Extracting information from spectral data.

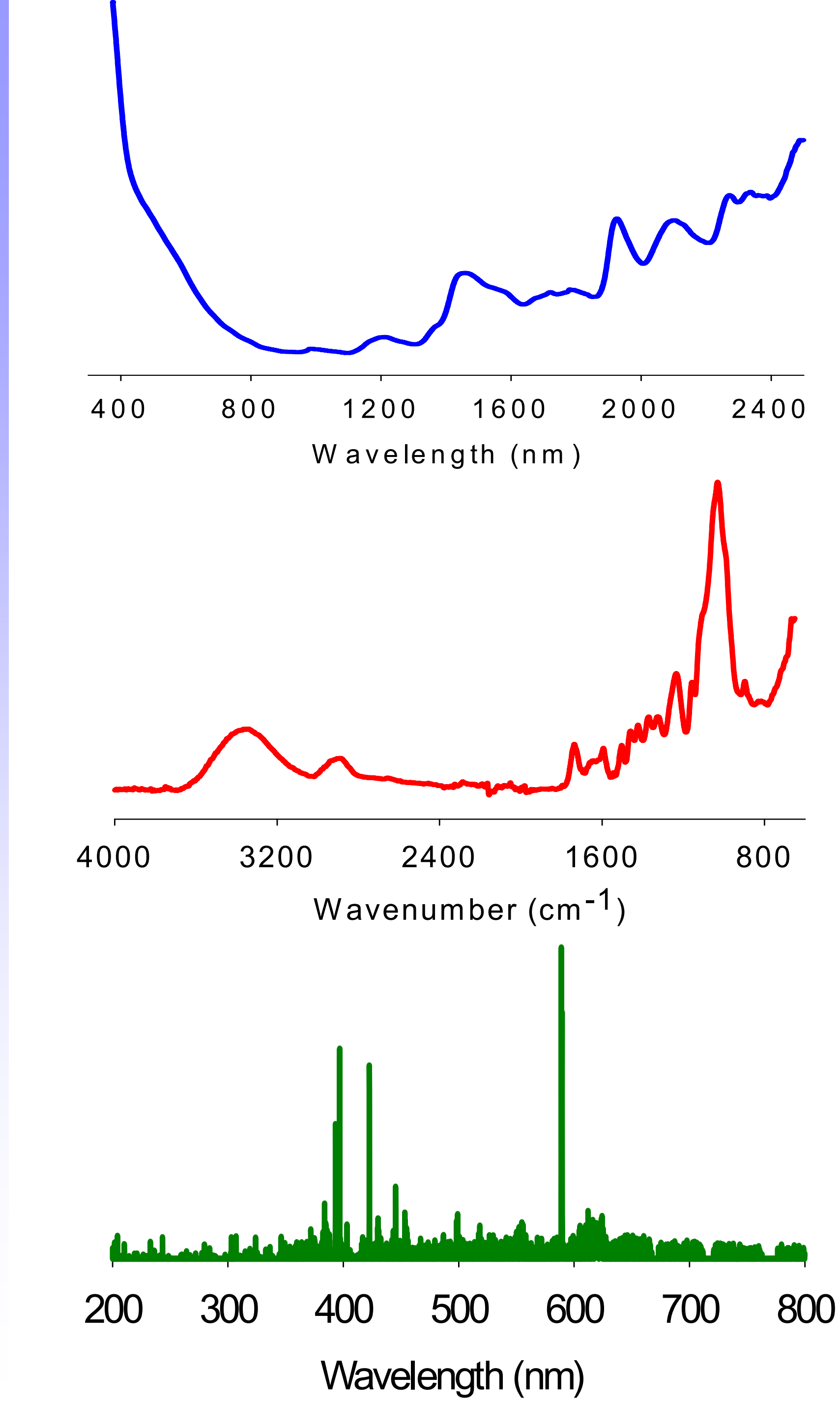
Nicole Labbé, University of Tennessee

Data collection

Near Infrared spectra
2150 data points, 350-2500 nm, 1 nm resolution, 8 scans

Mid Infrared spectra
3400 data points, 4000-600 cm-[1], 1 cm-[1] resolution, 4 scans

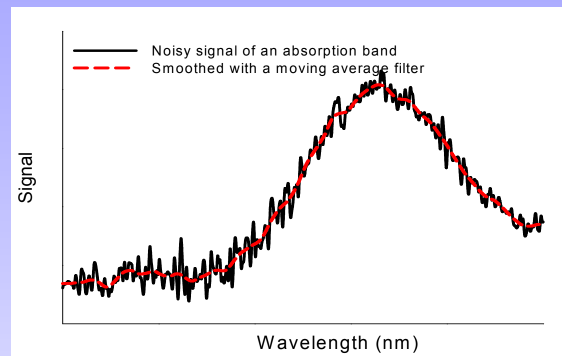
Laser Induced Breakdown spectra
30000 data points, 200-800 nm, 0.02 nm resolution, 10 scans



Prior to the extraction of the information.

Signal processing is used to transform spectral data prior to analysis

Data pretreatment



-Local filters

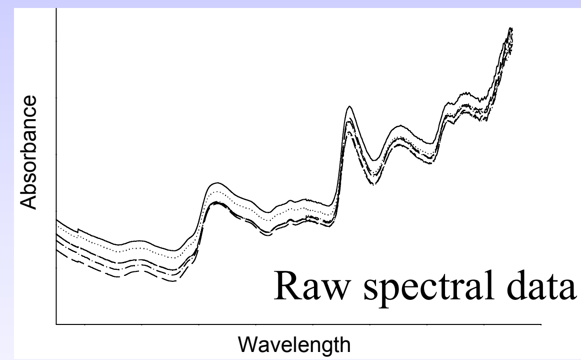
-Smoothing

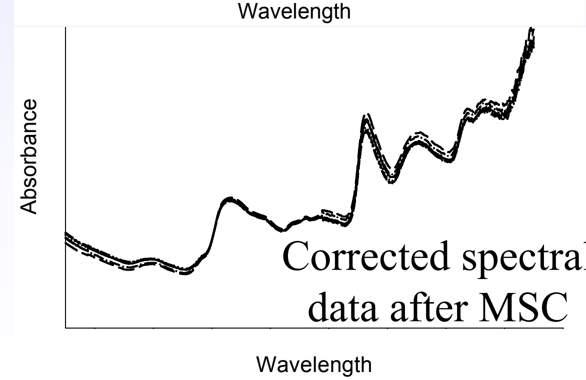
-Derivatives

-Baseline correction

-Multiplicative Scatter Correction (MSC)

-Orthogonal Scatter Correction (OSC)





What type of information?

1. Qualitative information

   Grouping and classification of spectral objects from samples into supervised and non-supervised learning methods.

2. Quantitative information

   Relationships between spectral data and parameter(s) of interest

How to extract the information?

1. Multivariate analysis (MVA)

   Principal Component Analysis (PCA), Projection to Latent Structures (PLS), PLS-Discriminant Analysis (PLS-DA), …

2. Two dimensional correlation spectroscopy

   Homo-correlation, Hetero-correlation

Multivariate data analysis

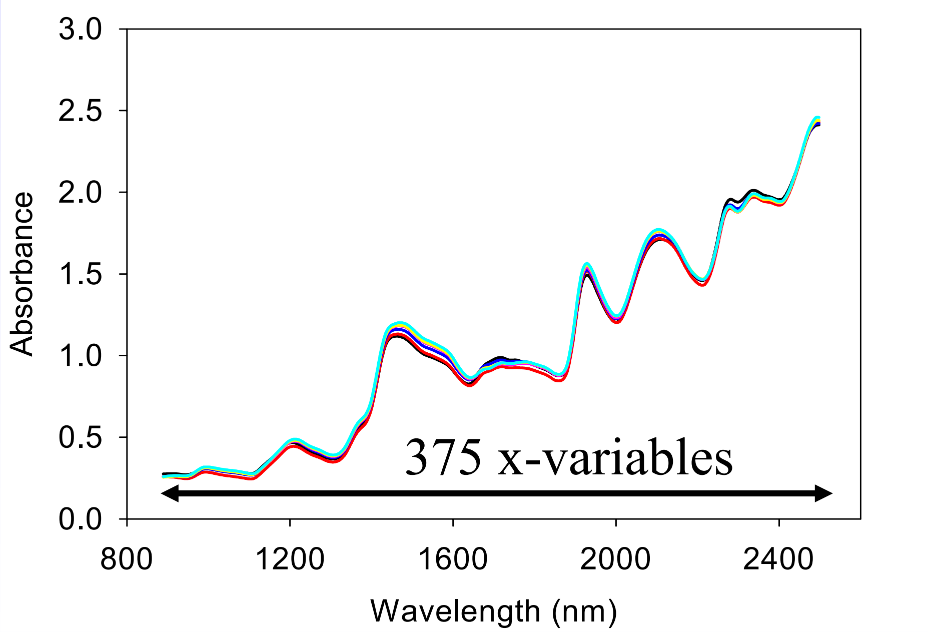
Separating the signal from the noise in data and presenting the results as easily interpretable plots.

Why are multivariate methods needed?
- -Large data sets
- -Problem of selectivity

Relationship between two variables: very simple, but does not always work

Many problems where several predictor variables used in combination give better results

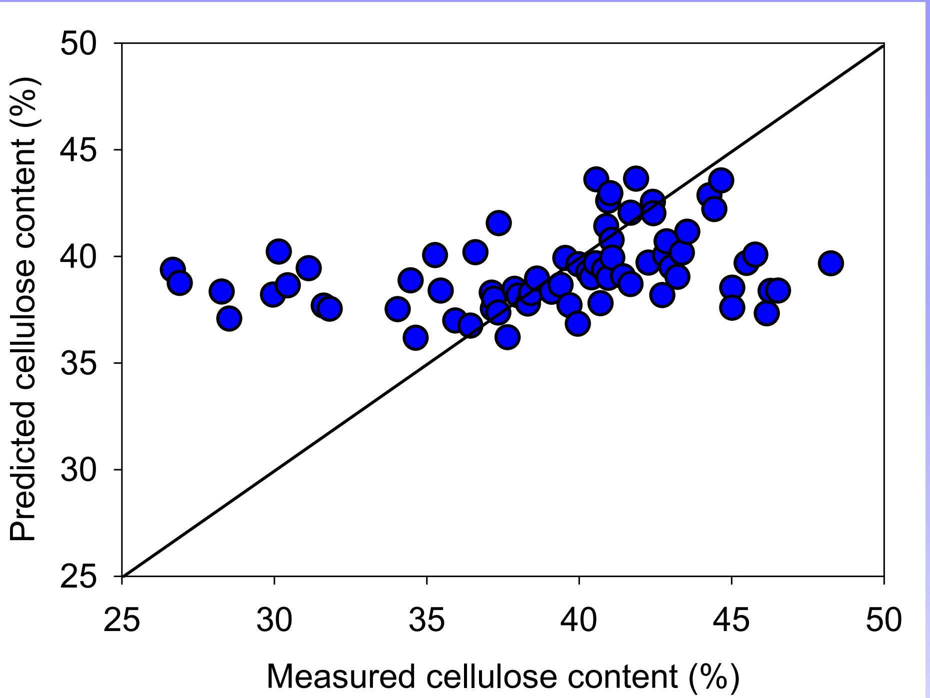
Two approaches:
- -Univariate analysis
- -Multivariate analysis



Near infrared spectra collected on 70 pine samples

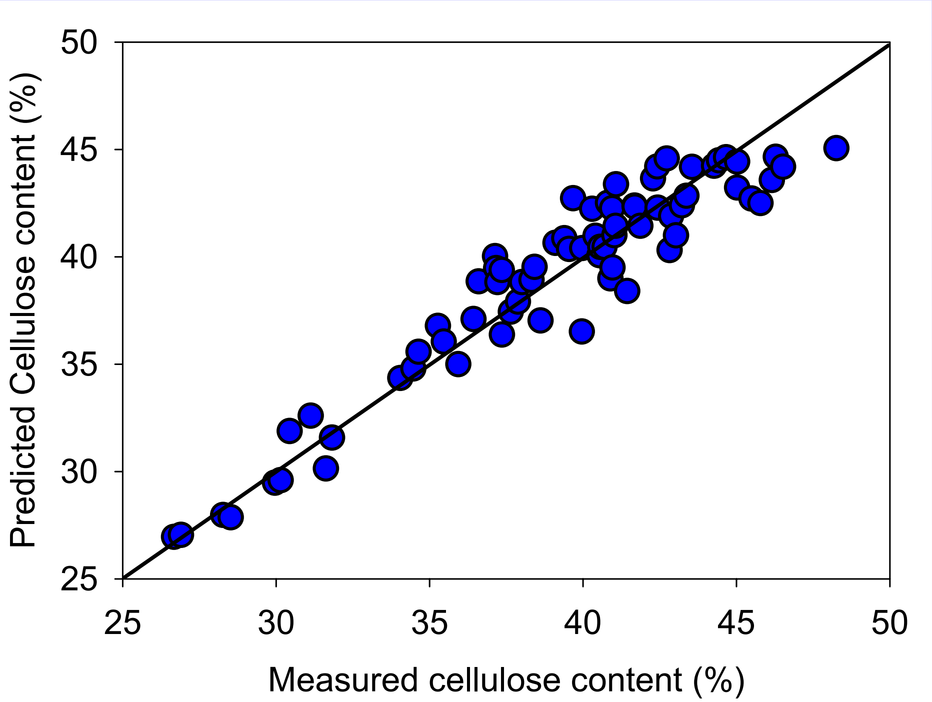
Univariate analysis

Measured cellulose content versus predicted cellulose content using one variable (1530 nm) as a predictor ($R^2 = 0.12$)



