



Chemometrics

a tool to analyze large datasets

Åsmund Rinnan, aar@life.ku.dk
KU-LIFE



Human pattern recognition uses all available data



Human pattern recognition uses all available data



Definitions

- Chemometrics:
“Application and development of mathematical and statistical methods to extract information from multivariate chemical data”
- Exploratory chemometric data analysis:
 - Seeking latent variables in data
 - Graphics
 - Hypothesis generating analysis

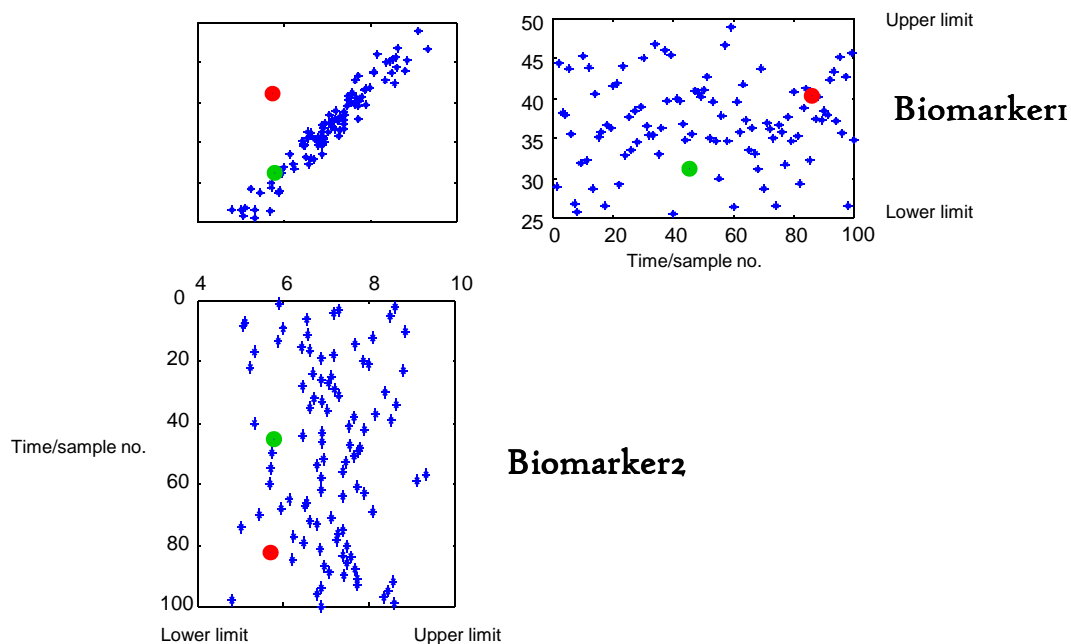


Multivariate "sensors"

- Fluorescence, ultraviolet-visual, nearinfrared (NIR), FT-Infrared (FT-IR), FT-IR Microscope, Raman, Raman Microscope, Nuclear Magnetic Resonance (NMR)
- GC-MS, MS-MS, HPLC-DAD
- Physical-chemical measurements
- Process parameters
- Or mixtures of the above-mentioned...



Co-variance – a central point



Important tools

- One data-structure, X:
 - Principal Component Analysis (PCA)
- Two data-structures, X & Y:
 - Partial Least Squares Regression (PLS)
- Classification, X & Class
 - Extended Canonical Variate Analysis (ECVA)

