Characterisation and selection of HPLC columns by chemometric techniques

Royal Flemish Chemical Society

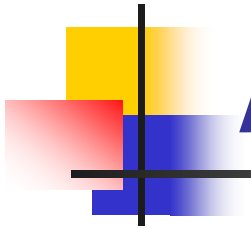
12 October 2006

Drs. Jan P.M. Andries

avans hogeschool

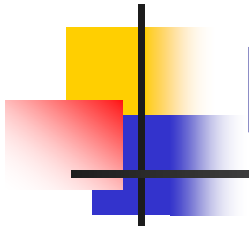
Research Group Separation Techniques

- Avans Hogeschool
  Breda
  The Netherlands
- School for Life Science and Technology
- Head: Dr. H.A. Claessens
- Project in cooperation with
  Chemometrics Research Department
  Prof. Dr. L.M.C. Buydens
  Radboud University Nijmegen



Aim

- Model retention of compounds
- Calculate
  - retention diagrams
  - resolution plots
- Predict chromatograms



Modelling retention

- On selected set of columns
- With Quantitative Structure Retention Relationships (QSRR)
- Using
  - small calibration set
  - molecular descriptors
  - chemometric techniques

Quantitative Structure Retention Relationships (QSRR)

- Mathematical model describes quantitatively relation between
  - retention of compound on RP-LC column
  - molecular and physical chemical properties of compound

QSRR modelling in HPLC

- Retention of a component on a LC column depends on:
  - stationary phase
  - mobile phase
  - physical and chemical properties of component

QSRR modelling

Y-block
- Log k_w
- S
- Log k

Modelling
MLR, PLS, PCR,
UVE-PLS/GA/MLR

X-block
- Set of predictors?
- Set of components?

