DOE AND MULTIVARIATE MODELING TECHNIQUES FOR MATERIAL SELECTION

JMP DISCOVERY SUMMIT CHINA 2019

BILL WORLEY
Sr. Systems Engineer
JMP Division

MAHMOUD HAMMOUD
Sr. Systems Engineer
JMP Division

SOME QUOTES

- Deming – “Statistics (Analytics) is too important to be left to the statisticians”
- Box – “All models are wrong, but how wrong do they have to be to not be useful?”
- Author Unknown – “Don’t fall in love with your models. They will NOT love you back” – Things change!
- Carly Fiorina – “The goal is to transform data into information and information into insight.”
- John Sall (the creator of JMP) - "Data rarely confesses all its information easily. You have to explore it. You have to use multiple tools. The process should be as easy as possible and the tools as powerful as possible.”

SOME CONTEXT

- Why about this concept is important – why should you care?
- Why would you do this and for whom would you do it?
- How can you be convinced like I have been?
- Will this approach allow you to be more successful in your role?
- Will this approach allow your company to be more successful?

INTRODUCTION

When testing options of different raw materials or formulation ingredients, a common practice is to vary them as multilevel categorical variable e.g. A, B, C….etc. In this scenario, identifying the best candidate material requires all options to be tested.

- A consequence of this is:
  - Time consuming physical testing and the resulting model is only applicable to predict the tested options but cannot predict options with different physical/chemical properties. Your model does not love you back!

INTRODUCTION

A much more efficient approach is to design the experiment based on the physical/chemical properties of each new material option.

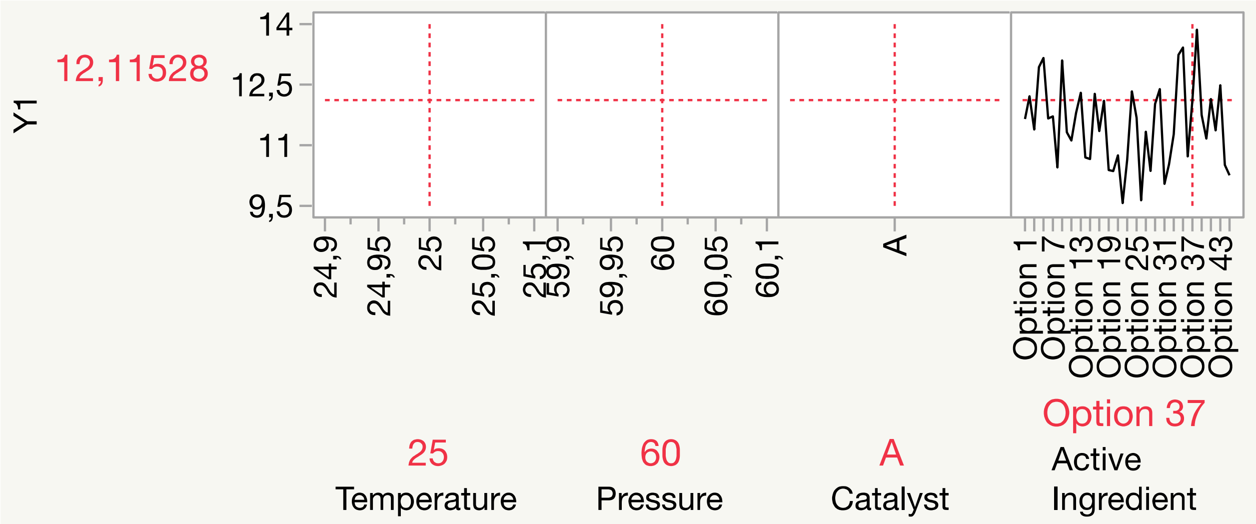
Activity = *f*(physiochemical properties and/or structural properties) + error

This can significantly decrease the number of required experimental conditions and results in a **sustainable** empirical model that can predict options not tested before.

Based on workflow presented by Silvio Miccio (Proctor & Gamble) at JMP Discovery Summit Europe 2017: https://community.jmp.com/t5/Discovery-Summit-Europe-2017/Increase-Efficiency-and-Model-Applicability-Domain-When-Testing/ta-p/36572

THE PROBLEM

	Temperature	Pressure	Catalyst	Active Ingredient	Y1	Y2
1	25	60	A	Option 1	•	•
2	25	60	A	Option 2	•	•
3	25	60	A	Option 3	•	•
4	25	60	A	Option 4	•	•
5	25	60	A	Option 5	•	•
6	25	60	A	Option 6	•	•
7	25	60	A	Option 7	•	•
	⋮			⋮	⋮	
39	25	60	A	Option 39	•	•
40	25	60	A	Option 40	•	•
41	25	60	A	Option 41	•	•
42	25	60	A	Option 42	•	•
43	25	60	A	Option 43	•	•
44	25	60	A	Option 44	•	•
45	25	60	A	Option 45	•	•



EXPRESSING EACH TESTED INGREDIENT BY ITS PHYSICOCHEMICAL PROPERTIES

	Active Ingredient
1	Option 1
2	Option 2
3	Option 3
4	Option 4
5	Option 5
6	Option 6
7	Option 7
8	Option 8
9	Option 9
10	Option 10
11	Option 11
12	Option 12
13	Option 13
14	Option 14
15	Option 15
16	Option 16
17	Option 17
18	Option 18
19	Option 19
20	Option 20
21	Option 21
22	Option 22
23	Option 23
24	Option 24
25	Option 25
26	Option 26
27	Option 27
28	Option 28
29	Option 29
30	Option 30
31	Option 31
32	Option 32
33	Option 33
34	Option 34
35	Option 35
36	Option 36
37	Option 37
38	Option 38
39	Option 39
40	Option 40
41	Option 41
42	Option 42
43	Option 43
44	Option 44
45	Option 45

