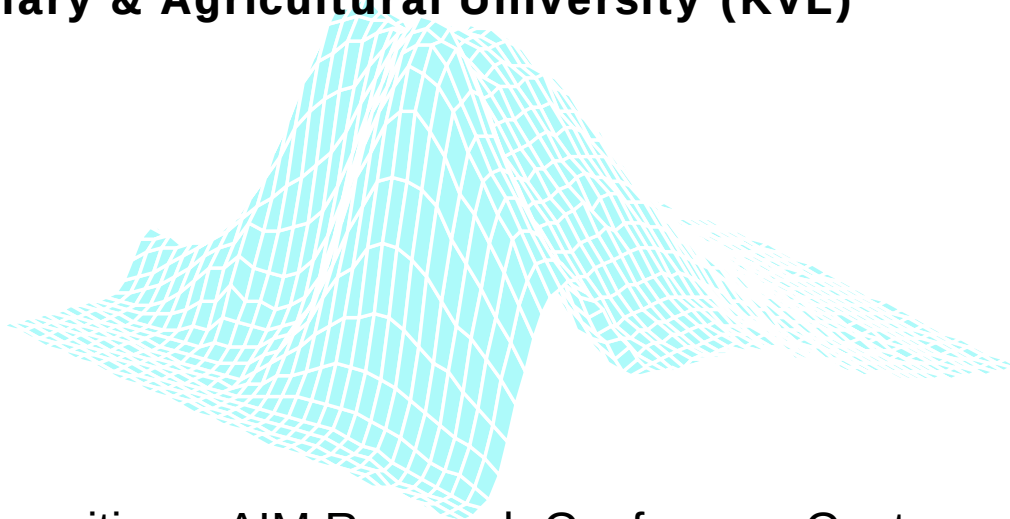
Practical Problems and Solutions in Applied Multiway Analysis

- **Rasmus Bro**

  Chemometrics & Spectroscopy Group

  Royal Veterinary & Agricultural University (KVL)

  rb@kvl.dk



ARCC Workshop on Tensor Decompositions, AIM Research Conference Center, Palo Alto, CA, July 19-23, 2004

Talk available at http://www.models.kvl.dk/users/rasmus/paloalto.zip

A motivating example:
Monitor pollution empirically from water samples

ELSEVIER Marine Chemistry 82 (2003) 239–254
www.elsevier.com/locate/marchem

Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy

Colin A. Stedmon[a,*], Stiig Markager[a], Rasmus Bro[b]

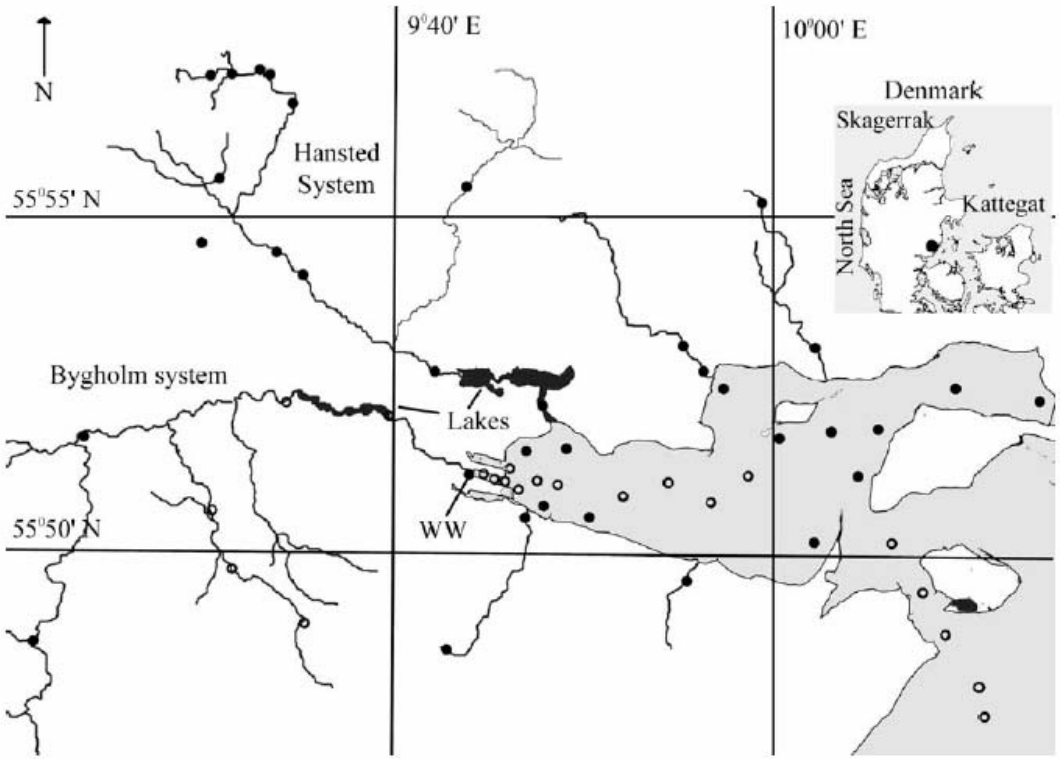
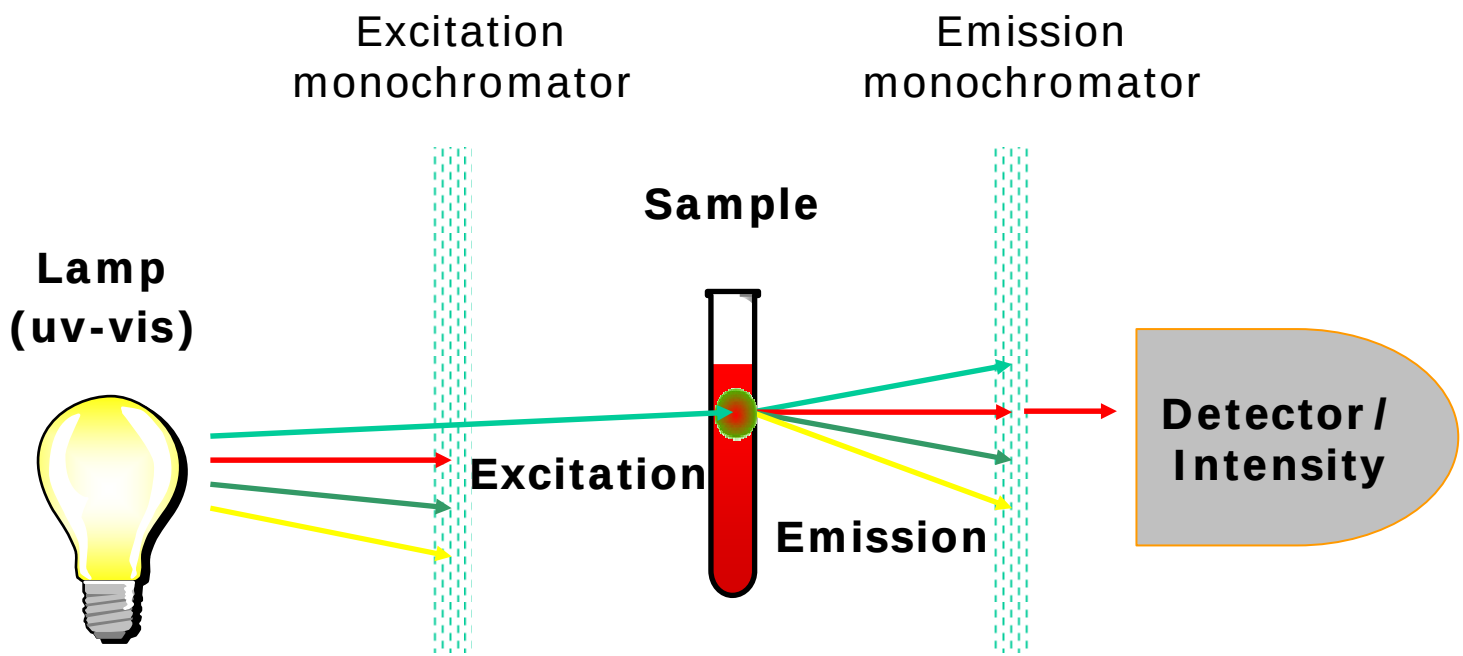
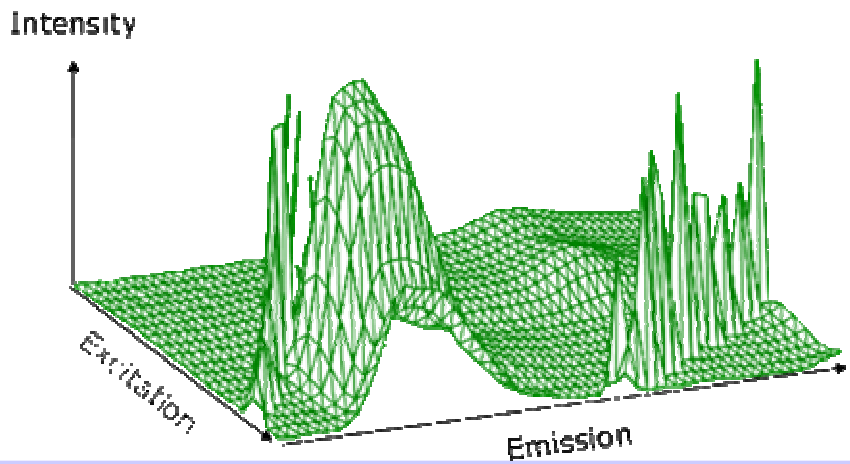