Multivariate analysis

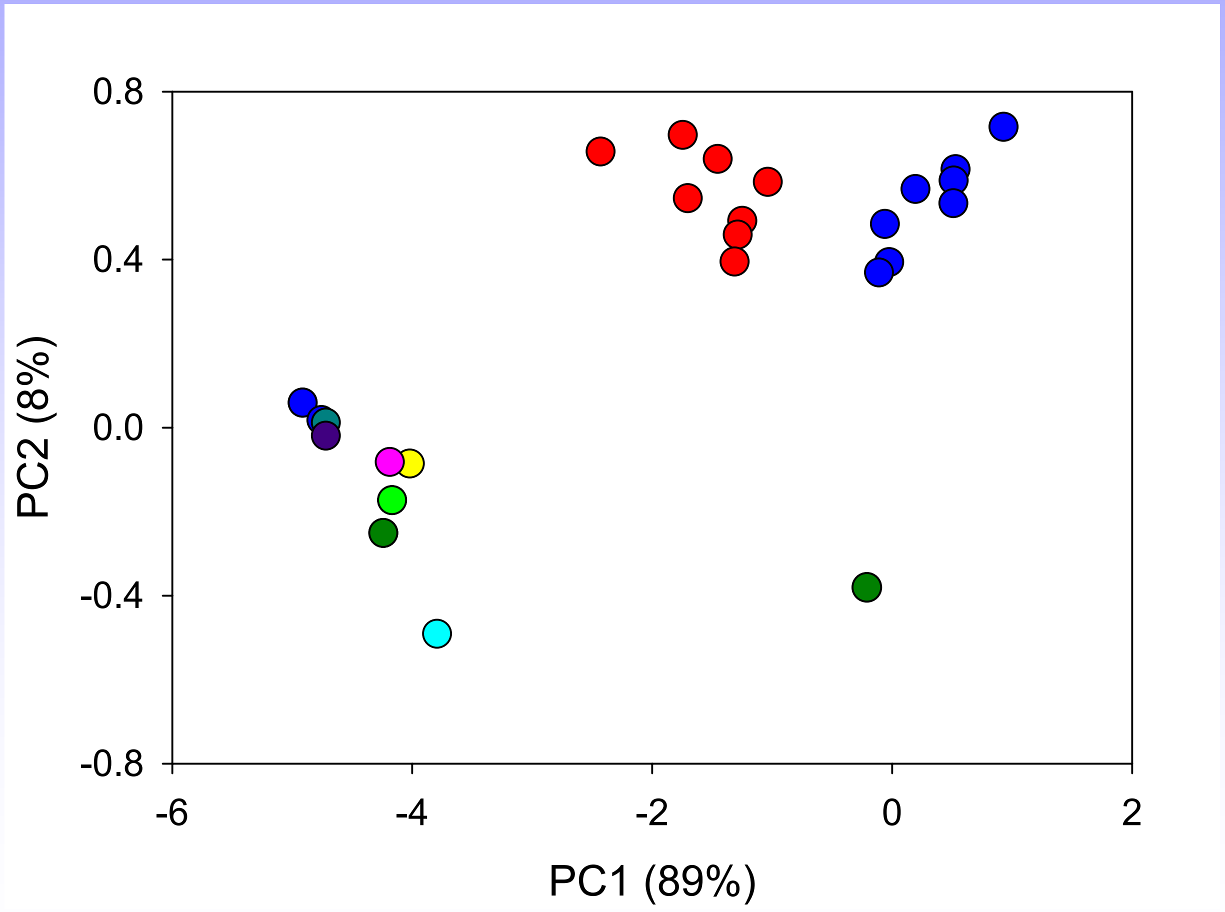
Measured cellulose content versus predicted cellulose content using multivariate method based on all variables from the spectral data ($R^2 = 0.87$)



Qualitative information

Principal Components Analysis (PCA)

Recognize patterns in data: outliers, trends, groups…



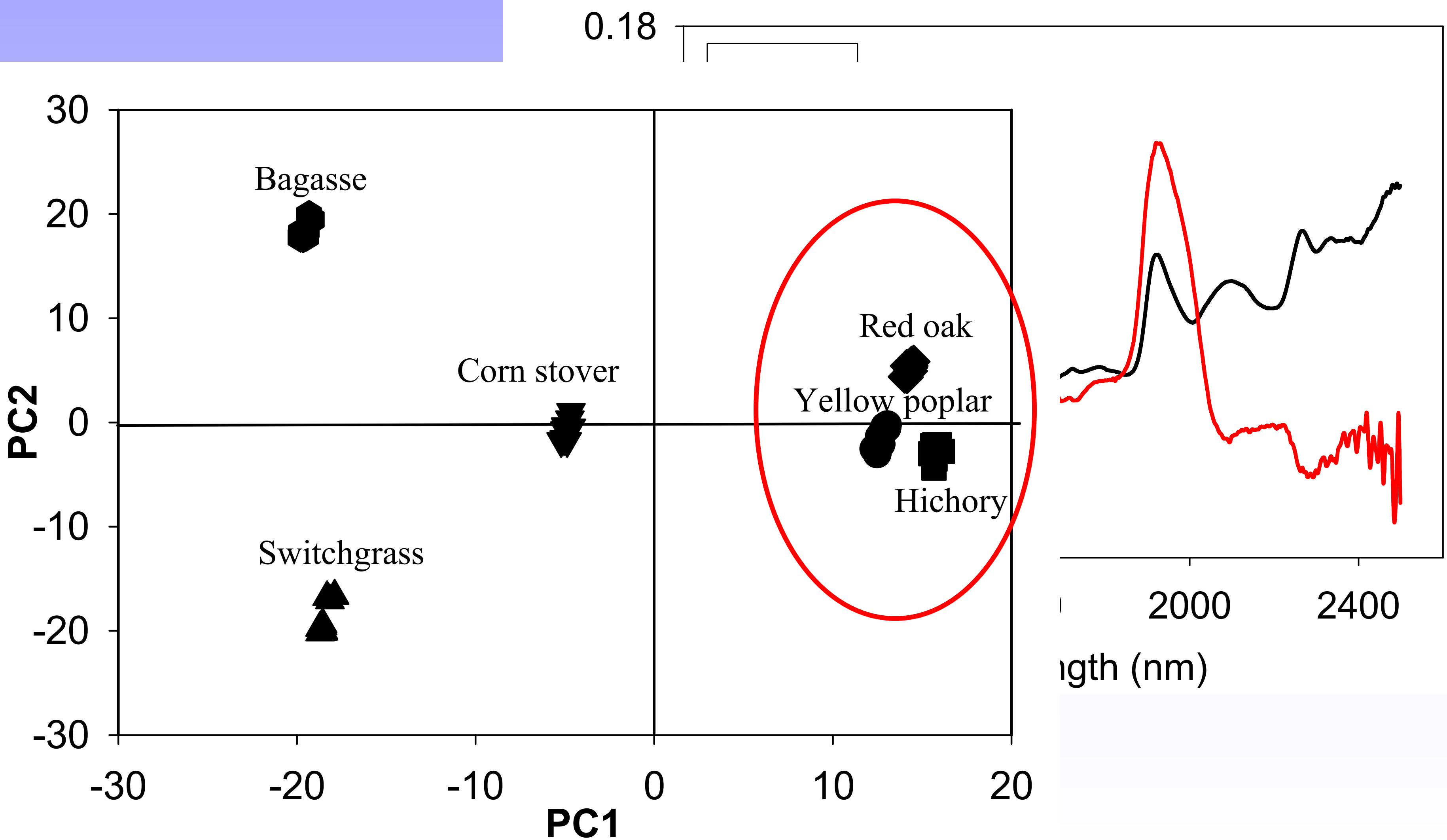
Biomass species and near infrared spectra

		355.0000	371.0000	387.0000	403.0000	419.0000	435.0000	451.0000	467.0000	483.0000	499.0000	515.0000	531.0000	547.0000	563.0000
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
yellow poplar	1	1.1167	1.0123	0.8762	0.7720	0.7169	0.6716	0.6003	0.5313	0.4847	0.4522	0.4256	0.4020	0.3804	0.3616
yellow poplar	2	1.1439	1.0495	0.9283	0.8443	0.8015	0.7591	0.6819	0.6071	0.5580	0.5221	0.4907	0.4623	0.4361	0.4129
yellow poplar	3	1.1666	1.0532	0.9043	0.7912	0.7327	0.6860	0.6118	0.5396	0.4911	0.4568	0.4280	0.4030	0.3804	0.3609
hickory	4	1.1209	1.0360	0.8866	0.7404	0.6379	0.5803	0.5456	0.5155	0.4898	0.4588	0.4264	0.3945	0.3612	0.3302
hickory	5	1.1453	1.0768	0.9599	0.8464	0.7629	0.7090	0.6710	0.6347	0.6018	0.5658	0.5282	0.4904	0.4517	0.4158
hickory	6	1.1168	1.0351	0.8851	0.7358	0.6310	0.5714	0.5352	0.5049	0.4788	0.4494	0.4186	0.3881	0.3568	0.3274
	7	1.3431	1.1793	0.9860	0.8484	0.7580	0.6920	0.6409	0.5981	0.5635	0.5328	0.5060	0.4825	0.4601	0.4394
	8	1.3868	1.2463	1.0623	0.9187	0.8176	0.7440	0.6880	0.6425	0.6063	0.5755	0.5491	0.5257	0.5031	0.4826
corn stover	9	1.4299	1.3007	1.1271	0.9895	0.8881	0.8132	0.7544	0.7059	0.6669	0.6334	0.6041	0.5778	0.5521	0.5277
switchgrass	10	1.3836	1.2323	1.0418	0.8948	0.7930	0.7194	0.6630	0.6248	0.5859	0.5475	0.5106	0.4838	0.4607	0.4420
switchgrass	11	1.3552	1.1965	0.9893	0.8326	0.7273	0.6575	0.6033	0.5651	0.5277	0.4902	0.4556	0.4293	0.4066	0.3877
switchgrass	12	1.4244	1.2840	1.0950	0.9490	0.8451	0.7700	0.7097	0.6680	0.6261	0.5834	0.5419	0.5111	0.4852	0.4643
bagasse	13	1.5106	1.3797	1.2624	1.1756	1.1140	1.0638	1.0189	0.9756	0.9349	0.8956	0.8577	0.8208	0.7844	0.7498
bagasse	14	1.5682	1.4386	1.3244	1.2375	1.1777	1.1281	1.0836	1.0402	0.9995	0.9601	0.9224	0.8859	0.8497	0.8152
bagasse	15	1.5556	1.4496	1.3407	1.2537	1.1943	1.1432	1.0970	1.0523	1.0103	0.9692	0.9299	0.8918	0.8542	0.8180
red oak	16	1.0661	0.9793	0.8138	0.6677	0.5873	0.5427	0.5120	0.4853	0.4594	0.4320	0.4024	0.3709	0.3385	0.3088
red oak	17	1.2186	1.1401	0.9805	0.8188	0.7160	0.6688	0.6449	0.6254	0.6055	0.5813	0.5509	0.5140	0.4711	0.4297
red oak	18	1.2165	1.1402	0.9858	0.8416	0.7542	0.7092	0.6813	0.6562	0.6310	0.6016	0.5673	0.5285	0.4858	0.4445

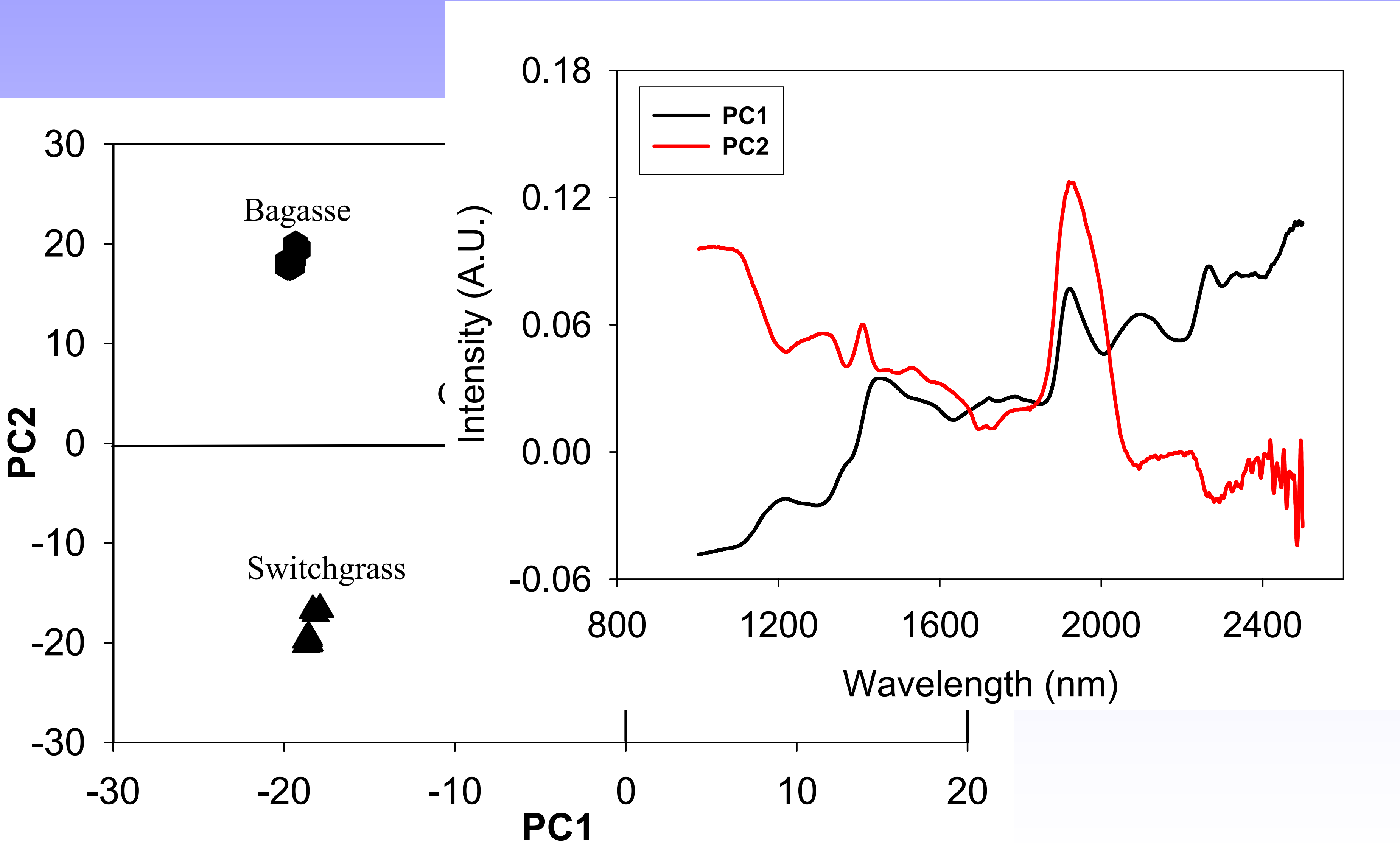
n Samples

Spectral data
x variables

Transformation of the numbers into pictures by using principal component analysis (scores)