20/0 Cols

	ID	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
1	1	-0,281	-0,739	0,659	0,113	1	2,318	0,327	0,279	1,212	-0,368	-0,097	0,601	0,114	-0,024	0,891	-2,209	-0,149	-0,343	0,116
2	2	1,023	-0,229	1,087	0,807	-1	-0,096	0,751	-0,796	-0,257	0,340	-0,700	1,200	0,054	-1,522	-0,270	0,568	-0,752	0,672	1,489
3	3	-1,044	0,364	0,395	-1,234	1	-0,967	-0,657	1,822	0,022	-2,830	-0,094	0,149	1,553	1,574	0,192	0,985	0,722	-1,627	-1,053
4	4	-0,236	0,592	-0,227	-0,467	-1	2,200	0,092	0,226	-0,384	0,340	-0,941	-0,258	-1,212	-0,235	1,533	-2,004	0,061	-0,949	-0,882
5	5	1,591	-0,543	-0,513	0,389	-1	0,937	0,452	-1,572	-1,488	0,340	0,332	-0,139	-1,305	-1,756	-0,430	-0,538	-1,522	1,669	1,833
6	6	-0,736	0,500	-1,402	0,536	1	-0,179	-0,844	0,720	0,195	0,340	-0,305	-1,403	-1,289	-0,332	-0,142	-0,149	0,839	0,109	-0,641
7	7	1,539	0,624	-0,056	1,065	1	-0,729	2,009	-1,421	-0,906	1,723	-0,099	0,260	1,258	-0,267	-1,005	-0,056	-1,051	1,251	1,933
8	8	2,556	-0,542	1,287	2,048	-1	-0,310	1,572	-2,605	-1,634	0,340	0,820	1,845	1,532	-0,790	-1,540	-1,155	-1,590	0,462	2,004
9	9	-1,284	1,274	-0,966	-1,479	-1	-0,690	-1,676	1,453	1,166	-1,520	0,097	-1,195	-0,244	1,283	0,625	0,950	1,371	-0,411	-1,025
10	10	0,414	0,299	0,169	1,134	1	-0,516	0,309	-0,607	-0,838	0,430	0,425	0,227	1,460	0,739	0,282	0,315	-0,646	-0,465	0,447
11	11	-1,144	-1,001	0,999	-0,758	1	-0,463	-1,705	1,326	1,026	0,340	-0,171	0,604	-1,632	0,002	-0,815	0,807	1,362	-1,475	-0,254
12	12	0,963	0,356	-0,772	0,641	1	-0,381	0,510	0,024	-0,783	0,340	0,829	-0,445	-1,276	-1,435	0,051	0,450	-1,170	0,619	0,985
13	13	-1,746	0,970	-1,098	-1,358	1	0,950	-1,581	1,205	0,912	-2,051	0,049	-1,406	-1,075	0,465	0,949	-1,365	1,003	0,093	-2,098
14	14	-1,268	1,079	-0,566	-0,671	1	-0,455	-1,700	1,028	1,574	-3,343	-0,802	-0,790	-0,158	1,541	1,116	0,372	1,428	-0,078	-0,677
15	15	0,800	-1,032	0,241	0,196	1	1,766	0,460	-0,558	-0,996	0,340	-0,714	0,438	-0,037	-1,173	-0,731	-1,785	-1,046	1,260	1,032
16	16	-0,091	0,959	-0,581	-0,793	-1	-0,927	-1,020	0,162	0,545	0,340	-0,868	-0,461	1,158	1,048	0,647	1,533	-0,101	0,552	0,374
17	17	-0,077	0,279	0,113	0,357	1	-1,131	-0,399	0,639	0,551	-0,135	0,009	0,066	-0,131	1,118	-0,463	0,867	-0,327	0,149	0,170
18	18	-2,861	1,763	-1,654	-2,679	1	-0,171	-3,043	2,471	3,337	0,340	-0,847	-2,118	0,286	1,326	1,551	-0,496	3,214	-2,817	-3,145
19	19	-0,675	-0,339	-0,579	-0,708	1	-0,097	-0,468	0,475	0,518	-1,981	0,065	-0,698	0,306	-0,378	-0,259	0,763	0,669	0,646	-0,811
20	20	-0,669	-0,181	0,423	-0,733	-1	0,234	-0,036	0,498	0,311	0,340	0,150	0,178	-0,269	1,841	0,172	-0,054	0,547	0,923	-0,592
21	21	-0,250	-3,043	2,902	-0,254	-1	-0,257	-0,392	-0,078	-0,289	0,340	-0,500	2,456	-1,184	-0,739	-2,159	0,652	-0,036	0,239	-0,011
22	22	0,460	0,480	-0,680	1,227	1	-0,159	0,841	-0,685	-0,630	0,340	2,621	-0,463	-1,168	-0,863	0,371	-0,146	-0,642	0,952	0,732
23	23	-0,186	-0,183	0,083	-0,409	1	-0,875	0,069	-0,114	0,506	-1,457	-0,893	0,093	0,181	0,793	-0,408	1,171	0,609	0,586	0,521
24	24	-0,564	-0,644	-0,805	-0,926	-1	-0,784	0,410	1,526	-0,113	-0,693	-0,561	-0,828	-1,354	-0,471	-1,069	0,799	1,189	-1,293	-0,685
25	25	0,404	1,435	-1,795	0,251	1	-0,621	0,816	0,113	-0,613	0,340	1,346	-1,376	1,256	0,065	0,679	0,398	-0,816	-2,514	0,799
26	26	0,988	-0,154	0,620	1,138	-1	-1,101	1,229	-0,910	-1,206	1,089	-0,341	0,764	1,204	-0,138	-0,597	1,716	-0,996	1,189	0,240
27	27	1,324	-0,541	1,301	2,271	-1	0,039	1,263	-1,763	-0,853	1,475	-0,001	1,471	1,345	-0,102	-0,500	-0,395	-1,702	1,677	0,898
28	28	-0,340	1,743	-0,718	-0,313	1	-0,245	0,101	0,062	1,228	-0,436	1,026	-0,800	1,548	0,991	1,125	-0,335	-0,024	-0,550	-0,400
29	29	-0,372	-0,431	0,595	-1,314	1	3,235	-0,811	-0,035	-0,007	-0,473	-0,126	0,421	-0,064	-0,668	2,049	-2,856	-0,196	0,392	-0,377
30	30	0,790	-1,448	1,266	0,938	1	-0,892	0,920	-0,915	-0,714	0,340	-0,179	1,224	0,066	-1,423	-2,473	0,057	-0,589	0,180	0,793
31	31	0,987	-0,540	1,156	1,270	-1	1,331	1,037	-1,098	-1,226	0,340	-0,105	1,234	1,677	-0,139	0,655	-0,345	-1,324	0,811	0,124
32	32	0,482	-0,672	-0,723	0,351	-1	-0,078	1,516	-0,369	0,407	0,340	0,564	-0,493	-1,516	-1,356	-0,603	-0,134	-0,753	0,068	-0,189
33	33	-0,161	0,014	-0,282	-0,383	-1	-0,703	-0,314	0,086	-0,563	-0,231	-0,508	-0,224	-0,066	1,013	-0,100	1,330	-0,204	0,244	-0,442
34	34	0,258	-0,955	1,335	0,410	-1	0,070	0,121	-0,312	-0,080	0,251	-0,698	1,243	0,011	0,744	-1,280	-0,281	0,467	-1,473	-0,084
35	35	0,659	0,416	-0,345	0,848	-1	1,853	0,330	0,117	-0,775	1,082	4,811	-0,091	0,060	-0,216	1,465	-1,279	-0,482	-0,043	0,319
36	36	-1,480	1,808	-2,135	-1,445	1	-0,634	-0,807	1,240	1,484	0,340	-0,946	-2,242	-0,050	0,176	1,508	0,671	1,411	-1,474	-0,528
37	37	0,590	-0,414	0,138	0,703	1	0,558	0,087	-0,673	-0,374	0,632	-0,097	0,255	-1,297	-1,234	-0,461	-0,626	-0,646	-0,322	0,347
38	38	0,114	-1,268	1,161	-0,877	-1	0,456	-0,010	-0,426	-0,466	0,340	-0,032	1,035	-1,492	-0,897	-0,511	-0,308	-0,351	0,572	0,505
39	39	-0,210	-1,130	1,590	-0,028	1	-0,008	-0,411	0,111	1,814	0,340	0,100	1,356	-0,021	0,943	-0,196	-0,017	0,996	-0,073	0,349
40	40	0,084	-0,608	0,059	-0,450	-1	-1,038	-0,143	-0,557	-0,355	0,340	-0,162	0,028	0,168	-1,006	-1,500	0,688	0,773	0,165	-0,010
41	41	-0,015	1,268	-0,146	-0,404	1	-1,101	-0,175	-0,197	-0,688	0,340	-0,673	-0,199	0,614	1,165	0,515	1,120	-0,094	-0,251	-0,023
42	42	-0,873	-0,234	-0,581	-0,499	1	-0,404	-0,672	0,927	-0,248	0,340	-0,194	-0,744	-0,096	0,119	-0,213	0,479	0,784	-0,438	-1,330
43	43	-0,214	0,195	-0,182	0,244	1	-0,326	-0,186	0,218	0,949	0,340	-0,944	-0,186	1,457	0,777	0,451	0,620	-0,162	0,456	-0,638
44	44	1,234	-1,091	-0,032	1,204	-1	0,403	1,200	-1,225	-1,028	0,340	-0,590	0,217	-0,069	-1,676	-0,227	-0,404	-0,689	1,012	0,957
45	45	-0,482	1,542	-0,739	0,043	1	-0,013	0,625	0,188	-0,242	0,340	-0,053	-0,807	-0,304	1,118	1,126	-0,374	0,020	-0,154	-1,070

PHYSICAL AND CHEMICAL PROPERTIES EXAMPLES

- Mw – Molecular Weight
- C – the number of carbon atoms on the hydrophobic end of the surfactant
- Griffin HLB – the hydrophylic-lypophilic balance (HLB) according to Griffin
- CMC – critical micellar concentration
- Number of chains……

- One thing to note is that many of these factors are likely highly collinear!