Fig. 1. Map of the Horsens catchment and its location on the east coast of Jutland, Denmark. Filled circles represent sampling stations and hollow circles represent the locations of the samples used in Fig. 7.

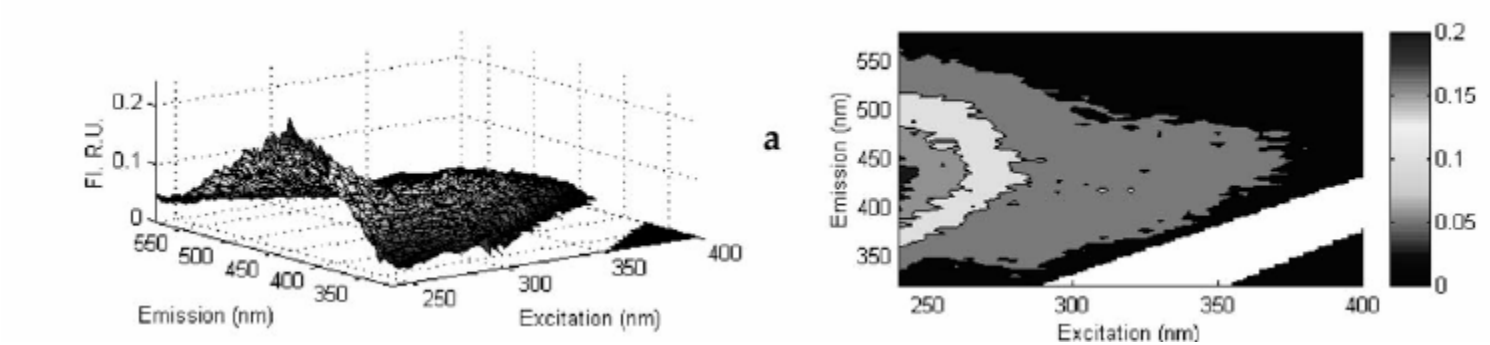
Fluorescence spectroscopy



Excitation-emission matrix – a chemical fingerprint



A practical example: Data from each sample

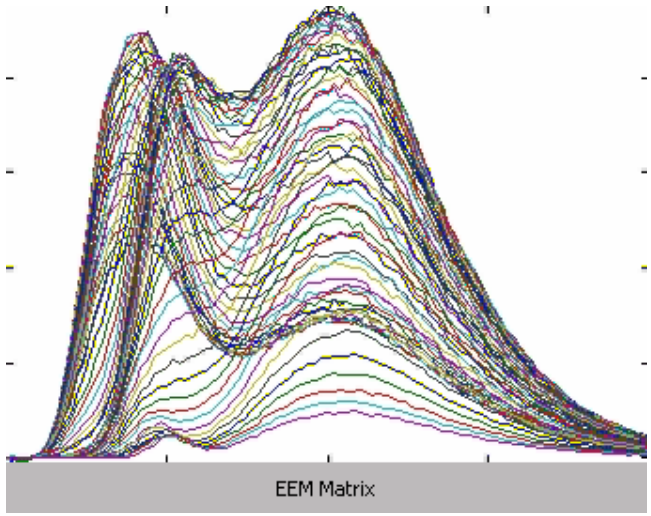


State of the art: Peak-spotting!!!

Practical example

What chemometrics and datamining had to offer

- Each sample yields 100 × 300 matrix
- Model with PCA (truncated SVD)

