

## Tiderne skifter...

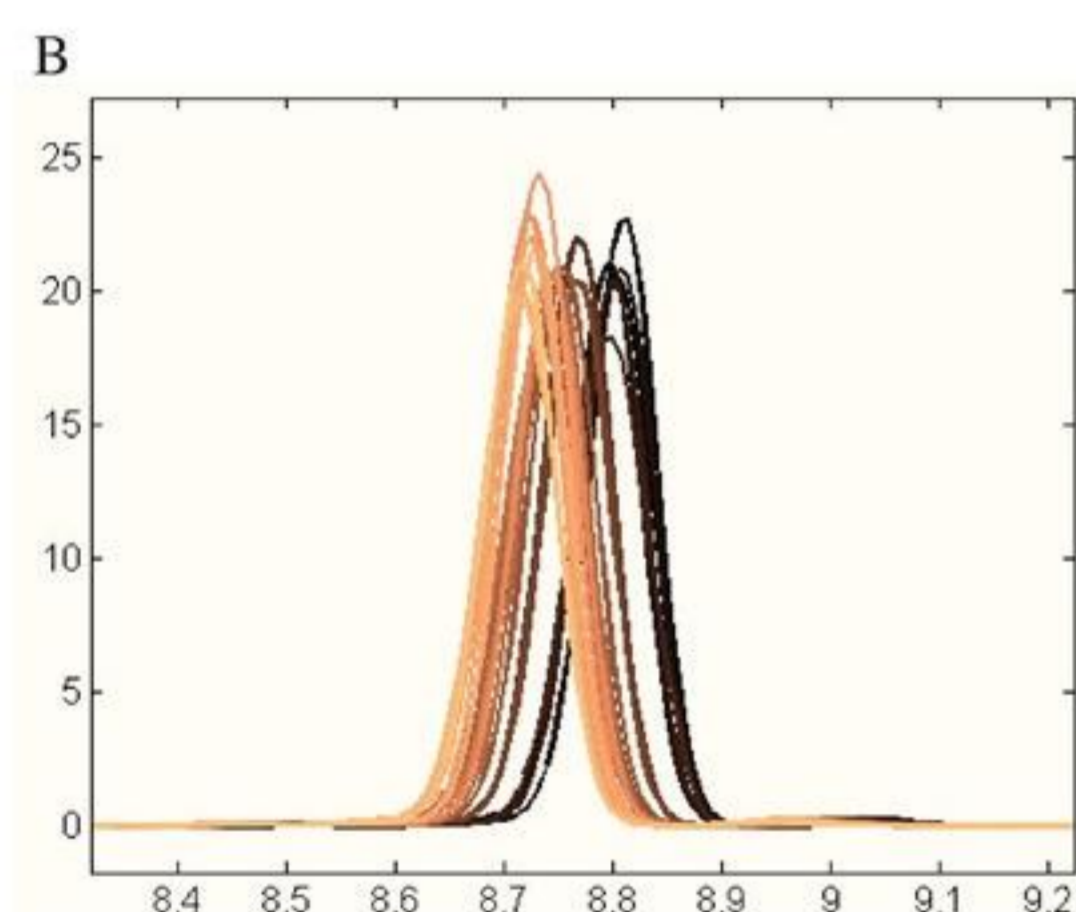
Til trods for moderne hightech-komponenter og instrumentdele har den omni-anvendte analyseteknik kromatografi stadig problemer med at reproducere retentionstiden for et bestemt kemisk stof. I denne klumme viser vi, hvordan dette uheldigt påvirker analysen af data, men også hvordan man relativt simpelt kan løse problemet

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

Ved måling og karakterisering af de enkelte kemiske komponenter i komplekse matricer som f.eks. fødevarer og kliniske prøver anvendes ofte kromatografiske analysemetoder. Lige siden opdagelsen af kromatografi og frem til nu har teknikken været afhængig af, at kunne reproducere den tid det tager for de enkelte kemiske komponenter at eluere fra kolonnen (såkaldt retentionstid).

