

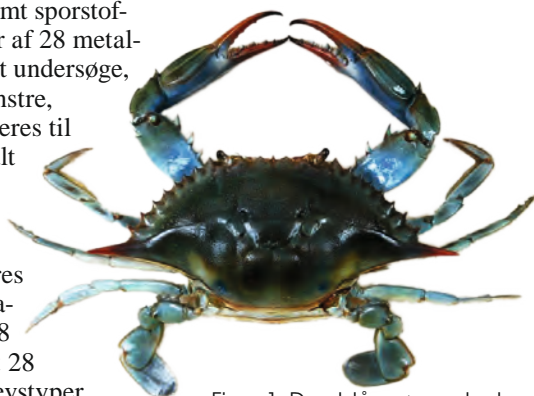


Tucker3-modellen er et stærkt redskab til eksplorativ analyse af multivejsdata. Modsat PARAFAC-modellen skal man ikke bekymre sig om numeriske problemer med algoritmen, eller om modellen giver matematisk mening. Her anvendes Tucker3 til at visualisere et trevejs datasæt fra miljøovervågning.

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

Den blå svømmekrabbe er af stor økonomisk betydning for fiskeriet i bl.a. staten North Carolina. Igennem nogle år var der imidlertid problemer med sygdom blandt disse krabber; bl.a. havde en stor andel af krabberne læsioner på skjoldet. Der blev spekuleret i, om disse problemer kunne tilskrives forurening fra landbrug, øget turisme og/eller minedrift [1,2]. En indsamling af syge og raske krabber blev gennemført i North Carolina. Man indsamlede 16 raske krabbeprovér fra Albemarle-sundet og 16 raske krabbeprovér fra Pamlico-floden samt 16 syge krabbeprovér fra Pamlico-floden.

Fra hver krabbeprové (én krabbeprové var i virkeligheden poolet fra tre krabber) blev der udtaget vævsprøver fra gæller, muskler og indre organer. Fra disse tre vævstyper blev der vha. ICP-AES bestemt sporstofkoncentrationer af 28 metaller. Målet var at undersøge, om der var mønstre, der kunne relateres til sygdom. Alt i alt var der tale om 3600 koncentrationer, som kunne arrangeres i en trevejs datastruktur med 48 krabbeprovér \times 28 metaller \times 3 vævstyper.



Figur 1. Den blå svømmekrabbe (*Callinectes sapidus*).

Forbehandling af data

Forbehandling af multivejsdata kan til tider være en kompliceret affære [3], men de fleste programmer til multivejsanalyse indeholder hjælpefunktioner. I dette tilfælde findes de individuelle metaller i meget forskellige koncentrationer i prøverne (tabel 1).

Hvis man ikke håndterer disse forskelle i niveau og spredning, vil en model af data kun afspejle de variable, der har meget stor spredning eller meget høje middelværdier. Dette håndteres ved at centrere og skalere data, så hver vektor af et metal for et væv (f.eks. 48 målinger af Zn i gællerne) fratrækkes sit gennemsnit. Dernæst skaleres hver matrix af metal (eksempel-

	Middelværdi	Spredning
Ag	0,4	0,3
Al	177,4	503,5
As	4,4	2,3
Ca	18,9	13,1
Cd	0,7	0,9
Co	0,3	0,3
Cr	0,3	0,5
Cu	97,5	95,7
Fe	415,6	1037,3
K	11,6	2,7
Li	0,6	1,5
Mg	36,6	13,2
Mn	80,4	188,3
Mo	0,3	0,2
Na	27,3	10
Ni	0,4	0,5
P	16,3	8,4
Pb	0,8	3,6
Se	1,1	1,1
Si	46,7	73
Sn	2,5	5,7
Ti	0,1	1
U	0,5	1,1
V	0,2	0,2
Zn	181,1	95,9

Tabel 1. Middelværdi og spredning på målte metalkoncentrationer – enheden er ppm.

vis alle Ag-målinger) med den samlede spredning på Ag. På denne vis har vi opnået noget, som vi konceptuelt kan sammenligne med autoskalering.

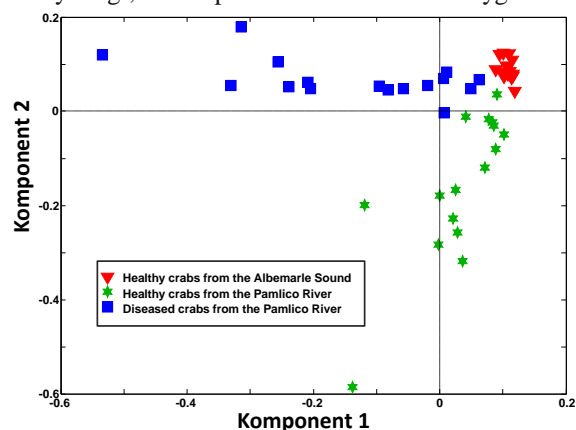
Antal komponenter

For at beregne en Tucker-model på de forbehandlede data, må antallet af komponenter besluttes. Da der er tale om en eksplorativ analyse, er det ikke kritisk at vælge det korrekte antal komponenter. Ønsket er at visualisere de mest væsentlige tendenser i data, så enhver model, der beskriver en rimelig andel af variationen, vil være anvendelig. Der vælges et antal komponenter, som er relativt lavt, men højt nok til at beskrive en rimelig del af variationen. I den oprindelige artikel søgte man at beskrive omkring 70% af variationen og noget lignende gøres her. Komponenter i den sidste retning, afspejlende de tre vævstyper, bidrager ikke meget til at beskrive data, så i denne retning vælges to komponenter. I de to andre retninger vælges tre komponenter. Dvs. at vi bestemmer en (3,3,2) Tucker3-model. Denne beskriver 68% af variationen i de forbehandlede data.

