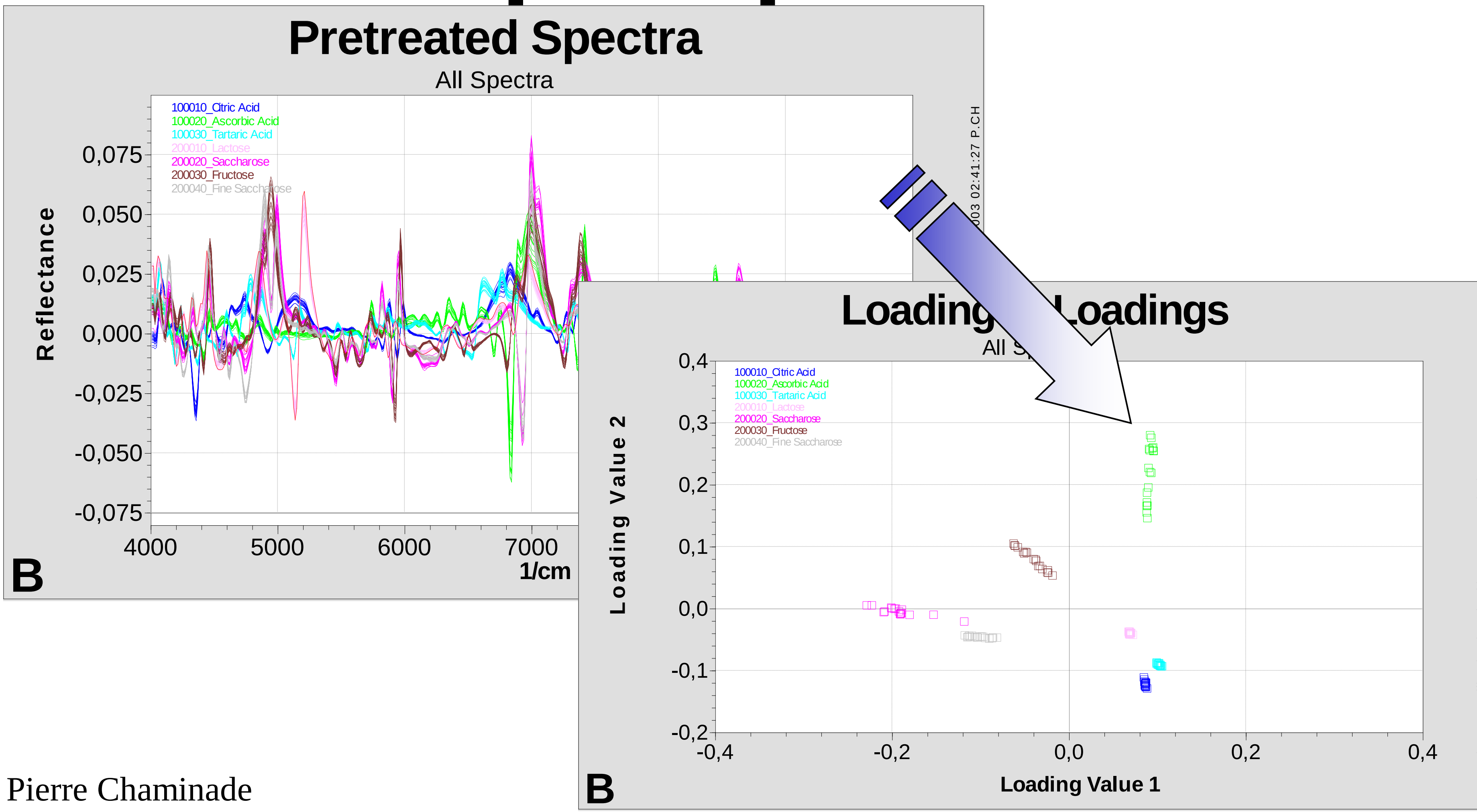
Analyse en composantes principales



Pierre Chaminade
M2 PHARMACOTECHNIE – BIOPHARMACIE
2020

Présentation

- L'ACP permet de réduire un système complexe de corrélations à un plus petit nombre de dimensions.
- Permet de générer des variables latentes non corrélées

Données de départ

- Des individus caractérisés par un certain nombre de variables
  - Laquelle est la plus importante pour caractériser les individus ?
  - Sont elles toutes pertinentes et/ou corrélées ?

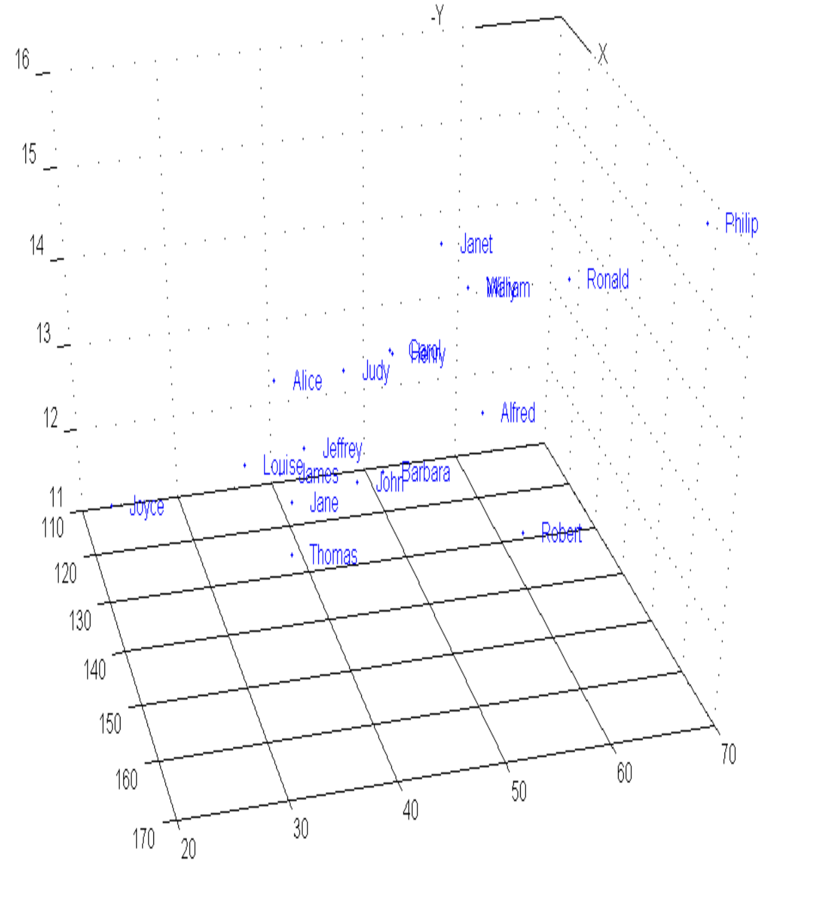
Prénom	Taille	Poids	Age
Alfred	155	51	14
Alice	124	38	13
Barbara	146	44	13
Carol	140	46	14
Henry	141	46	14
James	126	38	12
Jane	132	38	12
Janet	139	51	15
Jeffrey	139	38	13
John	130	45	12
Joyce	110	23	11
Judy	143	41	14
Louise	123	35	12
Mary	149	51	15
Philip	163	68	16
Robert	145	58	12
Ronald	150	60	15
Thomas	126	39	11
William	149	51	15