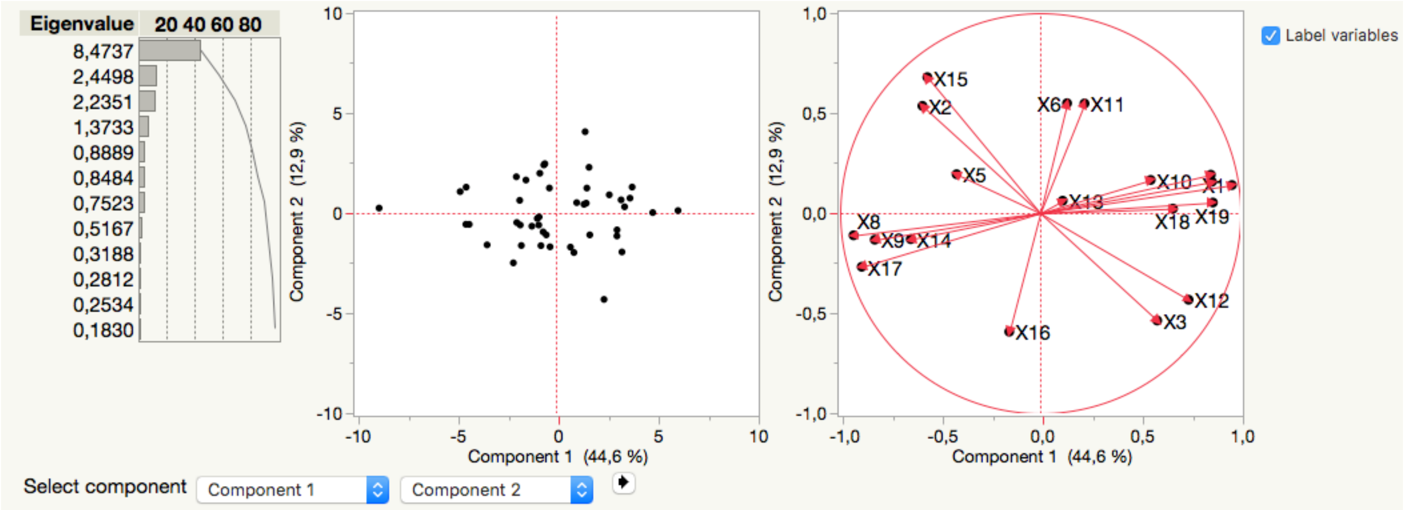
TABLE WITH CHEMICAL/BIOLOGICAL PROPERTIES

		Mw	C	redC	redC/C	Eow	Griffin	Davis	CPP	redCPP	CP	dCP	Chains	RMChain	F-alcohol	maxEO	w33EO	w66EO	CMC	logP	BATP	BABC
1	B-048	641	13	9	69.23	10	13.75	6.23	0.27	0.37	67	27.5	1	100	4.99	8	11	6	0.12	5.661	0.0469	0.0956
2	B-058	553	13	9	69.23	8	12.75	5.53	0.3	0.41	44	23	1	100	5.28	7	13	6	0.079	5.532	0.07	
3	B-065	684	16	16.1	100	10	12.87	4.75	0.27	0.27	76	4.5	2	95	4	8	16	9	0.001	7.768	-0.8842	-0.9175
4	B-09	669	15	8.5	56.67	10	13.17	5.28	0.27	0.44	54	29	1	100	0	10	9	5.7	0.31	6.109	-0.0788	0.0514
5	B-160	474	14	13.7	100	6	11.14	4.49	0.34	0.34	40	10	2	85	6.9	5	11	7	0.013	6.453	-0.7612	
6	B-267	581	15	8.5	56.67	8	12.13	4.58	0.3	0.5	16		1	100	0	7	8	5.2	0.031	5.98	-0.0611	0.0435
7	B-271	612	15	8.5	56.67	9	12.53	4.82	0.29	0.48	32	10	1	100	0	8	8.3	5.2	0.052	6.044	-0.0036	0.1141
8	B-535	393	11	10.9	98.91	5	11.22	5.43	0.38	0.38	27	10.5	2	85	11.34	5	9.4	5.7	0.11	4.801	-0.2745	
9	BOX-257	515	13	13.2	97.92	7	11.97	4.96	0.32	0.32	47	11.5	7	28	7.3	5	12.2	7.6	0.042	5.988	-0.6891	-0.4368
10	BOX-4511	707	15	14.1	96.23	11	13.7	5.82	0.25	0.26	87	22.5	5	36	0	9	18	10.2	0.23	6.775	-0.7401	-0.5716
11	BOX-915	377	10	9.8	98.69	5	11.68	5.95	0.37	0.38	36	10	6	54	12.99	4	9.2	7	0.3	4.272	-0.0009	
12	BOX-918	508	10	9.6	97.65	8	13.88	7.05	0.3	0.3	77	33	6	42	7.4	6	12	6.8	0.33	4.465	0.291	0.6124
13	GO-100	704	18	17.7	100	10	12.47	4	0.27	0.27	76	9	2	85	3.18	11	16.5	10.6	0.05	8.697	-0.8123	-0.8007
14	GT-110	730	16	16.2	100	11	13.28	5.06	0.25	0.25	90	7.5	2	90	2.77	11	12.5	8.8	0.005	7.833	-0.9397	-0.979
15	GUD-050	393	11	10.5	95.64	5	11.22	5.43	0.38	0.39			2	52	18.85	3	10.8	8.3	0.1	4.801	-0.1271	0.3537
16	GX-060	465	13	9	69.23	6	11.38	4.83	0.34	0.48			1	100	7.45	3	8.9	6.9	0.034	5.404	0.1088	0.0669
17	HDo-11	752	18	17.9	100	11	12.86	4.25	0.25	0.25	76	6	2	95	1.69	10	15.2	7.8	0.1	8.407	-1.094	-0.9471
18	HDo-7	575	18	17.9	100	7	10.68	2.85	0.32	0.32	30		2	95	4.9	6	13.2	8.2	0.005	8.149	-0.8649	-0.8148
19	HDo-9	664	18	17.9	100	9	11.91	3.55	0.28	0.28	58	7	2	95	2.97	8	14	7.8	0.03	8.278	-0.9087	-0.9327
20	Is-11	671	12	8	66.67	11	14.45	7.05	0.25	0.36	72	22	1	100	6.43	10	19	9.1	0.16	5.587	0.5646	0.5254
21	Is-8	539	12	8	66.67	8	13.08	6	0.3	0.43	16		1	100	14	8	15	11.7	0.08	5.393	0.2735	
22	Is-9	583	12	8	66.67	9	13.61	6.35	0.28	0.4	59	16	1	100	10.86	9	15.4	10.9	0.09	5.458	0.4146	0.5396
23	LON-50	379	10	7	69.5	5	11.64	5.9	0.37	0.52	34	9	2	95	13.3	5	11.3	8.9	0.899	3.882	0.8723	1.0786
24	LON-60	423	10	6.9	69	6	12.51	6.25	0.34	0.47	36	13	2	90	14.7	5	9.9	7.4	0.64	3.947	0.8547	0.9235
25	LON-70	647	10	7	69.5	7	13.22	6.6	0.32	0.44	61	18	2	95	12.04	7	11.3	7.6	0.18	4.011	0.9724	1.1068
26	LON-80	511	10	6.9	69	8	13.8	6.95	0.3	0.41	80	23	2	90	5.4	9	15.4	8.4	0.19	4.075	1.0936	1.0963
27	LTO-10	641	13	8.5	65.38	10	13.75	6.23	0.27	0.39	70	23	5	50	6.91	9	15.7	8.6	0.1	5.661	0.0263	-0.2006
28	LTO-8	553	13	8.6	66.15	8	12.75	5.53	0.3	0.43	61	12.5	4	60	7.01	8	15	8	0.07	5.532	-0.1726	0.0583
29	M-1618/10	684	16	16.1	100	10	12.87	4.75	0.27	0.27	69		2	95	3.79	9	15.6	8.8	0.004	7.768	-1.0837	-0.8868
30	M-24/60	477	14	13.9	100	6	11.08	4.4	0.34	0.34	32	3	2	95	10.49	5	10.3	7	0.04	6.453	-0.662	-0.1699
31	MO-11/50	393	11	11	99.55	5	11.22	5.43	0.38	0.38	36	14	2	95	10.71	6	12.5	9.5	0.24	4.801	0.492	
32	MO-13/100	641	13	8.6	66.15	10	13.75	6.23	0.27	0.39	78	16.5	4	60	4.85	10	17.2	8.7	0.07	5.661	0.0772	0.1942
33	MO-13/80	553	13	8.6	66.15	8	12.75	5.53	0.3	0.43	48	17	4	60	23.82	5	10.3	8	0.037	5.532	-0.1131	0.1933
34	So-10	631	12	7.2	58.13	10	13.96	6.56	0.27	0.43	74	29	2	85	0.21	10	12.2	7.2	0.19	5.432	-0.1461	0.2101
35	So-6	455	12	7.2	58.13	6	11.62	5.16	0.34	0.55	20		2	85	0.27	7	9.8	6.3	0.022	5.175	-0.2703	0.1913
36	So-9	587	12	7.2	58.13	9	13.51	6.21	0.28	0.45	59	32	2	85	0.23	9	12	6.9	0.02	5.368	-0.1702	0.2027

