

der også udført forsøg med nedbrydning i jord blandet med fæces. Den anvendte fæces var fra en kvie (estimeret vægt 450 kg), der 3 dage inden var blevet behandlet ved injektion med ivermectin 9 ml Ivomec® (10 mg/ml ivermectin opløsning, Merial Norden A/S, Skovlunde, Danmark). Fæces blev tilsat jorden i forholdet 3.5 og 7 % (m/m %, våd vægt) svarende til forventede mark koncentration, når der spredes 120 kg N/ha gylle og en kvie udgør 1.7 dyreenheder/ha. Prøver blev udtaget på følgende tidspunkter dag 0, 7, 14, 42 and 70 (n = 3).

Det ses af resultaterne, at nedbrydning forløber hurtigere i de aerobe prøver, når fæces blandes i jorden. Halveringstiden er henholdsvis $T_{1/2} = 12$ og 5 dage ved henholdsvis 3.5 og 7 % blandingerne, mens den er 37 dage i ren jord. I lignende forsøg med sulfadimethoxin (veterinært antibiotika) har man ligeledes observeret at nedbrydningshastigheden øges ved tilsætning af gylle [16]

Resultaterne samt data fra litteraturen indikerer at nedbrydning i jord og hvor jord blandes med møg/gylle/fæces ikke har samme nedbrydningshastighed. For bedre at kunne forudsige veterinære lægemidlers skæbne, er det derfor vigtigt at inddrage forsøg, hvor også jorden blandes i et naturligt forhold med møg/gylle/fæces. Eftersom matricerne møg/gylle/fæces er en meget heterogen størrelse kan det være svært at lave en standardiseret matrix. I forbindelse med CREAM, et forskningsprojekt finansieret af VELUX fonden, arbejdes der videre med dette forskningsområde.

ERAPharm

ERAPharm er et EU finansieret forskningsprojekt (EU nr. 511135) der var bevilliget inden for EU kommissionens 6. ramme-program. ERAPharm projektet bestod af 15 partner fordelt over hele Europa, heraf to partnere var fra Danmark (FARMA og DMU) samt en partner fra Canada. Projektet blev afsluttet efter 3 år i oktober 2007.

kak@farma.ku.dk

Referencer

- [1] EMEA. Guideline on the environmental impact assessment for veterinary medicinal products CVMP/ERA/418282/2005. European Medicine Agency, London, UK. <http://www.emea.europa.eu/pdfs/vet/era/41828205fin.pdf>
- [2] W.C.Campbell, M.H.Fischer, E.O.Stapley, G.Albers-Schönberg, T.A.Jacob. 1983 *Science* 221:823
- [3] B.A.Halley, T.A.Jacob, A.Y.H.Lu. 1989 *Chemosphere* 18(7-8):1543
- [4] R.Wall, L.Strong. 1987 *Nature* 327:418
- [5] M.Madsen, J.Grønvold, P.Nansen, P.Holter. 1988 *Acta Veterinaria Scandinavica* 29:515
- [6] C.Sommer, B.Steffansen, B.O.Nielsen, J.Grønvold, K.-M.V.Jensen, J.B.Jespersen, J.Springborg, P.Nansen. 1992 *Bull. Ent. Res.* 82:257
- [7] K.G.Wardhaugh, P.Holter, W.A.Whitby, K.Shelley. 1996 *Australian Veterinary Journal* 74(5):370
- [8] K.G.Wardhaugh, P.Holter, B.Longstaff. 2001 *Australian Veterinary Journal* 79(2):125
- [9] J.Jensen, P.H.Krogh, L.E.Sverdrup. 2003 *Chemosphere* 50(3):437
- [10] J.Garric, B.Vollat, K.Duis, A.Pery, T.Junker, M.Ramil, G.Fink, T.A.Ternes. 2007 *Chemosphere* 69(6):903
- [11] OECD. 2000. OECD Guideline for Testing of Chemicals No. 106. Adsorption - Desorption Using a Batch Equilibrium Method, Paris, OECD
- [12] OECD. 2002. OECD Guideline for Testing of Chemicals No. 307. Aerobic and anaerobic transformation in soil. Paris, OECD
- [13] K.A.Krogh, T.Søeborg, B.Brodin, B.Halling-Sørensen. Submitted to *J. Environ. Qual.*
- [14] K.A.Krogh, E.Bjorklund, D.Löffler, T.Ternes, B.Halling-Sørensen. (in prep)
- [15] CRUCIAL (<http://www.gf.life.ku.dk/Forsoegsgaarde/omfsg/arealer/vaerksted/crucial.aspx>)
- [16] Q.Q.Wang, S.A.Bradford, W.Zheng, S.R.Yates. 2006 *J. Environ. Qual.* 35(6):2162