Corrélations

- Les 3 variables sont dépendantes les unes des autres.

	TAILLE	POIDS	AGE
TAILLE	1		
POIDS	0,87779	1	
AGE	0,81143	0,74089	1

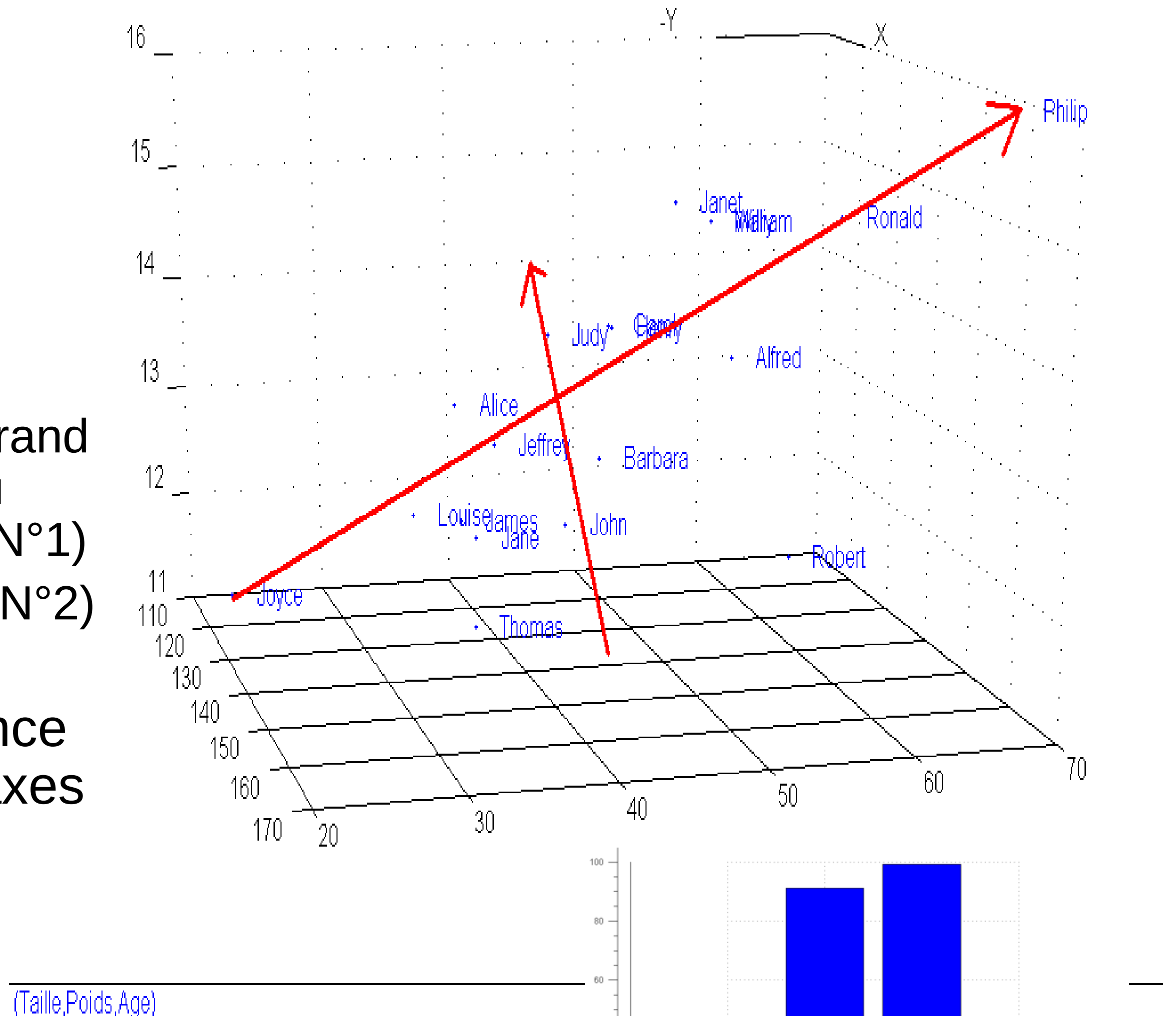
- La situation n’est peut être pas si compliquée.

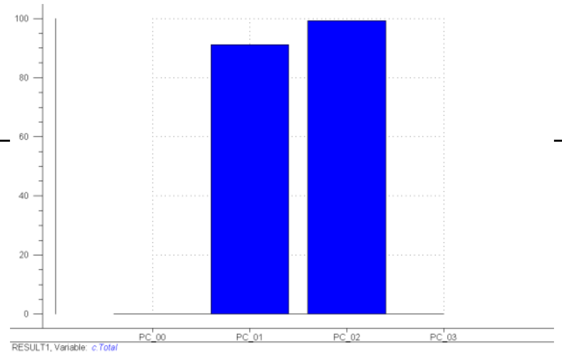


(Taille,Poids,Age)

Que fait l'ACP ?