Og et kig på modellen

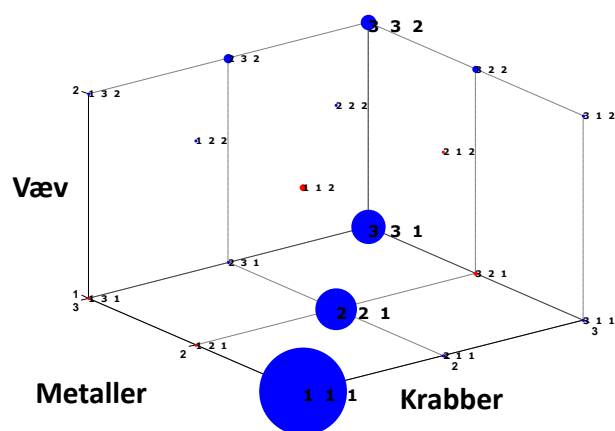
I figur 2 er vist et plot af score et og to for de 48 krabber. Som man kan se, så er der en meget tydelig gruppering.

Det er tydeligt, at komponent to er relateret til syge krabber.



Figur 2. Scoreplot for en (3,3,2) Tucker3-model.

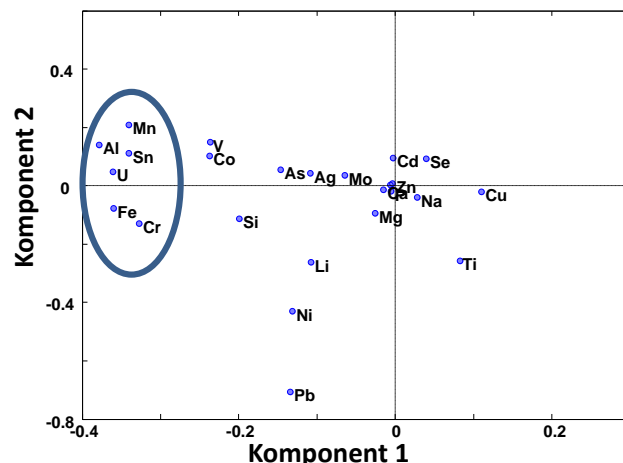
Ikke-syge krabber har meget lidt variation i denne retning. Det er derfor interessant at se, hvad denne retning repræsenterer. For at forstå det, må vi først igennem kernetensoren, som fortæller hvordan de tre scores, de tre metalloadings og de to vævsloadings spiller sammen. I figur 3 kan vi se kernen. Som det ofte er tilfældet, så er kernen rimelig simpel. Det



Figur 3. Kernen i en (3,3,2) Tucker3-model. De blå elementer er her positive.

er næsten kun komponent et i tredje vævsretning, der spiller ind, og for krabbe score to er det metalscore to, som er eneste væsentlige komponent. Denne kombination af komponenter er positivt korreleret, da kerne-elementet er positivt. Dvs. at de syge krabber er karakteriseret ved markant højere koncentrationer af de metaller, der ses som et styrende cluster i figur 4. Det er metaller som Al, Mn, Fe, Cr med flere.

Ydermere, så er loadings i komponent et i vævsretningen givet



Figur 4. Loadingplot for en (3,3,2) Tucker3-model for metaller.

ved værdierne [1,0 0,1 0,1], og det betyder, at dette fænomen primært er gældende for det første af de tre væv; nemlig gællerne. Denne iagttagelse bekræftes også efterfølgende i rådata.

Ifølge [1] kan disse forhøjede koncentrationer være en indikation på, at mineraler fra leret i den omgivende geografiske region absorberes af krabberne. Mineralerne kunne være opløseliggjort og frigivet fra leret pga. en kendt udledning af fluorid fra minedrift. Denne udledning stoppede nogle år efter den oprindelige artikel blev skrevet, og det samme gjorde forekomsten af syge krabber.

En række andre observationer kan let gøres i modellen, og det kan man se flere eksempler på i de oprindelige publikationer.

Outro

Som vist, så kan passende brug af Tucker-modellen være et stærkt redskab til at visualisere meget komplicerede data. På trods af en stor mængde kemiske koncentrationer målt på mange forskellige krabber og væv, kan en Tucker-model bruges til at lære nye sammenhænge om data, som så sidenhen kan danne baggrund for nye hypoteser og eksperimenter.

E-mail-adresser

Rasmus Bro: rb@life.dk.

Søren Balling Engelsen: se@life.ku.dk

Lars Nørgaard: LNO@foss.dk

Referencer

1. P. J. Gemperline, K. H. Miller, Terry L. West, J. E. Weinstein, J. C. Hamilton, J. T. Bray. Principal component analysis, trace elements, and blue crab shell disease. *Analytical Chemistry*. 64:523A-532A, 1992.
2. P. M. Kroonenberg, K. E. Basford, P. J. Gemperline. Grouping three-mode data with mixture methods: the case of the diseased blue crabs. *Journal of Chemometrics*, 18: 508-518, 2004.
3. R. Bro and A. K. Smilde. Centering and scaling in component analysis. *Journal of Chemometrics*. 17 (1):16-33, 2003.
4. Data tilgængelige fra <http://three-mode.leidenuniv.nl/> (13/8/2012).