A

Prøve	pH	Tryk	Vægt	Temperatur
1	7,1	1,0	110	34
2	7,4	1,1	121	45
3	7,6	0,9	113	54
4	8,0	1,1	123	37
5	6,9	1,43	0,9	38
6	6,5	1,0	154	43
7	6,6	1,0	154	41
8	6,6	1,0	32	143
9	8,9	0,9	167	39
10	7,0	0,9	120	40



Figur 1. Illustration af skift i variable. (A) For simple data hvor to sæt af værdier er ombyttet. (B) For kromatografiske data hvor retentionstiden for den viste kemiske komponent (retentionstid ca. 8,75 min) er skiftet mellem prøverne (markeret i forskellige farver efter kørsel over tid).

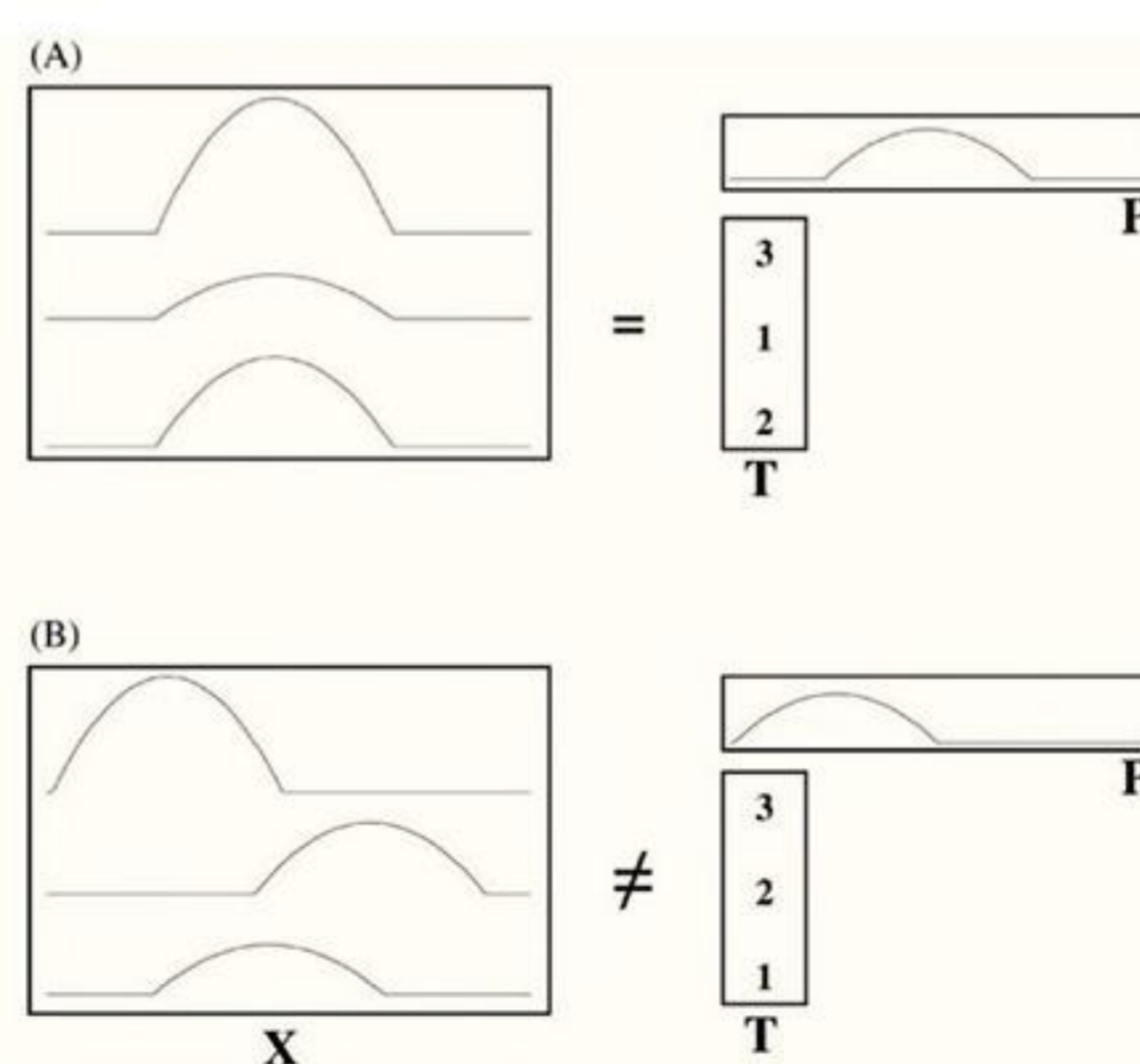
Dette er vigtigt for at kunne identificere et bestemt stof til en bestemt retentionstid, ved sammenligninger af koncentrationer af stoffer i flere prøver samt når bi-lineære kemometriske modeller som PCA (Dansk Kemi nr. 4, 2008) og PLS (Dansk Kemi nr. 11, 2008) anvendes. Her fokuseres på det sidste aspekt, og hvorfor det er vigtigt at kunne rette op på tingene, når *tiderne skifter!*