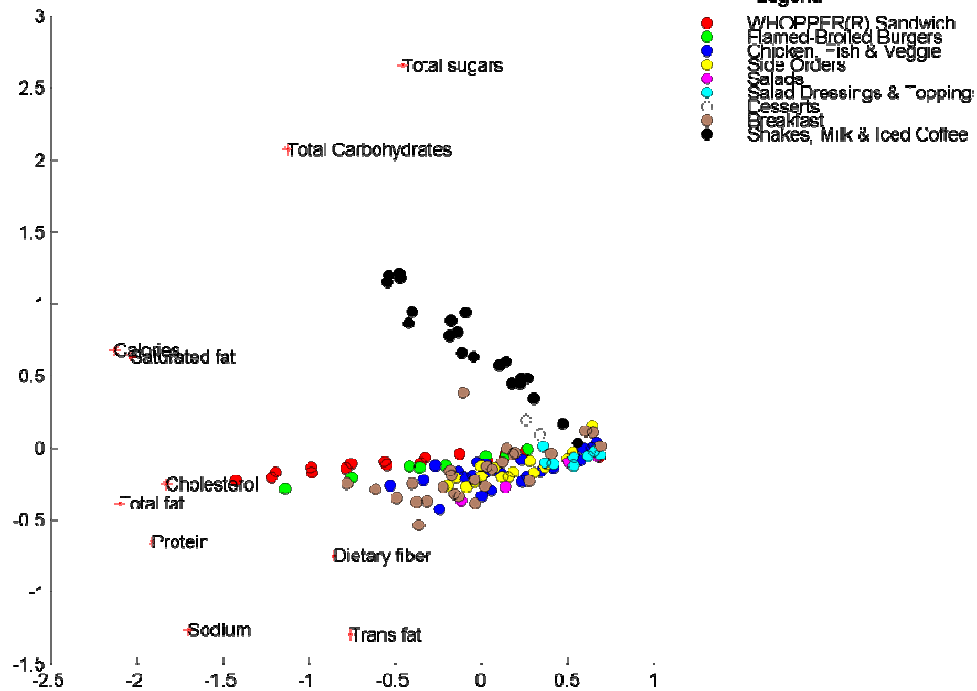


Burger King – Analyzed by PCA

		Calories	Total fat (g)	Saturated fat* (g)	Trans Fat (g)	Chol (mg)	Sodium (mg)	Total Carb (g)	Dietary Fiber(g)	Total Sugars	Protein (g)	Servin Size (g)
Chicken, Fish, & Veggie												
TENDERGRILL® Chicken Sandwich (with Mayo)		510	19	3.5	0.5	75	1180	49	4	7	37	258
	w/o Mayo	400	7	1.5	0	70	1090	49	4	7	36	244
TENDERCRISP® Chicken Sandwich		790	44	8	4	70	1640	88	5	9	33	294
Original Chicken Sandwich		660	40	8	2.5	70	1440	52	4	5	24	219
	w/o Mayo	450	17	4	2	60	1250	52	4	5	23	190
Spicy CHICKN CRISP™ Sandwich		480	31	5	2	45	870	36	1	4	15	144
	w/o Mayo	320	13	2.5	1.5	30	730	36	1	4	15	122
CHICKEN TENDERS® Kid's Meal	4 pc	170	10	2.5	1.5	25	480	11	0	0	9	82
CHICKEN TENDERS®	5 pc	210	12	3	2	35	600	13	0	0	12	77
CHICKEN TENDERS® Big Kid's Meal	6 pc	250	15	3.5	2.5	40	720	16	0	0	14	92
CHICKEN TENDERS®	8 pc	340	20	5	3	55	960	21	<1	1	19	123
Barbecue Dipping Sauce (1 oz)		40	0	0	0	0	310	11	0	10	0	28
Honey Mustard Dipping Sauce (1 oz)		90	6	1	0	10	180	8	0	7	0	28
Sweet and Sour Dipping Sauce (1 oz)		45	0	0	0	0	55	11	0	10	0	28
Ranch Dipping Sauce (1 oz)		140	15	2.5	0	5	95	1	0	1	1	28
BK™ CHICKEN FRIES	6 pc	260	15	3.5	3	35	650	18	2	1	12	85
	8 pc	360	23	5	4.5	50	980	26	3	1	18	128
	12 pc	520	31	7	6	65	1300	35	4	2	25	170
Buffalo Dipping Sauce (1 oz)		60	6	1.5	0	5	350	2	0	1	0	28
BK BIG FISH® Sandwich		640	32	6	2.5	65	1450	67	3	9	24	249
	w/o Tartar Sauce	470	13	3	2	60	1240	65	3	7	23	220
BK VEGGIE® Burger		420	16	2.5	0	10	1100	46	7	8	23	215
	w/ Cheese	470	20	5	0	20	1320	47	7	9	25	228
	w/o Mayo	340	8	1	0	0	1030	46	7	8	23	205
Side Orders												
MOTT'S® Strawberry Flavored Apple Sauce		90	0	0	0	0	0	23	<1	21	0	113
Onion Rings - Small		140	7	1.5	1	0	210	18	2	2	2	43
Onion Rings - Medium		310	15	3.5	2.5	0	440	37	3	4	4	81
Onion Rings - Large		440	22	4.5	4	0	620	53	5	6	6	130
Onion Rings - King		500	25	5	4.5	0	720	62	5	7	7	160
Zesty Onion Ring Dipping Sauce (1 oz)		150	15	2.5	0	15	210	3	<1	2	0	28
CHEESY TOTS™ Potatoes - (6 pc)		210	12	4.5	2	20	650	20	2	1	7	77
CHEESY TOTS™ Potatoes - (8 pc)		320	18	7	3	30	970	30	2	2	10	115
CHEESY TOTS™ Potatoes - (12 pc)		430	24	9	4	40	1300	40	3	2	14	153
French Fries - Small (Salted)		230	13	3	3	0	380	26	2	1	2	74
French Fries - Medium (Salted)		360	20	4.5	4.5	0	590	41	4	1	4	116
French Fries - Large (Salted)		500	28	6	6	0	820	57	5	1	5	160
French Fries - King (Salted)		600	33	8	7	0	990	69	6	2	6	194
French Fries - Small (Salt not added)*		230	13	3	3	0	240	26	2	1	2	74
French Fries - Medium (Salt not added)*		360	20	4.5	4.5	0	380	41	4	1	4	116
French Fries - Large (Salt not added)*		500	28	6	6	0	530	57	5	1	5	160
French Fries - King (Salt not added)*		600	33	8	7	0	640	69	6	2	6	194