Log k_w and 2 descriptor sets

Journal of Chromatography A, 855 (1999) 455

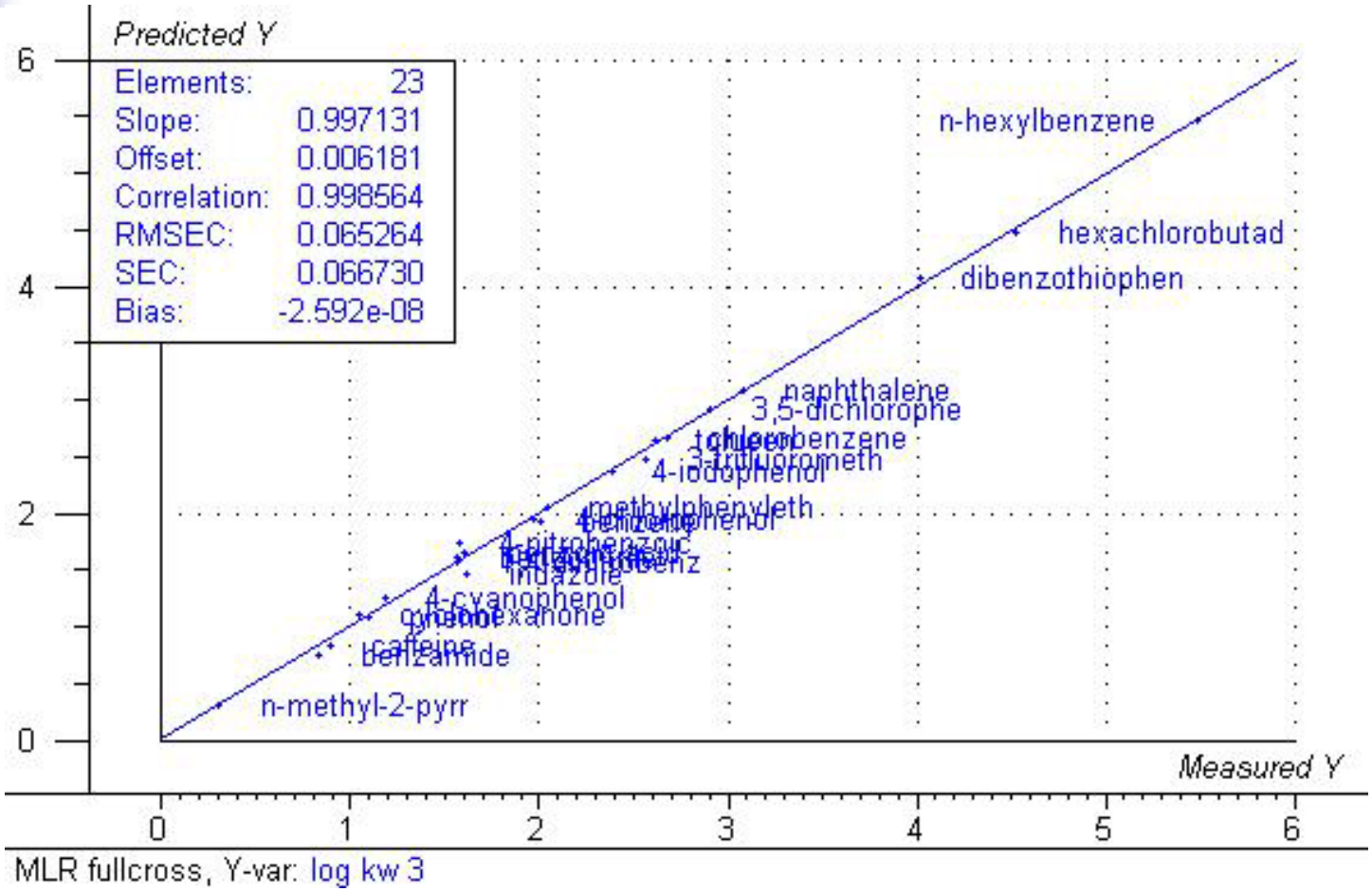
	log kw	Log P	R_2	π_2^H	α_2^H	β_2^H	V_x
1 n-Hexylbenzene	5,4892	5,52	0,591	0,50	0,00	0,15	1,562
3 1,4-Dinitrobenzene	1,5692	1,47	1,130	1,63	0,00	0,41	1,065
4 3-Trifluoromethylphenol	2,5650	2,95	0,425	0,87	0,72	0,09	0,969
5 3,5-Dichlorophenol	2,9099	3,62	1,020	1,10	0,83	0,00	1,020
6 4-Cyanophenol	1,1786	1,60	0,940	1,63	0,79	0,29	0,930
7 4-Iodophenol	2,3819	2,91	1,380	1,22	0,68	0,20	1,033
8 Methylphenylether	2,0436	2,11	0,708	0,75	0,00	0,29	0,916
9 Benzamide	0,8308	0,64	0,990	1,50	0,49	0,67	0,973
10 Benzene	2,0052	2,13	0,610	0,52	0,00	0,14	0,716
11 Chlorobenzene	2,6725	2,89	0,718	0,65	0,00	0,07	0,839
12 Cyclohexanone	1,0396	0,81	0,403	0,86	0,00	0,56	0,861
13 Dibenzothiophene	4,0185	4,38	1,959	1,31	0,00	0,18	1,379
14 Phenol	1,0938	1,47	0,805	0,89	0,60	0,30	0,775
15 Hexachlorobutadiene	4,5311	4,78	1,019	0,85	0,00	0,00	1,321
16 1ndazole	1,6175	1,77	1,180	1,25	0,54	0,34	0,905
17 Caffeine	0,8923	-0,07	1,500	1,60	0,00	1,35	1,363
18 4-Nitrobenzoic acid	1,5758	1,89	0,990	1,07	0,62	0,54	1,106
19 N-Methyl-2-pyrrolidinone	0,3015	-0,54	0,491	1,50	0,00	0,95	0,820
20 Naphthalene	3,0787	3,30	1,340	0,92	0,00	0,20	1,085
21 4-Chlorophenol	1,9694	2,39	0,915	1,08	0,67	0,20	0,898
22 Toluene	2,6125	2,73	0,601	0,52	0,00	0,14	0,716
23 Benzonitrile	1,5680	1,56	0,742	1,11	0,00	0,33	0,871
24 Benzoic acid	1,6027	1,87	0,730	0,90	0,59	0,40	0,932

Extended Linear solvation-energy relationship (LSER)

$$\log k_w = k_1' + k_2 \log P + k_2' R_2 + k_3' \pi_2^H + k_4' \alpha_2^H + k_5' \beta_2^H + k_6' V_x$$