Som diskuteret i tidligere klummer anvender f.eks. PCA den information, der ligger i de enkelte variable til at finde mønstre og sammenhænge mellem variablene på tværs af prøverne. En forudsætning for at dette er meningsfuldt er, at den samme information findes i de samme søjler (altså at variabel 10 er variabel 10 for alle prøver). Er dette ikke tilfældet falder

modellen til jorden, ligesom en tabel heller ikke vil give mening såfremt én variabel er ombyttet med en anden (figur 1). Nøjagtigt det samme gør sig gældende med kromatografi, hvor tiderne skifter for en enkelt kemisk komponent (figur 1).

### Problem

I de data, der er vist til højre i figur 1, bør hverken klassisk statistik (f.eks. standardafvigelse) eller multivariate metoder som PCA anvendes, før der er rettet op på de skiftede variable (tider). Er tiderne skiftet, kan PCA ikke beskrive data på en hensigtsmæssig måde, da der ikke kan opstilles en linearkombination af de skiftede variable, så systemet beskrives i en én-komponent model. Dette er ellers ønskeligt, da vi har samme mønster i alle prøver (samme kurveform), og den eneste forskel



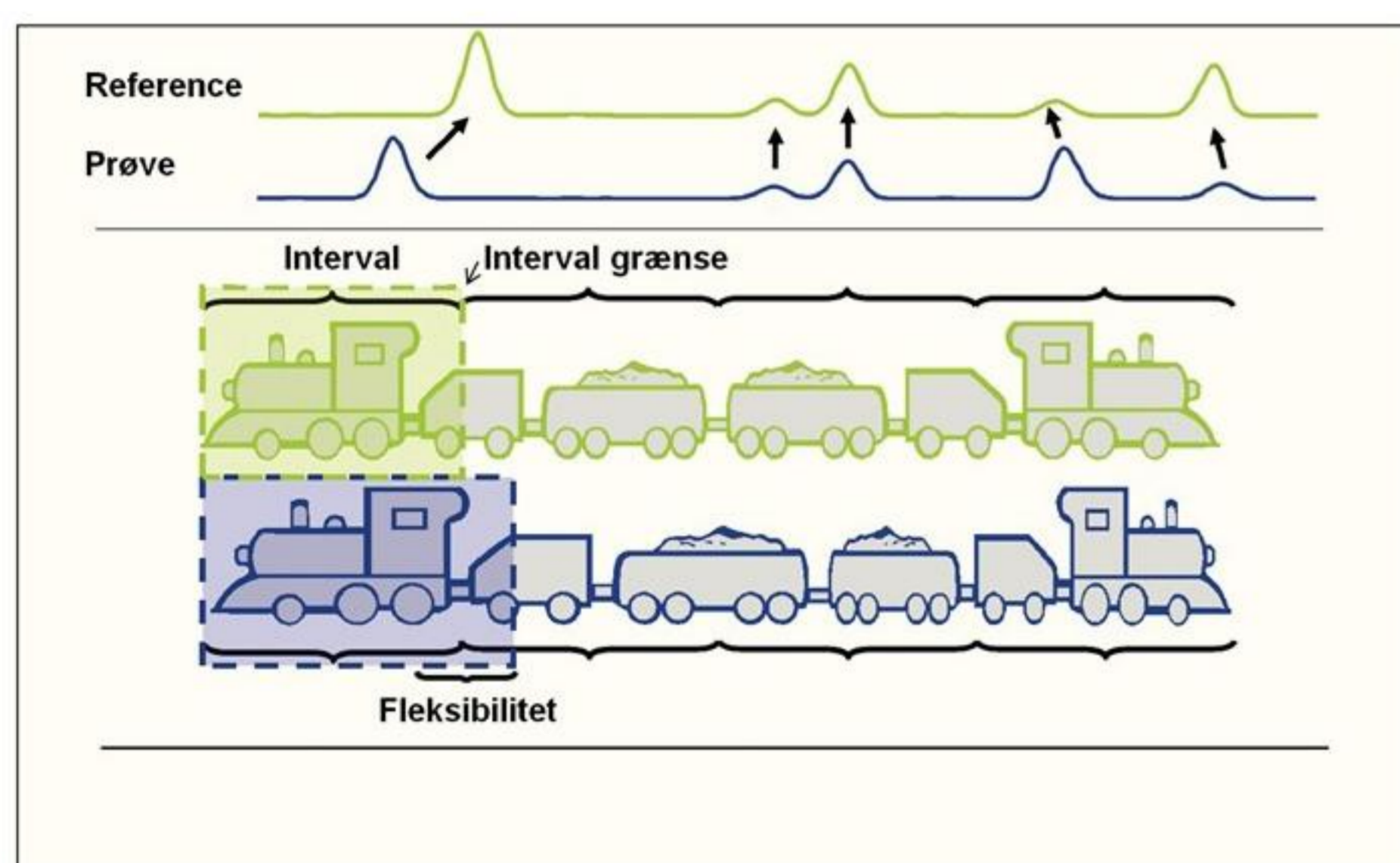
Figur 2. Illustration af brugen af PCA på data med skift i en enkelt kromatografisk top i tre prøver. (A) For data, hvor *tiderne ikke skifter*, kan en simpel én-komponent model opstilles, hvor fællesnævneren (lig loadingen, P eller en linearkombination af variablene) er topformen, og hvor indholdet