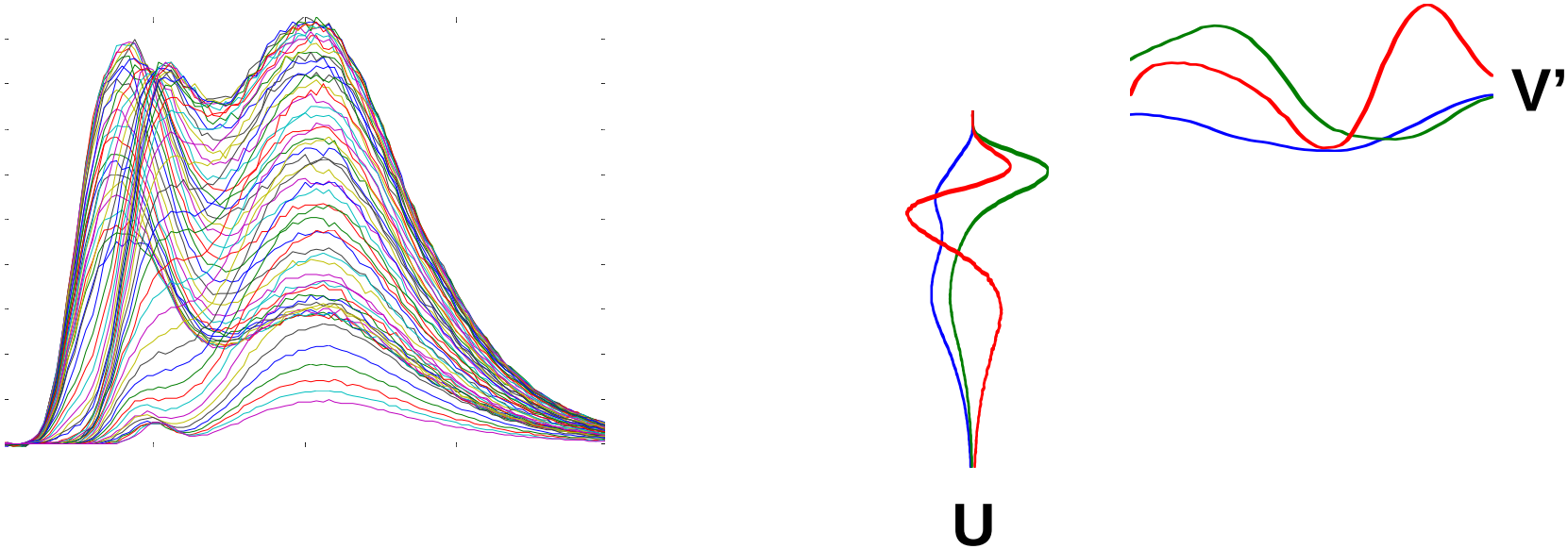


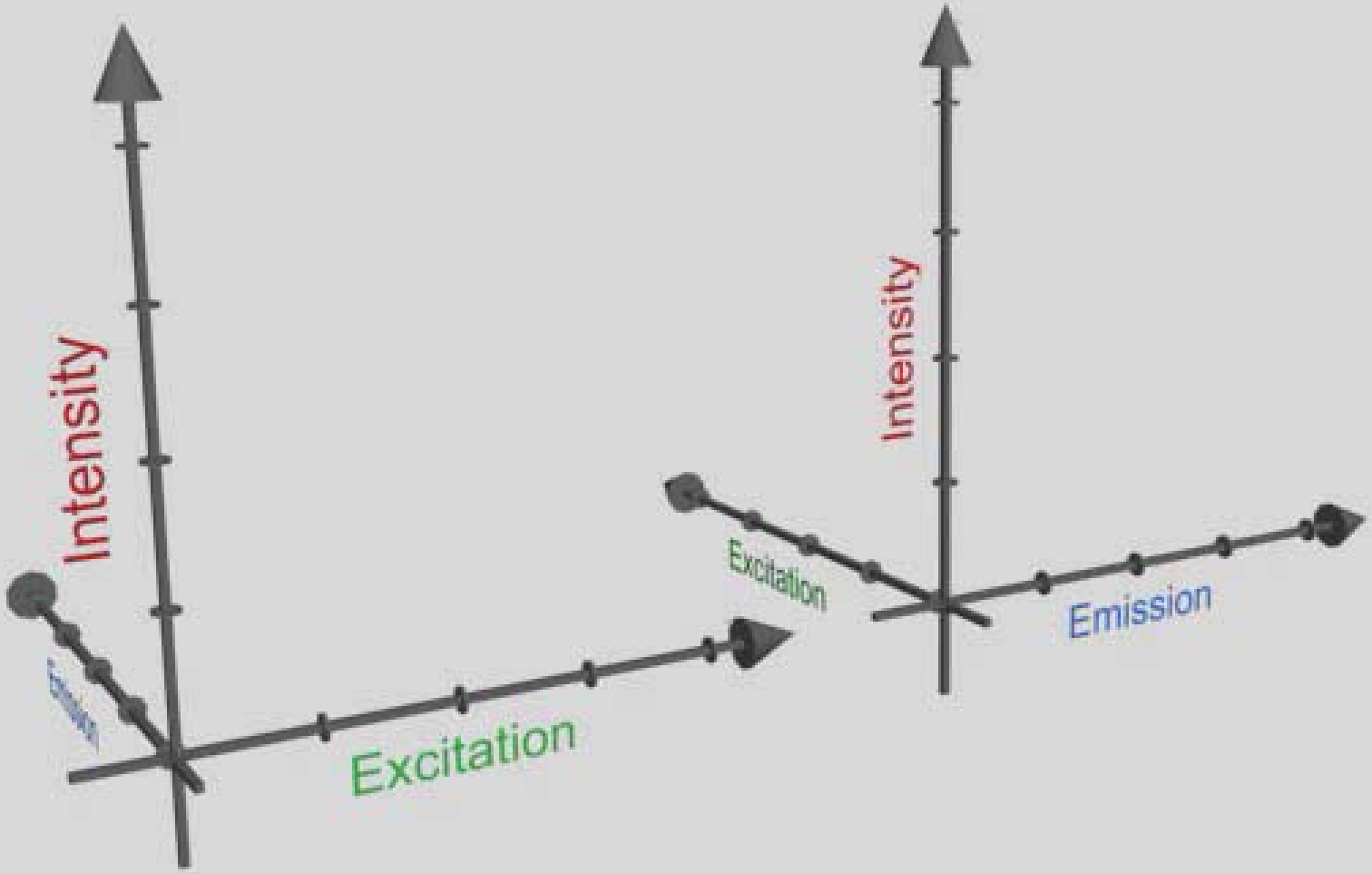
Practical example

What chemometrics and datamining had to offer

- Each sample yields 100 × 300 matrix
- Model with PCA (truncated SVD)

X = **USV'**




Intensity
Excitation
Intensity
Excitation
Emission

PARallel FACtor analysis

PARAFAC invented in 1970 by Harshman and independently by Carroll & Chang under the name CANDECOMP. Based on a principle of parallel proportional profiles suggested in 1944 by Cattell

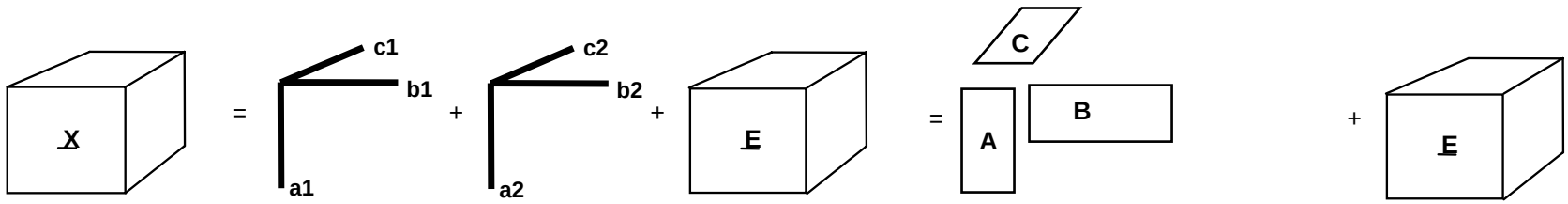
•R. A. Harshman. *UCLA working papers in phonetics* 16:1-84, 1970.
•J. D. Carroll and J. Chang. *Psychometrika* 35:283-319, 1970.
•R. B. Cattell. *Psychometrika* 9:267-283, 1944.

- PCA - bilinear model,

$$x_{ij} = \sum_{f=1}^{F} a_{if} b_{jf} + e_{ij} \qquad \mathbf{X} = \mathbf{A}\mathbf{B}^{T} + \mathbf{E}$$

- PARAFAC - trilinear model,

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \qquad \mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}^{T} + \mathbf{E} \qquad \mathbf{D}_k = \mathrm{diag}(\mathbf{c}_{(k,:)})$$



PARallel FACtor analysis

Low-rank:

- Khatri-Rao (Rao 73) rather than Kronecker

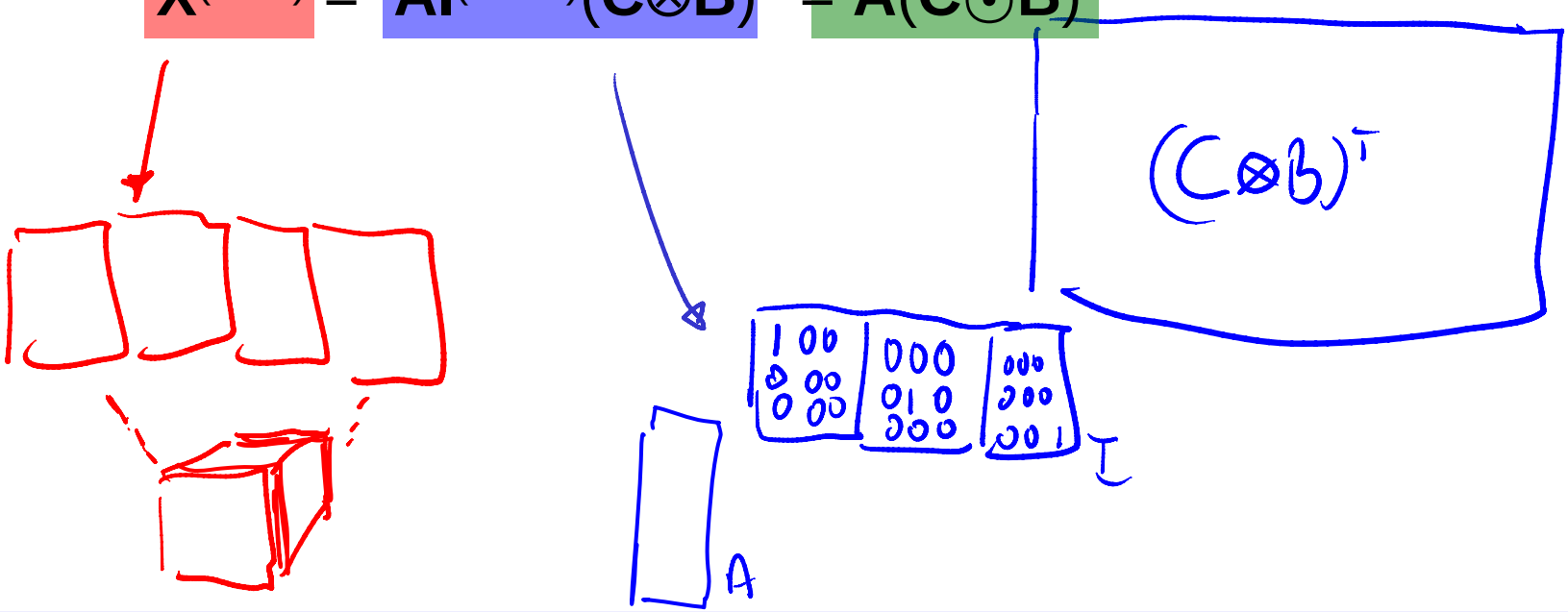
$$\mathbf{X}^{(I\times JK)} = \mathbf{A}\mathbf{I}^{(F\times FF)}(\mathbf{C}\otimes\mathbf{B})^T = \mathbf{A}(\mathbf{C}\odot\mathbf{B})^T$$