Vi tager udgangspunkt i et eksempel fra et tidligere projekt udført i samarbejde med Danisco Sugar. Formålet med projektet var bl.a. at undersøge om fysisk-kemiske kvalitetsmålinger af sukker kan afsløre, hvor sukkeret er produceret.

	Fabrik	Aske	Farve	Turbiditet	Kornstørrelse	Kornstørrelse	SO ₂	Invert	Flocc	Rest	Amino N
C02	0	0.01	27	0.37	0.57	0.21	2.7	5	0	7	1.92
C03	0	0.009	19	0.22	0.53	0.22	4.6	18	1	4	2.19
C04	0	0.011	18	0.21	0.54	0.2	5.4	18	0	1	2.77
C05	0	0.012	15	0.22	0.54	0.17	5.8	50	1	5	3.18
C06	0	0.01	19	0.21	0.59	0.19	4.8	39	0	3	3.2
C07	0	0.012	27	0.15	0.57	0.2	6.1	29	1	3	3.6
C08	0	0.012	21	0.27	0.58	0.2	6.1	57	0	5	4.59
C09	0	0.011	22	0.26	0.57	0.17	4.5	57	1	8	3.23
C10	0	0.01	25	0.31	0.61	0.15	2.7	35	0	4	3.28
C11	0	0.01	22	0.32	0.56	0.2	3.4	50	1	8	3.45
C12	0	0.011	22	0.25	0.57	0.15	4.3	79	0	3	3.68
C13	0	0.011	22	0.24	0.55	0.2	5.9	77	1	10	3.5
C14	0	0.012	21	0.2	0.61	0.17	6.5	81	0	6	3.23
C15	0	0.013	23	0.24	0.57	0.2	6.4	93	1	5	3.81
C16	0	0.013	25	0.38	0.60	0.18	3.6	61	0	7	3.4
D03	1	0.01	18	0.35	0.37	0.22	3.3	3	1	0	2.19
D04	1	0.012	19	0.27	0.56	0.21	4.3	19	0	4	2.43
D05	1	0.012	17	0.31	0.54	0.21	4.4	44	1	3	2.42
D06	1	0.012	17	0.35	0.58	0.21	6.6	11	0	2	2.7
D07	1	0.014	18	0.42	0.59	0.2	8.9	14	1	4	3.18
D08	1	0.013	18	0.36	0.60	0.18	8.9	41	0	1	3.18
D09	1	0.013	19	0.38	0.57	0.21	7.6	44	1	4	3.28
D10	1	0.013	19	0.38	0.59	0.19	6.3	40	0	3	3.18
D11	1	0.013	19	0.40	0.54	0.2	7.2	34	1	8	3.25
D12	1	0.014	20	0.52	0.55	0.19	6.2	58	0	9	3.43
D13	1	0.013	19	0.38	0.55	0.19	6.4	69	1	3	3.2

Tabel 1. Udsnit af måledata.

I alt 42 sukkerprøver er analyseret for 10 forskellige kvalitetsmål: aske-indhold, farve, turbiditet, to forskellige typer af kornstørrelser, SO₂, invert sukker, flokkulat, uopløselig rest og amino-N. For detaljerede oplysninger, se [1]. Det ønskes undersøgt om man ud fra denne prøve-information kan bestemme, hvilken fabrik sukkeret er produceret på (C, D eller F). Desuden kaldes to af prøverne S1 og S2; dem vil vi fokusere på i det efterfølgende.

Data fra dette og andre eksempler kan downloades fra www.models.life.ku.dk/dansk kemi. Programmet LatentiX (www.latentix.com) er anvendt til at plote rådata samt til beregning af PCA modeller og outlierdetektion.

Afvigende prøver – Outliers

Outliers kan defineres som prøver, der har et afvigende variabel mønster i forhold til andre prøver i datasættet. Når der er målt mange variable (f.eks. spektrale data) kan det være vanskeligt at finde sådanne mønstre ved direkte inspektion af måledata.