- Calcul de nouvelles variables: les composantes principales.
  - Géométriquement,
    - on cherche le plus grand sens d'élongation du nuage de point (CP N°1)
    - Puis, le second (CP N°2)
  - Et on regarde le pourcentage de variance « expliqué » par ces axes (variables)
    - PC 1 91%
    - PC 1+2 99%

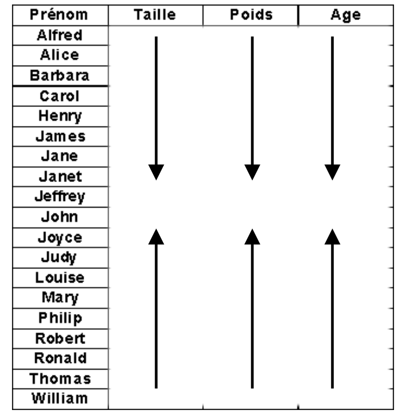




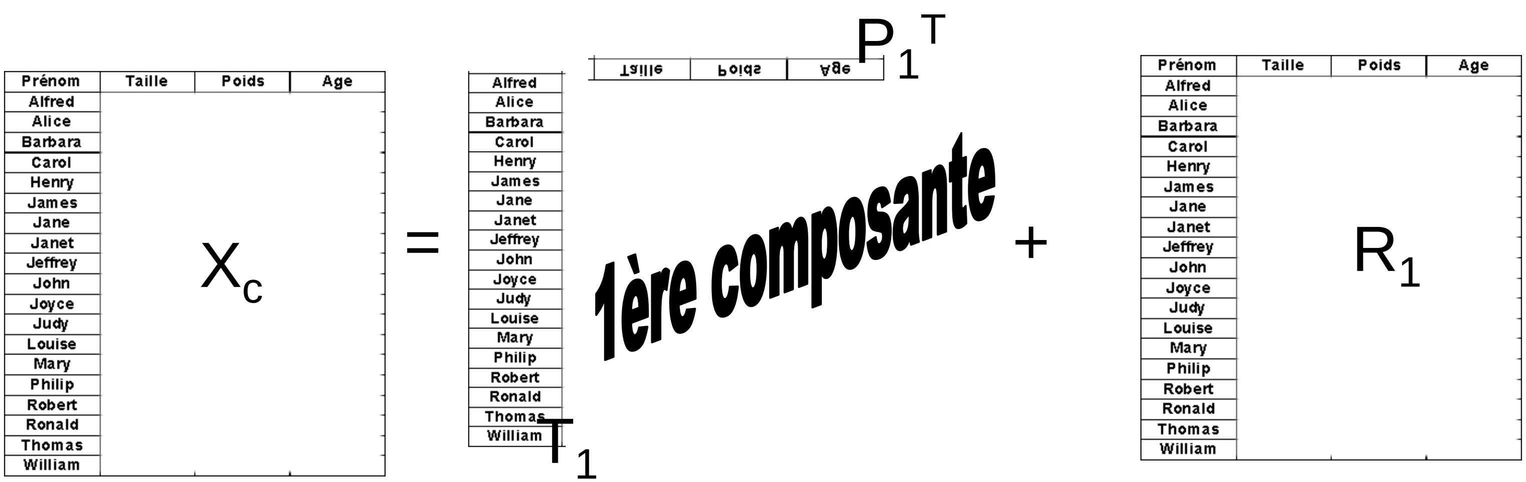
Le calcul…

Prénom	Taille	Poids	Age
Alfred	155	51	14
Alice	124	38	13
Barbara	146	44	13
Carol	140	46	14
Henry	141	46	14
James	126	38	12
Jane	132	38	12
Janet	139	51	15
Jeffrey	139	38	13
John	130	45	12
Joyce	110	23	11
Judy	143	41	14
Louise	123	35	12
Mary	149	51	15
Philip	163	68	16
Robert	145	58	12
Ronald	150	60	15
Thomas	126	39	11
William	149	51	15

