempel på en tofaktor-degenerering, hvor to komponenter er næsten identiske, men med omvendt fortegn.

I figur 1 ses det, at de to viste scores i første mode (oste) er negativt korrelerede, og hvad der måske er værre, så er loadings i sensorik-deskriptor-retningen (figur 1, til højre) modsat korrelerede. Det er vanskeligt rent visuelt at vurdere, hvad der sker imellem prøver og variable, og det er også vanskeligt at beskrive rent matematisk.

## Tucker virker

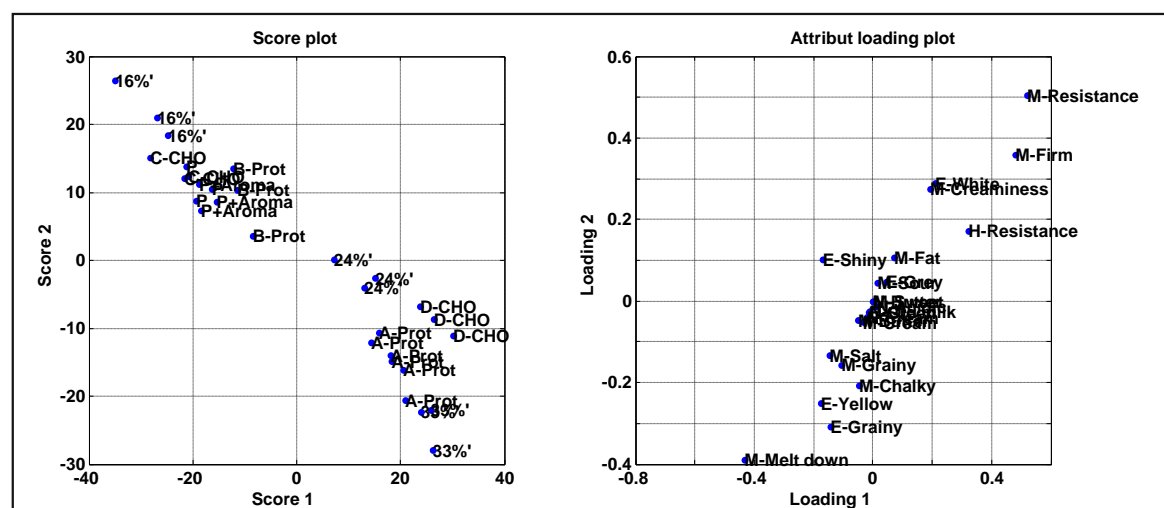
I stedet for PARAFAC forsøger vi at lave en Tucker3-model. I en sådan model skal vi beslutte antal komponenter for hver mode. Altså antal scorevektorer for oste-retningen, antal loadingvektorer for attribut/deskriptor-retningen og antal loadingvektorer for dommer-mode. I dette eksempel gør vi blot nogenlunde som i PARAFAC-modellen og anvender tre komponenter i hver retning. Det kalder man en (3,3,3)-komponent Tucker3-model. I figur 2 er vist Tucker3-scores og loadings svarende til PARAFAC-modellen i figur 1.

Det er tydeligt, at Tucker3-komponenterne ser meget anderledes ud. For det første så findes problemet med to-faktor-degenerering ikke. For det andet, så er Tucker3-modellen ikke unik som PARAFAC-modellen. I stedet har den egenskaber, der kan sammenlignes med PCA – inkl. kravet om at scores og loadings skal være ortogonale. Man ser, at prøver og attributter spreder sig langt bedre ud i de respektive plot pga. ortogonaliteten, og fordi der ikke er problemer med degenerering.