Burger King – Bi-plot/Overview



Burger King

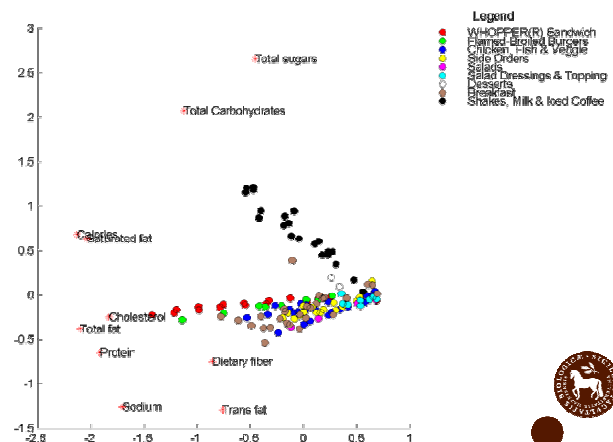
Chicken, Fish & Veggie									
	Calories	Total fat (g)	Saturated fat (g)	Trans fat (g)	Chol (mg)	Sodium (mg)	Total Carb (g)	Dietary Fiber (g)	Total Protein (g)
TENDERGRILL® Chicken Sandwich (with Mayo)	510	19	3.5	0.5	75	1180	46	4	27
TENDERCRISP® Chicken Sandwich	400	7	1.5	0	70	1060	46	4	7
Original Chicken Sandwich	730	44	8	4	70	1940	65	5	33
Spikey CHICKN CRISP™ Sandwich	480	17	4	2	50	1200	52	4	5
CHICKEN TENDER® Kid's Meal	480	31	5	2	45	870	36	1	4
CHICKEN TENDER® Big Kid's Meal	520	13	2.5	1.5	30	730	36	1	4
CHICKEN TENDER® 6 pc	170	10	2.5	1.5	25	480	11	0	0
CHICKEN TENDER® 8 pc	210	12	3	2	35	650	13	0	0
Salad Dressing Sauce (1 oz)	230	15	3.5	2.5	40	720	16	0	0
Honey Mustard Dipping Sauce (1 oz)	340	20	5	3	55	860	21	<1	1
Swiss and Sour Dipping Sauce (1 oz)	40	0	0	0	0	310	11	0	0
Ranch Dipping Sauce (1 oz)	90	0	0	0	0	150	8	0	0
BK CHICKEN FRIES	45	0	0	0	0	55	11	0	0
Buffalo Dipping Sauce (1 oz)	140	15	2.5	0	5	65	1	0	1
BK BIG FISH® Sandwich	280	15	3.5	3	35	650	18	2	1
BK VEGGIE® Burger	390	23	5	4.5	50	980	26	3	1
BK Brio® Sandwich	520	31	7	6	65	1300	35	4	2
BK Brio® Sandwich	80	8	1.5	0	5	350	2	0	1
BK Brio® Sandwich	940	32	8	2.5	65	1450	67	3	9
BK Brio® Sandwich	470	13	3	2	50	1240	65	3	7
BK Brio® Sandwich	420	16	2.5	0	10	1100	46	7	8
BK Brio® Sandwich	470	20	5	0	20	1520	47	7	9
BK Brio® Sandwich	340	8	5	0	0	1030	46	7	8
BK Brio® Sandwich	90	0	0	0	0	23	<1	21	0
BK Brio® Sandwich	140	7	1.5	1	0	210	18	2	2
BK Brio® Sandwich	310	15	3.5	2.5	0	440	37	3	4
BK Brio® Sandwich	440	22	4.5	4	0	620	53	5	6
BK Brio® Sandwich	580	25	5	4.5	0	720	62	5	7
BK Brio® Sandwich	150	15	2.5	0	15	210	3	<1	2
BK Brio® Sandwich	210	12	4.5	2	20	650	20	2	1
BK Brio® Sandwich	120	18	5	3	30	670	30	2	1
BK Brio® Sandwich	430	24	9	4	40	1300	40	3	2
BK Brio® Sandwich	230	13	3	0	30	45	2	2	74
BK Brio® Sandwich	360	20	4.5	4.5	0	590	41	4	1
BK Brio® Sandwich	600	28	6	6	0	820	67	5	1
BK Brio® Sandwich	500	33	8	7	0	990	69	6	2
BK Brio® Sandwich	230	13	3	0	240	28	2	1	2
BK Brio® Sandwich	360	20	4.5	4.5	0	380	41	4	1
BK Brio® Sandwich	600	28	6	6	0	840	69	6	2
BK Brio® Sandwich	600	33	8	7	0	940	69	6	2

TM © 2007 Burger King Brands, Inc. (USA only).
TM © 2007 Burger King Corporation (Outside USA). All rights reserved.

Nutrition - Web

Page 2
1/15/2008

Table or graphics?



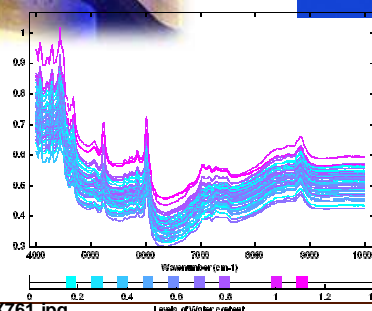
Important tools

- One data-structure, X:
 - Principal Component Analysis (PCA)
- Two data-structures, X & Y:
 - Partial Least Squares Regression (PLS)
- Classification
 - Extended Canonical Variate Analysis (ECVA)

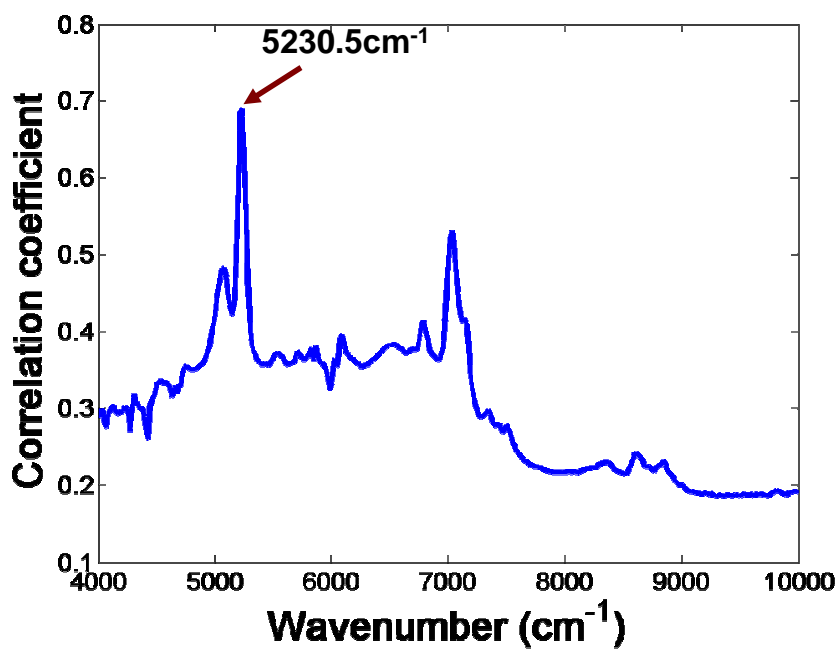


Tablets – Analyzed by PLSR

Concentration of water (y) in tablets by near infrared spectroscopy (X)