Matrice X

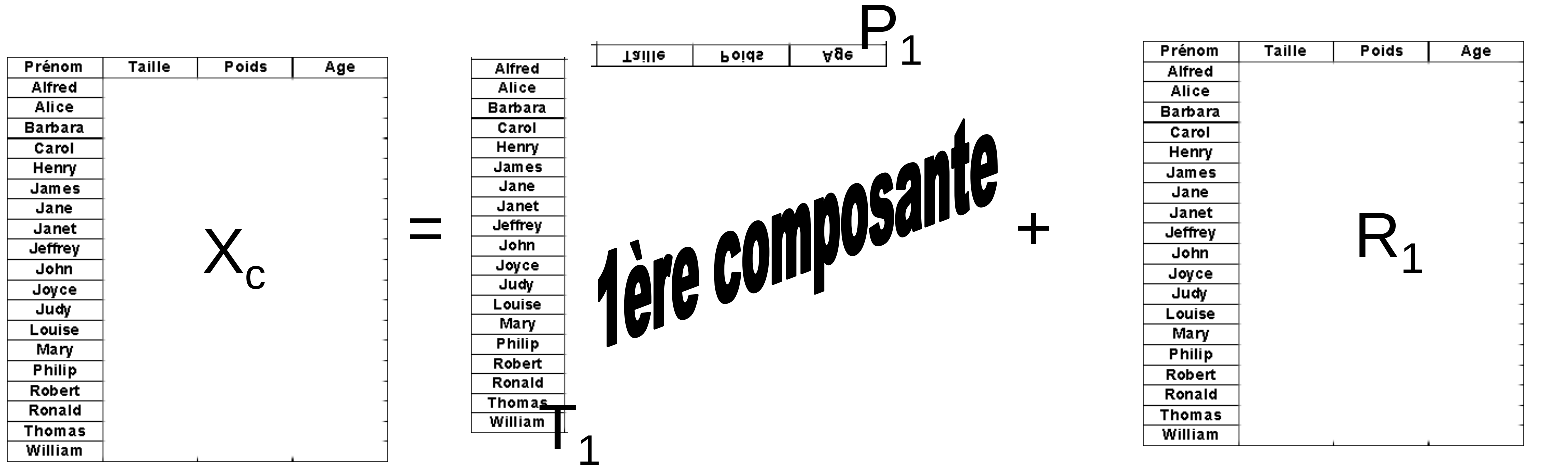


Matrice Xc centrée sur la moyenne variable par variable

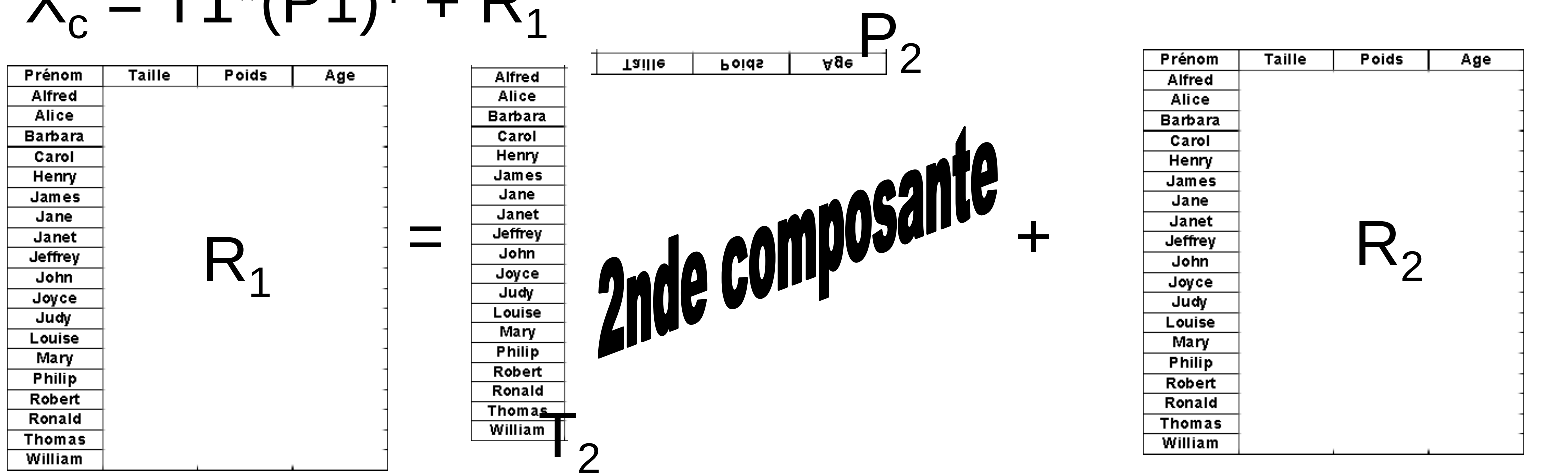


$$X_c = T1*(P1)^T + R_1$$

Le calcul est itératif



$X_c = T1*(P1)^T + R_1$



$R_1 = T_2.(P_2)^T + R_2$

$R_{n-1} = T_n.(P_n)^T + R_n$

P. CHAMINADE M2: Pharmacotechnie et Biopharmacie… Faculté de Pharmacie Univ. Paris-Saclay.

Le calcul conduit a:

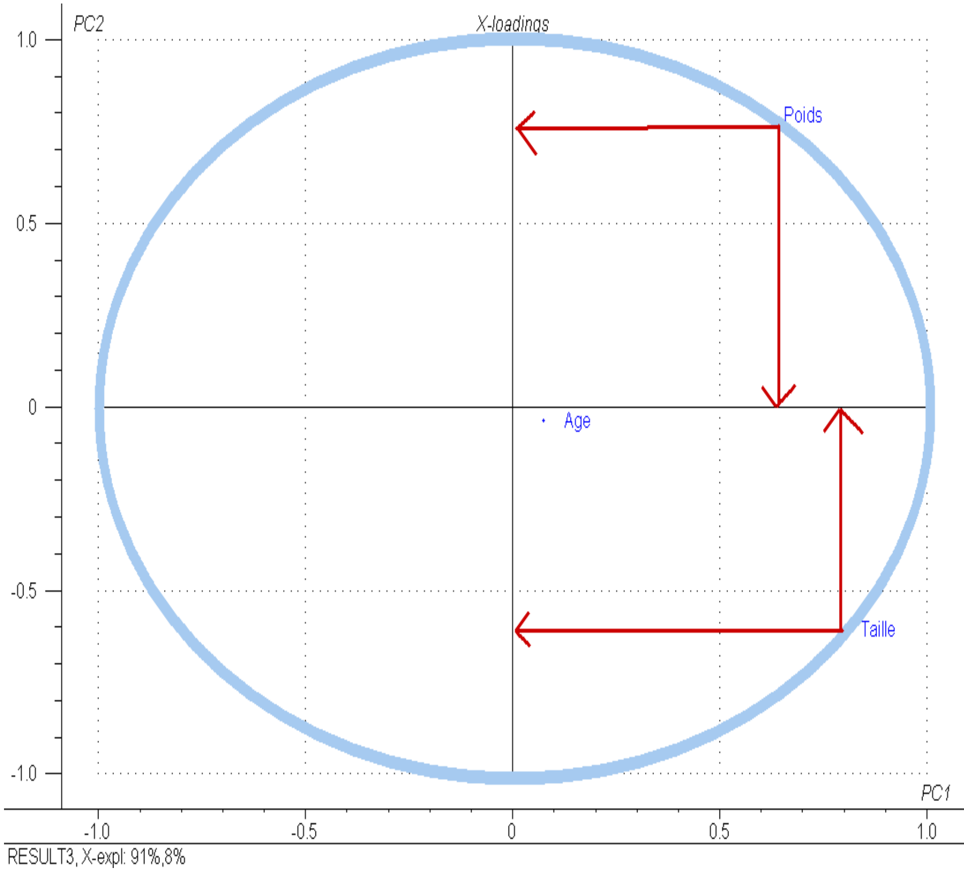
- Une modèlisation des données par
  - Un ensemble de CP (T et P) qui
    - contiennent une part décroissante d’information
      - Importante pour les premières composantes
      - Mineure pour les dernières
    - Ne sont pas corrélées (sont indépendantes les unes des autres)
- Les loadings représentent le « poids » des variables de départ dans chaque composante
- Les scores représentent la position des individus sur chacune des composantes

Dans la pratique

- Le calcul s'effectue par minimisation de la SCE pour calculer scores et loadings à chaque étape.
- Une fraction des individus doit être laissée en dehors du processus de calcul pour vérifier la convergence de l'algorithme « validation set »
- Les individus qui servent au calcul correspondent au « calibration set »
- Face à un petit nombre d'individus, on peut utiliser la technique de « cross validation » ou « leave one out ».