- Log P = log octanol-water partition coefficient
- R2 = excess molar refraction
- π_2^H = dipolarity/polarizability
- α_2^H = hydrogen-bond acidity
- β_2^H = hydrogen-bond basicity
- Vx = characteristic volume of McGowan

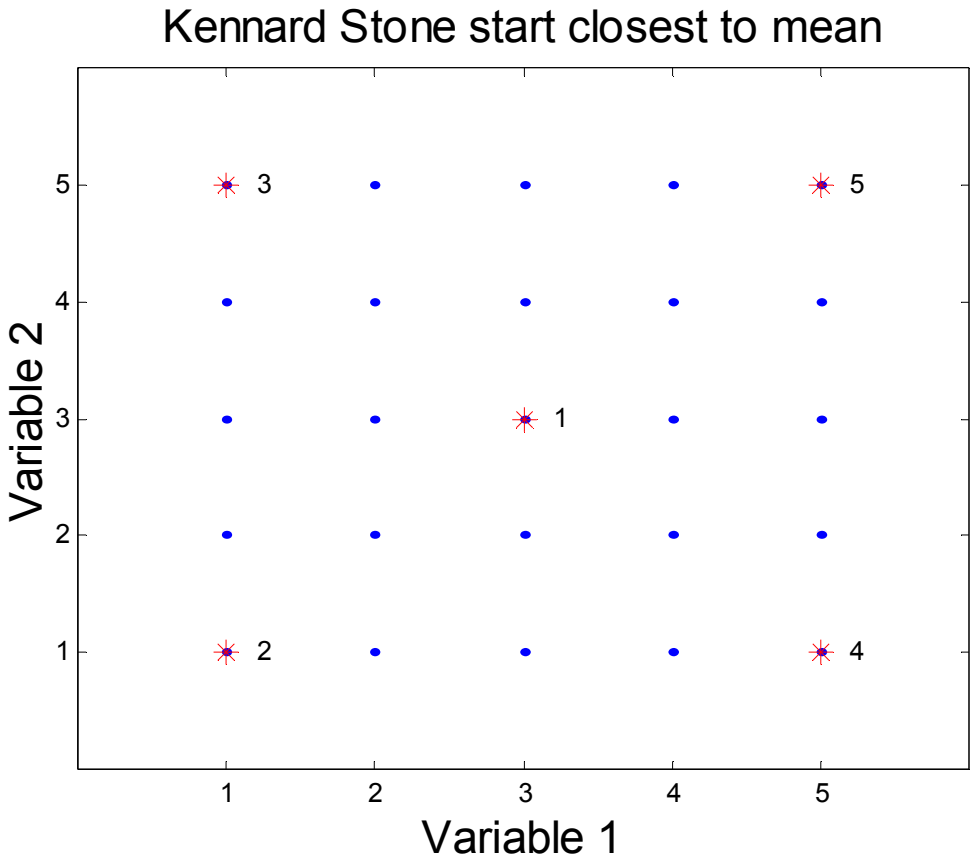
Model Extended LSER



Reduction calibration set
Selection procedure : max min

max/min selection	3					
	MeOH-buff	Log P	R_2	π_2^H	α_2^H	β_2^H
1 n-Hexylbenzene	5,4892	5,52	0,591	0,50	0,00	0,15
3 1,4-Dinitrobenzene	1,5692	1,47	1,130	1,63	0,00	0,41
4 3-Trifluoromethylphenol	2,5650	2,95	0,425	0,87	0,72	0,09
5 3,5-Dichlorophenol	2,9099	3,62	1,020	1,10	0,83	0,00
6 4-Cyanophenol	1,1786	1,60	0,940	1,63	0,79	0,29
7 4-Iodopheno1	2,3819	2,91	1,380	1,22	0,68	0,20
8 Methylphenylether	2,0436	2,11	0,708	0,75	0,00	0,29
9 Benzamide	0,8308	0,64	0,990	1,50	0,49	0,67
10 Benzene	2,0052	2,13	0,610	0,52	0,00	0,14
11 Chlorobenzene	2,6725	2,89	0,718	0,65	0,00	0,07
12 Cyclohexanone	1,0396	0,81	0,403	0,86	0,00	0,56
13 Dibenzothiophene	4,0185	4,38	1,959	1,31	0,00	0,18
14 Phenol	1,0938	1,47	0,805	0,89	0,60	0,30
15 Hexachlorobutadiene	4,5311	4,78	1,019	0,85	0,00	0,00
16 1ndazole	1,6175	1,77	1,180	1,25	0,54	0,34
17 Caffeine	0,8923	-0,07	1,500	1,60	0,00	1,35
18 4-Nitrobenzoic acid	1,5758	1,89	0,990	1,07	0,62	0,54
19 N-Methyl-2-pyrrolidinone	0,3015	-0,54	0,491	1,50	0,00	0,95
20 Naphthalene	3,0787	3,30	1,340	0,92	0,00	0,20
21 4-Chlorophenol	1,9694	2,39	0,915	1,08	0,67	0,20
22 Toluene	2,6125	2,73	0,601	0,52	0,00	0,14
23 Benzonitrile	1,5680	1,56	0,742	1,11	0,00	0,33
24 Benzoic acid	1,6027	1,87	0,730	0,90	0,59	0,40
	max	5,520	1,959	1,630	0,830	1,350
	min	-0,540	0,403	0,500	0,000	0,000