af denne top er angivet i scoreværdien, T. (B) For data hvor *tiderne skifter*, er det ikke muligt at opstille en simpel én-komponent model, men flere komponenter må inddrages for at kunne karakterisere de skiftede tider.

burde derfor ligge i koncentrationsniveauet (lig toparealet) i de enkelte prøver. Dette kan simpelt visualiseres for en PCA på en enkelt top (figur 2), men princippet er det samme, når der findes langt flere toppe i kromatogrammerne.

### Løsning

For at kunne anvende PCA på data, hvor tiderne skifter, må man ty til en forbehandling af kromatogrammerne. Denne forbehand-

ling er en korrektion af de skiftede tider og benævnes alignment (eller warping). Princippet er, at man først udvælger et kromatogram som referencekromatogram og dernæst korrigerer alle de andre kromatogrammer mod dette referencekromatogram, således at de kromatografiske toppe (lig den kemiske information) kommer til at ligge som i referencekromatogrammet. For at kunne sammenligne og rykke rundt i kromatogrammerne opdeles disse i intervaller. Intervallerne flyttes rundt, indtil de passer sammen med det tilsvarende stykke i referencekromatogrammet (se mere detaljeret beskrivelse af princippet i figur 3). Det måles simpelt ved at beregne korrelationen mellem kurveformen for intervallet i prøven og den tilsvarende kurveform for intervallet i referencen. Når intervallet er flyttet passende, vil korrelationen være højest. Dette er illustreret i figur 3.