Loading plot

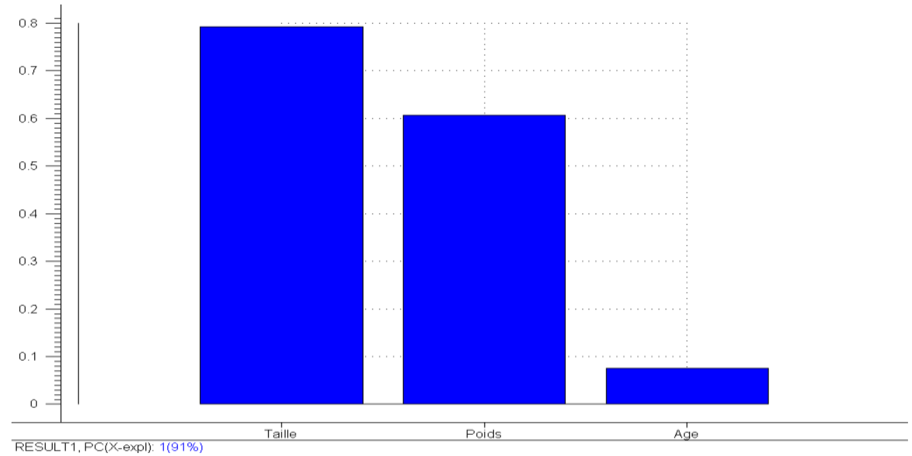
- Positionne les variables d'origine dans le nouveau système
  - Ici, on « voit » 99% des données
  - La taille est la variable qui corrèle le plus avec le nouvel axe. L'axe 1 contient aussi l'information « poids ».
  - L'axe 2 oppose taille/poids
  - L'age ne semble pas être un critère pertinent pour décrire nos individus

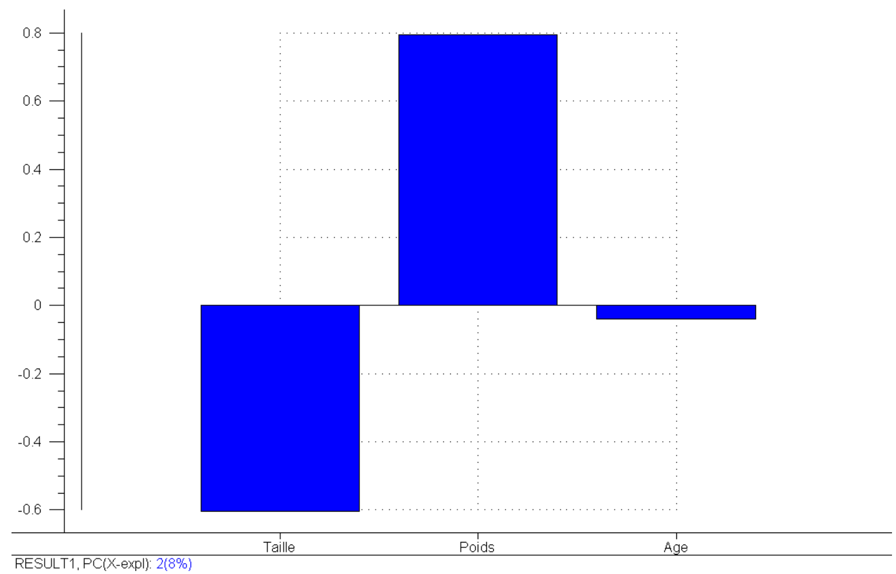


Corrélations.

	TAILLE	POIDS	AGE
TAILLE	1		
POIDS	0,87779	1	
AGE	0,81143	0,74089	1

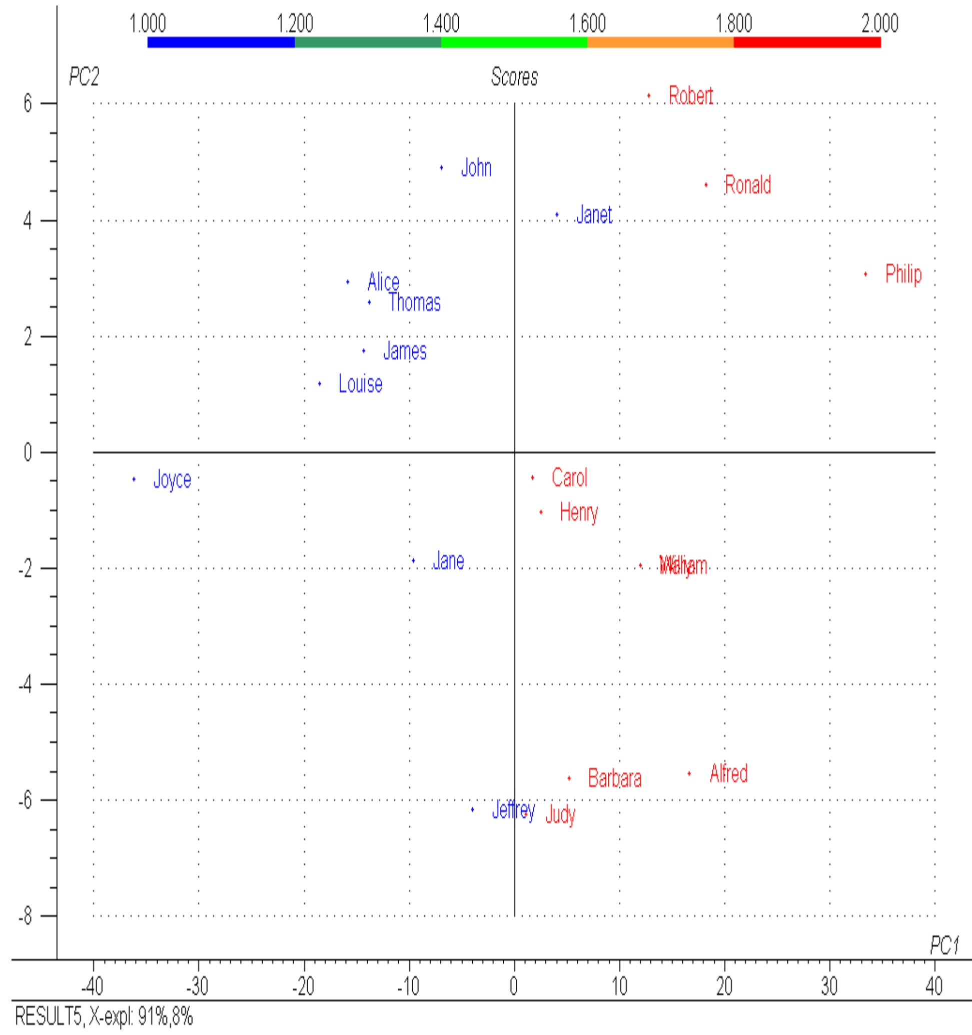
- On retrouve la corrélation taille-poids sur le loading 1
- Sur le 2 ces variables sont anticorrélées