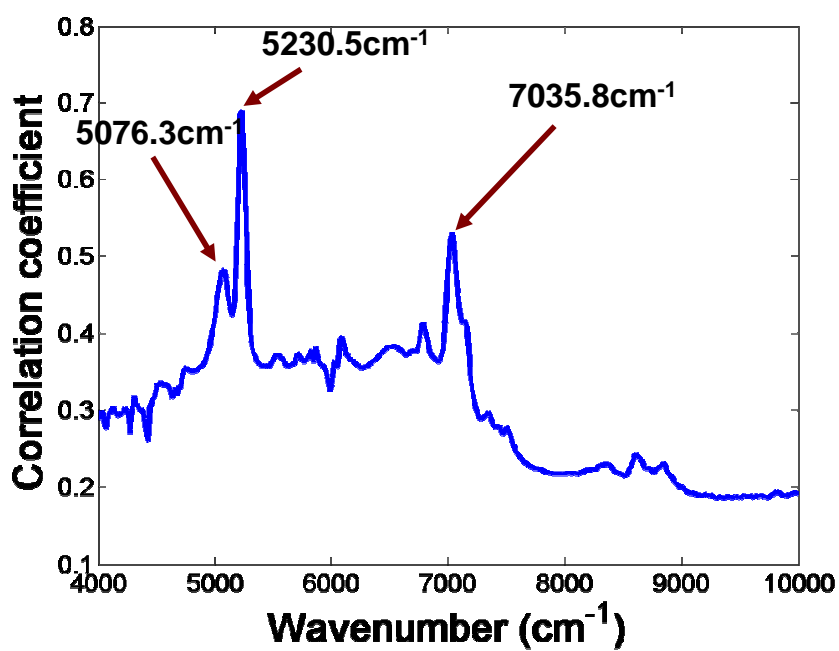
Tablets - Options



Univariate



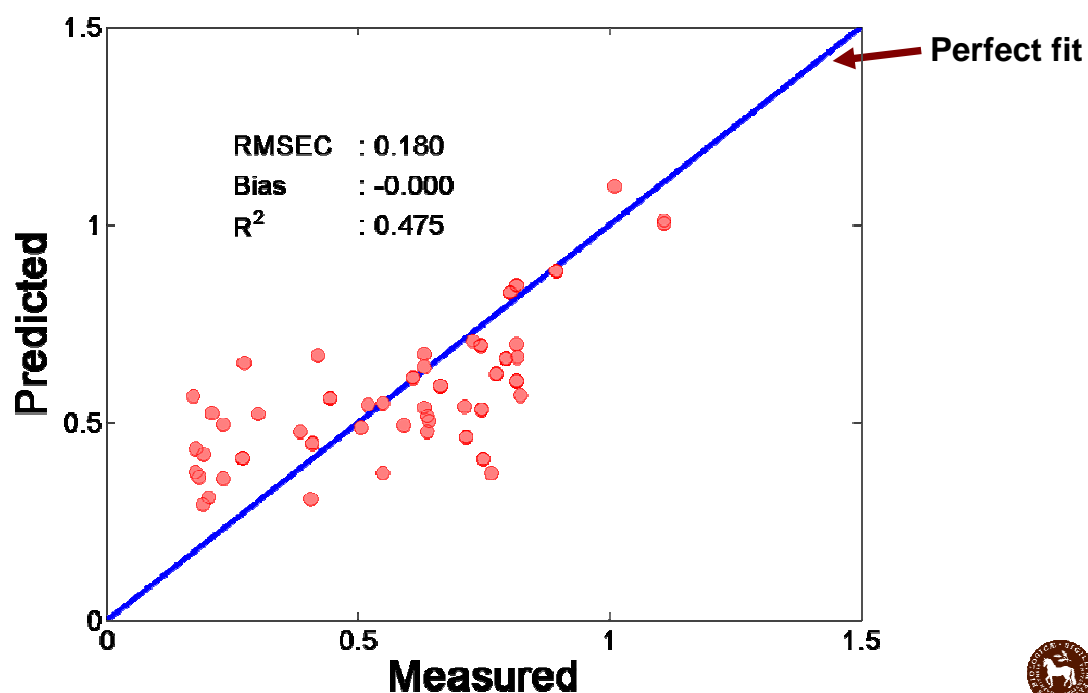
Tablets - Options



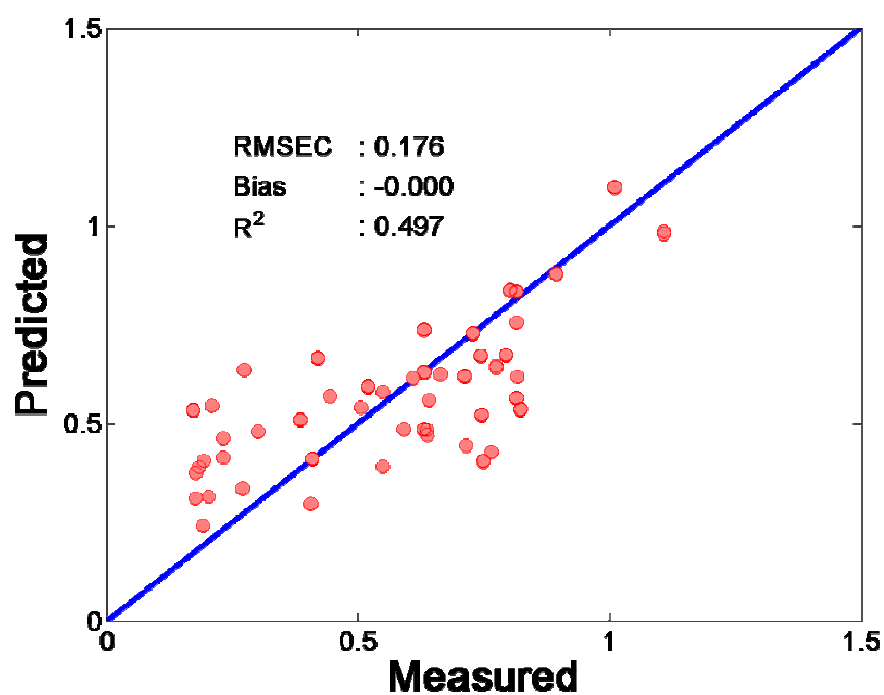
Multivariate