Mw = molecular weight; C = the number of carbon atoms in the hydrophobic part of the surfactant; red C = the number of carbon atoms in the longest chain of the hydrophobic part of the surfactant; redC/C = the ratio between the longest chain and the total number of carbon atoms in the hydrophobic part of the surfactant; Eow = the wanted moles of ethylene oxide per fatty acid alcohol; Griffin = The hydrophilic-lipophilic balance according to Griffin; Davis = the hydrophilic-lipophilic balance according to Davis; CPP = the critical packing parameter according to Israelachvili; redCPP = the critical packning parameter with respect to whether the hydrophobic part is branched or not; CP = the cloud point; dCP = the highest derivative of the transmittance-temperature curve; Chains = the number of different carbon chains in the hydrophobic part; RMChain = the molar ratio between the dominating type of carbon chain and the total carbon chain in the hydrophobic part of the surfactant; F-alcohol = the ratio of non-ethoxylated fatty alcohol in the surfactant product; maxEO = the position of the peak in the ethylene-oxide distribution chromatogram; w33EO = the width of the digitized chromatogram at 33% peak height; w66EO = the width of the digitized chromatogram at 66% peak height; CMC = the critical micellar concentration; log P = the logarithm of the octanol/water partition coefficient; BATP = relative toxicity scale for *Thamnocephalus platyurus* (low values imply high toxicity); BABC = relative toxicity scale for *Brachionus calyciflorus* (low values imply high toxicity).

References: (1) Åsa Lindgren, PhD Thesis, Umeå University, 1995; (2) Lise-Lott Uppgård, Graduate Thesis, Umeå University, 1995.

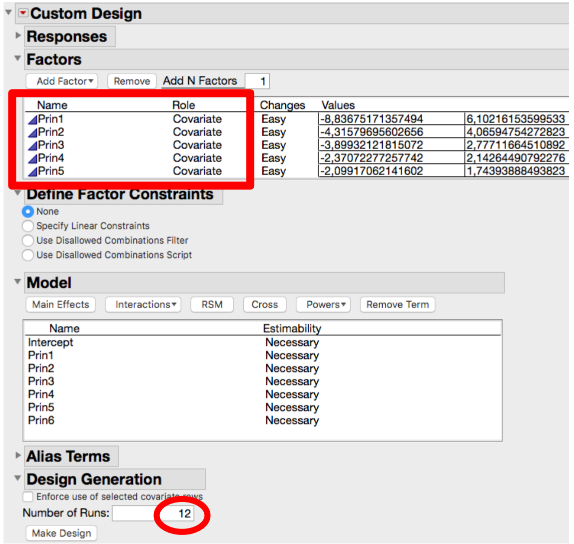
DIMENSIONALITY REDUCTION USING PRINCIPAL COMPENENT ANALYSIS



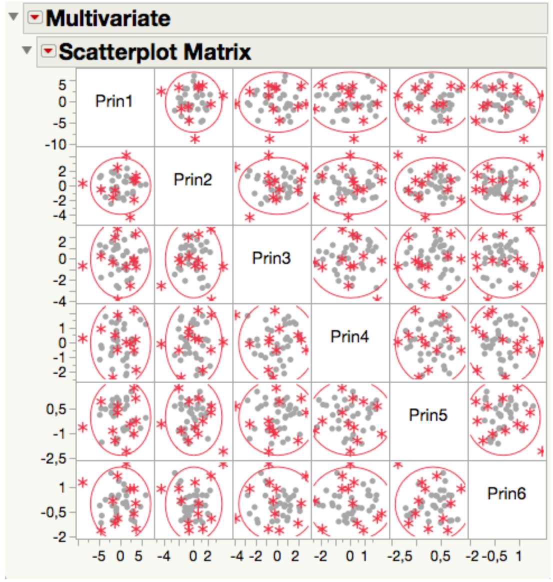
Eigenvalues

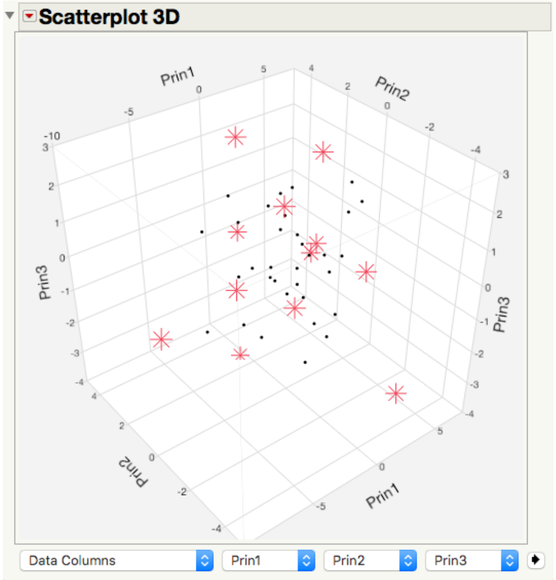
Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	8,4737	44,599		44,599
2	2,4498	12,894		57,492
3	2,2351	11,764		69,256
4	1,3733	7,228		76,484
5	0,8889	4,679		81,162
6	0,8484	4,465		85,628
7	0,7523	3,960		89,588
8	0,5167	2,719		92,307
9	0,3188	1,678		93,985
10	0,2812	1,480		95,465
11	0,2534	1,334		96,799
12	0,1830	0,963		97,762
13	0,1585	0,834		98,596
14	0,1021	0,537		99,133
15	0,0768	0,404		99,538
16	0,0541	0,285		99,822
17	0,0179	0,094		99,917
18	0,0151	0,080		99,996
19	0,0007	0,004		100,000