Score plot

- Les coordonnées de nos individus dans le système.
  - En rouge individus de taille ≥ 1,40 m
  - En bleu ≤ 1,40 m



Prénom	Taille	Poids	Age
Philip	163	68	16
Robert	145	58	12
Jeffrey	139	38	13
Joyce	110	23	11

Exemple 1

- Développement de nanoparticules* par voie orale pour un API faiblement hydrosoluble (telmisartan).
- *Nanosuspension stabilisée par des surfactants
- Combinaison d'un plan factoriel et d'une ACP

- Plan: 3²
  - X1 stabilisant %w/v
  - X2 agent de broyage en g
- Réponses:
  - Taille de particules
  - indice de polydispersité (PI)
  - Solubilité à saturation
  - Potentiel Zeta

Table 1
Design layout of 3^2 full factorial design batches for TLM loaded nanosuspensions.

Batch code	Transformed values	
	X_1[a]	X_2[b]
TLM-NS-F1	−1	−1
TLM-NS-F2	−1	0
TLM-NS-F3	−1	1
TLM-NS-F4	0	−1
TLM-NS-F5	0	0
TLM-NS-F6	0	1
TLM-NS-F7	1	−1
TLM-NS-F8	1	0
TLM-NS-F9	1	1

Jaydeep Patel, Anjali Dhingani, Kevin Garala, Mihir Raval, Navin Sheth,
Design and development of solid nanoparticulate dosage forms of telmisartan for bioavailability enhancement by integration of experimental design and principal component analysis,
Powder Technology, Volume 258, May 2014, Pages 331-343

Table 2
Responses of 3^2 full factorial design batches of TLM loaded nanosuspensions.

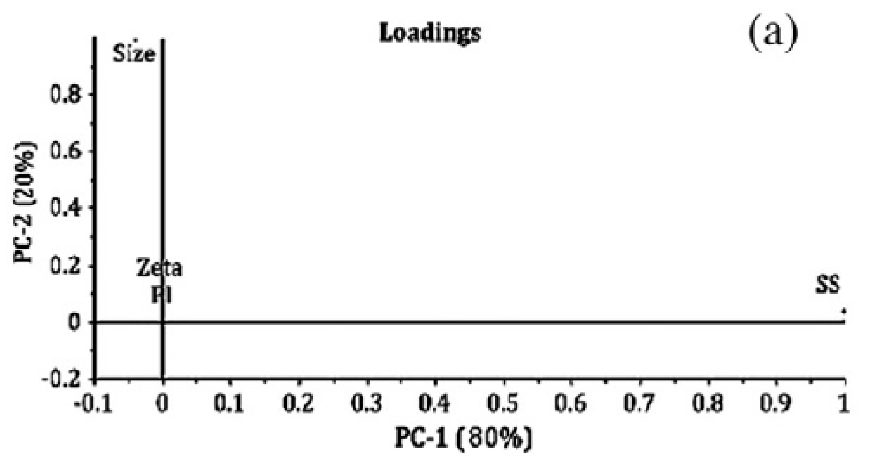
Batch code	Particle size (nm)	Saturation solubility (μg/mL)	Polydispersibility index	Zeta potential (m
TLM-NS-F1	470.67 ± 10.60	1211.23 ± 54.73	0.18 ± 0.07	−27.55 ± 2.67
TLM-NS-F2	423.67 ± 8.43	2190.14 ± 76.56	0.18 ± 0.03	−29.90 ± 3.78
TLM-NS-F3	406.00 ± 4.65	2837.47 ± 98.54	0.13 ± 0.04	−31.55 ± 5.61
TLM-NS-F4	390.33 ± 6.87	3056.09 ± 101.66	0.18 ± 0.03	−32.73 ± 3.67
TLM-NS-F5	356.67 ± 8.40	4431.10 ± 114.70	0.10 ± 0.07	−34.85 ± 6.52
TLM-NS-F6	330.33 ± 6.88	4438.26 ± 109.78	0.12 ± 0.02	−35.76 ± 1.55
TLM-NS-F7	424.00 ± 5.32	2630.13 ± 71.00	0.16 ± 0.02	−30.63 ± 3.70
TLM-NS-F8	386.00 ± 8.43	3073.91 ± 96.12	0.15 ± 0.04	−32.46 ± 6.54
TLM-NS-F9	367.33 ± 9.11	3876.25 ± 89.80	0.14 ± 0.02	−32.10 ± 1.43

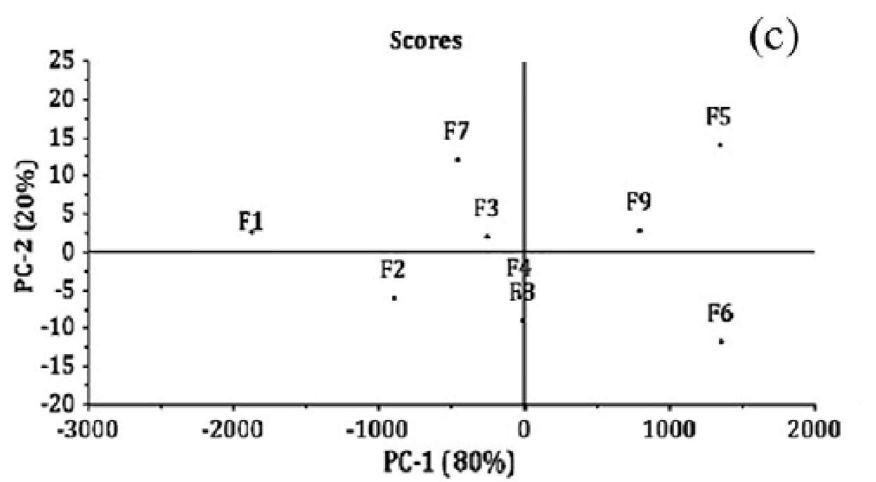
The results are mean ± SD (n = 3).