Transformation of the numbers into pictures by using principal component analysis (loadings)

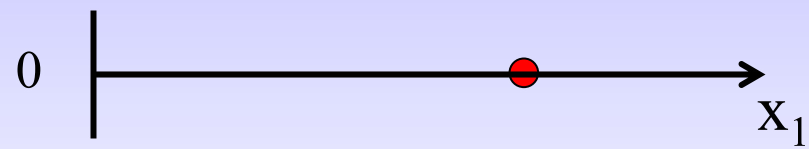


Principal Components Analysis (PCA)

PCA is a projection method, it decomposes the spectral data into a "structure" part and a "noise" part

X is an n samples (observations) by x variables (spectral variables) matrix

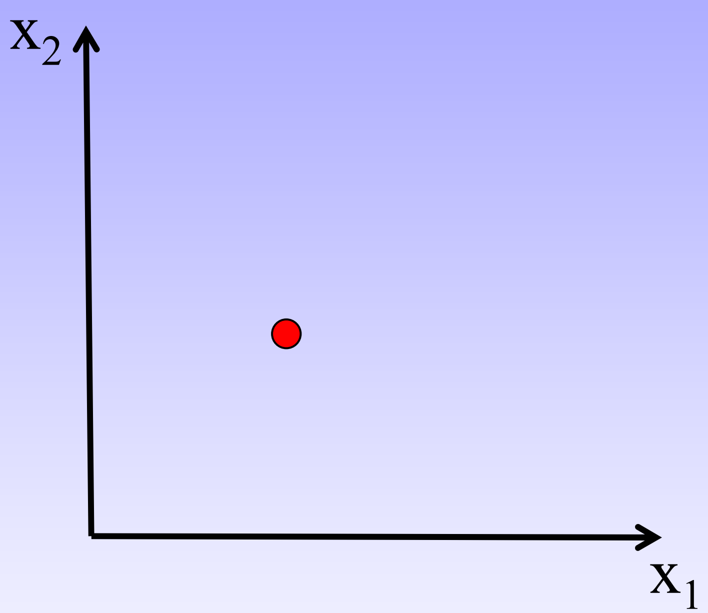
		355.0000	371.0000	387.0000	403.0000	419.0000	435.0000	451.0000	467.0000	483.0000	499.0000	515.0000	531.0000	547.0000	563.0000
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
yellow poplar	1	1.1167	1.0123	0.8762	0.7720	0.7169	0.6716	0.6003	0.5313	0.4847	0.4522	0.4256	0.4020	0.3804	0.3616
yellow paplar	2	1.1439	1.0495	0.9283	0.8443	0.8015	0.7591	0.6819	0.6071	0.5580	0.5221	0.4907	0.4623	0.4361	0.4129
yellow paplar	3	1.1666	1.0532	0.9043	0.7912	0.7327	0.6860	0.6118	0.5396	0.4911	0.4568	0.4280	0.4030	0.3804	0.3609
hichory	4	1.1209	1.0360	0.8866	0.7404	0.6379	0.5803	0.5456	0.5155	0.4898	0.4588	0.4264	0.3945	0.3612	0.3302
hichory	5	1.1453	1.0768	0.9599	0.8464	0.7629	0.7090	0.6710	0.6347	0.6018	0.5658	0.5282	0.4904	0.4517	0.4158
hichory	6	1.1168	1.0351	0.8851	0.7358	0.6310	0.5714	0.5352	0.5049	0.4788	0.4494	0.4186	0.3881	0.3568	0.3274
corn stover	7	1.3431	1.1793	0.9860	0.8484	0.7580	0.6920	0.6409	0.5981	0.5635	0.5328	0.5060	0.4825	0.4601	0.4394
corn stover	8	1.3868	1.2463	1.0623	0.9187	0.8176	0.7440	0.6880	0.6425	0.6063	0.5755	0.5491	0.5257	0.5031	0.4826
corn stover	9	1.4299	1.3007	1.1271	0.9895	0.8881	0.8132	0.7544	0.7059	0.6669	0.6334	0.6041	0.5778	0.5521	0.5277
switchgrass	10	1.3836	1.2323	1.0418	0.8948	0.7930	0.7194	0.6630	0.6248	0.5859	0.5475	0.5106	0.4838	0.4607	0.4420
switchgrass	11	1.3552	1.1965	0.9893	0.8326	0.7273	0.6575	0.6033	0.5651	0.5277	0.4902	0.4556	0.4293	0.4066	0.3877
switchgrass	12	1.4244	1.2840	1.0950	0.9490	0.8451	0.7700	0.7097	0.6680	0.6261	0.5834	0.5419	0.5111	0.4852	0.4643
bagasse	13	1.5106	1.3797	1.2624	1.1756	1.1140	1.0638	1.0189	0.9756	0.9349	0.8956	0.8577	0.8208	0.7844	0.7498
bagasse	14	1.5682	1.4386	1.3244	1.2375	1.1777	1.1281	1.0836	1.0402	0.9995	0.9601	0.9224	0.8859	0.8497	0.8152
bagasse	15	1.5556	1.4496	1.3407	1.2537	1.1943	1.1432	1.0970	1.0523	1.0103	0.9692	0.9299	0.8918	0.8542	0.8180
red oak	16	1.0661	0.9793	0.8138	0.6677	0.5873	0.5427	0.5120	0.4853	0.4594	0.4320	0.4024	0.3709	0.3385	0.3088
red oak	17	1.2186	1.1401	0.9805	0.8188	0.7160	0.6688	0.6449	0.6254	0.6055	0.5813	0.5509	0.5140	0.4711	0.4297
red oak	18	1.2165	1.1402	0.9858	0.8416	0.7542	0.7092	0.6813	0.6562	0.6310	0.6016	0.5673	0.5285	0.4858	0.4445

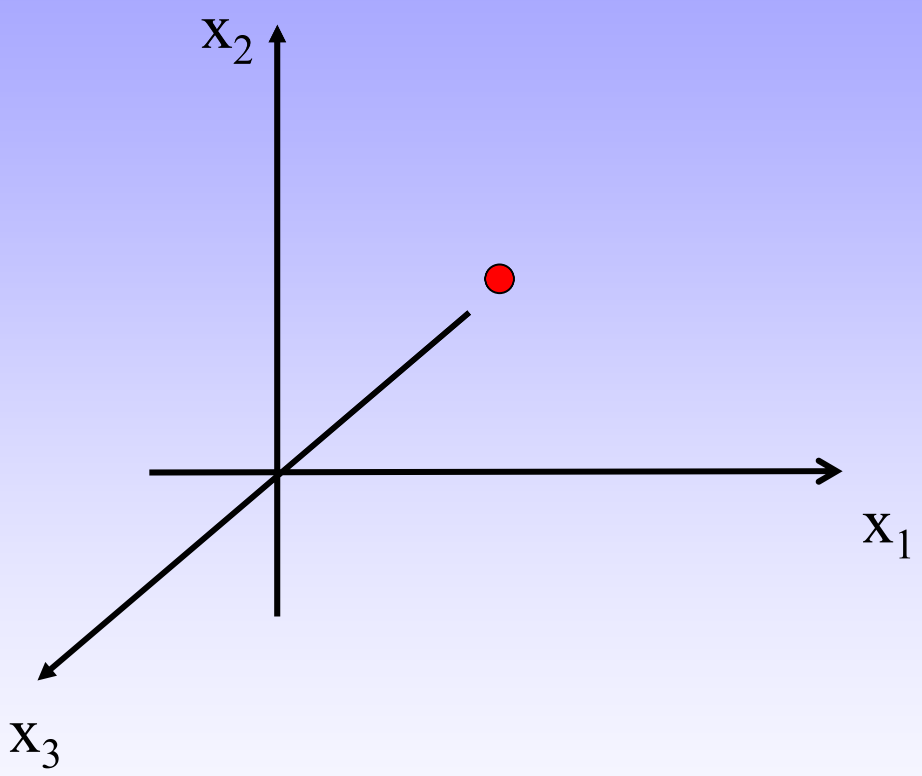


1 variable = 1 dimension

Principal Components Analysis (PCA)

X is an n samples (observations) by x variables (spectral variables) matrix



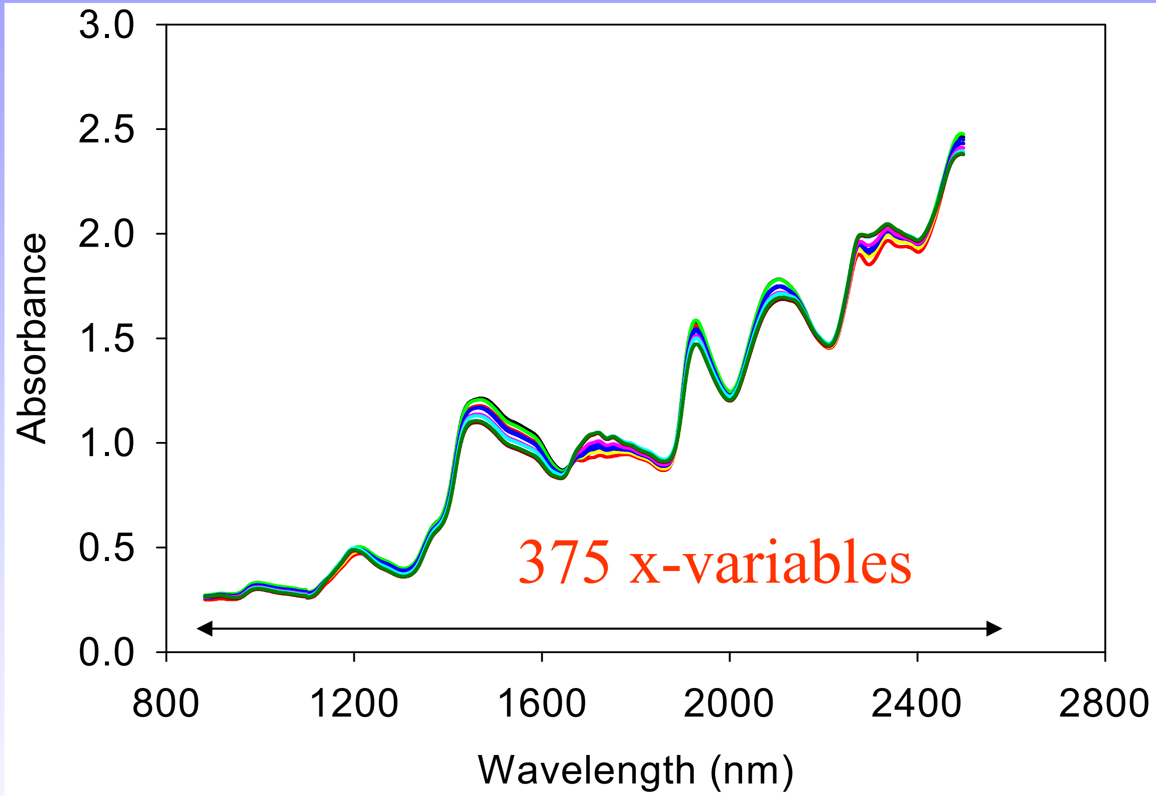


2 variables = 2 dimensions

3 variables = 3 dimensional space

Principal Components Analysis (PCA)

Beyond 3 dimensions, it is very difficult to visualize what's going on.