PARallel FACtor analysis

Low-rank:

- Khatri-Rao rather than Kronecker

$$\mathbf{X}^{(I\times JK)} = \mathbf{A}\mathbf{I}^{(F\times FF)}(\mathbf{C}\otimes\mathbf{B})^T = \mathbf{A}(\mathbf{C}\odot\mathbf{B})^T$$

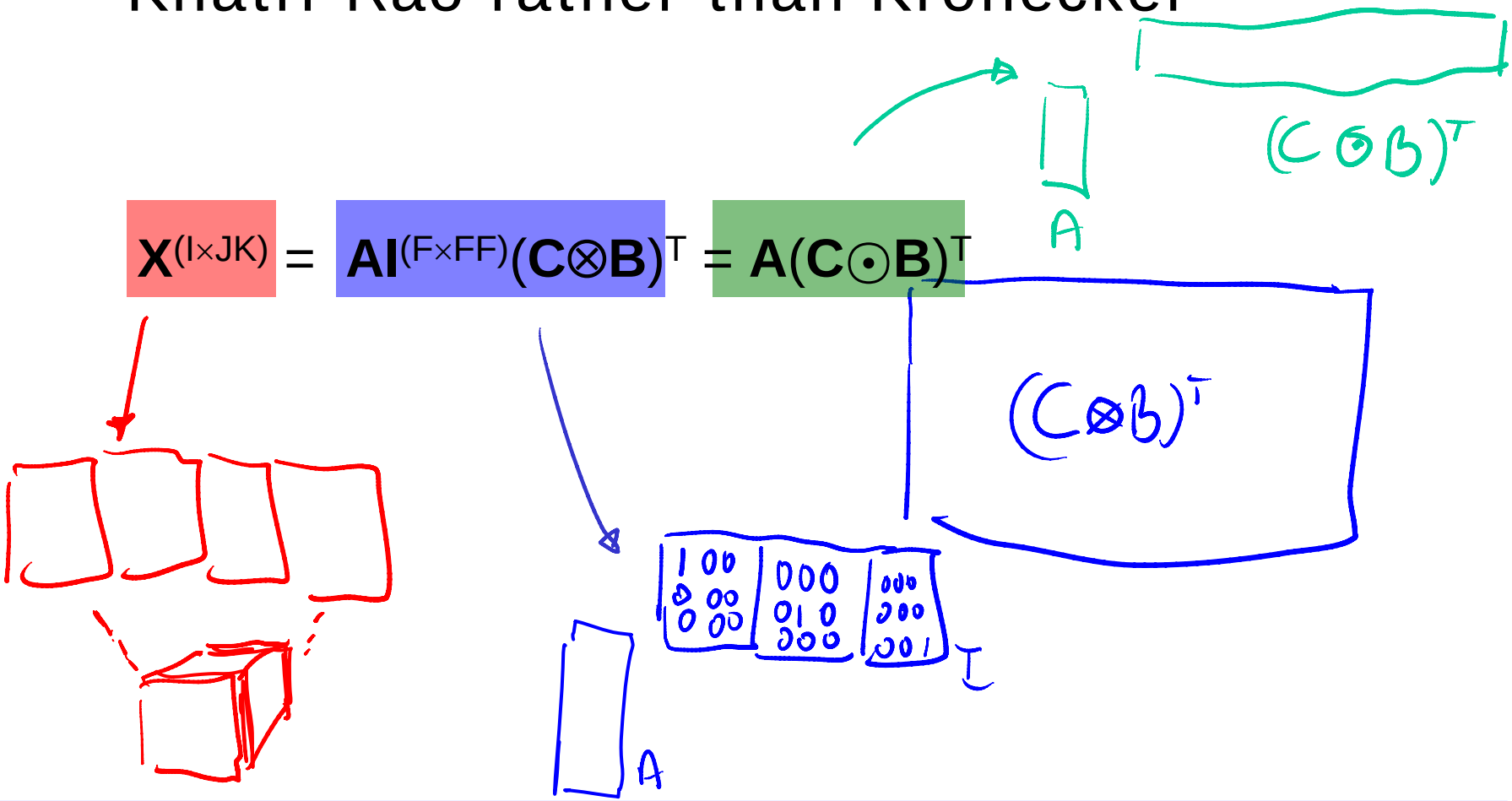


PARallel FACtor analysis

Low-rank:

- Khatri-Rao rather than Kronecker

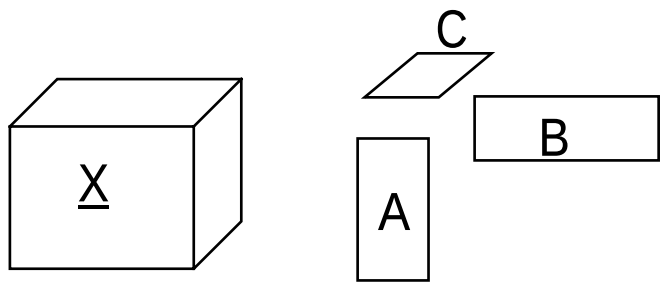
$$\mathbf{X}^{(I\times JK)} = \mathbf{A}\mathbf{I}^{(F\times FF)}(\mathbf{C}\otimes\mathbf{B})^T = \mathbf{A}(\mathbf{C}\odot\mathbf{B})^T$$



PARAFAC - uniqueness

- **Uniqueness - conditions**

  A PARAFAC model is unique when



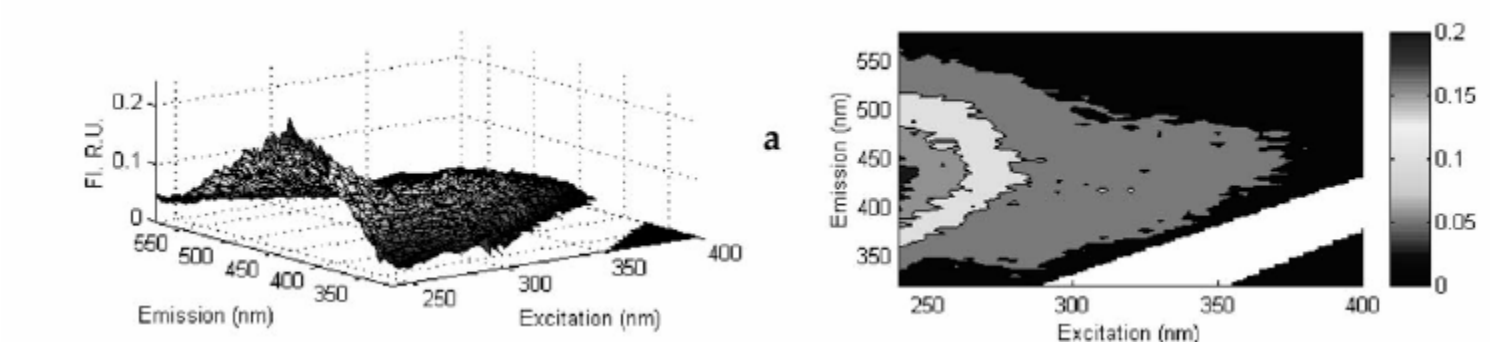
$$k_A + k_B + k_C \geq 2F + 2$$

F is the number of components and k_A is the k-rank of loading **A** : maximal number of randomly chosen columns which will have full rank

J. B. Kruskal, 1977.
N. D. Sidiropoulos and R. Bro, 2000.
ten Berge & Sidiropoulos, 2002

- Uniqueness – repercussions
  - An 8-component PARAFAC model of a 6×6×6 array unique
  - i.e. six samples – eight components!