PCA og mønstre

PCA modellen kan i dette tilfælde skrives

$$\mathbf{X}_{auto} = \mathbf{T}_a \mathbf{P}'_a + \mathbf{E}_a$$

Hvor \mathbf{X}_{auto} repræsenterer de autoskalerede data (hver måling er fratrukket variabelens middelværdi og divideret med dens spredning). \mathbf{X}_{auto} har samme dimensioner, som den oprindelige \mathbf{X} matrix, altså en tabel med 42 objekter og 10 variable. Indeks a angiver antal principale komponenter, der er beregnet i modellen.

\mathbf{T}_a og \mathbf{P}'_a indeholder henholdsvis scores og loadings for a -komponent modellen, og \mathbf{E}_a indeholder residualerne; dvs. den del af data, der ikke er beskrevet af PCA modellen.

I PCA analyseres mønstret eller sammenhængen mellem alle variable hvilket håndteres gennem de beregnede loadings (\mathbf{P}). Den fundamentale antagelse i PCA er at alle prøver kan beskrives med det samme sæt af loadings. Prøver, for hvilket

Outliers i Principal Component Analysis – fokus på individet

En af styrkerne ved PCA er muligheden for at undersøge den enkelte prøve samtidig med at grupperinger kan observeres. Her beskrives det, hvordan man i PCA kan finde afvigende prøver – outliers – ved hjælp af relevante grafiske afbildninger, der leder direkte tilbage til måledata

Af Lars Norgaard, Søren Balling Engelsen, Rasmus Bro og Thomas Skov, Københavns Universitet

dette ikke gælder, vil have et variabel mønster, der afviger fra alle andre prøver i datasættet.

I PCA er der to *afstandsmål*, der på forskellig måde angiver, hvor meget en prøve afviger fra resten af datasættet: residualt og Hotelling's T^2 .

Residualt

En prøves residual kan beregnes direkte fra E_a -matricen: for en given prøve beregnes kvadratsummen af alle elementerne i den tilsvarende række i E_a . Hver prøve vil i eksemplet have 10 elementer i residualt; et for hver variabel.

For prøve **S1** er residualt efter to komponenter:

$$\text{Res}(\mathbf{S1}) = (0,64; 2,15; -0,51; -0,63; -0,16; 0,18; -0,14; 0,54; -0,62; -1,51)$$

Dvs. residualt for aske indhold er 0,64, for farve er det 2,15 osv.

Residual variansen beregnes således:

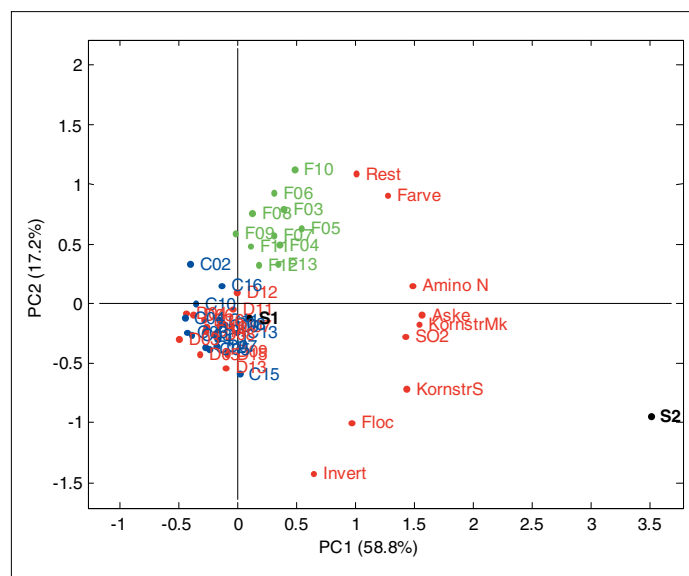
$$\text{Res-var}(\mathbf{S1}) = 0,64^2 + 2,15^2 + (-0,51)^2 + (-0,63)^2 + (-0,16)^2 + 0,18^2 + (-0,14)^2 + 0,54^2 + (-0,62)^2 + (-1,51)^2 = 8,7$$

Alle andre prøver har residual varians på 5,2 eller derunder. En prøve med højere residual varians (**S1**) vil have et mønster eller en variation i de oprindelige data, som ikke ligner de resterende prøver. I eksemplet er residualt beregnet for to komponenter; det giver generelt god mening at se på residualerne for de første 1-5 komponenter.