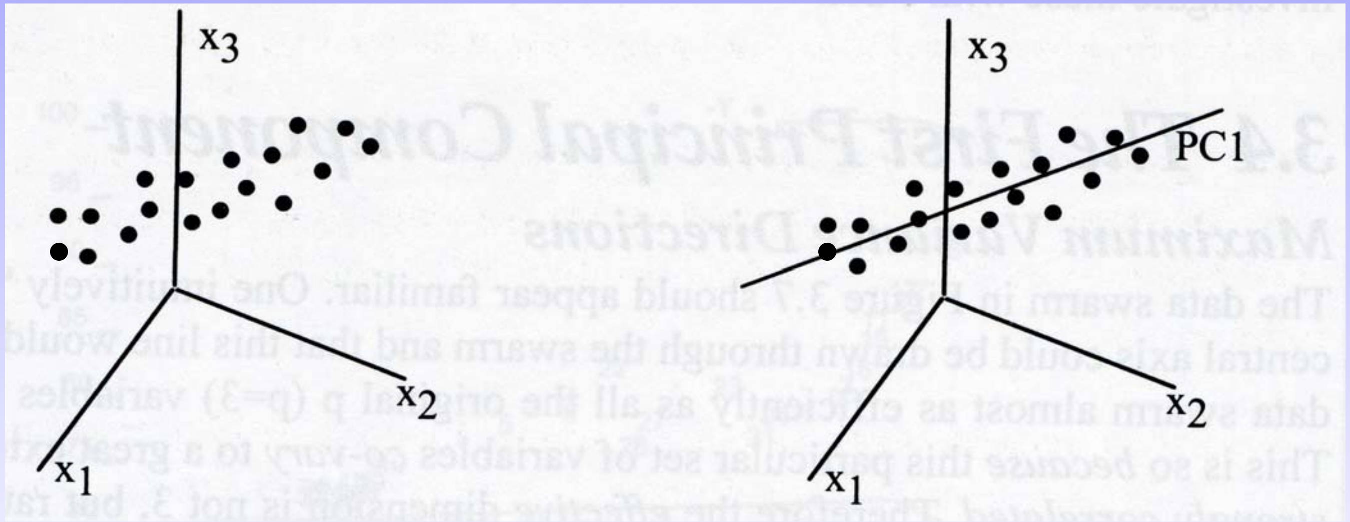
18 Near infrared spectra of wood samples

Each sample is represented as a co-ordinate axis in 375-dimensional space

Principal Components Analysis (PCA)

X has only 3 variables (wavelengths x1, x2 and x3)

The sample (n = 18) are represented in a 3D space



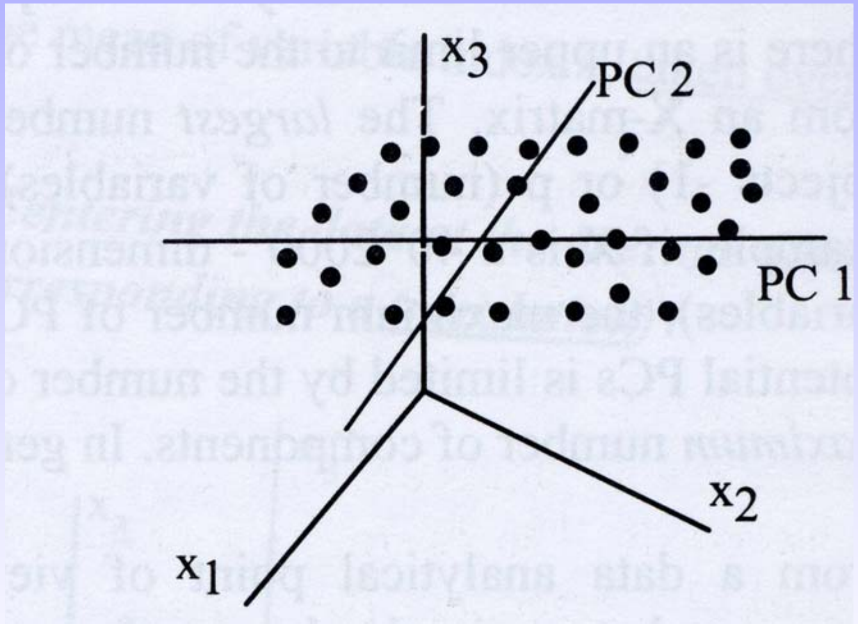
<u>The first principal component</u>

New co-ordinate axis representing the direction of maximum variation through the data.

Higher-order Principal Components (PC2, PC3,…)

After PC1, next best direction for approximating the original data

The second PC lies along a direction orthogonal to the first PC



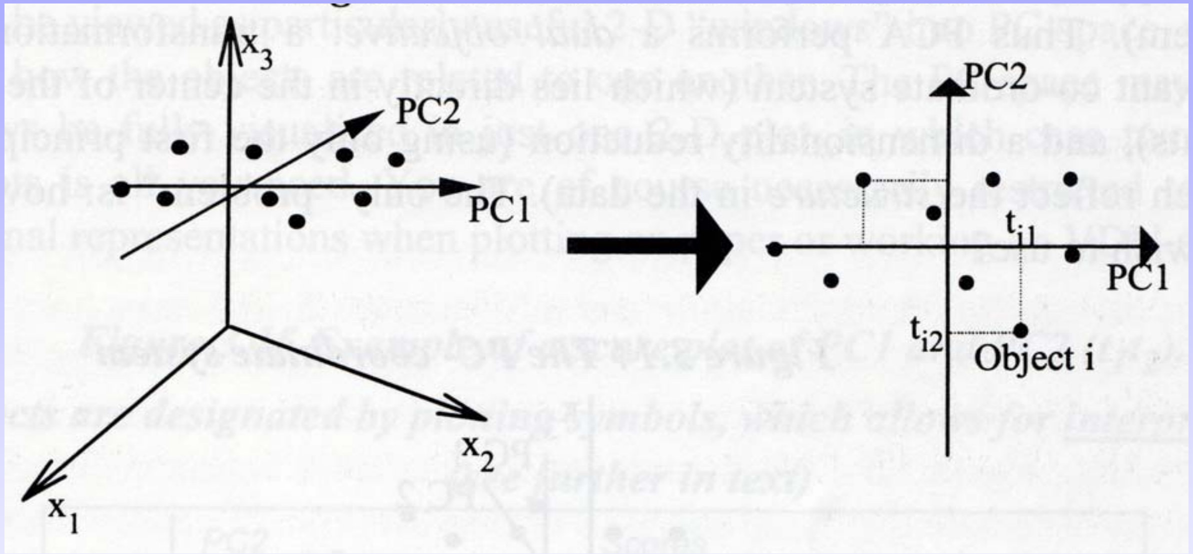
PC3 will be orthogonal to both PC1 and PC2 while simultaneously lying along the direction of the third largest variation.

The new variables (PCs) are uncorrelated with each other (orthogonal)

Scores (T) = Coordinates of samples in the PC space

Representation of the samples in the PC space
There is a set of scores for each PC (score vector)

Original variable space



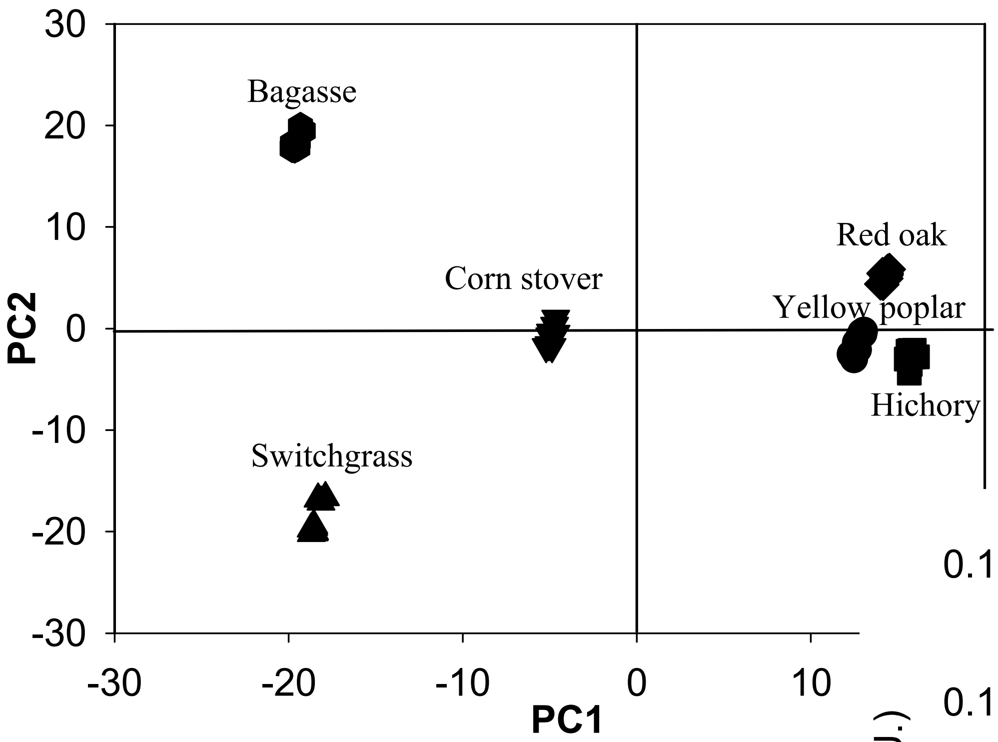
PC space

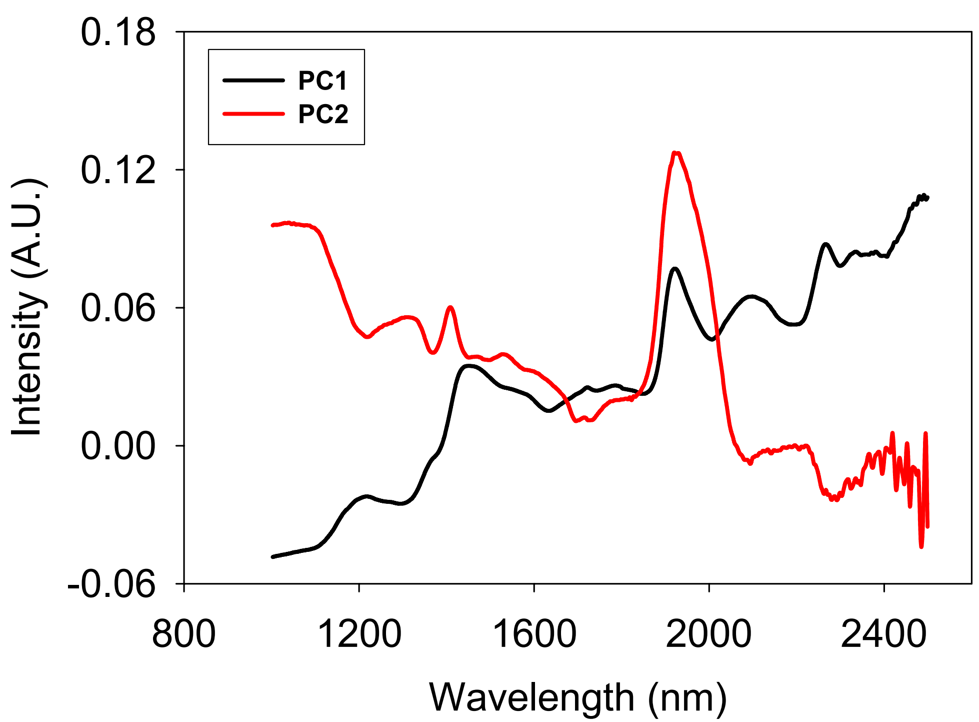
Loadings (P) = Relations between X and PCs

Relationship between the original variable space and the new PCs space

There is a set of loadings for each PC (loading vector)

Transformation of the numbers into pictures by using PCA

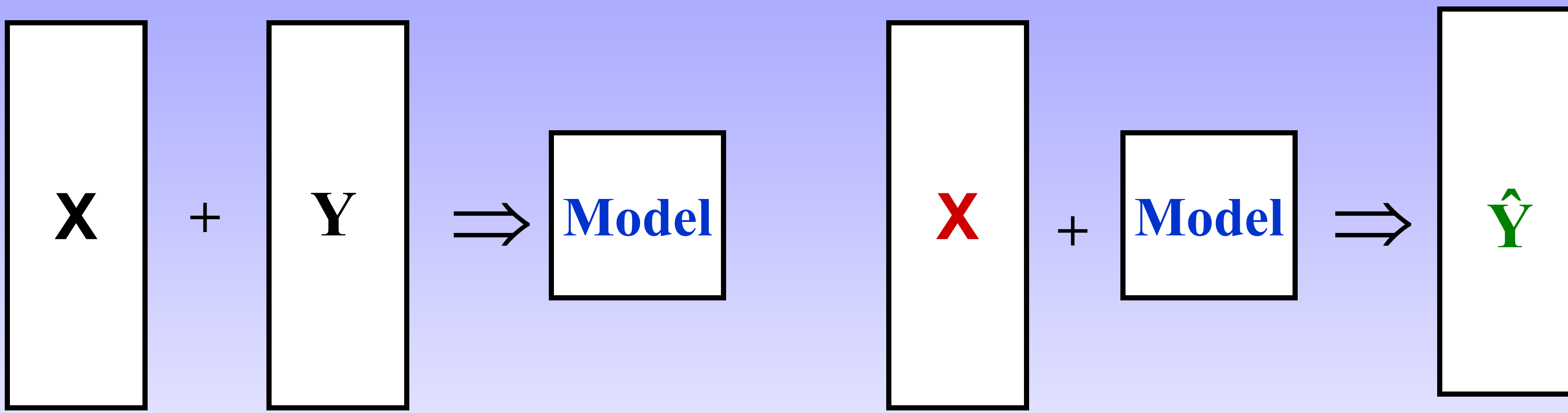




Quantitative information

Projection to Latent Structures or Partial Least Squares Regression (PLS)

Establish relationships between input and output variables, creating predictive models.