A practical example: Data from each sample

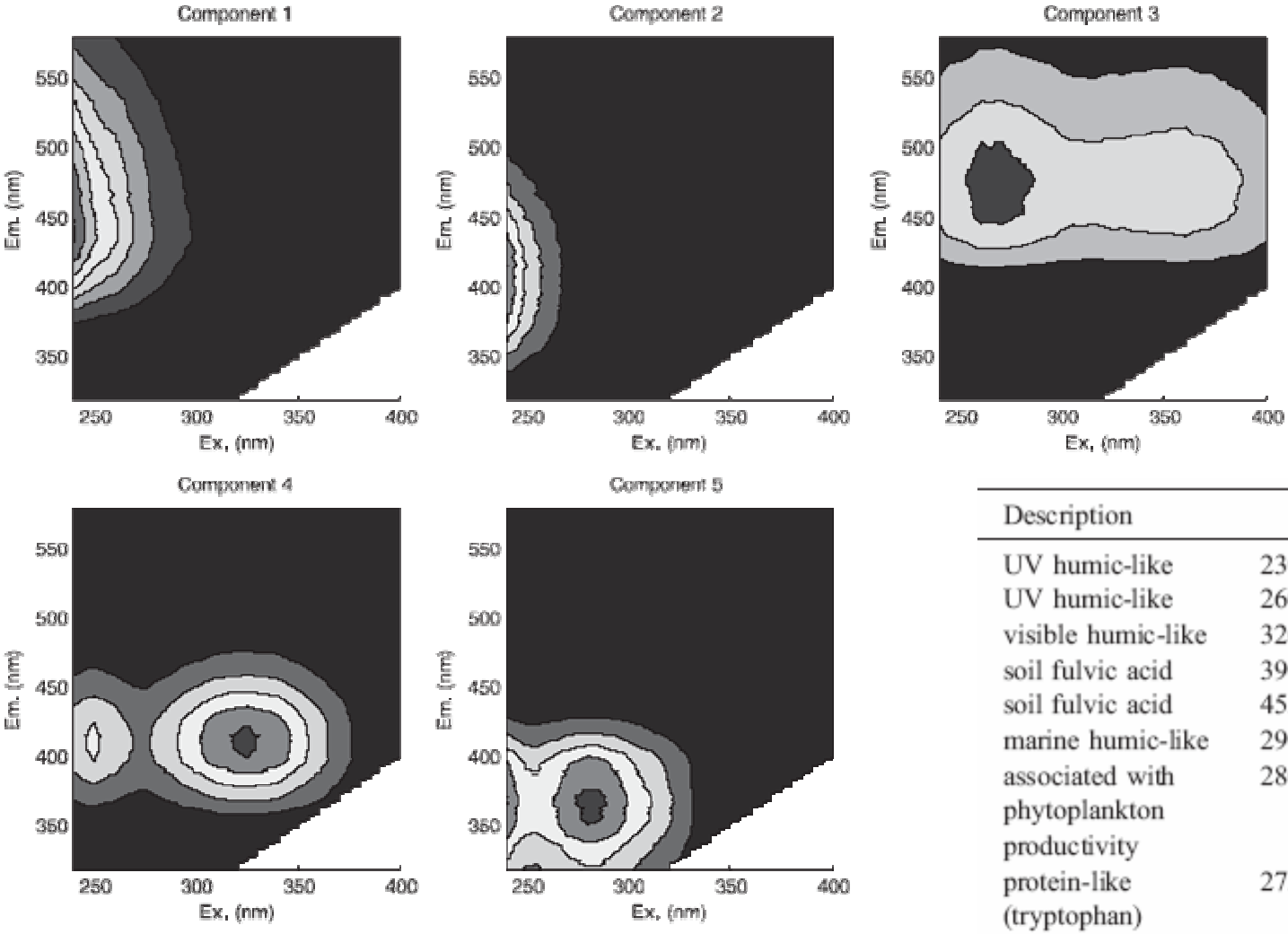


State of the art: Peak-spotting!!!

Practical example

PARAFAC on environmental samples

C.A. Stedmon et al. / Marine Chemistry 82 (2003) 239–254 247



Description		
UV humic-like	230	430
UV humic-like	260	380–460
visible humic-like	320–360	420–460
soil fulvic acid	390	509
soil fulvic acid	455	521
marine humic-like	290–310	370–410
associated with phytoplankton productivity	280	370
protein-like (tryptophan)	275	340

PARAFAC - algorithm

Efficient ALS algorithm

(Harshman 70)

1. Initialize **B** and **C**

2. $\mathbf{A} = \left(\sum_{k=1}^{K} \mathbf{X}_k \mathbf{B} \mathbf{D}_k \right) \left\{ \left(\mathbf{B}'\mathbf{B} \right) * \left(\mathbf{C}'\mathbf{C} \right) \right\}^{-1}$

2. $\mathbf{B} = \left(\sum_{k=1}^{K} \mathbf{X}'_k \mathbf{A} \mathbf{D}_k \right) \left\{ \left(\mathbf{A}'\mathbf{A} \right) * \left(\mathbf{C}'\mathbf{C} \right) \right\}^{-1}$

3. $diag\mathbf{D}_k = \left\{ \left(\mathbf{B}'\mathbf{B} \right) * \left(\mathbf{A}'\mathbf{A} \right) \right\}^{-1} \mathrm{diag}\left(\mathbf{A}'\mathbf{X}_k\mathbf{B} \right), k = 1,..,K$

4. Step 2 until relative change in fit is small

– * Hadamard (elementwise product)

Why ALS?

Simple

Extends to N-way

Handles missing

Handles ML fitting

Constraints:

- Nonnegativity
- Unimodality
- Orthogonality
- Linear constraints
- Fixed parameters
- Smoothness
- Functional
- etc

Important notes

- PARAFAC is a *low rank* (least squares) model
  - Identical to candecomp
  - Alternative: Sequential rank-one deflation-based low-rank power method (Bro, 1996),
    - no least squares properties
    - no physically meaningful uniqueness properties
    - not low rank!!!
  - Tri-/multi-linear PARAFAC model has no orthogonality properties
    - Imposing orthogonality would lead to loss of fit (unlike in two-way decompositions)

Important notes

- Alternative to low-rank approximation
  - Tucker3 model (HO-SVD) is a subspace based method
  - Retains many properties of SVD (Tucker 67, Kroonenberg 83, de Lathauwer 97)
  - No (physical) uniqueness
  - No ”rank properties”
  - Useful for
    - Datamining (Henrion et al. 97)
    - Compression (Bro & Andersson 98)
    - Exploratory analysis (Kroonenberg 83)
  - Suggestion: Forget the rank-aspect of SVD and focus on subspace properties -> Tucker HO-SVD!
    - Tucker: Subspaces
    - PARAFAC: Low-rank

Some PARAFAC problems

Algorithmic
Statistical

PARAFAC algorithm

The algorithm is slow, sometimes non-convergent and intrinisically just difficult!

Remedies:

- **Initialization**
  - Random or SVD,
  - GRAM (eigenproblem)
- **Speed-up**
  - "Optimal" compression — Not for difficult data (Alsberg 93, Bro & Andersson 98)
  - "Ad hoc" compression — More general (Bro 97)
  - Line-search — Works well (Harshman 70)
- **Alternative algorithms**
  - GRAM/DTLD — Not LS, not constraints, not missing
  - (damped) Gauss-Newton — Not big problems, good >70% missing (Paatero 97, Tomasi & Bro, 04)
  - Conjugate gradients etc. — Nice, not big problems, not nway, not constraints
  - Separable least squares — Feasible but constraints not possible (Bro & Sidiropoulos)
  - Grid computing — General matlab toolbox under development!