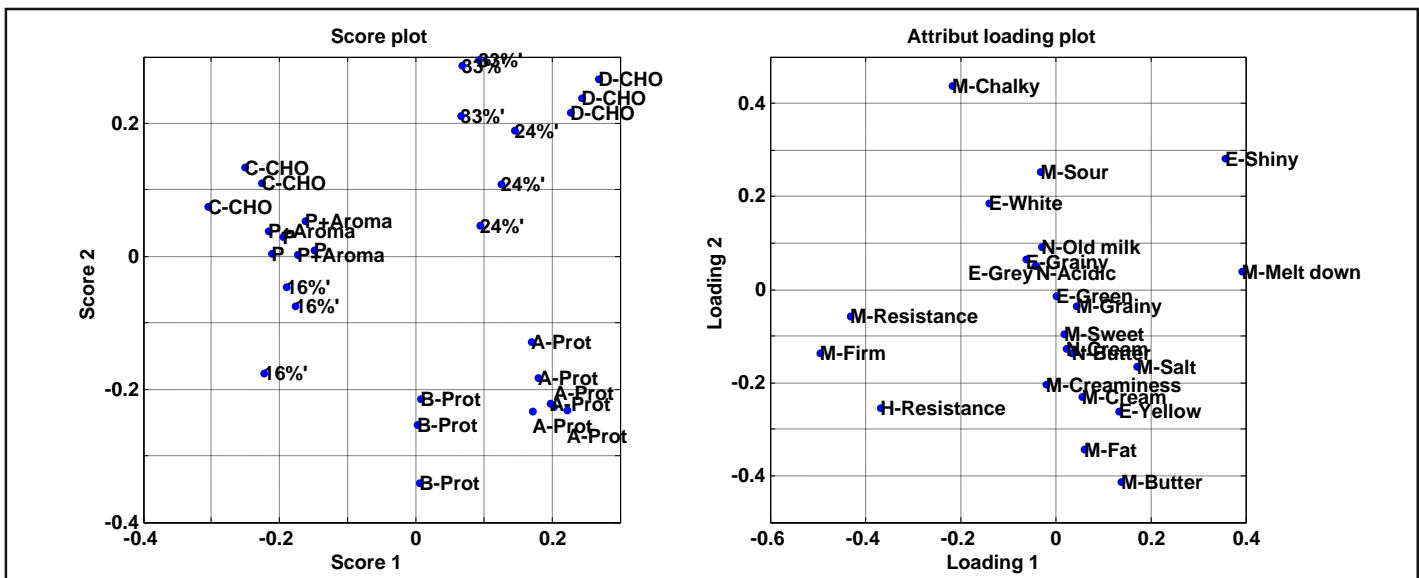
## Fortolkning af enkel-plot er let

Vi kan fortolke scoreplottet fuldstændig analogt til et scoreplot i PCA. Vi kan f.eks. se, at triplikater ligger tæt, hvilket fortæller, at de er meget ens relativt til de øvrige prøver. Vi kan også se, at prøverne mærket med D-CHO ligger ret langt fra prøverne mærket med 16% og dermed er forskellige fra disse relativt set og mht. den variation, de to komponenter beskriver.

På samme måde kan vi fortolke plottet af attributter og se, at **M-Resistance** og **M-Firm** er tæt på hinanden og derfor følges



Figur 1. Score- og loadingplot af trekomponent PARAFAC-model for sensoriske data. I dette tilfælde er 30 forskellige oste blevet bedømt af otte dommere mht. til 23 forskellige deskriptorer. Komponent et er plottet mod komponent to (figur fra Dansk Kemi 5, 2012).

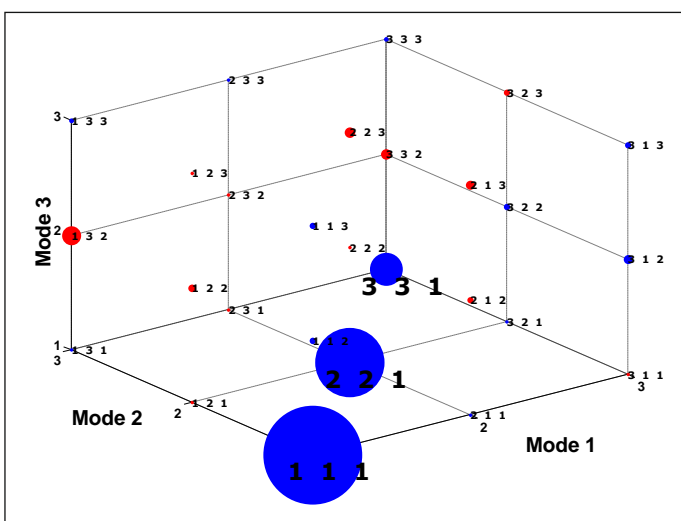


Figur 2. Score- og loadingplot af (3,3,3)-komponent Tucker3-model for de sensoriske data. Komponent et er plottet mod komponent to.

ad mht. variationen i disse komponenter. Vi kan også se, at **M-Chalky** er modsat korreleret **M-Butter**.