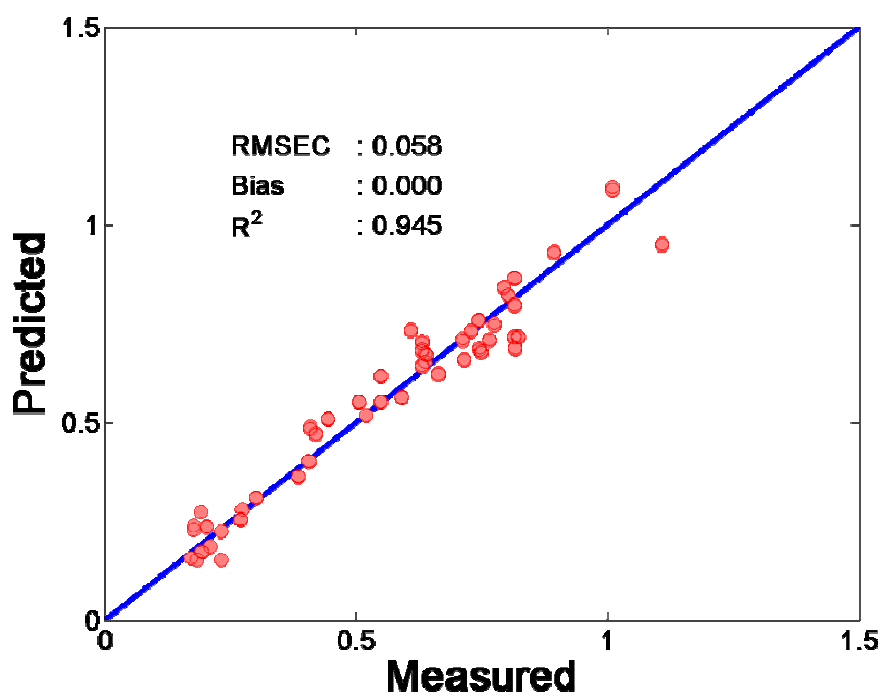
Tablets - MLR (1 variable)



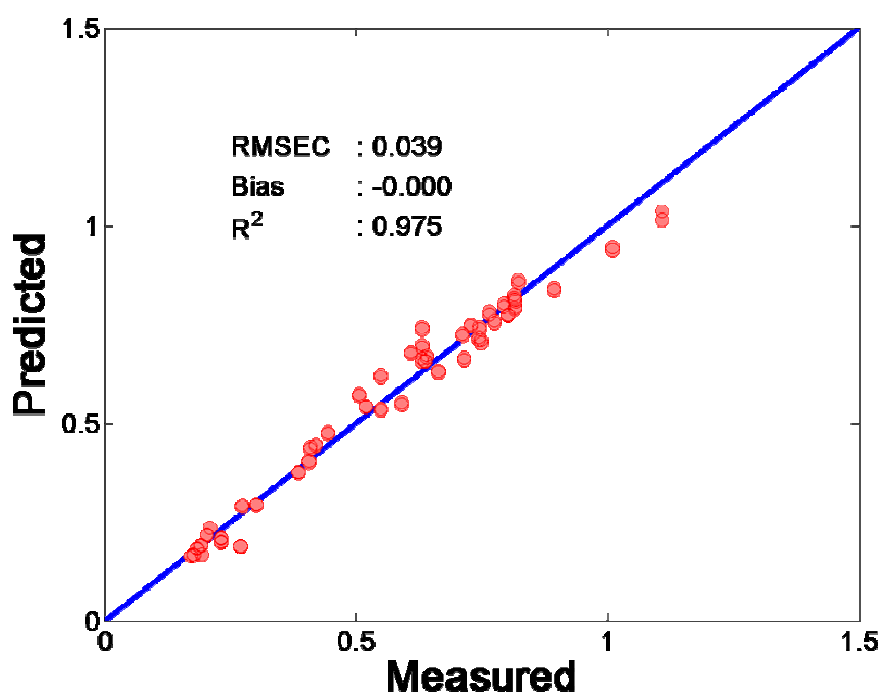
Tablets - MLR (2 variables)



Tablets - MLR (3 variables)



Tablets - PLS (3 factors out of 780 variables)