CUSTOM DOE USING PRINCIPAL COMPONENTS AS COVARIATES



27/0 Cols		X11	X12	X13	X14	X15	X16	X17	X18	X19	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
	1	-0,097	0,601	0,114	-0,024	0,891	-2,...	0,4...	-0,...	0,116	-0,325997034	1,2518977344	-2,770075613	1,7961583135	0,995466043	0,5082502745
	2	-0,700	1,200	0,054	-1,522	-0,270	0,568	-0,...	0,...	1,489	3,0535209286	-0,841318059	0,0249610045	-0,124774579	-0,301057182	-0,619342589
	3	-0,094	0,149	1,553	1,574	0,192	0,985	0,7...	-1,...	-1,...	-3,439784551	-1,589490745	1,0869622073	1,7411269522	0,4624911396	0,2230863097
	4	-0,941	-0,258	-1,212	-0,235	1,533	-2,...	0,0...	-0,...	-0,...	-0,80425916	1,9892225691	-2,581271762	0,1239370589	-1,341751453	-0,61985213
*	5	0,332	-0,139	-1,305	-1,756	-0,430	-0,...	-1,...	1,...	1,833	3,8091479075	1,2998116948	-0,804376792	-1,869687815	-0,256490583	-1,551752689
	6	-0,305	-1,403	-1,289	-0,332	-0,142	-0,...	0,8...	0,...	-0,...	-1,809374936	0,6362947703	-0,07175246	-1,830172724	0,7390587874	-0,134872702
*	7	-0,099	0,260	1,258	-0,267	-1,005	-0,...	-1,...	1,...	1,933	3,7015219305	0,7460906559	2,307331369	0,1864308731	1,119848785	0,0079475137
	8	0,820	1,845	1,532	-0,790	-1,540	-1,...	-1,...	0,...	2,004	6,102161536	0,1252954397	0,7921989913	1,2223587817	-0,061331979	0,5060539044
	9	0,097	-1,195	-0,244	1,283	0,625	0,950	1,3...	-0,...	-1,...	-4,333202166	-0,564348577	0,5846049106	-0,382348372	-1,553667624	-0,7362684
	10	0,425	0,227	1,460	0,739	0,282	0,315	-0,...	-0,...	0,447	1,0386723454	0,5207844745	1,7900661934	1,2412635329	0,4844206614	0,8041224119
	11	-0,171	0,604	-1,632	0,002	-0,815	0,807	1,3...	-1,...	-0,...	-2,127501759	-2,489163811	-1,496968906	-0,870372862	0,6471527792	1,85387372
	12	0,829	-0,445	-1,276	-1,435	0,051	0,450	-1,...	0,...	0,985	1,5460119286	1,2486333245	0,6947861092	-2,103763056	1,023651698	-0,08341615
	13	0,049	-1,406	-1,075	0,465	0,949	-1,...	1,0...	0,...	-2,...	-4,485775673	1,2940685475	-1,795635102	-0,214182333	0,5037518167	-1,176081616
*	14	-0,802	-0,790	-0,158	1,541	1,116	0,372	1,4...	-0,...	-0,...	-4,50054635	-0,56212671	0,2383048437	0,9097698673	0,9251682304	-1,634065244
	15	-0,714	0,438	-0,037	-1,173	-0,731	-1,...	-1,...	1,...	1,032	2,6660507624	0,9121493227	-1,876597745	0,204340391	1,5300525818	-0,976272206
*	16	-0,868	-0,461	1,158	1,048	0,647	1,533	-0,...	0,...	0,374	-1,20257038	-0,655226303	1,9993707712	0,4909625079	-1,453255563	-1,203088746
	17	0,009	0,066	-0,131	1,118	-0,463	0,867	-0,...	0,...	0,170	-0,638520537	-0,940435719	1,2620346608	0,116008633	0,7016911325	0,3451661773
*	18	-0,847	-2,118	0,286	1,326	1,551	-0,...	3,2...	-2,...	-3,...	-8,836751714	0,249583158	-0,691566026	0,2673268744	-0,645251926	1,2808620507
	19	0,065	-0,698	0,306	-0,378	-0,259	0,763	0,6...	0,...	-0,...	-1,782589872	-0,59719035	0,0383592702	-0,35101441	1,0352072596	-1,049974984
	20	0,150	0,178	-0,269	1,841	0,172	-0,...	0,5...	0,...	-0,...	-0,858022677	-0,597843135	-0,509292715	0,7855711403	-1,631348315	-0,281512805
*	21	-0,500	2,456	-1,184	-0,739	-2,159	0,652	-0,...	0,...	-0,...	2,4040498406	-4,315796956	-2,571011227	-0,115823078	-0,446329632	0,8879652473
	22	2,621	-0,463	-1,168	-0,863	0,371	-0,...	-0,...	0,...	0,732	1,6548511036	2,2874854629	0,6036412599	-1,600385839	0,6929162238	0,7913421204
	23	-0,893	0,093	0,181	0,793	-0,408	1,171	0,6...	0,...	0,521	-0,748234318	-1,627924035	0,9740984927	0,3428544364	1,1673776848	-1,033890978
*	24	-0,561	-0,828	-1,354	-0,471	-1,069	0,799	1,1...	-1,...	-0,...	-1,727993189	-1,621166923	-0,346432959	-2,370722773	-0,852921091	0,1672882103
*	25	1,346	-1,376	1,256	0,065	0,679	0,398	-0,...	-2,...	0,799	-0,605447639	2,406317433	2,7771166451	-0,51779934	0,2567701113	1,671138701
	26	-0,341	0,764	1,204	-0,138	-0,597	1,716	-0,...	1,...	0,240	3,0493633986	-1,144739386	2,179191217	0,2217784457	-1,080737293	-0,464543769
	27	-0,001	1,471	1,345	-0,102	-0,500	-0,...	-1,...	1,...	0,898	4,8385940847	0,0252401978	0,7111422725	1,5360385904	-0,689631296	-0,195543583
	28	1,026	-0,800	1,548	0,991	1,125	-0,...	-0,...	-0,...	-0,...	-1,978117789	1,8144628791	1,5795801784	1,3629076927	0,2120718312	0,4802264302
*	29	-0,126	0,421	-0,064	-0,668	2,049	-2,...	-0,...	0,...	-0,...	-0,553736263	2,4670055713	-3,899321218	1,8456303949	0,6415031822	-0,910930546
*	30	-0,179	1,224	0,066	-1,423	-2,473	0,057	-0,...	0,...	0,793	3,3004376698	-1,944299485	-0,057609077	-0,53561084	1,7439388849	0,8941543356
*	31	-0,105	1,234	1,677	-0,139	0,655	-0,...	-1,...	0,...	0,124	3,2574993875	0,6592770154	-0,190933234	2,1426449079	-1,041604902	-0,623999198
	32	0,564	-0,493	-1,516	-1,356	-0,603	-0,...	-0,...	0,...	-0,...	1,4006126328	0,4384797642	-0,547189626	-2,229029539	-0,622560556	0,0904455831
	33	-0,508	-0,224	-0,066	1,013	-0,100	1,330	-0,...	0,...	-0,...	-0,482988058	-1,088909358	1,033639679	-0,265075701	-1,348078455	-0,855834283
	34	-0,698	1,243	0,011	0,744	-1,280	-0,...	0,4...	-1,...	-0,...	0,8999301828	-1,971634818	-0,877145431	0,9016379638	-0,846502999	1,178230122
*	35	4,811	-0,091	0,060	-0,216	1,465	-1,...	-0,...	-0,...	0,319	1,4597739283	4,0659475427	-0,766882971	0,0173309171	-2,099170621	2,4042302526
	36	-0,946	-2,242	-0,050	0,176	1,508	0,671	1,4...	-1,...	-0,...	-4,785556133	1,0719415622	1,2814106781	-1,163819723	0,1107383591	-0,127942241
	37	-0,097	0,255	-1,297	-1,234	-0,461	-0,...	-0,...	-0,...	0,347	1,5160402184	0,4938222895	-1,083354612	-1,035285995	1,1627663867	0,4818961034
	38	-0,032	1,035	-1,492	-0,897	-0,511	-0,...	-0,...	0,...	0,505	1,6923842405	-1,090506113	-2,106794896	-0,766889277	-0,724967387	-0,223232851
	39	0,100	1,356	-0,021	0,943	-0,196	-0,...	0,9...	-0,...	0,349	-0,293603033	-1,686808953	-1,049036056	1,5260929023	0,7736137399	1,365872289