Selection procedure reduction calibration set: Kennard Stone



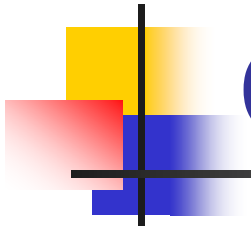
max/min selection	3 MeOH-buff	Log P	R_2	π_2^H	α_2^H	β_2^H	Kennard Stone selection furthest from mean
1 n-Hexylbenzene	5,4892	5,52	0,591	0,50	0,00	0,15	1 n-Hexylbenzene
3 1,4-Dinitrobenzene	1,5692	1,47	1,130	1,63	0,00	0,41	3 1,4-Dinitrobenzene
4 3-Trifluoromethylphenol	2,5650	2,95	0,425	0,87	0,72	0,09	4 3-Trifluoromethylphenol
5 3,5-Dichlorophenol	2,9099	3,62	1,020	1,10	0,83	0,00	5 3,5-Dichlorophenol
6 4-Cyanophenol	1,1786	1,60	0,940	1,63	0,79	0,29	6 4-Cyanophenol
7 4-Iodopheno1	2,3819	2,91	1,380	1,22	0,68	0,20	7 4-Iodopheno1
8 Methylphenylether	2,0436	2,11	0,708	0,75	0,00	0,29	8 Methylphenylether
9 Benzamide	0,8308	0,64	0,990	1,50	0,49	0,67	9 Benzamide
10 Benzene	2,0052	2,13	0,610	0,52	0,00	0,14	10 Benzene
11 Chlorobenzene	2,6725	2,89	0,718	0,65	0,00	0,07	11 Chlorobenzene
12 Cyclohexanone	1,0396	0,81	0,403	0,86	0,00	0,56	12 Cyclohexanone
13 Dibenzothiophene	4,0185	4,38	1,959	1,31	0,00	0,18	13 Dibenzothiophene
14 Phenol	1,0938	1,47	0,805	0,89	0,60	0,30	14 Phenol
15 Hexachlorobutadiene	4,5311	4,78	1,019	0,85	0,00	0,00	15 Hexachlorobutadiene
16 1ndazole	1,6175	1,77	1,180	1,25	0,54	0,34	16 1ndazole
17 Caffeine	0,8923	-0,07	1,500	1,60	0,00	1,35	17 Caffeine
18 4-Nitrobenzoic acid	1,5758	1,89	0,990	1,07	0,62	0,54	18 4-Nitrobenzoic acid
19 N-Methyl-2-pyrrolidinone	0,3015	-0,54	0,491	1,50	0,00	0,95	19 N-Methyl-2-pyrrolidinor
20 Naphthalene	3,0787	3,30	1,340	0,92	0,00	0,20	20 Naphthalene
21 4-Chlorophenol	1,9694	2,39	0,915	1,08	0,67	0,20	21 4-Chlorophenol
22 Toluene	2,6125	2,73	0,601	0,52	0,00	0,14	22 Toluene
23 Benzonitrile	1,5680	1,56	0,742	1,11	0,00	0,33	23 Benzonitrile
24 Benzoic acid	1,6027	1,87	0,730	0,90	0,59	0,40	24 Benzoic acid
	max	5,520	1,959	1,630	0,830	1,350	
	min	-0,540	0,403	0,500	0,000	0,000	