Table 3
Summary of ANOVA study for TLM loaded nanosuspensions.

Particle size (Y_1)				
	DF[a]	*SSR*[b]	*MS*[c]	$R^2 = 0.9962$ Fcal = 0.91
Regression				Fcritical = 10.13 DF = (1, 3)
FM[d]	5	13,965.58	2793.12	
RM[e]	4	13,953.33	3488.33	
Residual				
FM	3	40.42	13.47	
RM	4	52.67	13.17	
Saturation solubility (Y_2)				
	DF	*SSR*	*MS*	$R^2 = 0.9560$ Fcal = 0.80
Regression				Fcritical = 9.55 DF = (2, 3)
FM	5	8,597,271	1,719,454	
RM	3	8,461,416	2,820,472	
Residual				
FM	3	253,091.1	84363.7	
RM	5	388,946.7	77,789.33	

[a] DF: Degree of freedom.
[b] SSR: Sum of square residuals.
[c] MS: Mean of squares.
[d] FM: Full model.
[e] RM: Reduced model.





Kevin C. Garala, Jaydeep M. Patel, Anjali P. Dhingani, Abhay T. Dharamsi, Preparation and evaluation of agglomerated crystals by crystallo-co-agglomeration: An integrated approach of principal component analysis and Box–Behnken experimental design, International
Journal of Pharmaceutics, Volume 452, Issues 1–2, 16 August 2013, Pages 135-156

CCA: Cristallisation de l'API et agglomération avec les excipients.

3 variables sélectionnées après screening:

Ratio CH2Cl2 / eau
Concentration du PEG6000
Vitesse d'agitation

Table 1
Box–Behnken design batches.

Batch code	Independent variables		
	X_1 [a]	X_2 [b]	X_3 [c]
R1	−1	−1	0
R2	1	−1	0
R3	−1	1	0
R4	1	1	0
R5	−1	0	−1
R6	1	0	−1
R7	−1	0	1
R8	1	0	1
R9	0	−1	−1
R10	0	1	−1
R11	0	−1	1
R12	0	1	1
R13	0	0	0

Factor	Level		
	−1	0	1
X_1 (DCM: water)	0.04	0.07	1
X_2 (Concentration of PEG 6000, %w/w)	2	3.5	5
X_3 (Speed, rpm)	600	800	1000

Kevin C. Garala, Jaydeep M. Patel, Anjali P. Dhingani, Abhay T. Dharamsi, Preparation and evaluation of agglomerated crystals by crystallo-co-agglomeration: An integrated approach of principal component analysis and Box–Behnken experimental design, International
Journal of Pharmaceutics, Volume 452, Issues 1–2, 16 August 2013, Pages 135-156

Results of evaluation parameters of pristine and agglomerated RCD.

Batch*	RCD	R1	R2	R3	R4	R5	R6
dg[a] (mm)	0.137 ± 0.103	0.697 ± 0.142	0.892 ± 0.139	0.725 ± 0.127	0.973 ± 0.116	1.009 ± 0.168	1.214 ± 0.193
CS[b] (gm)	–	46.27 ± 3.66	35.41 ± 1.51	51.07 ± 2.69	42.18 ± 2.38	49.32 ± 2.75	39.24 ± 1.98
SF[c]	0.415 ± 0.052	0.826 ± 0.083	0.809 ± 0.067	0.901 ± 0.065	0.896 ± 0.073	0.938 ± 0.071	0.857 ± 0.053
CI[d]	39.48 ± 2.05	14.57 ± 1.29	17.84 ± 2.67	17.09 ± 1.68	15.43 ± 1.34	14.55 ± 1.06	17.73 ± 1.25
%Y[e]	–	94.59 ± 2.07	88.17 ± 2.66	90.43 ± 1.92	96.25 ± 1.43	98.16 ± 2.61	91.64 ± 2.75
AoR[f] (°)	45.63 ± 3.65	23.81 ± 1.38	22.94 ± 2.07	19.07 ± 1.76	20.18 ± 1.59	18.59 ± 1.67	22.45 ± 0.98
HR[g]	1.613 ± 0.132	1.141 ± 0.011	1.176 ± 0.013	1.165 ± 0.014	1.154 ± 0.012	1.150 ± 0.013	1.205 ± 0.027
CF[h]	0.354 ± 0.038	1.015 ± 0.271	0.973 ± 0.184	0.991 ± 0.143	0.987 ± 0.156	1.089 ± 0.117	1.107 ± 0.128
IF[i]	2.017 ± 0.073	2.936 ± 0.115	2.975 ± 0.129	3.024 ± 0.134	3.066 ± 0.128	2.891 ± 0.119	3.032 ± 0.083
Φ[j]	0.075 ± 0.012	0.856 ± 0.107	0.832 ± 0.083	0.907 ± 0.074	0.883 ± 0.051	0.940 ± 0.087	0.895 ± 0.075
DC[k] (%)	–	98.54 ± 1.62	99.35 ± 1.75	95.67 ± 1.23	100.23 ± 1.07	97.84 ± 1.68	97.17 ± 1.64
MC[l] (%)	0.324 ± 0.101	0.735 ± 0.134	0.864 ± 0.153	1.113 ± 0.203	1.011 ± 0.196	0.996 ± 0.173	0.965 ± 0.135
a[m]	17.455 ± 1.06	5.154 ± 0.638	9.457 ± 0.921	5.181 ± 0.315	5.378 ± 0.423	4.735 ± 0.437	6.348 ± 0.539
$1/b$[m]	1.375 ± 0.073	0.752 ± 0.064	0.528 ± 0.027	0.441 ± 0.092	0.541 ± 0.078	0.994 ± 0.083	0.645 ± 0.035
K[n]	0.079 ± 0.015	0.974 ± 0.093	0.953 ± 0.084	0.931 ± 0.102	0.987 ± 0.113	1.025 ± 0.132	0.983 ± 0.064
A[o]	1.375 ± 0.073	0.752 ± 0.064	0.528 ± 0.027	0.441 ± 0.092	0.541 ± 0.078	0.994 ± 0.083	0.645 ± 0.035
k[o]	0.079 ± 0.015	0.974 ± 0.093	0.953 ± 0.084	0.931 ± 0.102	0.987 ± 0.113	1.025 ± 0.132	0.983 ± 0.064
Py[o] (tons)	29.24 ± 1.83	1.711 ± 0.09	1.739 ± 0.12	2.020 ± 0.22	1.594 ± 0.08	1.516 ± 0.09	1.639 ± 0.17
σ_0[o]	9.746 ± 1.02	0.570 ± 0.16	0.579 ± 0.19	0.673 ± 0.17	0.531 ± 0.15	0.505 ± 0.18	0.546 ± 0.12
ER[p] (%)	4.453 ± 0.497	0.815 ± 0.165	0.731 ± 0.234	0.725 ± 0.186	0.697 ± 0.127	0.554 ± 0.201	0.735 ± 0.179
TS[q] (kg/cm^2)	1.253 ± 0.131	5.215 ± 0.652	4.942 ± 0.341	5.697 ± 0.763	6.015 ± 0.434	4.684 ± 0.372	5.248 ± 0.421