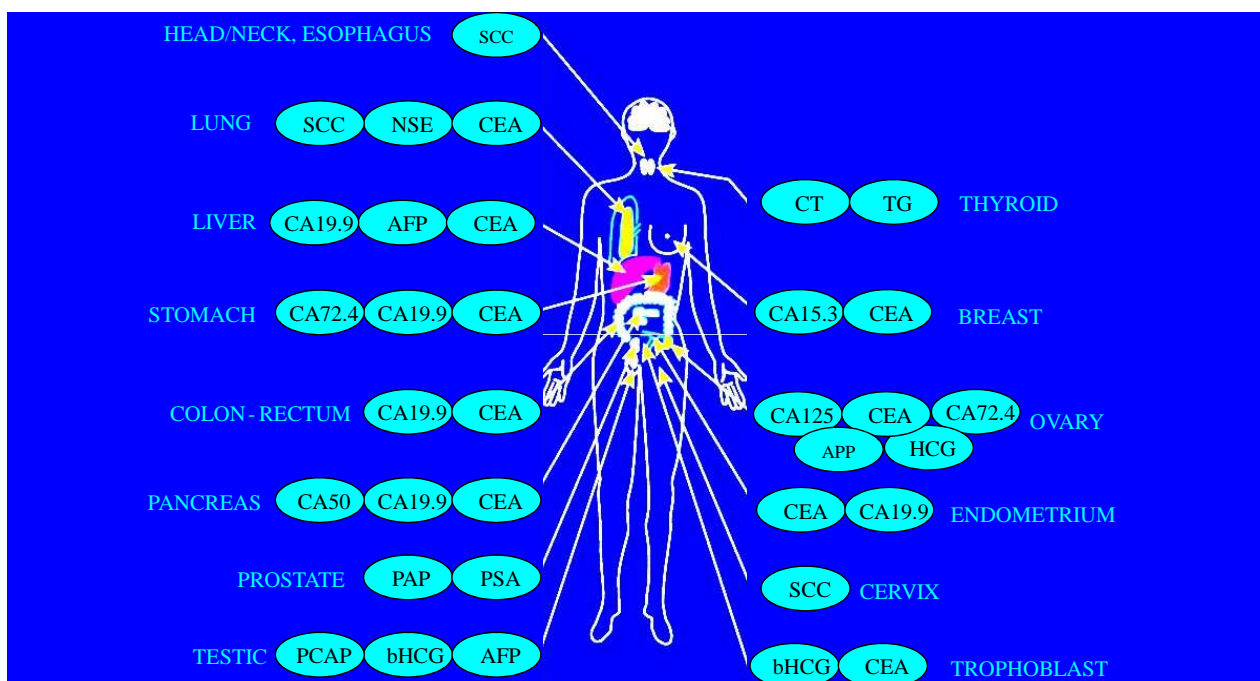


Important tools

- One data-structure, X:
 - Principal Component Analysis (PCA)
- Two data-structures, X & Y:
 - Partial Least Squares Regression (PLS)
- Classification
 - Extended Canonical Variate Analysis (ECVA)

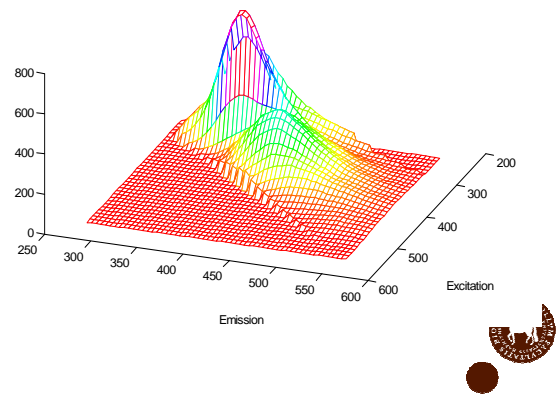
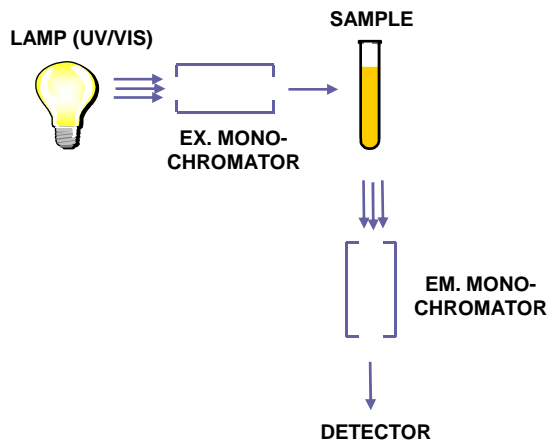


Cancer diagnosis – Analyzed by ECVA



Cancer diagnosis - Overview

- A fluorescence landscape measured directly on the serum sample (undiluted or diluted) yields a multivariate spectroscopic fingerprint that contains information about health status of the individual



Cancer diagnosis

Control Group

- Group 1: 13 healthy females

Patient Group (26 females)

- Group 3: 11 females with solitary metastases
- Group 5: 15 females with progressive disease

Cancer diagnosis - Options

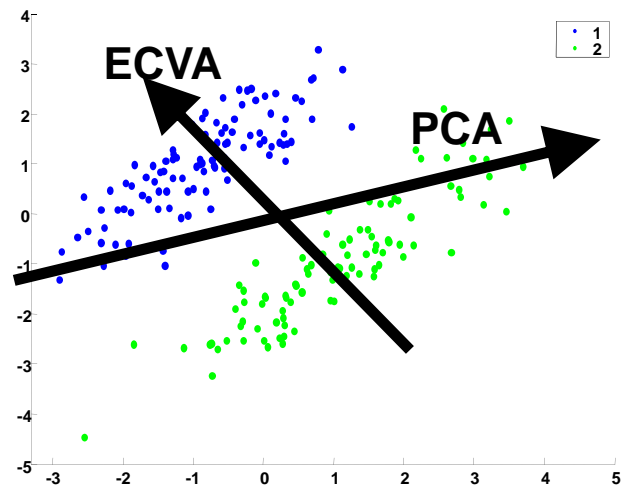
ECVA - Extended Canonical Variance Analysis

Find w for which:

“between class scatter”

“within class scatter”

is maximized



Cancer diagnosis - Results

	CA 15-3, Cutoff value* 30 kU/L	CEA Cutoff value* 7.5 µg/L	Fluorescence spectroscopy & chemometrics
False positives	0	0	0
False negatives	8 7 in group 3 1 in group 5	9 7 in group 3 2 in group 5	1 1 in group 3
Outliers	Not possible	Not possible	1 (in group 5)