Reduced calibration set Results

reduction calibration set with max min method and Kennard Stone

		3					
		MeOH-buff	Log P	R_2	π_2^H	α_2^H	β_2^H
1	n-Hexylbenzene	5,4892	5,52	0,591	0,50	0,00	0,15
3	1,4-Dinitrobenzene	1,5692	1,47	1,130	1,63	0,00	0,41
5	3,5-Dichlorophenol	2,9099	3,62	1,020	1,10	0,83	0,00
12	Cyclohexanone	1,0396	0,81	0,403	0,86	0,00	0,56
13	Dibenzothiophene	4,0185	4,38	1,959	1,31	0,00	0,18
17	Caffeine	0,8923	-0,07	1,500	1,60	0,00	1,35
19	N-Methyl-2-pyrrolidinone	0,3015	-0,54	0,491	1,50	0,00	0,95
21	4-Chlorophenol	1,9694	2,39	0,915	1,08	0,67	0,20

	n	p	R	R^2	R_a^2	RMSEC	RMSEP	slope
fit calibration Set	23	5	0,998	0,997	0,996	0,071	0,096	0,997
fit calibration Set	8	5	1,000	1,000	1,000	0,019		1,000
fit test set	15	5	0,991	0,982	0,975		0,136	0,945



Conclusions

- Reduced calibration set (8 components) has good prediction properties
- Selected subset is representative

Molecular descriptors from Dragon

- Dragon
  - Milano Chemometrics and QSAR Research Group
  - Developed for the calculation of molecular descriptors

Dragon
Molecular descriptors

Molecular Descriptors	*No.*	*Molecular Descriptors*	*No.*
Constitutional descriptors	48	Randic molecular profiles	41
Topological descriptors	119	Geometrical descriptors	74
Walk and path counts	47	RDF descriptors	150
Connectivity indices	33	3D-MoRSE descriptors	160
Information indices	47	WHIM descriptors	99
2D autocorrelations	96	GETAWAY descriptors	197
Edge adjacency indices	107	Functional group counts	121
Burden eigenvalues	64	Atom-centred fragments	120
Topological charge indices	21	Charge descriptors	14
Eigenvalue-based indices	44	Molecular properties	28