PARAFAC algorithm – statistical problem

Three-Way vs Two-Way Arrays - Differences

- Two-way ($I \times J$): row-rank = column-rank = rank $\leq \min(I, J)$;
- Three-way: row-rank $\neq$ column-rank $\neq$ "tube"-rank $\neq$ rank
- Two-way: rank(randn(I,J))=min(I,J) w.p. 1;
- Three-way: rank(randn(2,2,2)) is a RV (2 w.p. 0.3, 3 w.p. 0.7)
- 2-way: rank insensitive to whether or not underlying field is open or closed ($\mathbb{R}$ versus $\mathbb{C}$); 3-way: rank sensitive to $\mathbb{R}$ versus $\mathbb{C}$
- 3-way: Except for loose bounds and special cases [Kruskal; J.M.F. ten Berge], general results for maximal rank and typical rank sorely missing for decomposition over $\mathbb{R}$; theory more developed for decomposition over $\mathbb{C}$ [Burgisser, Clausen, Shokrollahi, *Algebraic complexity theory*, Springer, Berlin, 1997]

PARAFAC algorithm – statistical problem

”Any number without associated uncertainty is meaningless.”

To calculate confidence intervals etc. degrees-of-freedom are needed

- Standard deviation uses denominator N-1 (average eats one DoF)
- Multiple linear regression with J variables uses J DoFs
- A 2×2 random matrix has 4 DoF.
  - It is rank 2 with probability 1
  - Hence a 2 comp PCA/SVD uses at least 4 DoF
- A 2×2×2 random array has 8 DoF?
  - It is rank 2 or rank 3 with positive probability
  - Hence sometimes a 2 comp PARAFAC uses 8 DoF but sometimes not!

=> Statistics don’t work here. Even though the distribution of the residuals are known they can not be used for calculating coverage, confidence etc.

=> Rank is not a function of dimension of array!!!

PARAFAC – Some applications

Slicing – exponential fitting

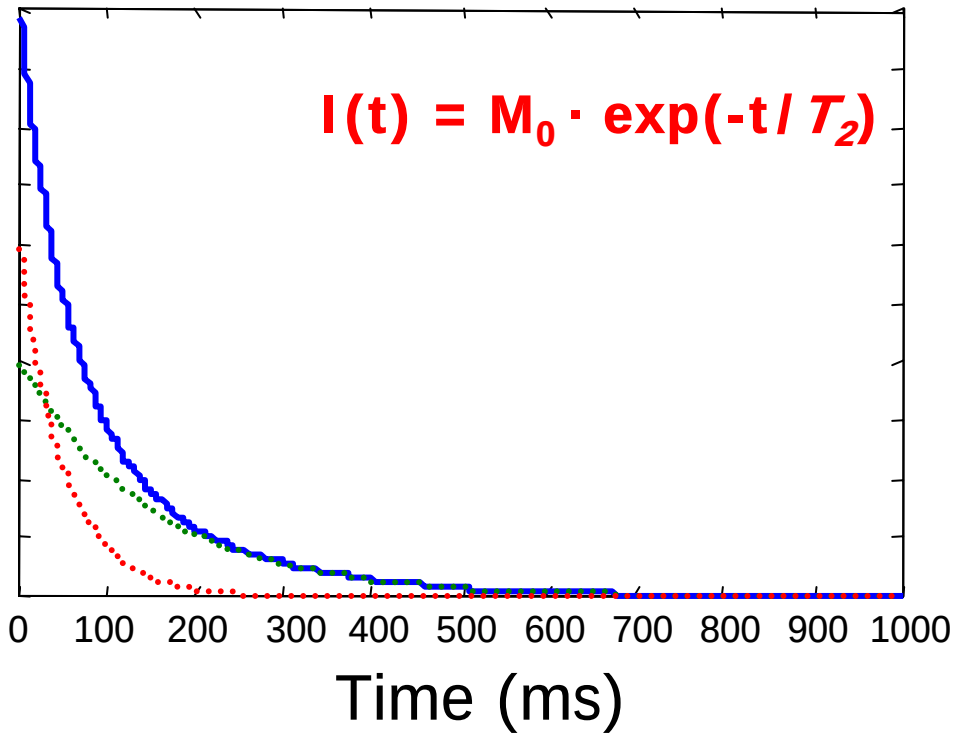
Using a general maximum likelihood fitting for handling bias

Other applications

Resolving two-way exponentials

Two-way data but with very special structure in loadings

$$X = AB' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 27 & 9 & 3 & 1 \\ 16 & 8 & 4 & 2 \end{bmatrix}$$



Sample 1	Sample 2
27	16
9	8
3	4
1	2

Pedersen et al 01

SLICING - Pseudo Three-way Data

$$X_1 = AD_1B' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 9 & 3 & 1 \\ 8 & 4 & 2 \end{bmatrix}$$

Sample 1	Sample 2
27	16
9	8
3	4

SLAB 1 (X_1)

Sample 1	Sample 2
27	16
9	8
3	4

$$X_2 = AD_2B' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 & 1 \\ 8 & 4 & 2 \end{bmatrix}$$

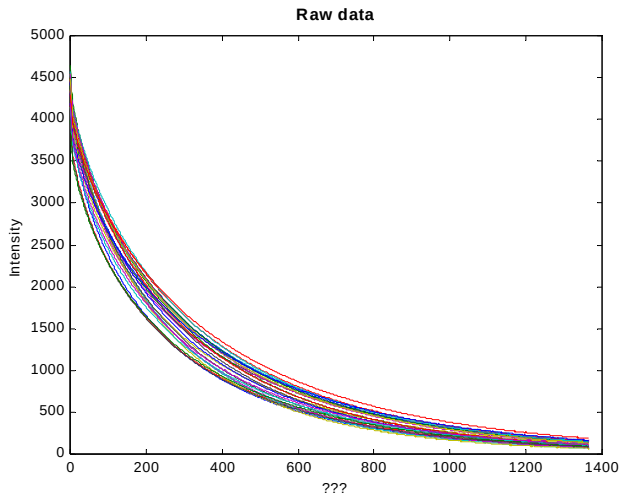
SLAB 2 (X_2)

Sample 1	Sample 2
9	8
3	4
1	2

Slicing principle

Example: Sensory quality

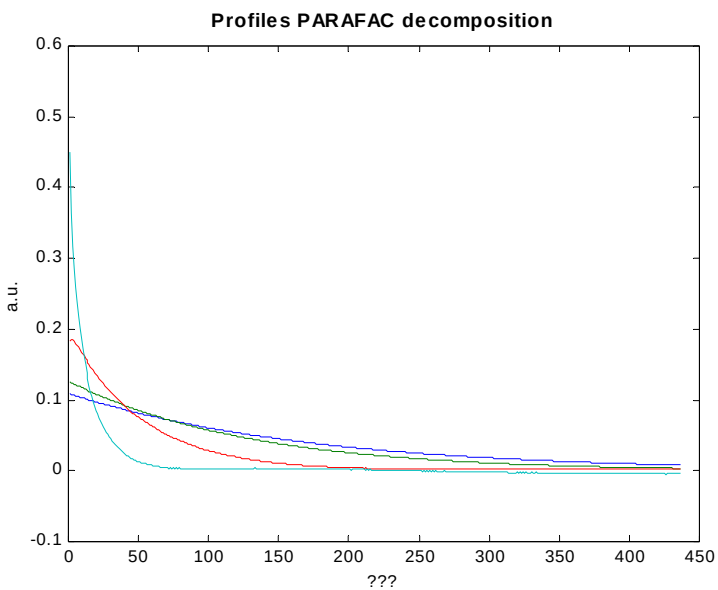
- **Predicting cooked potato quality from NMR on raw data**
  - Potatoes cooked and served hot. The assessors (10) evaluate texture profile, e.g. mealiness
  - *Raw* (!) potato measured by NMR (CPMG pulse sequence)