Hotelling's T^2

Det andet og næstvigtigste afstandsmål er baseret på score-værdierne (T). I bi-plottet figur 1 er prøve **S2** placeret langt fra de andre prøver. Afstanden til centrum for en prøve i scores-rummet kan beregnes ved hjælp af Hotelling's T^2 (der tager højde for kovariansen i data).

I to dimensioner kan vi direkte observere at en prøve er placeret langt fra andre prøver, men i f.eks. et 5-dimensionalt score-rum er det vanskeligt; i stedet kan Hotelling's T^2 beregnes for alle fem dimensioner. For prøve **S2** er Hotelling's T^2 36,6



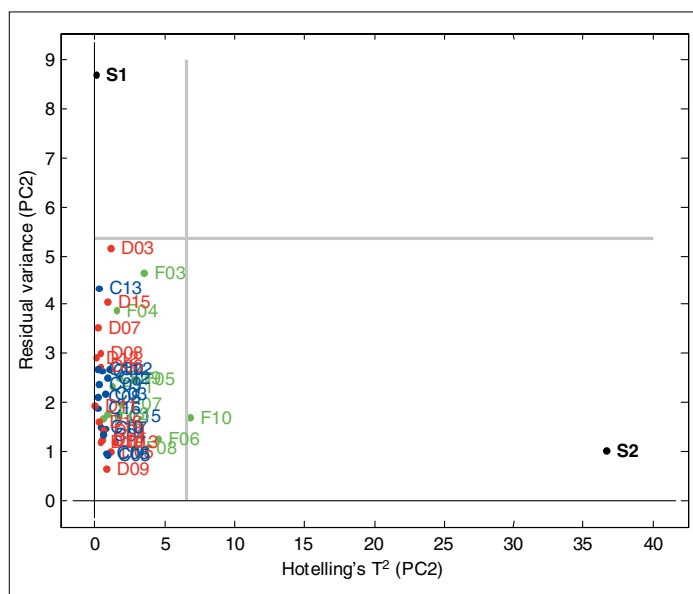
Figur 1. Bi-plot fra en PCA model af kvalitetsmålinger på sukker. Prøver fra de tre fabrikker er i forskellige farver (C: blå, D: rød, F: grøn). Loadings er angivet i rød.

efter to komponenter; alle øvrige prøver har værdier under 6,9 (figur 2).

Residual mod Hotelling's T^2

Det vigtigste diagnostiske plot i PCA er residual varians mod Hotelling's T^2 . Dette er illustreret i figur 2, hvor plottet ses for en to-komponent PCA på de 42 sukker prøver. To prøver, **S1** og **S2**, adskiller sig fra mængden. Den ene, **S1**, er karakteriseret ved højt residual, der betyder stærkt afvigende variabel mønster, og den anden er karakteriseret ved høj Hotelling's T^2 , der betyder stærkt afvigende indhold eller »koncentration« (stor afstand i PCA modellen indtil centrum). Vi har tidligere vist effektiviteten af scatterplottet til at få et overblik over prøverne i et scoreplot; her benyttes den samme grafiske teknik til at identificere outliers.

Ved at inspicere bi-plottet kan umiddelbart aflæses at prøve **S2** har ekstremt høje værdier for de fleste målinger. Derimod ser det ikke ud til at variabel mønsteret for denne prøve afviger



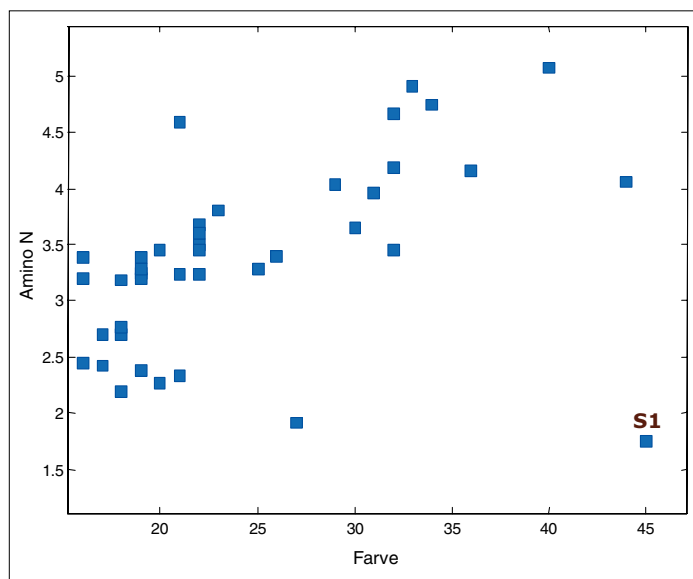
Figur 2. Residual mod Hotelling's T^2 . Approximative 95 %-konfidensgrænser er angivet som vertikal og horisontal linie. Bemærk prøverne S1 og S2.