Establishing calibration model from known X and Y data

Using calibration model to predict new Y-values from new set of X-measurement

PLS can be seen as two interconnected PCA analyses, PCA(**X**) and PCA(**Y**)

PLS uses the Y-data structure (variation) to guide the decomposition of **X**

The X-data structure also influences the decomposition of **Y**

Biomass composition and near infrared spectra

		cellulose	Hemis	lignin	extractives	400	402	404	406	408	410	412	414	416	418	420
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2-1 A-2	1	26.6700	22.0000	22.1600	26.9000	0.5305	0.5479	0.5643	0.5769	0.5854	0.5929	0.5993	0.6031	0.6049	0.6063	0.6065
3-1 A-2	2	26.9000	21.8000	23.0900	25.9000	0.5316	0.5493	0.5658	0.5781	0.5867	0.5949	0.6022	0.6064	0.6084	0.6101	0.6108
2-5 A-2	3	28.2700	22.1000	23.5100	23.5000	0.5333	0.5507	0.5673	0.5801	0.5885	0.5956	0.6021	0.6059	0.6073	0.6084	0.6087
3-2 A-2	4	28.5200	20.1000	23.8900	26.2000	0.5105	0.5283	0.5431	0.5541	0.5627	0.5696	0.5736	0.5761	0.5783	0.5793	0.5781
2-2 A-2	5	29.9600	20.5000	27.1400	18.5000	0.5242	0.5407	0.5555	0.5663	0.5735	0.5795	0.5836	0.5858	0.5867	0.5869	0.5855
2-1 A-5	6	30.1500	22.4000	24.5200	20.9000	0.5232	0.5408	0.5567	0.5684	0.5763	0.5831	0.5887	0.5922	0.5938	0.5946	0.5943
1-[illegible]	7	30.4400	23.4000	26.9000	16.3000	0.5347	0.5517	0.5682	0.5817	0.5906	0.5980	0.6050	0.6099	0.6117	0.6128	0.6136
3-	8	31.1300	21.1000	25.7500	17.7000	0.5235	0.5411	0.5568	0.5680	0.5759	0.5832	0.5887	0.5914	0.5926	0.5937	0.5934
2-	9	31.6200	23.1000	25.5300	18.3000	0.5280	0.5454	0.5619	0.5741	0.5821	0.5893	0.5955	0.5986	0.5998	0.6010	0.6012
2-	10	31.8200	23.3000	26.9400	16.0000	0.5235	0.5407	0.5562	0.5674	0.5753	0.5824	0.5876	0.5899	0.5910	0.5920	0.5917
1-	11	34.0500	24.7000	29.5400	9.9000	0.5214	0.5389	0.5548	0.5666	0.5749	0.5823	0.5879	0.5907	0.5922	0.5936	0.5935
3-	12	34.4700	22.9000	27.0400	13.2000	0.5210	0.5385	0.5539	0.5650	0.5728	0.5795	0.5841	0.5865	0.5880	0.5888	0.5878
3-	13	34.6400	24.8000	27.7000	10.0000	0.5069	0.5243	0.5387	0.5493	0.5571	0.5629	0.5667	0.5693	0.5708	0.5709	0.5697
1-[illegible]	14	35.2700	23.7000	28.1200	9.9000	0.5252	0.5423	0.5585	0.5708	0.5788	0.5855	0.5909	0.5940	0.5952	0.5959	0.5954
1-2 A-2	15	35.4600	23.8000	29.7400	8.8000	0.5191	0.5368	0.5525	0.5638	0.5714	0.5779	0.5825	0.5848	0.5859	0.5863	0.5852
1-3 A-2	16	35.9400	24.7000	28.5800	8.1000	0.5073	0.5239	0.5377	0.5476	0.5543	0.5590	0.5621	0.5643	0.5649	0.5637	0.5618
1-5 A-2	17	36.4300	23.6000	29.4200	6.9000	0.4869	0.5014	0.5130	0.5215	0.5258	0.5274	0.5284	0.5283	0.5263	0.5235	0.5202
1-1 A-9	18	36.6000	23.4000	28.8500	8.2000	0.5182	0.5354	0.5512	0.5627	0.5704	0.5767	0.5811	0.5836	0.5851	0.5857	0.5847
2-3 A-5	19	37.1300	23.5000	27.6100	9.4000	0.5035	0.5209	0.5349	0.5454	0.5531	0.5582	0.5610	0.5629	0.5636	0.5632	0.5618
3-2 A-5	20	37.1700	22.8000	27.5500	11.2000	0.4860	0.5011	0.5138	0.5237	0.5291	0.5318	0.5343	0.5358	0.5354	0.5347	0.5337
2-2 A-5	21	37.2100	23.1000	27.0900	10.6000	0.4924	0.5079	0.5204	0.5296	0.5349	0.5377	0.5403	0.5421	0.5417	0.5403	0.5382
3-4 A-5	22	37.3500	22.3000	27.4400	9.8000	0.4906	0.5064	0.5192	0.5289	0.5345	0.5376	0.5404	0.5425	0.5424	0.5411	0.5391
1-1 A-32	23	37.3600	23.8000	33.2000	4.6000	0.5425	0.5597	0.5763	0.5907	0.6006	0.6079	0.6151	0.6210	0.6238	0.6250	0.6260
3-5 A-2	24	37.6500	26.5000	27.8300	8.9000	0.5001	0.5171	0.5307	0.5407	0.5475	0.5520	0.5555	0.5584	0.5593	0.5584	0.5568
3-1 A-32	25	37.8800	23.7000	33.0000	5.8000	0.5157	0.5340	0.5498	0.5611	0.5691	0.5760	0.5811	0.5844	0.5864	0.5876	0.5874
1-4 A-5	26	38.0100	24.2000	28.9300	4.7000	0.4979	0.5140	0.5271	0.5373	0.5442	0.5481	0.5499	0.5510	0.5509	0.5498	0.5483

Samples

y variables

Spectral data x variables

If y variables are not correlated $\Rightarrow$ PLS1

Biomass composition and near infrared spectra

Samples

		cellulose	Hemis	lignin	extractives	400	402	404	406	408	410	412	414	416	418	420
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2-1 A-2	1	26.6700	22.0000	22.1600	26.9000	0.5305	0.5479	0.5643	0.5769	0.5854	0.5929	0.5993	0.6031	0.6049	0.6063	0.6065
3-1 A-2	2	26.9000	21.8000	23.0900	25.9000	0.5316	0.5493	0.5658	0.5781	0.5867	0.5949	0.6022	0.6064	0.6084	0.6101	0.6108
2-5 A-2	3	28.2700	22.1000	23.5100	23.5000	0.5333	0.5507	0.5673	0.5801	0.5885	0.5956	0.6021	0.6059	0.6073	0.6084	0.6087
3-2 A-2	4	28.5200	20.1000	23.8900	26.2000	0.5105	0.5283	0.5431	0.5541	0.5627	0.5696	0.5736	0.5761	0.5783	0.5793	0.5781
2-2 A-2	5	29.9600	20.5000	27.1400	18.5000	0.5242	0.5407	0.5555	0.5663	0.5735	0.5795	0.5836	0.5858	0.5867	0.5869	0.5855
2-1 A-5	6	30.1500	22.4000	24.5200	20.9000	0.5232	0.5408	0.5567	0.5684	0.5763	0.5831	0.5887	0.5922	0.5938	0.5946	0.5943
1-	7	30.4400	23.4000	26.9000	16.3000	0.5347	0.5517	0.5682	0.5817	0.5906	0.5980	0.6050	0.6099	0.6117	0.6128	0.6136
3-	8	31.1300	21.1000	25.7500	17.7000	0.5235	0.5411	0.5568	0.5680	0.5759	0.5832	0.5887	0.5914	0.5926	0.5937	0.5934
2-	9	31.6200	23.1000	25.5300	18.3000	0.5280	0.5454	0.5619	0.5741	0.5821	0.5893	0.5955	0.5986	0.5998	0.6010	0.6012
2-	10	31.8200	23.3000	26.9400	16.0000	0.5235	0.5407	0.5562	0.5674	0.5753	0.5824	0.5876	0.5899	0.5910	0.5920	0.5917
1-	11	34.0500	24.7000	29.5400	9.9000	0.5214	0.5389	0.5548	0.5666	0.5749	0.5823	0.5879	0.5907	0.5922	0.5936	0.5935
3-	12	34.4700	22.9000	27.0400	13.2000	0.5210	0.5385	0.5539	0.5650	0.5728	0.5795	0.5841	0.5865	0.5880	0.5888	0.5878
3-	13	34.6400	24.8000	27.7000	10.0000	0.5069	0.5243	0.5387	0.5493	0.5571	0.5629	0.5667	0.5693	0.5708	0.5709	0.5697
1-	14	35.2700	23.7000	28.1200	9.9000	0.5252	0.5423	0.5585	0.5708	0.5788	0.5855	0.5909	0.5940	0.5952	0.5959	0.5954
1-2 A-2	15	35.4600	23.8000	29.7400	8.8000	0.5191	0.5368	0.5525	0.5638	0.5714	0.5779	0.5825	0.5848	0.5859	0.5863	0.5852
1-3 A-2	16	35.9400	24.7000	28.5800	8.1000	0.5073	0.5239	0.5377	0.5476	0.5543	0.5590	0.5621	0.5643	0.5649	0.5637	0.5618
1-5 A-2	17	36.4300	23.6000	29.4200	6.9000	0.4869	0.5014	0.5130	0.5215	0.5258	0.5274	0.5284	0.5283	0.5263	0.5235	0.5202
1-1 A-9	18	36.6000	23.4000	28.8500	8.2000	0.5182	0.5354	0.5512	0.5627	0.5704	0.5767	0.5811	0.5836	0.5851	0.5857	0.5847
2-3 A-5	19	37.1300	23.5000	27.6100	9.4000	0.5035	0.5209	0.5349	0.5454	0.5531	0.5582	0.5610	0.5629	0.5636	0.5632	0.5618
3-2 A-5	20	37.1700	22.8000	27.5500	11.2000	0.4860	0.5011	0.5138	0.5237	0.5291	0.5318	0.5343	0.5358	0.5354	0.5347	0.5337
2-2 A-5	21	37.2100	23.1000	27.0900	10.6000	0.4924	0.5079	0.5204	0.5296	0.5349	0.5377	0.5403	0.5421	0.5417	0.5403	0.5382
3-4 A-5	22	37.3500	22.3000	27.4400	9.8000	0.4906	0.5064	0.5192	0.5289	0.5345	0.5376	0.5404	0.5425	0.5424	0.5411	0.5391
1-1 A-32	23	37.3600	23.8000	33.2000	4.6000	0.5425	0.5597	0.5763	0.5907	0.6006	0.6079	0.6151	0.6210	0.6238	0.6250	0.6260
3-5 A-2	24	37.6500	26.5000	27.8300	8.9000	0.5001	0.5171	0.5307	0.5407	0.5475	0.5520	0.5555	0.5584	0.5593	0.5584	0.5568
3-1 A-32	25	37.8800	23.7000	33.0000	5.8000	0.5157	0.5340	0.5498	0.5611	0.5691	0.5760	0.5811	0.5844	0.5864	0.5876	0.5874
1-4 A-5	26	38.0100	24.2000	28.9300	4.7000	0.4979	0.5140	0.5271	0.5373	0.5442	0.5481	0.5499	0.5510	0.5509	0.5498	0.5483

y variables

Spectral data x variables

If y variables are correlated ⇒ PLS2

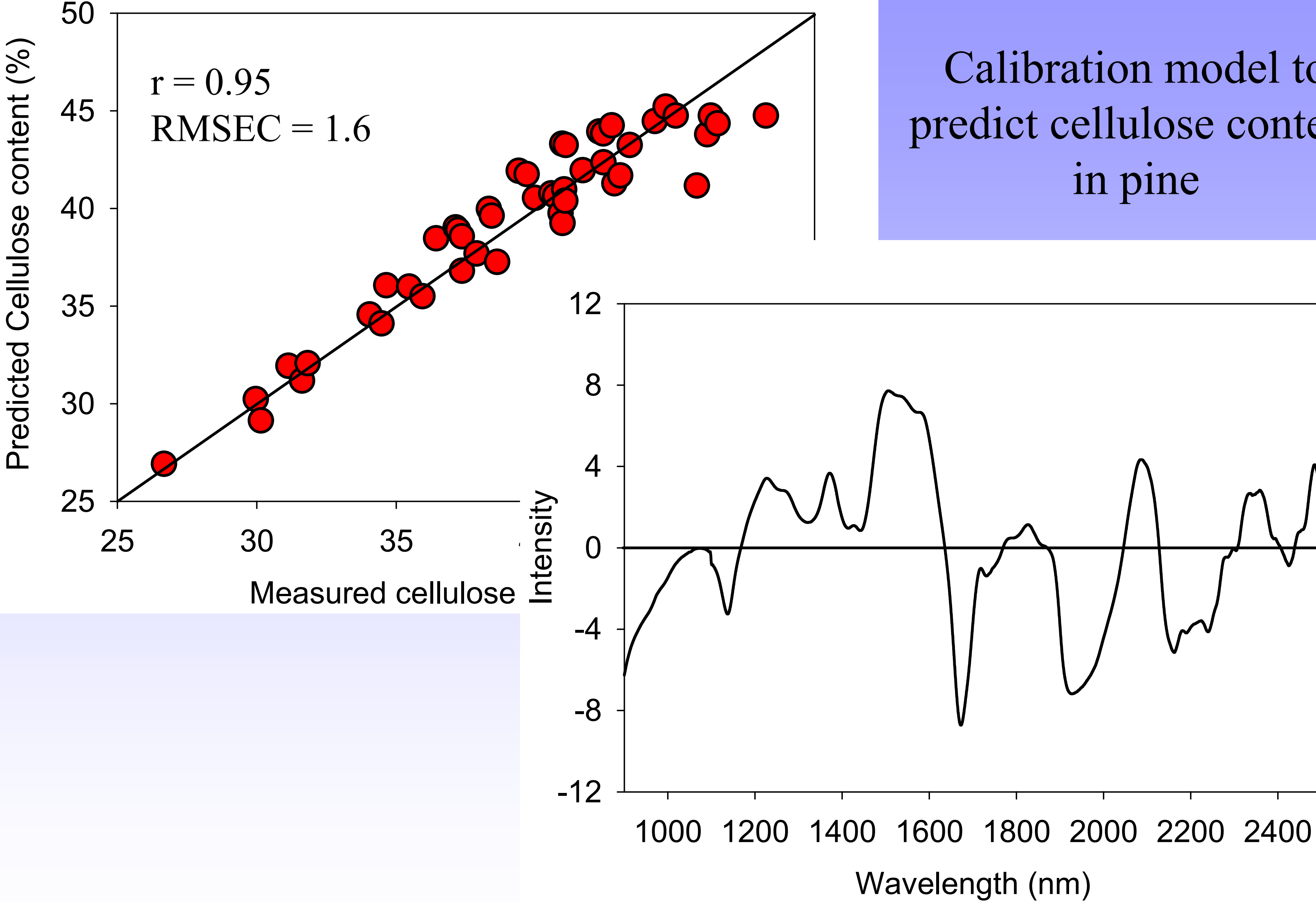
Biomass composition and near infrared spectra