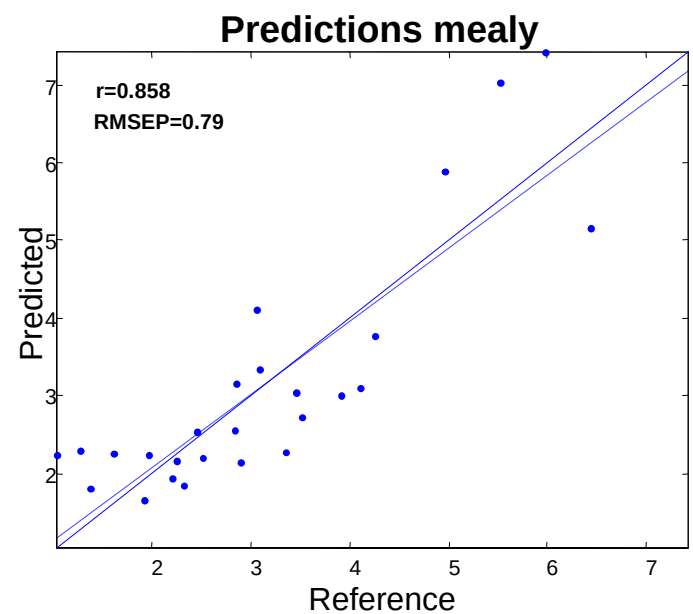


A. K. Thybo et. al., *Food Science and Technology,* 33 (2):103-111, 2000.

Example: Sensory quality

- **Decomposing NMR into meaningful latent variables**
  - PARAFAC on lagged data enables meaningful decomposition
  - Four components adequate for describing NMR data
- **Predictions of mealiness from NMR**
  - Predicting sensory quality of *cooked* potatoes from amount of latent variables (i.e. directly from NMR of *raw* potatoes)





MILES – maximum likelihood

- Algorithm **MILES** (Maximum likelihood via Iterative Least squares EStimation)
- Enables weighted least squares and maximum likelihood fitting of any model which has a least squares algorithm

1. Initialize model, m_0, with LS, set c := 0;
2. $\mathbf{q} = \mathbf{m}_c + 1/\beta\ \mathbf{W}^T\mathbf{W}(\mathbf{x} - \mathbf{m}_c)$
3. $\mathbf{m}_{c+1} = \underset{\mathbf{m}\in\Upsilon}{\arg\min} \|\mathbf{m} - \mathbf{q}\|_F^2$
4. c := c+1; go to step 2 until convergence

Calculate **q**

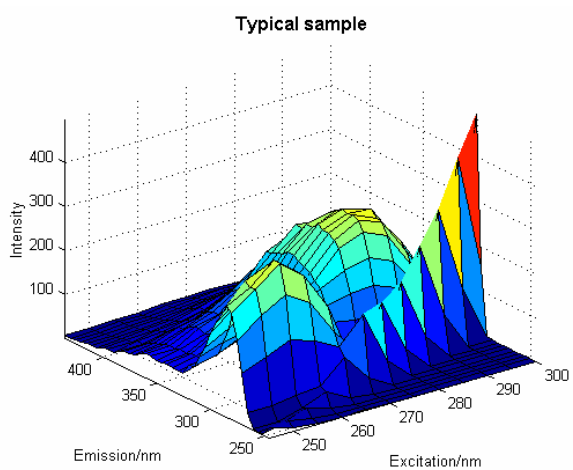
Fit LS model to **q** instead of to data

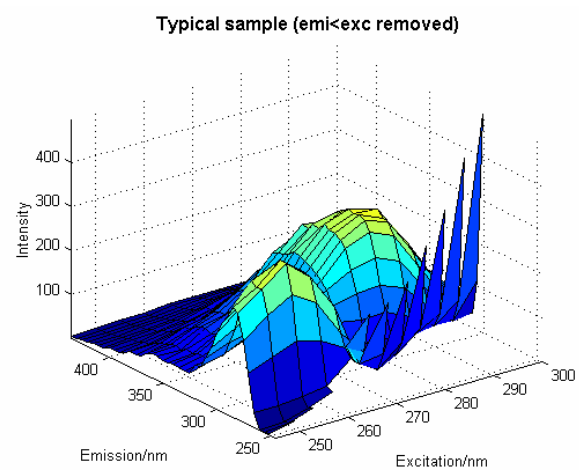
Example: fluorescence

- 21 samples containing L-phenylalanine, L-3,4-dihydroxy-phenyl-alanine (DOPA), 1,4-dihydroxy-benzene & L-tryptophan

- Three types of unwanted variation
  - Measurement error (~iid Gaussian)
  - Rayleigh and Raman scatter
  - Non-chemical area

ε

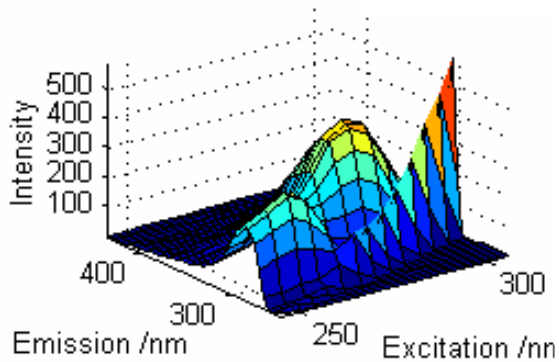




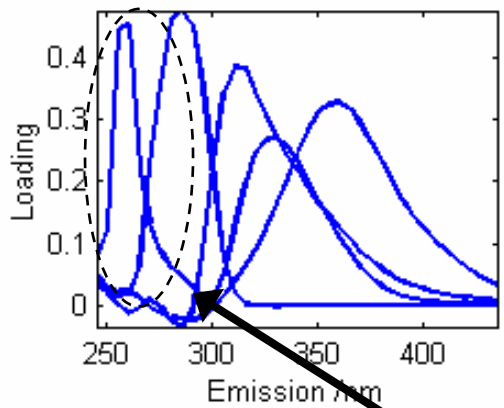
Baunsgaard D, Factors affecting 3-way modelling (PARAFAC) of fluorescence landscapes, The Royal Veterinary & Agricultural University, 1999

Example: fluorescence

RAW DATA



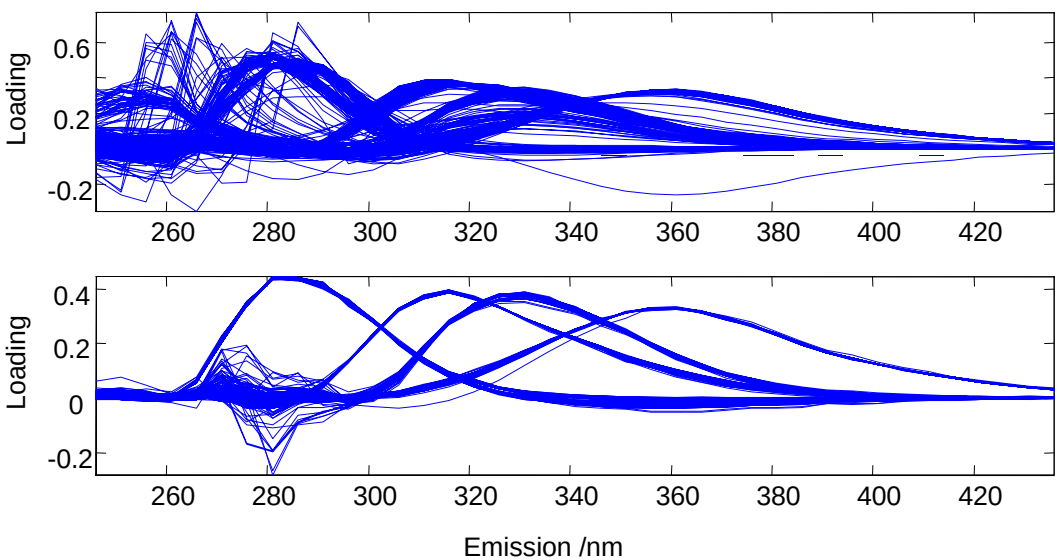
Least squares PARAFAC



Artifact

Example: fluorescence

- **Emission spectra from 100 resamplings**



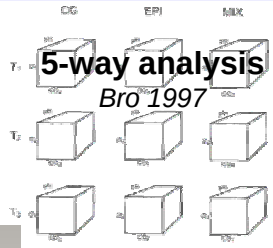
Least squares missing

"Maximum likelihood"