*recommended cutoff values





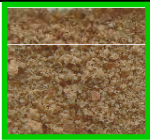




Biomarkers in food

Metabolomics / metabonomics

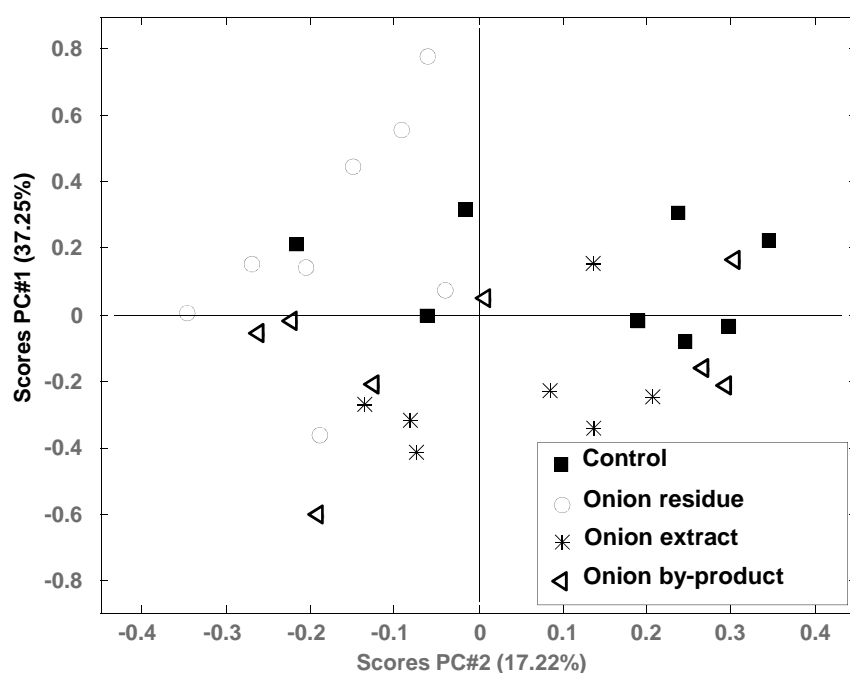


Biomarker for onion intake

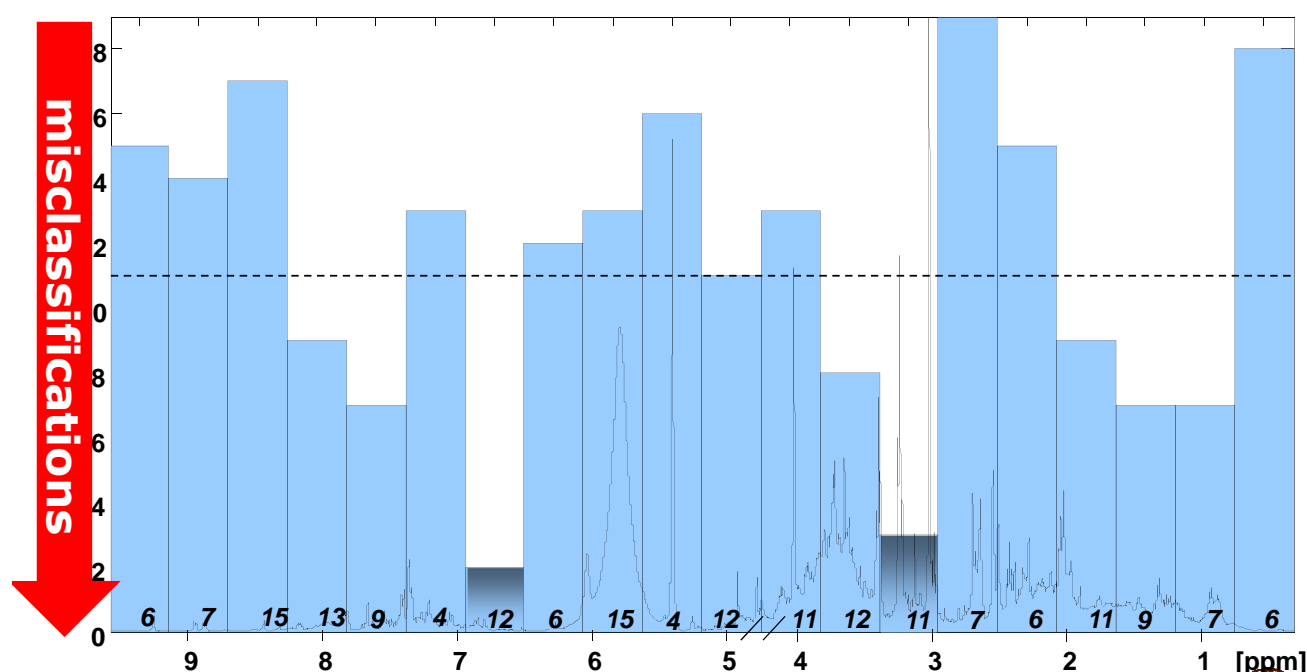
Composition of rat feed	 <i>Control</i>	 <i>Onion powder</i>	 <i>Onion extract</i>	 <i>Onion residual</i>
Control feed	1000g	900g	930g	970g
Powder 10% 	0	100g	0	0
Extract 7% 	0	0	70g	0
Residual 3% 	0	0	0	30g
Total feed	1000g	1000g	1000g	1000g