		cellulose	Hemis	lignin	extractives	400	402	404	406	408	410	412	414	416	418	420
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2-1 A-2	1	26.6700	22.0000	22.1600	26.9000	0.5305	0.5479	0.5643	0.5769	0.5854	0.5929	0.5993	0.6031	0.6049	0.6063	0.6065
3-1 A-2	2	26.9000	21.8000	23.0900	25.9000	0.5316	0.5493	0.5658	0.5781	0.5867	0.5949	0.6022	0.6064	0.6084	0.6101	0.6108
2-5 A-2	3	28.2700	22.1000	23.5100	23.5000	0.5333	0.5507	0.5673	0.5801	0.5885	0.5956	0.6021	0.6059	0.6073	0.6084	0.6087
3-2 A-2	4	28.5200	20.1000	23.8900	26.2000	0.5105	0.5283	0.5431	0.5541	0.5627	0.5696	0.5736	0.5761	0.5783	0.5793	0.5781
2-2 A-2	5	29.9600	20.5000	27.1400	18.5000	0.5242	0.5407	0.5555	0.5663	0.5735	0.5795	0.5836	0.5858	0.5867	0.5869	0.5855
2-1 A-5	6	30.1500	22.4000	24.5200	20.9000	0.5232	0.5408	0.5567	0.5684	0.5763	0.5831	0.5887	0.5922	0.5938	0.5946	0.5943
1-	7	30.4400	23.4000	26.9000	16.3000	0.5347	0.5517	0.5682	0.5817	0.5906	0.5980	0.6050	0.6099	0.6117	0.6128	0.6136
3-	8	31.1300	21.1000	25.7500	17.7000	0.5235	0.5411	0.5568	0.5680	0.5759	0.5832	0.5887	0.5914	0.5926	0.5937	0.5934
2-	9	31.6200	23.1000	25.5300	18.3000	0.5280	0.5454	0.5619	0.5741	0.5821	0.5893	0.5955	0.5986	0.5998	0.6010	0.6012
2-	10	31.8200	23.3000	26.9400	16.0000	0.5235	0.5407	0.5562	0.5674	0.5753	0.5824	0.5876	0.5899	0.5910	0.5920	0.5917
1-	11	34.0500	24.7000	29.5400	9.9000	0.5214	0.5389	0.5548	0.5666	0.5749	0.5823	0.5879	0.5907	0.5922	0.5936	0.5935
3-	12	34.4700	22.9000	27.0400	13.2000	0.5210	0.5385	0.5539	0.5650	0.5728	0.5795	0.5841	0.5865	0.5880	0.5888	0.5878
3-	13	34.6400	24.8000	27.7000	10.0000	0.5069	0.5243	0.5387	0.5493	0.5571	0.5629	0.5667	0.5693	0.5708	0.5709	0.5697
1-	14	35.2700	23.7000	28.1200	9.9000	0.5252	0.5423	0.5585	0.5708	0.5788	0.5855	0.5909	0.5940	0.5952	0.5959	0.5954
1-2 A-2	15	35.4600	23.8000	29.7400	8.8000	0.5191	0.5368	0.5525	0.5638	0.5714	0.5779	0.5825	0.5848	0.5859	0.5863	0.5852
1-3 A-2	16	35.9400	24.7000	28.5800	8.1000	0.5073	0.5239	0.5377	0.5476	0.5543	0.5590	0.5621	0.5643	0.5649	0.5637	0.5618
1-5 A-2	17	36.4300	23.6000	29.4200	6.9000	0.4869	0.5014	0.5130	0.5215	0.5258	0.5274	0.5284	0.5283	0.5263	0.5235	0.5202
1-1 A-9	18	36.6000	23.4000	28.8500	8.2000	0.5182	0.5354	0.5512	0.5627	0.5704	0.5767	0.5811	0.5836	0.5851	0.5857	0.5847
2-3 A-5	19	37.1300	23.5000	27.6100	9.4000	0.5035	0.5209	0.5349	0.5454	0.5531	0.5582	0.5610	0.5629	0.5636	0.5632	0.5618
3-2 A-5	20	37.1700	22.8000	27.5500	11.2000	0.4860	0.5011	0.5138	0.5237	0.5291	0.5318	0.5343	0.5358	0.5354	0.5347	0.5337
2-2 A-5	21	37.2100	23.1000	27.0900	10.6000	0.4924	0.5079	0.5204	0.5296	0.5349	0.5377	0.5403	0.5421	0.5417	0.5403	0.5382
3-4 A-5	22	37.3500	22.3000	27.4400	9.8000	0.4906	0.5064	0.5192	0.5289	0.5345	0.5376	0.5404	0.5425	0.5424	0.5411	0.5391
1-1 A-32	23	37.3600	23.8000	33.2000	4.6000	0.5425	0.5597	0.5763	0.5907	0.6006	0.6079	0.6151	0.6210	0.6238	0.6250	0.6260
3-5 A-2	24	37.6500	26.5000	27.8300	8.9000	0.5001	0.5171	0.5307	0.5407	0.5475	0.5520	0.5555	0.5584	0.5593	0.5584	0.5568
3-1 A-32	25	37.8800	23.7000	33.0000	5.8000	0.5157	0.5340	0.5498	0.5611	0.5691	0.5760	0.5811	0.5844	0.5864	0.5876	0.5874
1-4 A-5	26	38.0100	24.2000	28.9300	4.7000	0.4979	0.5140	0.5271	0.5373	0.5442	0.5481	0.5499	0.5510	0.5509	0.5498	0.5483

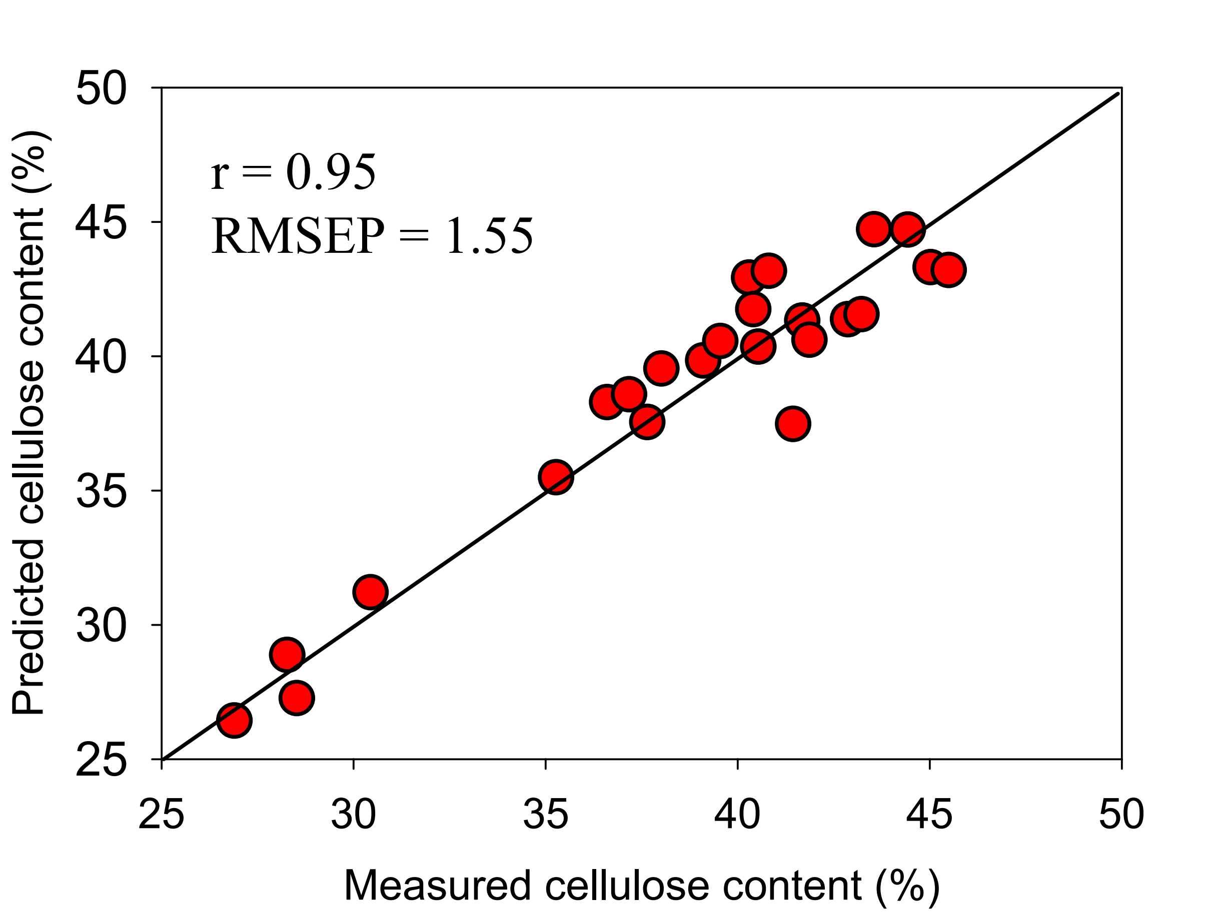
Samples

2/3 of the samples for calibration model
1/3 of the samples for validation model
Random selection

Calibration model to predict cellulose content in pine



Validation of the model to predict cellulose content in pine



PLS-Discriminant Analysis (PLS-DA)

A powerful method for classification. The aim is to create a predictive model which can accurately classify future unknown samples.

		value	species	350.0000	351.0000	352.0000	353.0000	354.0000	355.0000	356.0000	357.0000	358.0000	359.0000	360.0000	361.0000
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Spectrum00001.asd	1	0.0000	yellow poplar	1.1225	1.1276	1.1314	1.1176	1.1091	1.1111	1.1287	1.1031	1.0970	1.1070	1.0859	1.0771
Spectrum00002.asd	2	0.0000	yellow poplar	1.1701	1.1587	1.1481	1.1631	1.1488	1.1378	1.1437	1.1074	1.1060	1.1325	1.1130	1.1113
Spectrum00003.asd	3	0.0000	yellow poplar	1.1693	1.1690	1.1712	1.1622	1.1326	1.1200	1.1415	1.1224	1.1242	1.1431	1.1168	1.1001
Spectrum00004.asd	4	0.0000	yellow poplar	1.1832	1.1880	1.1786	1.1496	1.1443	1.1588	1.1987	1.1438	1.1395	1.1766	1.1397	1.1214
Spectrum00005.asd	5	0.0000	yellow poplar	1.2268	1.2087	1.1795	1.1968	1.1894	1.1736	1.1547	1.1270	1.1286	1.1520	1.1308	1.1331
Spectrum00006.asd	6	0.0000	yellow poplar	1.1206	1.1509	1.1812	1.1781	1.1522	1.1287	1.1159	1.0945	1.1040	1.1382	1.1304	1.1380
Spectrum00007.asd	7	0.0000	yellow poplar	1.1811	1.1602	1.1715	1.1921	1.1941	1.1952	1.2035	1.1604	1.1491	1.1636	1.1507	1.1531
Spectrum00008.asd	8	0.0000	yellow poplar	1.1755	1.2124	1.2154	1.2063	1.1816	1.1642	1.1633	1.1623	1.1705	1.1839	1.1608	1.1490
Spectrum00009.asd	9	0.0000	yellow poplar	1.2445	1.2238	1.1904	1.2108	1.2141	1.2042	1.1812	1.1599	1.1654	1.1888	1.1496	1.1279
Spectrum00017.asd	10	1.0000	hickory	1.1442	1.1427	1.1217	1.1361	1.1446	1.1513	1.1581	1.1206	1.1194	1.1474	1.1316	1.1264
Spectrum00018.asd	11	1.0000	hickory	1.1548	1.1410	1.1047	1.1284	1.1091	1.0956	1.1081	1.0876	1.0907	1.1147	1.1165	1.1137
Spectrum00010.asd	12	1.0000	hickory	1.1035	1.1160	1.1204	1.1350	1.1424	1.1412	1.1294	1.1030	1.1064	1.1325	1.1130	1.1092
Spectrum00011.asd	13	1.0000	hickory	1.1855	1.1616	1.1319	1.1733	1.1760	1.1656	1.1516	1.1363	1.1378	1.1511	1.1362	1.1287
Spectrum00012.asd	14	1.0000	hickory	1.0923	1.0956	1.0964	1.1466	1.1389	1.1263	1.1307	1.1086	1.1017	1.1077	1.1041	1.0950
Spectrum00013.asd	15	1.0000	hickory	1.1615	1.1579	1.1378	1.1287	1.1169	1.1110	1.1154	1.1179	1.1376	1.1686	1.1350	1.1216
Spectrum00014.asd	16	1.0000	hickory	1.2318	1.2369	1.2074	1.1909	1.1882	1.1848	1.1745	1.1527	1.1549	1.1756	1.1645	1.1569
Spectrum00015.asd	17	1.0000	hickory	1.1148	1.1160	1.1225	1.1477	1.1426	1.1384	1.1487	1.1360	1.1424	1.1622	1.1234	1.1225
Spectrum00016.asd	18	1.0000	hickory	1.1301	1.1217	1.1003	1.1591	1.1226	1.0963	1.1212	1.0976	1.1052	1.1407	1.1376	1.1092