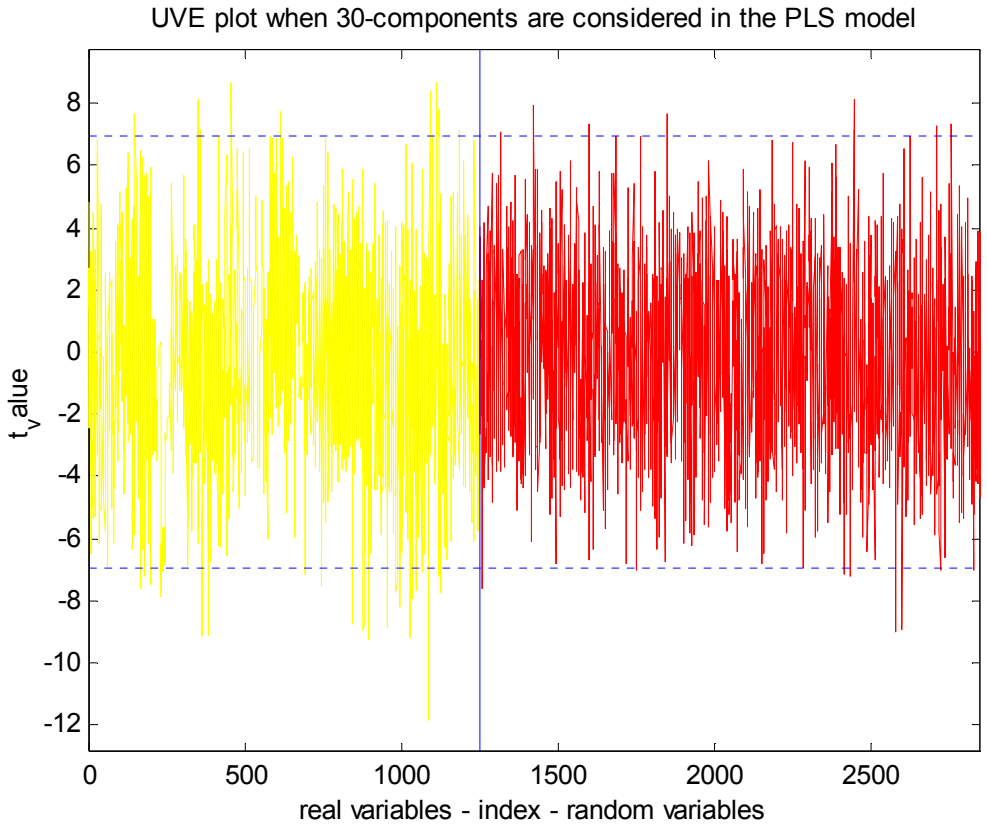
Example: Modelling log k_w

- 25 compounds on separation system 3
  - Column : Zorbax RX-C18
  - Supplier : Hewlett-Packard
  - Eluent : MeOH-buffer
  - Length : 150 mm
  - Int. Diameter : 4.6 mm

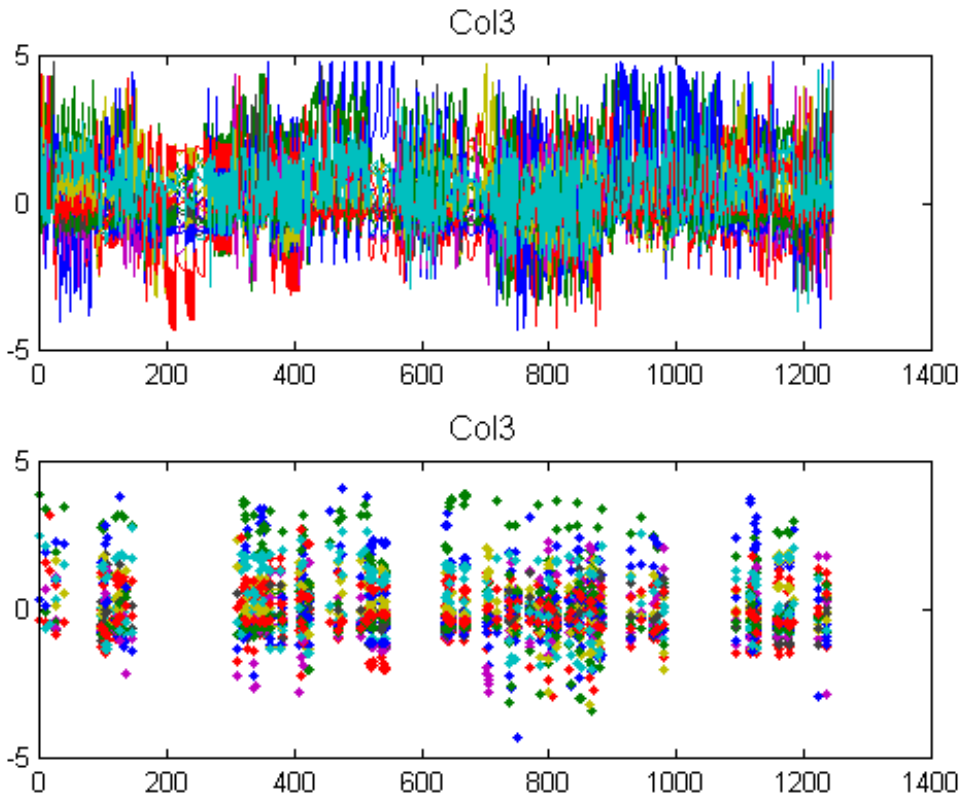
QSRR modelling with Dragon molecular descriptors

1. Input of 25 mdl-mol files of components in Dragon
2. Calculation of 1249 molecular descriptors in Dragon
3. Reduction of variables with UVE-PLS to less then 200 variables
4. Selection of variables with Genetic algoritm (GA)
5. Model building with stepwise MLR, PCR or PLS