Biomarkers – Analyzed by PCA

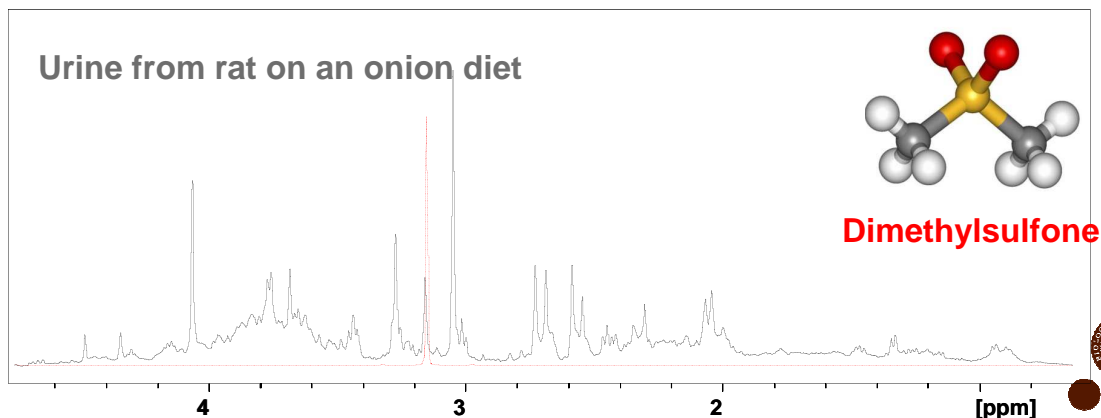
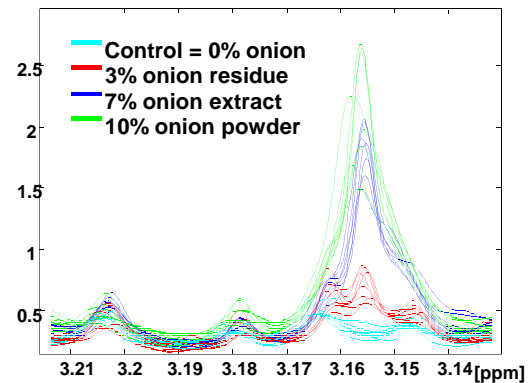


Biomarkers – interval ECVA



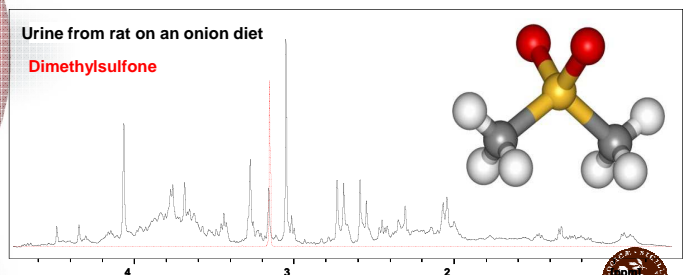
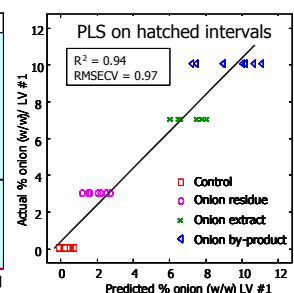
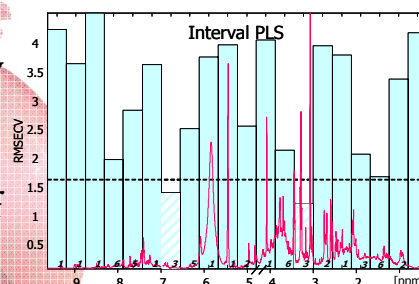
Biomarkers - Results

- Onion contains many sulfoxides
- Oxidation product of dimethylsulfoxide (DMSO)
- Respiration of dimethylsulfone has recently been linked with the occurrence of skin cancer

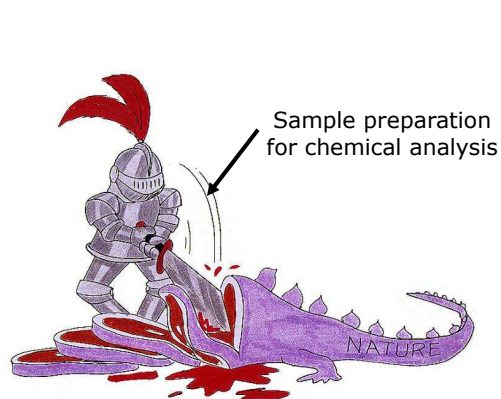


Biomarkers - Results

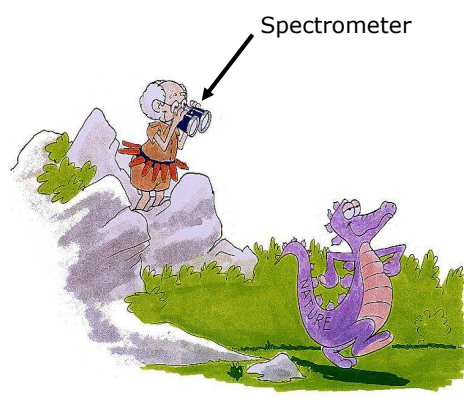
- **FACT:** Onions have potentially beneficial effects on health
- **AIM:** Evaluate the *in vivo* metabolome following intake of onion by-products
- **RESULT:** HR NMR spectroscopy combined with adv. multivariate data techniques discovers exclusive biomarkers in urine followed by onion feed



Why chemometrics & spectroscopy?



Invasive
Destructive
Slow
Environmentally harmful
Univariate
Breaking co-variance



Remote
Non-destructive
Rapid
Environmentally friendly
Multivariate
Taking care of co-variance



Summary

- **State-of-the-art**
 - Exploratory data analysis by Principal Component Analysis
 - Multivariate classification tools, i.e. Extended Canonical Variate Analysis
- **Hypotheses**
 - There should be a covariate underlying structure in data
- **Results achieved**
 - Successfully applied to a many scientific disciplines
- **Future work**
 - Dissemination of the methods to other scientific disciplines
 - Better and more user-friendly methods, i.e. less need for an expert user for the analysis