y variable

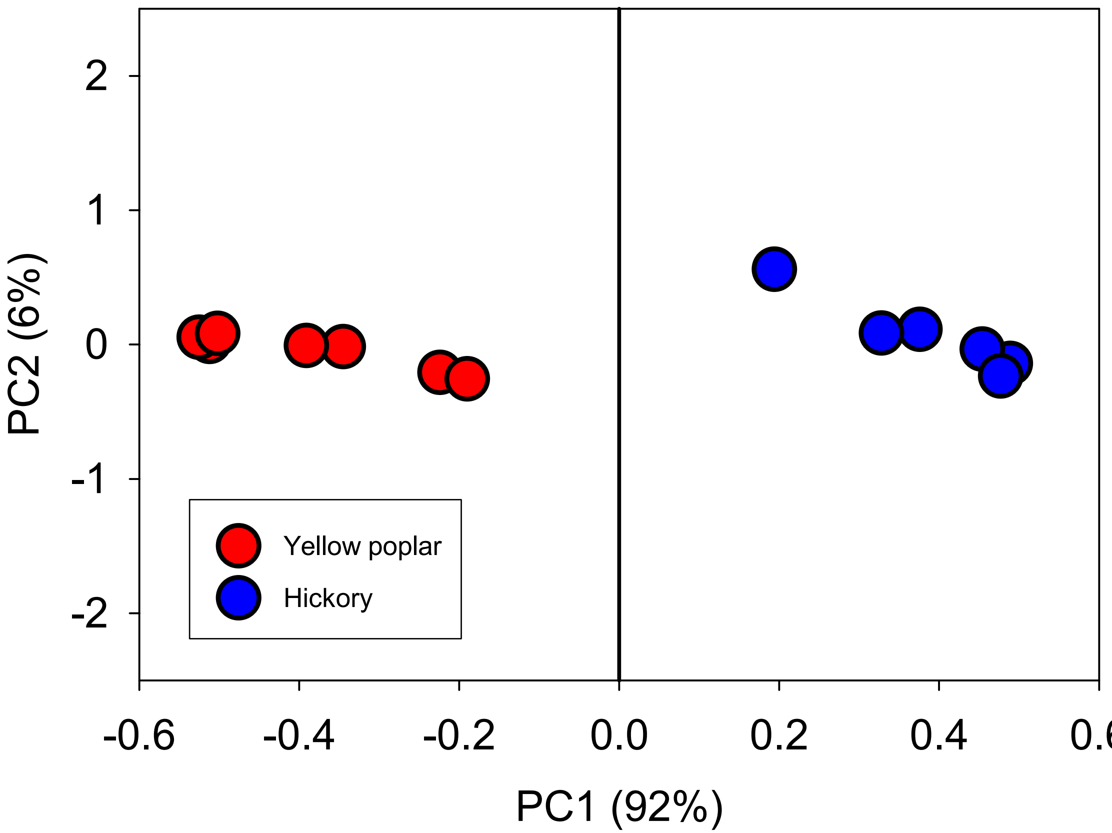
Spectral data
x variables

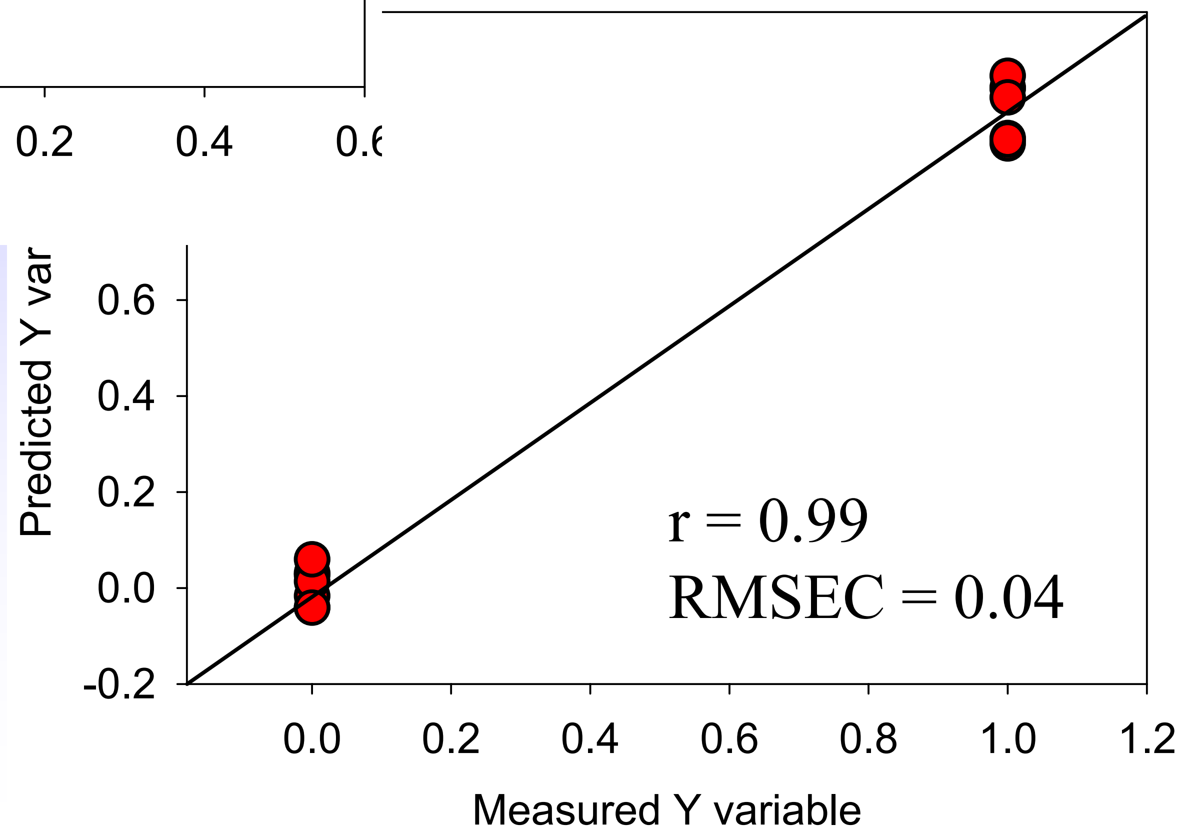
PLS-Discriminant Analysis (PLS-DA)

Development of PLS-DA calibration model

		value	species	350.0000	351.0000	352.0000	353.0000	354.0000	355.0000	356.0000	357.0000	358.0000	359.0000	360.0000	361.0000
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Spectrum00001.asd	1	0.0000	yellow poplar	1.1225	1.1276	1.1314	1.1176	1.1091	1.1111	1.1287	1.1031	1.0970	1.1070	1.0859	1.0771
Spectrum00002.asd	2	0.0000	yellow poplar	1.1701	1.1587	1.1481	1.1631	1.1488	1.1378	1.1437	1.1074	1.1060	1.1325	1.1130	1.1113
Spectrum00003.asd	3	0.0000	yellow poplar	1.1693	1.1690	1.1712	1.1622	1.1326	1.1200	1.1415	1.1224	1.1242	1.1431	1.1168	1.1001
Spectrum00004.asd	4	0.0000	yellow poplar	1.1832	1.1880	1.1786	1.1496	1.1443	1.1588	1.1987	1.1438	1.1395	1.1766	1.1397	1.1214
Spectrum00005.asd	5	0.0000	yellow poplar	1.2268	1.2087	1.1795	1.1968	1.1894	1.1736	1.1547	1.1270	1.1286	1.1520	1.1308	1.1331
Spectrum00006.asd	6	0.0000	yellow poplar	1.1206	1.1509	1.1812	1.1781	1.1522	1.1287	1.1159	1.0945	1.1040	1.1382	1.1304	1.1380
Spectrum00007.asd	7	0.0000	yellow poplar	1.1811	1.1602	1.1715	1.1921	1.1941	1.1952	1.2035	1.1604	1.1491	1.1636	1.1507	1.1531
Spectrum00008.asd	8	0.0000	yellow poplar	1.1755	1.2124	1.2154	1.2063	1.1816	1.1642	1.1633	1.1623	1.1705	1.1839	1.1608	1.1490
Spectrum00009.asd	9	0.0000	yellow poplar	1.2445	1.2238	1.1904	1.2108	1.2141	1.2042	1.1812	1.1599	1.1654	1.1888	1.1496	1.1279
Spectrum00017.asd	10	1.0000	hickory	1.1442	1.1427	1.1217	1.1361	1.1446	1.1513	1.1581	1.1206	1.1194	1.1474	1.1316	1.1264
Spectrum00018.asd	11	1.0000	hickory	1.1548	1.1410	1.1047	1.1284	1.1091	1.0956	1.1081	1.0876	1.0907	1.1147	1.1165	1.1137
Spectrum00010.asd	12	1.0000	hickory	1.1035	1.1160	1.1204	1.1350	1.1424	1.1412	1.1294	1.1030	1.1064	1.1325	1.1130	1.1092
Spectrum00011.asd	13	1.0000	hickory	1.1855	1.1616	1.1319	1.1733	1.1760	1.1656	1.1516	1.1363	1.1378	1.1511	1.1362	1.1287
Spectrum00012.asd	14	1.0000	hickory	1.0923	1.0956	1.0964	1.1466	1.1389	1.1263	1.1307	1.1086	1.1017	1.1077	1.1041	1.0950
Spectrum00013.asd	15	1.0000	hickory	1.1615	1.1579	1.1378	1.1287	1.1169	1.1110	1.1154	1.1179	1.1376	1.1686	1.1350	1.1216
Spectrum00014.asd	16	1.0000	hickory	1.2318	1.2369	1.2074	1.1909	1.1882	1.1848	1.1745	1.1527	1.1549	1.1756	1.1645	1.1569
Spectrum00015.asd	17	1.0000	hickory	1.1148	1.1160	1.1225	1.1477	1.1426	1.1384	1.1487	1.1360	1.1424	1.1622	1.1234	1.1225
Spectrum00016.asd	18	1.0000	hickory	1.1301	1.1217	1.1003	1.1591	1.1226	1.0963	1.1212	1.0976	1.1052	1.1407	1.1376	1.1092

PLS-DA
Calibration model





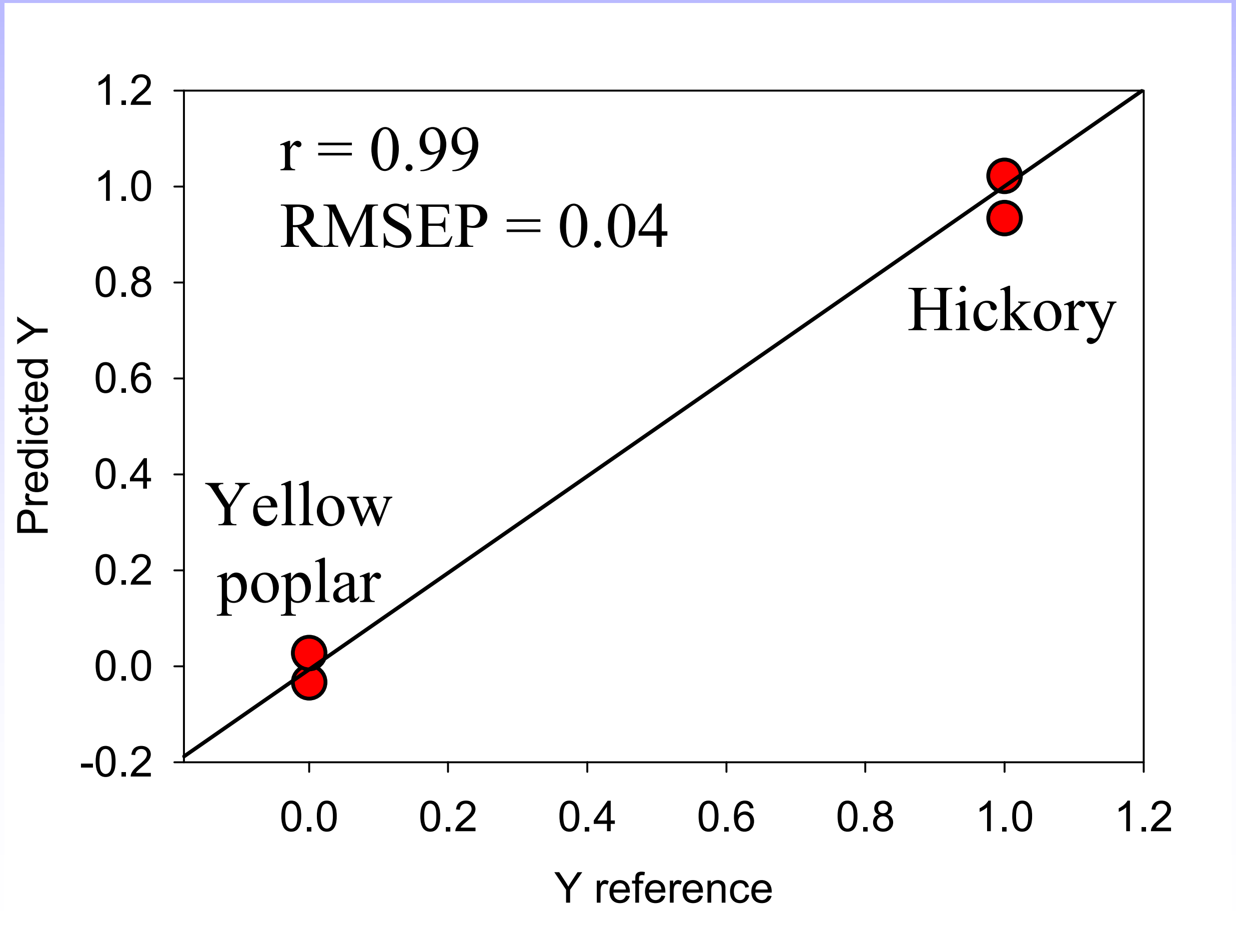
PLS-Discriminant Analysis (PLS-DA)