fra de øvrige prøver, da residualen er på niveau med andre prøvers residualer.

I modsætning til prøve S2, har prøve S1 et meget højt residual; dvs. afvigende mønster, men ikke ekstreme måleværdier, da Hotelling's T^2 er sammenlignelig med de øvrige prøver.

Ved at sammenligne residualen for prøve S1 med de øvrige prøvers residualer for hver enkelt variabel, kan man få et overblik over, hvilke variable prøven har afvigende værdier for (ikke vist).

For S1 er det især kombinationen af målingerne af amino-N og farve. Igen kan scatterplottet benyttes til at visualisere sammenhængen mellem de to enkelt variable (figur 3). Af figuren fremgår det, at S1 bryder mønsteret sammenlignet med alle øvrige prøver; ingen anden prøve har både lav amino-N og høj farve måling.



Figur 3. Scatterplot af amino-N mod farve for alle prøver. Prøve S1 bryder sammenhængen mellem disse i forhold til de resterende prøver. Bemærk: S2 er ikke med på figuren (farve=57, amino-N=9,84);

Outliers har indflydelse på modellen

Outliers er uønskede fordi de indgår i estimationen af loadings. PCA modellen må således genberegnes, når en prøve er karakteriseret som outlier og kasseret. Det er altså en iterativ proces at karakterisere og eliminere outliers. Dette foregår let i moderne kemometrisk software: prøven markeres i et residual varians mod Hotelling's T^2 plot, og dernæst genberegnes modellen uden den valgte prøve.

Der findes ikke en generel metode til outlier genkendelse og fjernelse. Dette skyldes blandt andet at Hotelling's T^2 outliere i visse tilfælde kan være ønskelige som ekstreme prøver, der spænder modellen ud samt det faktum, at man definerer en outlier i forhold til de andre prøver. Kort sagt: »alt er relativt«, og det gælder især i kemometrisk empirisk modellering.

Outro

Outliers er uhyre vigtige at håndtere; det er forkert at inkludere sådanne i sin model! På den anden side er der heller IKKE frit slag til at ekskludere prøver, der ikke lige »passer ind«. Ved at bruge de her beskrevne diagnostiske værktøjer kan man beskrive **hvordan** en prøve afviger. Derudover er det en absolut nødvendighed at forbinde en outlier tilbage til rådata, laboratorie-journaler og anden ekstern information for at forstå **hvorfor** prøven er afvigende.

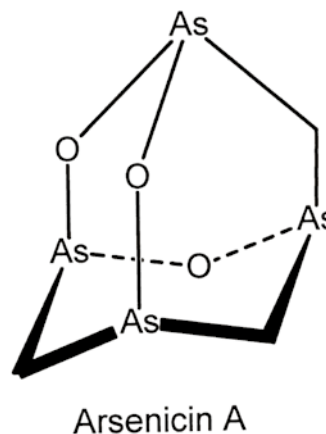
Lan@life.ku.dk

Kilder

[1] Nørgaard L. *Classification and Prediction of Quality and Process Parameters of Thick Juice and Beet Sugar by Fluorescence Spectroscopy and Chemometrics*. Zuckerindustrie 120 (11):970-981, 1995.

Nyt om...

... marin arsenforbindelse med 77% As



Fra en havsvamp *Echinochallina bargibani* fundet ud for Ny Caledoniens kyst er isoleret en arsenholdig forbindelse Arsenicin A, der har vist sig at have udtalt såvel fungicid som bactericid virkning. Forbindelsen har en adamantanlignende struktur.

Carl Th.

On the First Polyarsenic Organic Compound from Nature *Chemistry a European Journal*, 2006, side 8989.