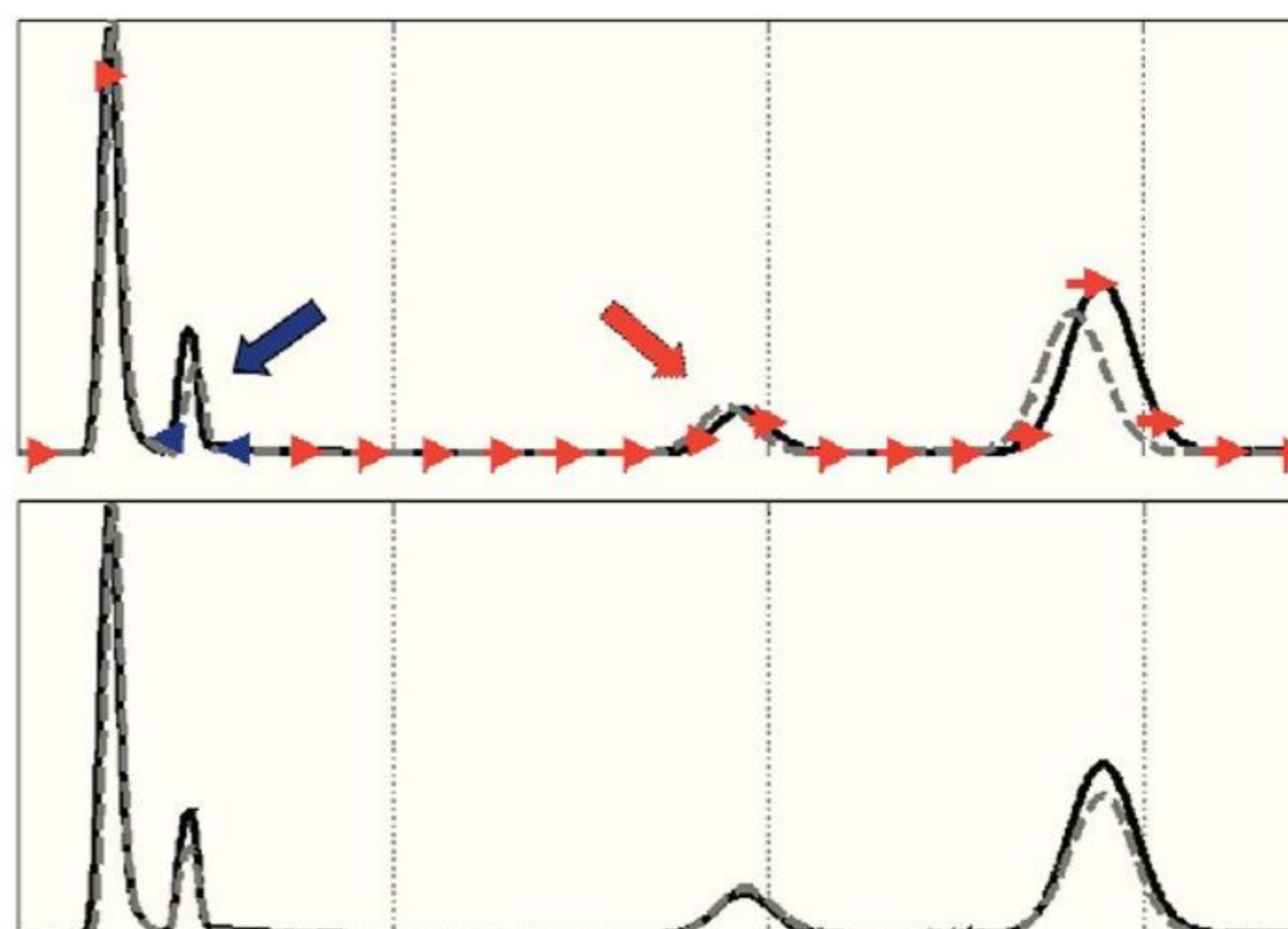


Figur 3. Illustration af alignment. Det grønne signal er referenceprøven mens det blå er prøvesignalet, der ønskes korrigeret mod det grønne. For det første interval i det blå togsæt rykkes grænsen til højre, hvorved indholdet bliver det samme som i referenceprøven; et lokomotiv og lidt togvogn (dog ses et lidt større lokomotiv, men dette er helt analogt til en større top i kromatogrammet og altså helt legalt). Gøres dette for alle intervaller i signalet, vil toppene i prøvechromatogrammet blive korrigeret bedst muligt (under de givne valg af interval- og fleksibilitetsstørrelse) og ideelt som vist øverst i figuren. Teknisk detalje: faktisk flyttes intervallerne ikke i sig selv, men korrektionen sker gennem ryk af intervalgrænser iht. den valgte fleksibilitet efterfulgt af et interpolationstrin og til sidst en kombinatorisk sammenligning af alle mulige ryk for at give det bedst korrigerede signal [1-3].

Den beskrevne og viste alignment-metode kaldes for Correlation Optimized Warping (COW) [1,2] og er den mest anvendte til at kromatografiske data, når tiderne skifter. I figur 4 er vist, hvordan COW virker globalt over hele kromatogrammet, og hvordan samspillet mellem intervalgrænser bevirker, at enkelte toppe kan korrigeres mod venstre, mens andre korrigeres mod højre. Det sidste er især en af årsagerne til populariteten af COW, da selv meget usystematiske og store skift kan korrigeres.

## Faldgruber

Når alignment-metoder anvendes, skal man være opmærksom på flere forhold for at sikre sig mod uønskede korrektioner. Først og fremmest skal man vælge et godt referencekromatogram, da alle andre kromatogrammer korrigeres mod dette. Dernæst skal intervalstørrelse og fleksibilitet vælges. Man bør undgå for små intervaller og for høj fleksibilitet, da det kan give ændrede kurveformer (da en kurve splittes op i flere små inter-