Validation of PLS-DA model

		value	species	350.0000	351.0000	352.0000	353.0000	354.0000	355.0000	356.0000	357.0000	358.0000	359.0000	360.0000	361.0000
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Spectrum00001.asd	1	0.0000	yellow poplar	1.1225	1.1276	1.1314	1.1176	1.1091	1.1111	1.1287	1.1031	1.0970	1.1070	1.0859	1.0771
Spectrum00002.asd	2	0.0000	yellow poplar	1.1701	1.1587	1.1481	1.1631	1.1488	1.1378	1.1437	1.1074	1.1060	1.1325	1.1130	1.1113
Spectrum00003.asd	3	0.0000	yellow poplar	1.1693	1.1690	1.1712	1.1622	1.1326	1.1200	1.1415	1.1224	1.1242	1.1431	1.1168	1.1001
Spectrum00004.asd	4	0.0000	yellow poplar	1.1832	1.1880	1.1786	1.1496	1.1443	1.1588	1.1987	1.1438	1.1395	1.1766	1.1397	1.1214
Spectrum00005.asd	5	0.0000	yellow poplar	1.2268	1.2087	1.1795	1.1968	1.1894	1.1736	1.1547	1.1270	1.1286	1.1520	1.1308	1.1331
Spectrum00006.asd	6	0.0000	yellow poplar	1.1206	1.1509	1.1812	1.1781	1.1522	1.1287	1.1159	1.0945	1.1040	1.1382	1.1304	1.1380
Spectrum00007.asd	7	0.0000	yellow poplar	1.1811	1.1602	1.1715	1.1921	1.1941	1.1952	1.2035	1.1604	1.1491	1.1636	1.1507	1.1531
Spectrum00008.asd	8	0.0000	yellow poplar	1.1755	1.2124	1.2154	1.2063	1.1816	1.1642	1.1633	1.1623	1.1705	1.1839	1.1608	1.1490
Spectrum00009.asd	9	0.0000	yellow poplar	1.2445	1.2238	1.1904	1.2108	1.2141	1.2042	1.1812	1.1599	1.1654	1.1888	1.1496	1.1279
Spectrum00017.asd	10	1.0000	hickory	1.1442	1.1427	1.1217	1.1361	1.1446	1.1513	1.1581	1.1206	1.1194	1.1474	1.1316	1.1264
Spectrum00018.asd	11	1.0000	hickory	1.1548	1.1410	1.1047	1.1284	1.1091	1.0956	1.1081	1.0876	1.0907	1.1147	1.1165	1.1137
Spectrum00010.asd	12	1.0000	hickory	1.1035	1.1160	1.1204	1.1350	1.1424	1.1412	1.1294	1.1030	1.1064	1.1325	1.1130	1.1092
Spectrum00011.asd	13	1.0000	hickory	1.1855	1.1616	1.1319	1.1733	1.1760	1.1656	1.1516	1.1363	1.1378	1.1511	1.1362	1.1287
Spectrum00012.asd	14	1.0000	hickory	1.0923	1.0956	1.0964	1.1466	1.1389	1.1263	1.1307	1.1086	1.1017	1.1077	1.1041	1.0950
Spectrum00013.asd	15	1.0000	hickory	1.1615	1.1579	1.1378	1.1287	1.1169	1.1110	1.1154	1.1179	1.1376	1.1686	1.1350	1.1216
Spectrum00014.asd	16	1.0000	hickory	1.2318	1.2369	1.2074	1.1909	1.1882	1.1848	1.1745	1.1527	1.1549	1.1756	1.1645	1.1569
Spectrum00015.asd	17	1.0000	hickory	1.1148	1.1160	1.1225	1.1477	1.1426	1.1384	1.1487	1.1360	1.1424	1.1622	1.1234	1.1225
Spectrum00016.asd	18	1.0000	hickory	1.1301	1.1217	1.1003	1.1591	1.1226	1.0963	1.1212	1.0976	1.1052	1.1407	1.1376	1.1092

PLS-DA
Validation model

	Y-reference	Predicted Y
Spectrum 00008	0.0000	-0.0338
Spectrum 00009	0.0000	0.0270
Spectrum 00015	1.0000	0.9340
Spectrum 00016	1.0000	1.0220



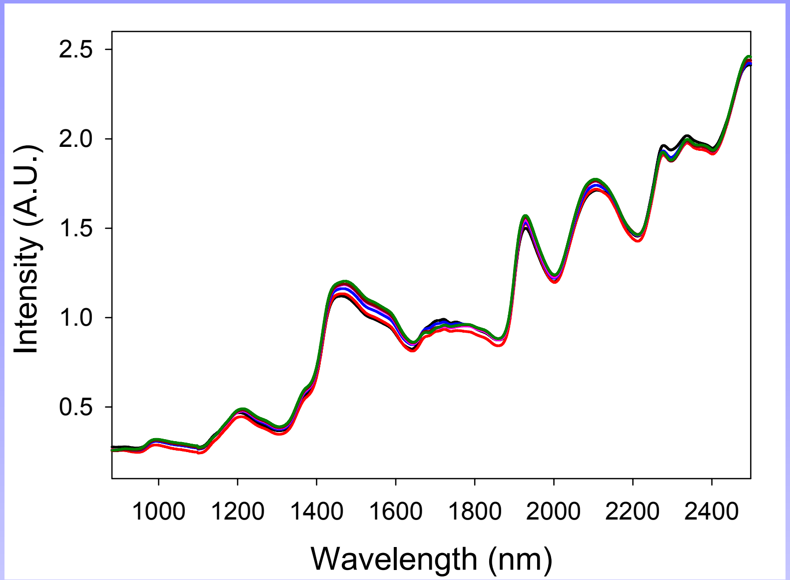
Two-dimensional correlation spectroscopy

2D correlation tools spread spectral peaks over a second dimension → simplify the visualization of complex spectra

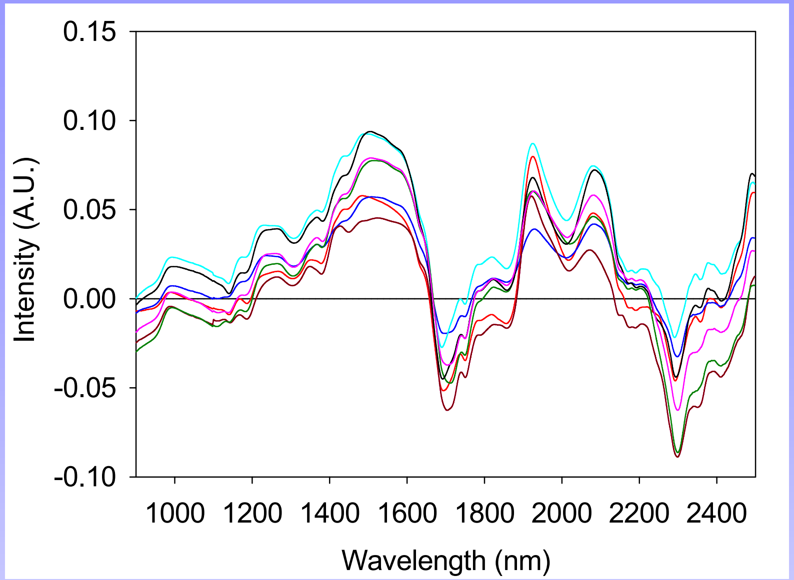
Perturbation { Mechanical, electrical, chemical, magnetic optical, thermal, …

↓

Electro-magnetic Probe (eg, IR, UV, LIBS,…) → System → 2D correlation maps



Spectral data collected under perturbation



Dynamic spectra [DYN]

$[S]_{(mxn)}$: m samples by n variables (wavelength) matrix

$$[DYN]_{(m\times n)} = [S]_{(1\times n)}^{1\to m} - [\overline{S}]_{(1\times n)}$$

Reference spectrum

Synchronous matrix

$$[SYNC]_{(n\times n)} = [DYN]_{(n\times m)}^{T} \times [DYN]_{(m\times n)}$$

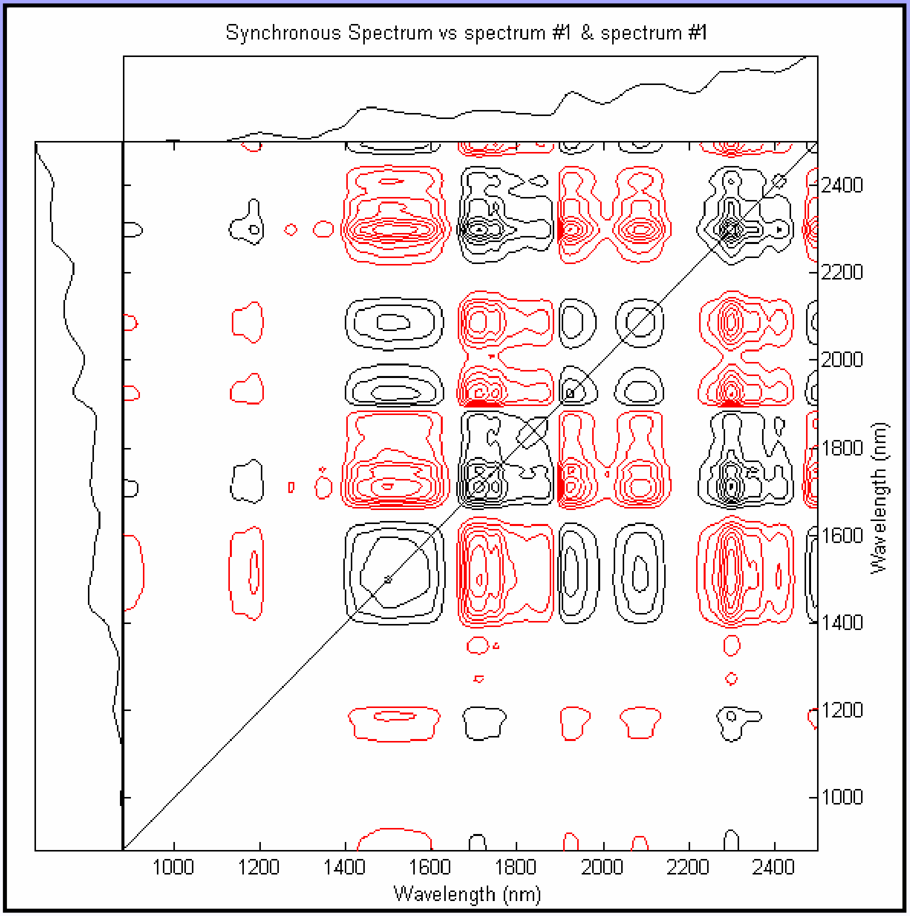
Asynchronous matrix

$$[ASYN]_{(n\times n)} = [DYN]_{(n\times m)}^{T} \times [N]_{(m\times m)} \times [DYN]_{(m\times n)}$$

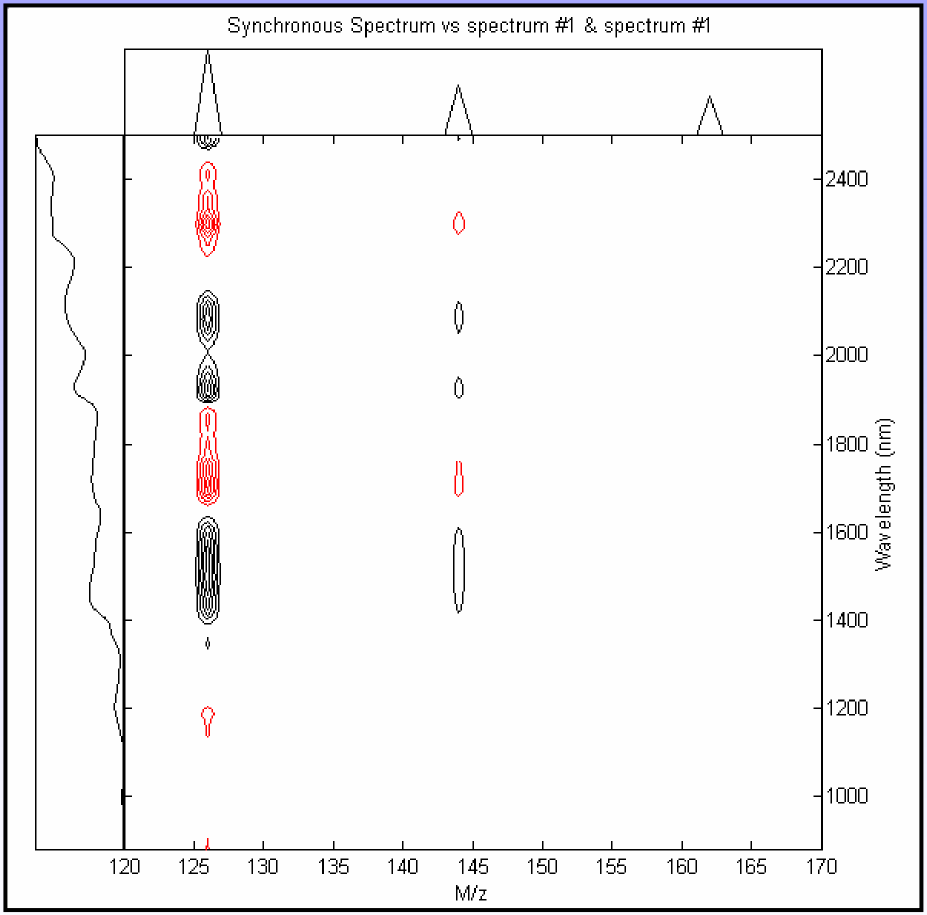
Noda-Hilbert matrix

Generation of orthogonal components: synchronous and asynchronous 2D correlation intensities

Homo-correlation
NIR/NIR



Hetero-correlation
NIR/MBMS



Perturbation: cellulose content

Other techniques to extract information

Classification

- Soft Independence Modeling of Class Analogy (SIMCA)

  The Unscrambler, User manual. CAMO,1998

- Kernel Principal Component Analysis (k-PCA)

Schölkopf B., Smola AJ., Müller K. (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10: 1299-1319

- Artificial Neural Networks (ANN)

Demuth H., Beale M. and Hagan M., Neural Network Toolbox 5 User's Guide – Matlab

non-linear data

Other techniques to extract information

Regression

- Orthogonal Projections to Latent Structures (O-PLS)

Trygg J., Wold S. (2002) Orthogonal projections to latent structures (O-PLS). J. Chemo. 16:119-128

- Artificial Neural Networks (ANN)

Demuth H., Beale M. and Hagan M., Neural Network Toolbox 5 User's Guide – Matlab

- Kernel Projection to Latent Structures (k-PLS)

Rosipal R., Trejo L.J. (2001) Kernel partial least squares regression in reproducing Kernel Hilbert space. J. Machine Learning Res. 2: 97–123

- Supervised Probabilistic Principal Component Analysis (SPPCA)

Yu S., Yu K., Tresp V., Kriegel H., Wu M. (2006) Supervised probabilistic principal component analysis. Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining (SIGKDD):464-473

non-linear data

Softwares

- www.camo.com
- www.umetrics.com
- www.infometrix.com
- www.mathworks.com

References

•A user-friendly guide to Multivariate Calibration and Classification; T. Næs, T. Isaksson, t. Fearn, T. Davies, NIR Publications, Chichester, UK, 2002

•Multivariate calibration, H. Martens and T. Næs, John Wiley & Sons, Chichester, UK, 1989

•Chemometric techniques for quantitative analysis, Marcel Dekker, New York, 1998

•Two-dimensional correlation spectroscopy, I. Noda and Y. Ozaki, John Wiley & Sons, Chichester, UK, 2004

•Neural Network Toolbox 5 User's Guide – Matlab, H. Demuth, M. Beale and M. Hagan.

Questions?