BUILDING A MODEL BASED ON SELECTED TEST OPTIONS

29/2 Cols

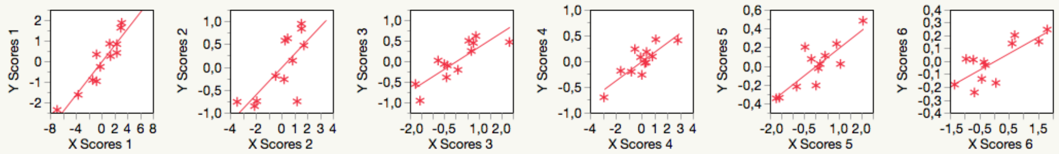
	ID	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	Y1	Y2
1	5	1,591	-0,543	-0,513	0,389	-1	0,937	0,452	-1,572	-1,488	0,340	0,332	-0,139	-1,305	-1,756	-0,430	-0,538	-1,522	1,669	1,833	0,0658907455	0,8989433551
2	7	1,539	0,624	-0,056	1,065	1	-0,729	2,009	-1,421	-0,906	1,723	-0,099	0,260	1,258	-0,267	-1,005	-0,056	-1,051	1,251	1,933	0,1995699882	0,9575689511
3	14	-1,268	1,079	-0,566	-0,671	1	-0,455	-1,700	1,028	1,574	-3,343	-0,802	-0,790	-0,158	1,541	1,116	0,372	1,428	-0,078	-0,677	0,7370807632	0,2901251701
4	16	-0,091	0,959	-0,581	-0,793	-1	-0,927	-1,020	0,162	0,545	0,340	-0,868	-0,461	1,158	1,048	0,647	1,533	-0,101	0,552	0,374	0,439409134	0,7240612587
5	18	-2,861	1,763	-1,654	-2,679	1	-0,171	-3,043	2,471	3,337	0,340	-0,847	-2,118	0,286	1,326	1,551	-0,496	3,214	-2,817	-3,145	1	0,3030029159
6	21	-0,250	-3,043	2,902	-0,254	-1	-0,257	-0,392	-0,078	-0,289	0,340	-0,500	2,456	-1,184	-0,739	-2,159	0,652	-0,036	0,239	-0,011	0,238962114	0,4601022211
7	24	-0,564	-0,644	-0,805	-0,926	-1	-0,784	0,410	1,526	-0,113	-0,693	-0,561	-0,828	-1,354	-0,471	-1,069	0,799	1,189	-1,293	-0,685	0,4746126052	0,2738733864
8	25	0,404	1,435	-1,795	0,251	1	-0,621	0,816	0,113	-0,613	0,340	1,346	-1,376	1,256	0,065	0,679	0,398	-0,816	-2,514	0,799	0,3432874573	0,3750541026
9	29	-0,372	-0,431	0,595	-1,314	1	3,235	-0,811	-0,035	-0,007	-0,473	-0,126	0,421	-0,064	-0,668	2,049	-2,856	-0,196	0,392	-0,377	0,6184603327	0,4159466391
10	30	0,790	-1,448	1,266	0,938	1	-0,892	0,920	-0,915	-0,714	0,340	-0,179	1,224	0,066	-1,423	-2,473	0,057	-0,589	0,180	0,793	0,1023113142	0,5322913497
11	31	0,987	-0,540	1,156	1,270	-1	1,331	1,037	-1,098	-1,226	0,340	-0,105	1,234	1,677	-0,139	0,655	-0,345	-1,324	0,811	0,124	0,3896049234	0,6958359505
12	35	0,659	0,416	-0,345	0,848	-1	1,853	0,330	0,117	-0,775	1,082	4,811	-0,091	0,060	-0,216	1,465	-1,279	-0,482	-0,043	0,319	0,2845934245	0,7504114302

Leave-One-Out Cross Validation With Method=NIPALS

Number of factors	Root Mean PRESS	van der Voet T²	Prob > van der Voet T²	Q
0	1,090909	6,396821	0,0010*	-0,19008
1	0,713674	3,249198	0,2240	0,52995
2	0,745938	2,547872	0,3450	0,49053
3	0,902765	4,067969	0,1250	0,26393
4	0,851251	3,060542	0,2620	0,34788
5	0,824663	4,789625	0,0510	0,38429
6	0,804600	5,410359	0,0310*	0,41031
7	0,774265	4,333486	0,0620	0,45291
8	0,665789	4,849188	0,0490*	0,59493
9	0,558446	0,000000	1,0000	0,71460
10	0,614300	6,727733	0,0090*	0,64596

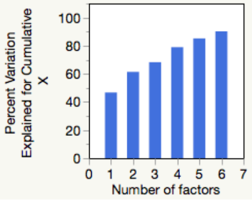
NIPALS Fit with 6 Factors

X-Y Scores Plots

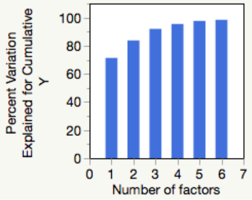


Percent Variation Explained

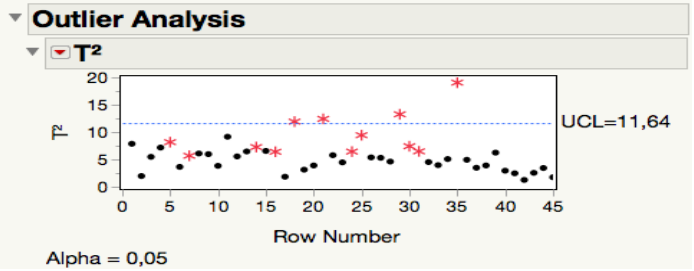
Number of factors	X Effects	Cumulative X
1	46,7477	46,748
2	14,6531	61,401
3	6,8787	68,280
4	10,6799	78,959
5	6,2101	85,170
6	5,0286	90,198

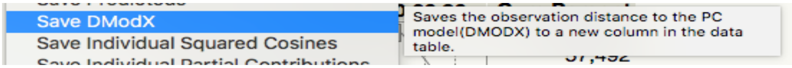


Y Responses	Cumulative Y
71,3206	71,3206
12,4168	83,7374
8,2345	91,9718
3,4643	95,4362
2,1807	97,6169
0,7835	98,4004