### Kobling af plot – Tucker-kernen

Fortolkning af Tucker-score- og loadingplot er helt analogt til PCA og derfor ikke det store problem for en erfaren kemometriker. Det er derimod noget mere vanskeligt, når man vil koble score og loadingplot. Hvis man f.eks. gerne vil vide, hvilke osteprøver der har høje værdier af attributten **M-Butter**, så kræver det, at man kigger på kernen, **G**, i Tucker3-modellen. Det er nemlig kernen, der viser, hvordan loadings i mode to hænger sammen med scores i mode et. I PCA og PARAFAC er denne del let, da score tre hænger sammen med loading tre og så videre. I Tucker3 må vi igennem kernen og resten af model-



Figur 3. Tucker-kerne fra (3,3,3)-Tucker3-model. Cirkler indikerer størrelsen af element, og blå farve angiver en positiv værdi, mens rød angiver en negativ.

len for at finde denne sammenhæng. I figur 3 er der vist en visualisering af kernen. Kernen har dimensionen  $3 \times 3 \times 3$ , hvilket giver 27 elementer. Element nummer  $i, j, k$  definerer, hvor vigtig komponenten med scorevektor  $i$ , anden-mode loadingvektor  $j$  og tredje-mode loadingvektor  $k$  er. Det lyder voldsomt at skulle

fortolke 27 kombinationer mellem de tre modes sammenlignet med, at vi i PARAFAC kun behøver at se på tre: 1,1,1; 2,2,2 og 3,3,3. Der er imidlertid hjælp at hente ved at kikke på kerneelementernes størrelse. I figur 3 er hvert kerneelements størrelse indikeret af størrelsen på dens cirkel. Som man kan se, så er *langt* de fleste elementer meget små, og det betyder eksplicit, at vi ikke behøver at bekymre os om dem. De beskriver så lidt variation, at det er uvæsentligt, når vi skal fortolke modellen.

I dette tilfælde kan man se, at vi kun behøver at fortolke komponent (1,1,1), (2,2,1) og (3,3,1). De resterende elementer er så små, at vi kan 'lade som om' de er nul. Det betyder så også, at vi måske burde beregne en (3,3,1)-model fremfor en (3,3,3)-model. Kan det give mening? Egentlig passer det fint, for den tredje retning er dommer-retningen og ideelt er dommerne replikater. Så selvom der findes tre forskellige fænomener i oste- og attribut-retningen, så manifesterer disse tre fænomener sig på samme vis i dommer/replikat-retningen. Tucker3-kerneelementernes størrelse antyder således blot, at der er tale om et godt og veltrænet sensorisk panel.

Da vi kan se, at også (1,1,1)- og (2,2,1)-komponenterne er vigtige og har positive kerneværdier, kan vi nu nemt fortolke og kombinere de to plot i figur 2. De kan i dette tilfælde fortolkes næsten ligesom i en PCA-model. Den eneste lille finthe er, at vi skal huske at tage dommer-loadings med i fortolkningen. Alle dommere har negative elementer i loading et (ikke vist). Derfor bliver fortolkningen, at prøver med store **M-Butter**-værdier (negativ loading to) må være prøver med høje positive score to-værdier. Det vil eksempelvis sige, at prøverne mærket **C-CHO** har relativt høje **M-Butter**-værdier, hvilket efterfølgende bør og kan bekræftes i rådata.

### Outro

Tucker3 virker altid og er et meget stærkt eksplorativt redskab. Tucker3-modellen er let at fortolke, når man holder sig til et mode ad gangen. Når man skal koble forskellige modes, så må man være varsom med at se på kernen og de øvrige retninger i modellen.

### Referencer

1. R. Bro, E. M. Qannari, H. A. L. Kiers, Tormod Næs, M. B. Frøst. Multi-way models for sensory profiling data. J.Chemom. 22:36-45, 2008.