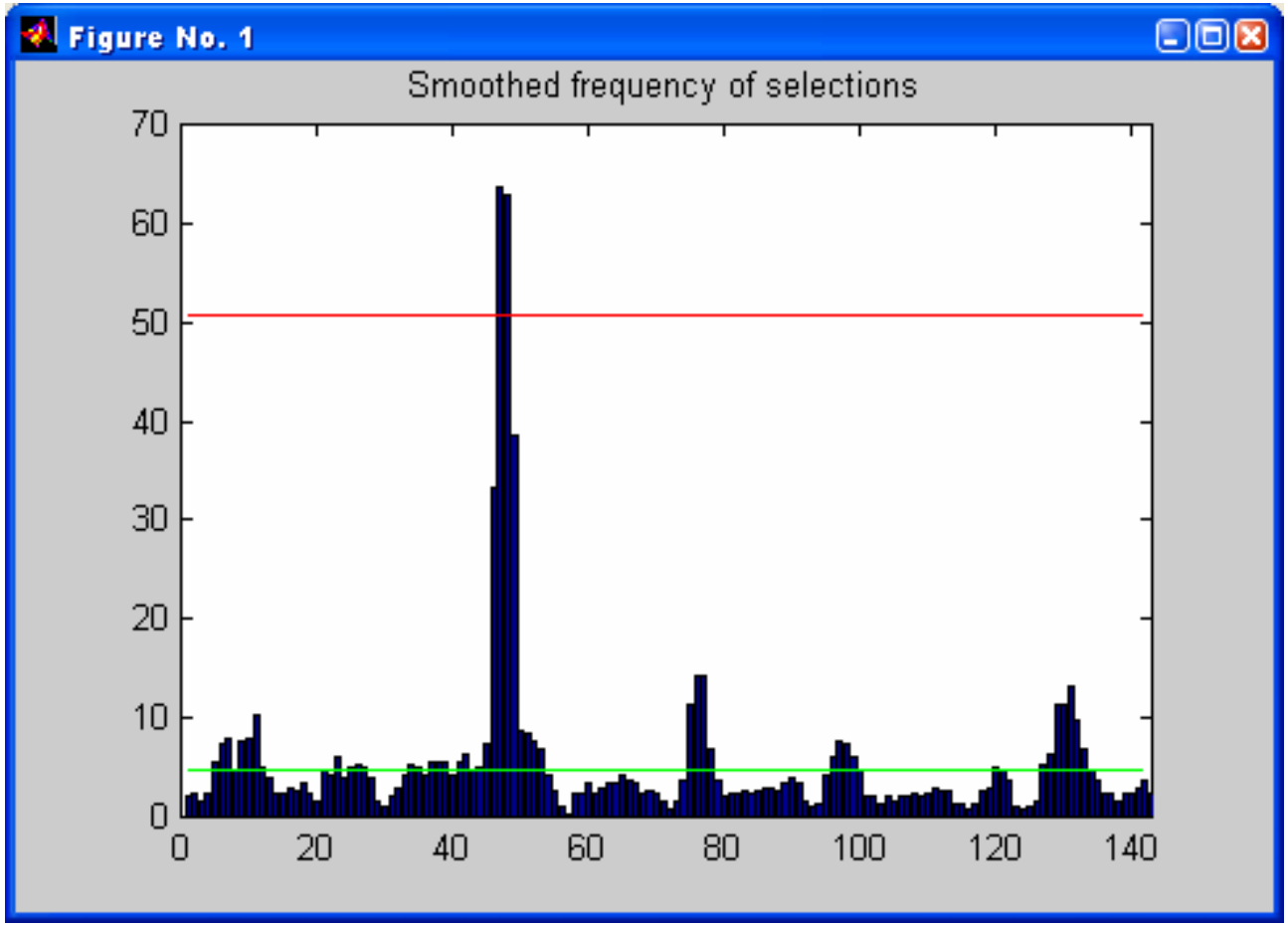
UVE-PLS: elimination of variables



UVE-PLS: elimination of variables



Selection of variables by GA



Most selected Dragon descriptors by GA for log k_w

- 412 MLOGP
- 413 MLOGP2
- 414 ALOGP
- 415 ALOGP2

Estimated MLR model

- Estimated model auto scaled data
  - R^2 =0.983
  - R_a^2 =0.981
  - n=25; s=0.140; F=620; sign(F)=0.000

$$\log k_w = 0.543 \cdot A\log P + 0.481 \cdot M\log P2$$

Columns selected with KS
http://www.pharm.kuleuven.ac.be