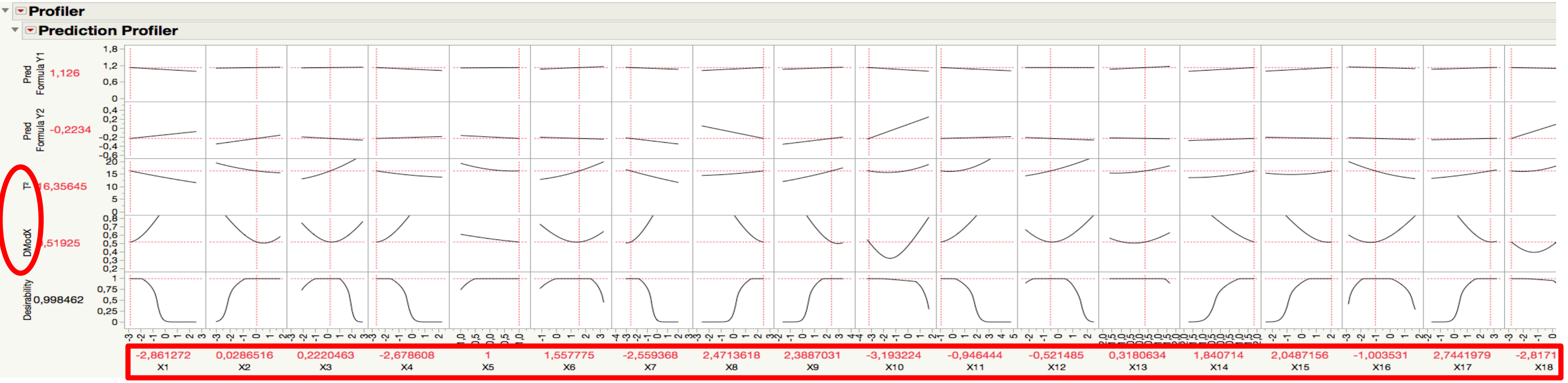


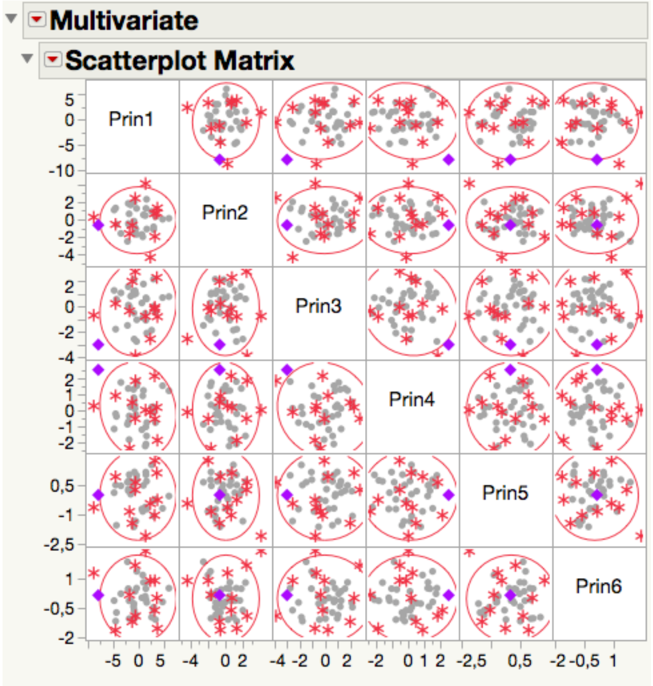
FINDING THE OVERALL OPTIMUM

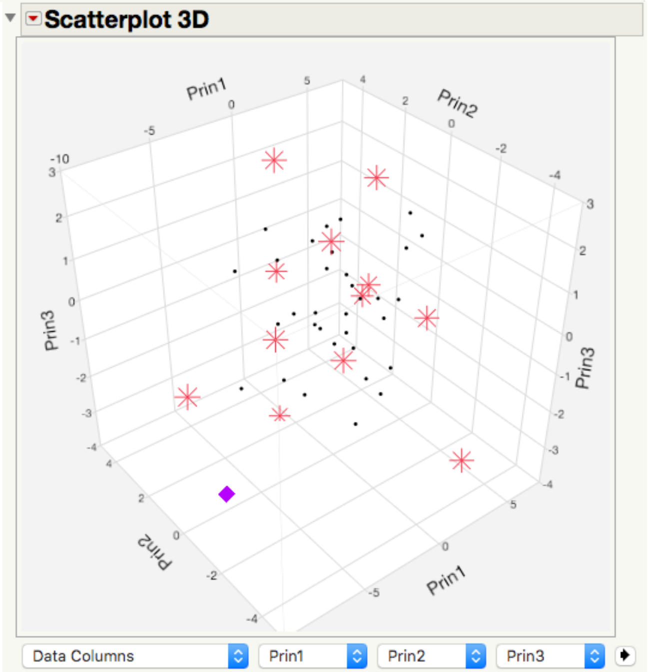




Pred Formula Y1	Pred Formula Y2	T²	DModX
0,7351610797	-0,032393137	5,4629707393	0,7684744689
0,5344112521	0,4877765954	7,1619160874	0,5704743206
0,0764336668	0,8914452586	8,1489966541	0,3397830922
0,4138066671	0,5997023892	3,62881969	0,4457523504
0,1359962934	0,9467463514	5,6621797828	0,5015983403
0,0579064197	0,8087819062	6,0755934688	0,5568613724
0,6645008925	0,4372006124	5,9596125557	0,371494562
0,3777355438	0,5668495764	3,8196978135	0,255683891
0,4630466463	0,3365258268	9,1394656637	0,3440309916
0,166204154	0,6679132604	5,544158677	0,3393660643
0,7321635432	0,3172880214	6,4499475931	0,4363231841



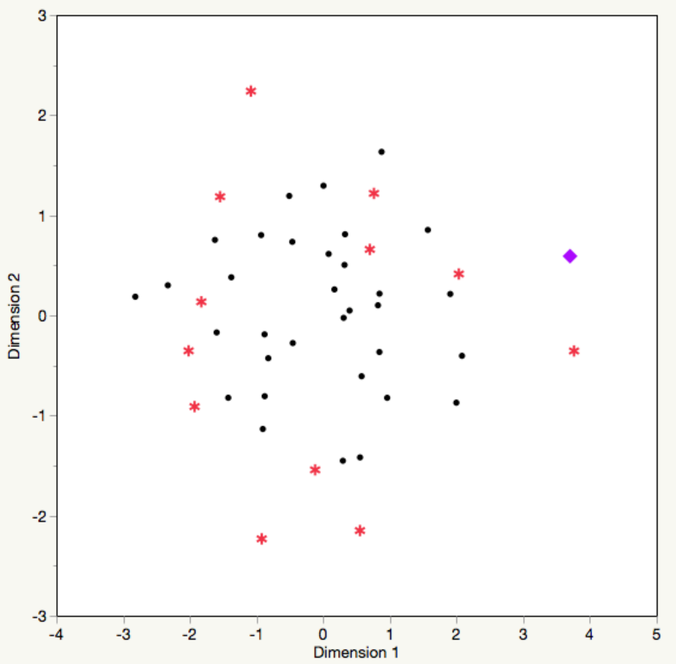




FIND CLOSEST COMMERCIALLY AVAILABLE INGREDIENT

2D Multidimensional Scaling

Multidimensional Scaling Plot



Solution Proximity

From Object	To Object	Actual Proximity	Predicted Proximity
1	2	2,2456	2,4814
1	3	2,3457	2,6339
1	4	1,3127	0,2594
1	5	2,5889	2,2886
1	6	1,9786	1,2163
1	7	2,7421	2,5607
1	8	3,0748	3,5177
1	9	2,6896	2,3169
1	10	1,9937	1,3952
1	11	2,1565	3,1396
1	12	2,2756	1,3380
1	13	2,0565	1,7977
1	14	2,4203	2,5504
1	15	1,5753	1,8304
1	16	2,5094	2,1493
1	17	2,0041	1,7160
1	18	3,7279	3,6361
1	19	1,8470	1,7569
1	20	1,7010	1,9561
1	21	2,7194	3,9409
1	22	2,1636	1,2417
1	23	2,1217	2,2623
1	24	2,4829	2,7109

	From Object	To Object	Actual Proximity	Predicted Proximity
1	18	47	1,5199	0,9486
2	13	47	2,1075	2,2467
3	14	47	2,2005	1,6770
4	9	47	2,5033	1,8354
5	3	47	2,5420	2,1485
6	36	47	2,7407	1,9015
7	11	47	3,0075	3,0122

Model Comparison

Predictors

Measures of Fit for Y1

Label	Predictor	Creator	.2,4,6,8	RSquare	RASE	AAE	Freq
Training	Pred Formula Y1	Partial Least Squares		0,9801	0,0367	0,0309	12
Validation	Pred Formula Y1	Partial Least Squares		0,8702	0,0673	0,0547	33

Measures of Fit for Y2

Label	Predictor	Creator	.2,4,6,8	RSquare	RASE	AAE	Freq
Training	Pred Formula Y2	Partial Least Squares		0,9879	0,0254	0,0211	12
Validation	Pred Formula Y2	Partial Least Squares		0,8281	0,0847	0,0642	33

HOW MANY PRINCIPLE COMPONENTS/RUNS?

How Many Principle Components?

- Most critical is to extract enough principle components
- In practice it has proven to be on the safe side when extracting enough principle components to explain >85% of the variation in the data.

How Many Runs?

- Number of runs depends to a certain extent on the number of principle components that have to be extracted
- The number of runs is by far not as critical as extracting the right number of principle components
- In practice 20 % – 25 % of the original number of options is typically sufficient

UBIQUITOUS METHOD WITH WIDE-RANGING APPLICATIONS

- QSAR inspired
- Extends to pretty much any high-dimensional problem
- Best applicable when factors and responses continuous
- Potential for saving significant testing effort/time/costs by combining covariate-based custom DOE and PCA.
- Modeling not limited to PLS: Neural Networks yield good results for this class of problems -> as always, try and compare!
- Models highly generalizable within the model space and can be used to explore chemical compositions that might not even exist yet
- Examples: Material selection, Physical property prediction (e.g., biodegradability, melting point, etc.), Spectral data (spectrometry/chemometrics), ...

JMP Demo

A TWIST ON ANALYZING SPECTRAL DATA

- In the recent past PLS has been used to analyze and build predictive models for spectral data. (ex. gasoline.jmp in help)
- We want to apply some of the same multivariate techniques used for material selection to show how to build good predictive models with far few conditions.