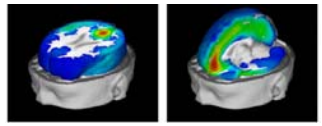
R. Bro, N. D. Sidiropoulos, and A. K. Smilde. Maximum likelihood fitting using simple least squares algorithms. Journal of Chemometrics, 2002

Other Examples

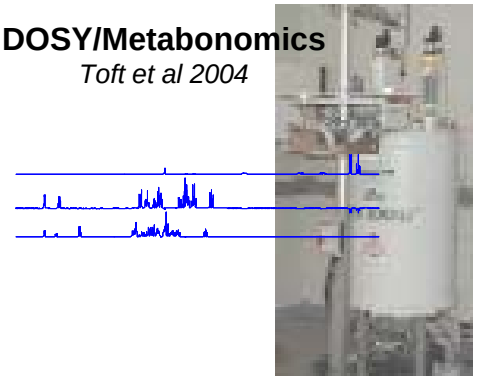
5-way analysis
Bro 1997



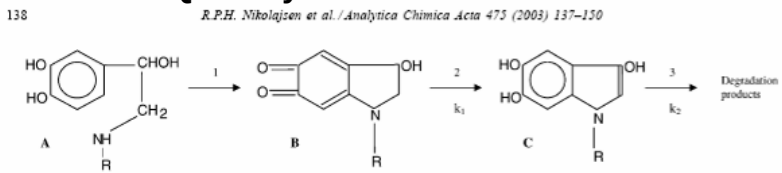
EEG
Miwakeichi et al 2004



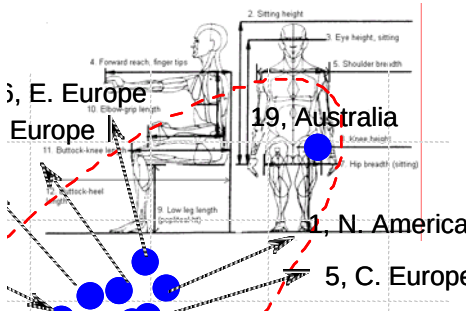
DOSY/Metabonomics
Toft et al 2004



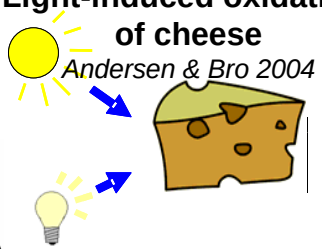
Quantify catecholamine in urine



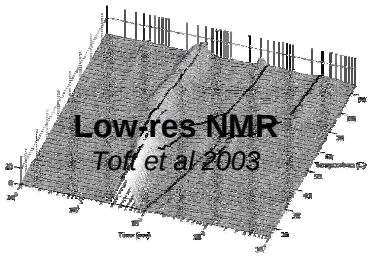
Anthropometry
S Lee & R Bro, 2004



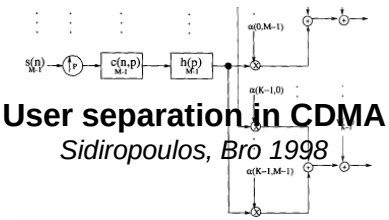
Light-induced oxidation of cheese
Andersen & Bro 2004



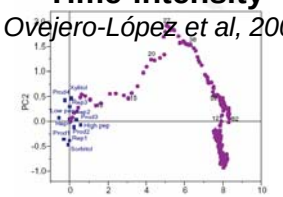
Low-res NMR
Toft et al 2003



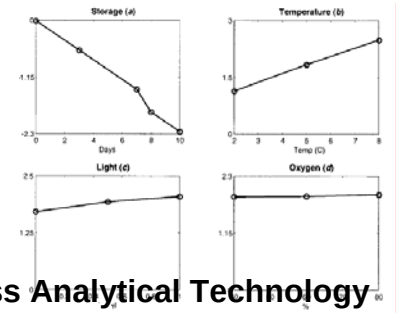
User separation in CDMA
Sidiropoulos, Bro 1998



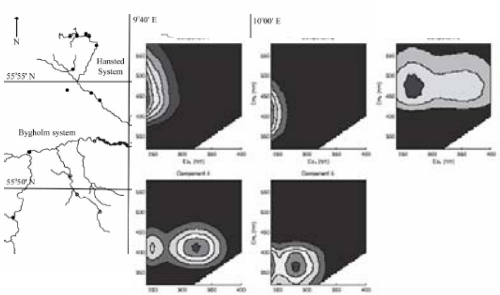
Time-intensity
Ovejero-López et al, 2004



Generalized ANOVA
Bro & Jakobsen 1996, 2002

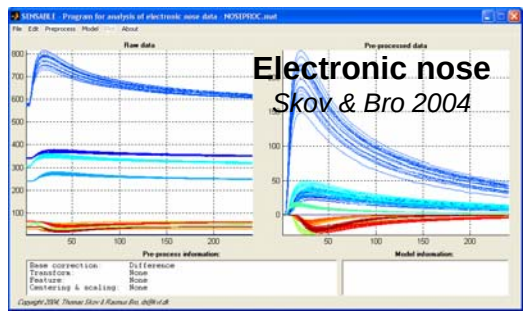


Tracing DOM
Stedmon, Markager, Bro 2003



Process Analytical Technology
Datamining
Bro 1998

Electronic nose
Skov & Bro 2004



Concluding remarks

- **Uniqueness** (mainly PARAFAC)

- **Better structural model**
  - Robustness/noise reduction
  - Simpler model
  - Interpretation
  - Better predictions

Concluding remarks

What I would have liked to talk about!

PARAFAC2

N-PLS, Multilinear partial least squares regression

Jack-knifing

References

References

- R. A. Harshman. Foundations of the PARAFA
  'explanatory' multi-modal factor analysis. UC
  1970 + all others by Harshman
- R. Bro. Multi-way Analysis in the Food Indus
  University of Amsterdam (NL), Ph.D,
  http://www.mli.kvl.dk/staff/foodtech/brothe
- P. M. Kroonenberg. Three-mode Principal Co
  Applications, Leiden:DSWO Press, 1983
- L. R. Tucker. Some mathematical notes on t
  31:279-311, 1966
- A. Smilde, R. Bro, P. Geladi, Multi-way Analysis, Wiley & Sons, 2004
- **H. A. L. Kiers. Towards a standardized notation and terminology in multiway analysis. J.Chemom. 14 (3):105-122, 2000.**

JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 105–122

Towards a standardized notation and terminology in multiway analysis

Henk A. L. Kiers*†

Heymans Institute (PA), University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands

SUMMARY

This paper presents a standardized notation and terminology to be used for three- and multiway analyses, especially when these involve (variants of) the CANDECOMP/PARAFAC model and the Tucker model. The notation also deals with basic aspects such as symbols for different kinds of products, and terminology for three- and higher-way data. The choices for terminology and symbols to be used have to some extent been based on earlier (informal) conventions. Simplicity and reduction of the possibility of confusion have also played a role in the choices made. Copyright © 2000 John Wiley & Sons, Ltd.

Software & info

- www.models.kvl.dk (free code,courses,papers,theses)
- www.eigenvector.com (Matlab Toolbox)
- publish.uwo.ca/~harshman/ Hard to get papers by Harshman
- TRICAP 2006, Technical Univ. of Crete, Greece,
  - contact: Nikos Sidiropoulos, nikos@telecom.tuc.gr;nikos@ece.umn.edu
- www.models.kvl.dk/users/rasmus/paloalto.zip Talk