Figur 4. Illustration af intervalgrænseryk (vist øverst) over hele prøvechromatogrammet (det grå stiplede signal) mod referencekromatogrammet (sort fuldt optrukket signal) og effekten heraf (vist nederst). Ved at anvende COW vil enkelte toppe blive korrigeret mod venstre (markeret med rødt) og andre mod højre (markeret med blå) helt i tråd med, hvordan tiderne i virkeligheden kan skifte i kromatografi.

valler) og derved arealer (lig koncentrationer). For store intervaller og for lav fleksibilitet er heller ikke at foretrække, da der så ikke kan korrigeres ret meget. En pragmatisk mellemting vil ofte give en god og PCA-brugbar løsning. En tommelfingerregel er altid at plote de korrigerede kromatogrammer, og tjekke om resultatet er som ønsket (alternativt kan matematiske termer anvendes, men disse omtales ikke i denne klumme – for mere information se [3]).

## Outro

Vi har i denne klumme vist, hvordan PCA (og andre kemometriske metoder) ikke har det godt med data, hvor tiderne skifter. Heldigvis findes der flere forbehandlingsmetoder, som kan korrigerer for dette, og disse kan med fordel anvendes for kromatografiske data, hvor skift i retentionstiden er normalen.

## E-mail-adresser

Thomas Skov: thsk@life.ku.dk

Søren Balling Engelsen: se@life.ku.dk

Rasmus Bro: rb@life.ku.dk

Lars Nørgaard: lan@life.ku.dk

## Referencer

1. Nielsen, N.P.V., Carstensen, J.M., and Smedsgaard, J. 1998. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A*, 805 (1-2): 17-35.
2. Tomasi, G., van den Berg, F., and Andersson, C. 2004. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18 (5): 231-241.
3. Skov, T., van den Berg, F., Tomasi, G., and Bro, R. 2006. Automated alignment of chromatographic data. *Journal of Chemometrics*, 20 (11-12): 484-497.