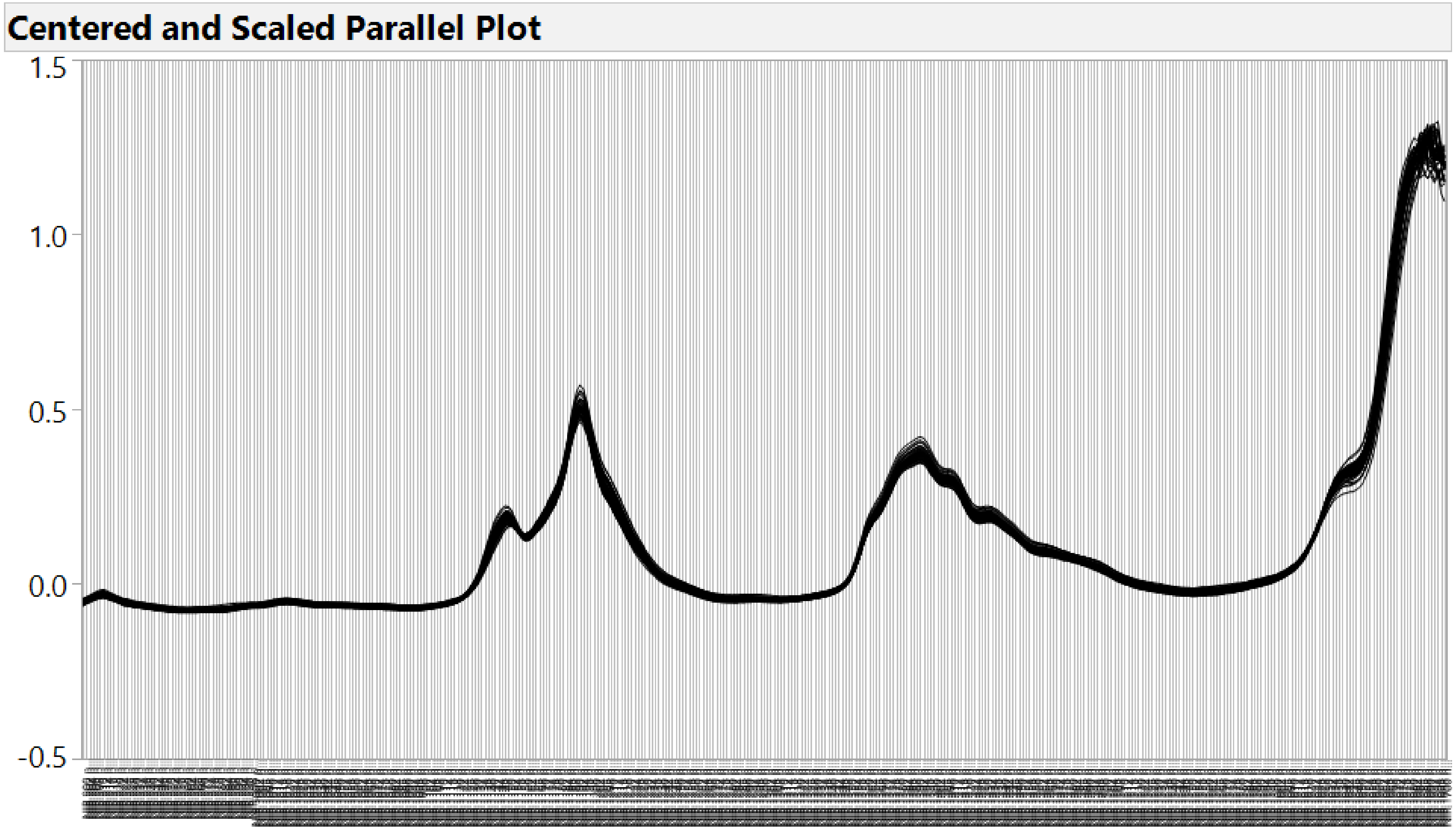
- Why would you want to do the analysis differently?
  - Time savings
  - Getting as good or better predictive model with about 25 percent of the data.
  - Using penalized regression helps determine the important wavelengths.

Steps

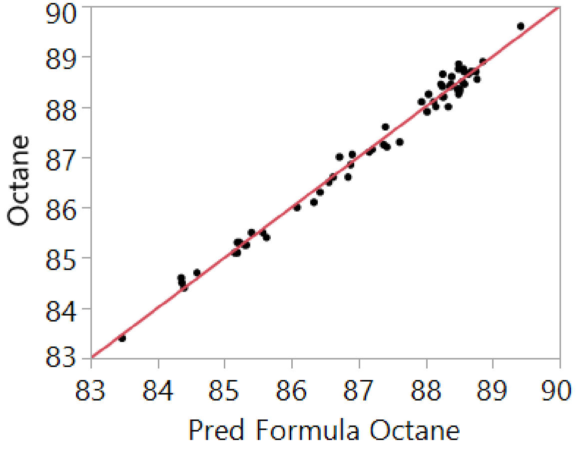
1. Take spectra of the desired samples. No need for output (Y) information at this point.
2. Identify most prominent dimensions of variation in the data with Functional Principle Components, f(PC's), from Functional Data Explorer – not for use in the traditional sense!
3. Create an experimental design using the f(PC's) as factors (covariates)
4. Run the experiment
5. Model the results via PLS, and/or Generalized Regression (or other methods able to handle correlated factors)
6. Determine the overall optimum solution
7. Use this sustainable model to determine the "active" for all future samples*

* - the model will hold true for samples analyzed using the same calibrated instrument

60 GASOLINE NIR SPECTRA OVERLAID



Bivariate Fit of Octane By Pred Formula Octane



Linear Fit

Linear Fit

Octane = -1.14e-13 + 1*Pred Formula Octane

Summary of Fit

RSquare	0.987963
RSquare Adj	0.987756
Root Mean Square Error	0.16931
Mean of Response	87.1775
Observations (or Sum Wgts)	60

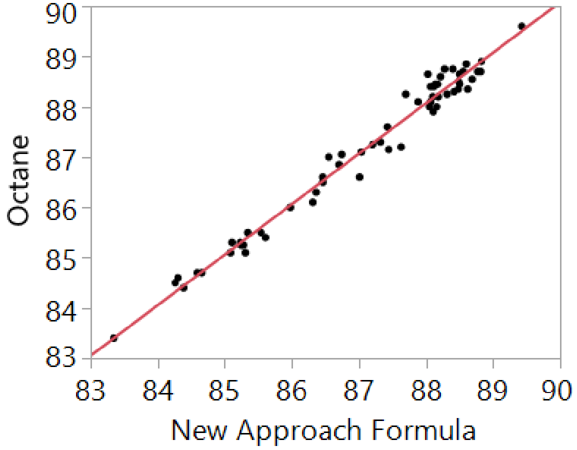
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	136.46451	136.465	4760.529
Error	58	1.66262	0.029	Prob > F
C. Total	59	138.12712		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.14e-13	1.263694	-0.00	1.0000
Pred Formula Octane	1	0.014493	69.00	<.0001*

Bivariate Fit of Octane By New Approach Formula



Linear Fit

Linear Fit

Octane = -0.242858 + 1.0035597*New Approach Formula

Summary of Fit

RSquare	0.979139
RSquare Adj	0.978779
Root Mean Square Error	0.222891
Mean of Response	87.1775
Observations (or Sum Wgts)	60

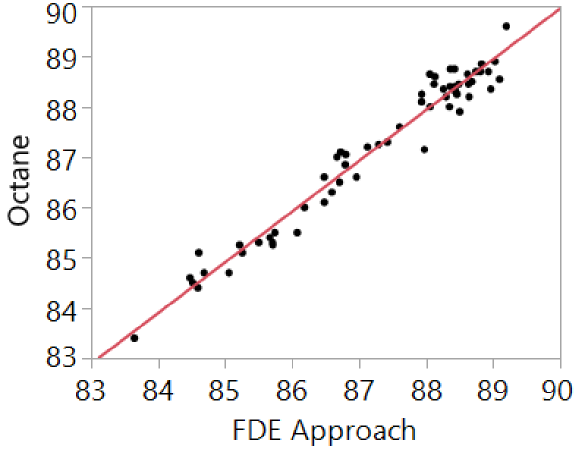
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	135.24567	135.246	2722.325
Error	58	2.88145	0.050	Prob > F
C. Total	59	138.12712		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.242858	1.67574	-0.14	0.8853
New Approach Formula	1.0035597	0.019234	52.18	<.0001*

Bivariate Fit of Octane By FDE Approach



Linear Fit

Linear Fit

Octane = -0.913466 + 1.0096427*FDE Approach

Summary of Fit

RSquare	0.960984
RSquare Adj	0.960311
Root Mean Square Error	0.304822
Mean of Response	87.1775
Observations (or Sum Wgts)	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	132.73797	132.738	1428.574
Error	58	5.38915	0.093	Prob > F
C. Total	59	138.12712		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.913466	2.330998	-0.39	0.6966
FDE Approach	1.0096427	0.026713	37.80	<.0001*

JMP Demo

CONCLUSION

- Compress the available information regarding spectral wavelengths of different options via functional principal components
- Use covariate DOE to select the “corners of the box” for testing representative sample spectra based on Design of Experiments
- Model the data via PLS – Generalized Regression is an excellent option
- Find the overall optimum solution
- As new samples are tested the spectral data can be input into the data table to determine level of active or desired component.

→ **Highly efficient experimentation**

→ **Sustainable empirical model based on spectral data/wavelengths**