Batch*	R7	R8	R9	R10	R11	R12	R13
dg[a] (mm)	0.351 ± 0.083	0.723 ± 0.091	1.102 ± 0.104	1.175 ± 0.157	0.437 ± 0.113	0.566 ± 0.115	0.834 ± 0.138
CS[b] (gm)	47.01 ± 2.65	38.44 ± 2.83	58.89 ± 3.61	64.35 ± 3.89	43.73 ± 2.37	62.98 ± 3.95	61.58 ± 3.52
SF[c]	0.992 ± 0.084	0.970 ± 0.077	0.792 ± 0.052	0.876 ± 0.047	0.841 ± 0.061	0.912 ± 0.043	0.959 ± 0.051
CI[d]	16.33 ± 1.78	15.61 ± 0.92	14.05 ± 0.83	15.84 ± 1.29	15.56 ± 1.15	14.24 ± 0.83	14.74 ± 0.73
%Y[e]	91.07 ± 2.94	96.65 ± 1.32	97.35 ± 2.73	92.17 ± 1.61	89.83 ± 1.36	93.55 ± 2.83	97.23 ± 2.97
AoR[f] (°)	17.84 ± 1.32	17.06 ± 2.17	17.51 ± 1.38	21.23 ± 1.52	22.35 ± 1.13	18.97 ± 1.39	18.41 ± 1.26
HR[g]	1.154 ± 0.016	1.144 ± 0.015	1.096 ± 0.011	1.155 ± 0.014	1.143 ± 0.012	1.136 ± 0.011	1.128 ± 0.013
CF[h]	0.999 ± 0.135	0.989 ± 0.113	0.982 ± 0.126	0.974 ± 0.149	0.987 ± 0.134	0.993 ± 0.148	1.054 ± 0.117
IF[i]	3.112 ± 0.201	2.959 ± 0.124	2.872 ± 0.139	2.983 ± 0.097	3.013 ± 0.133	3.047 ± 0.116	2.981 ± 0.124
Φ[j]	0.999 ± 0.054	0.981 ± 0.037	0.728 ± 0.056	0.868 ± 0.048	0.876 ± 0.074	0.912 ± 0.063	0.965 ± 0.055
DC[k] (%)	98.42 ± 2.09	99.68 ± 1.48	97.75 ± 1.96	98.23 ± 1.93	101.03 ± 1.61	99.58 ± 1.37	98.24 ± 1.46
MC[l] (%)	0.935 ± 0.158	0.918 ± 0.127	0.815 ± 0.139	0.997 ± 0.162	0.834 ± 0.137	1.017 ± 0.193	0.973 ± 0.175
a[m]	9.366 ± 0.832	2.854 ± 0.261	7.656 ± 0.733	5.335 ± 0.342	2.806 ± 0.303	2.847 ± 0.376	3.739 ± 0.294
$1/b$[m]	0.814 ± 0.054	0.969 ± 0.032	0.796 ± 0.051	0.640 ± 0.072	0.514 ± 0.061	0.624 ± 0.049	0.756 ± 0.051
K[n]	1.214 ± 0.083	0.995 ± 0.097	0.953 ± 0.108	0.941 ± 0.093	1.106 ± 0.112	0.967 ± 0.094	0.985 ± 0.082
A[o]	0.814 ± 0.054	0.969 ± 0.032	0.796 ± 0.051	0.640 ± 0.072	0.514 ± 0.061	0.624 ± 0.049	0.756 ± 0.051
k[o]	1.214 ± 0.083	0.995 ± 0.097	0.953 ± 0.108	0.941 ± 0.093	1.106 ± 0.112	0.967 ± 0.094	0.985 ± 0.082
Py[o] (tons)	1.971 ± 0.09	1.437 ± 0.07	1.806 ± 0.15	1.462 ± 0.06	1.421 ± 0.07	1.208 ± 0.06	1.367 ± 0.08
σ_0[o]	0.656 ± 0.13	0.479 ± 0.11	0.604 ± 0.17	0.487 ± 0.15	0.473 ± 0.14	0.402 ± 0.16	0.455 ± 0.17
ER[p] (%)	0.836 ± 0.193	0.565 ± 0.152	0.468 ± 0.103	0.578 ± 0.118	0.549 ± 0.134	0.605 ± 0.156	0.765 ± 0.149
TS[q] (kg/cm^2)	5.367 ± 0.587	6.164 ± 0.832	5.654 ± 0.426	4.964 ± 0.319	5.548 ± 0.571	5.218 ± 0.385	5.584 ± 0.463

[a] Mean geometric diameter.
[b] Crushing strength.
[c] Shape factor.
[d] Carr's Index
[e] Percent yield.
[f] Angle of repose.
[g] Hausner ratio.
[h] Circularity factor
[i] Irregularity factor.
[j] Aspect ratio.
[k] Drug content.
[l] Moisture content.
[m] Kawakita parameters.
[n] Kuno's constant.
[o] Heckel parameters.
[p] Elastic recovery.
[q] Tensile strength.
[*] Results are mean of three determinations ±SD.

Kevin C. Garala, Jaydeep M. Patel, Anjali P. Dhingani, Abhay T. Dharamsi, Preparation and evaluation of agglomerated crystals by crystallo-co-agglomeration: An integrated approach of principal component analysis and Box–Behnken experimental design, International
Journal of Pharmaceutics, Volume 452, Issues 1–2, 16 August 2013, Pages 135-156

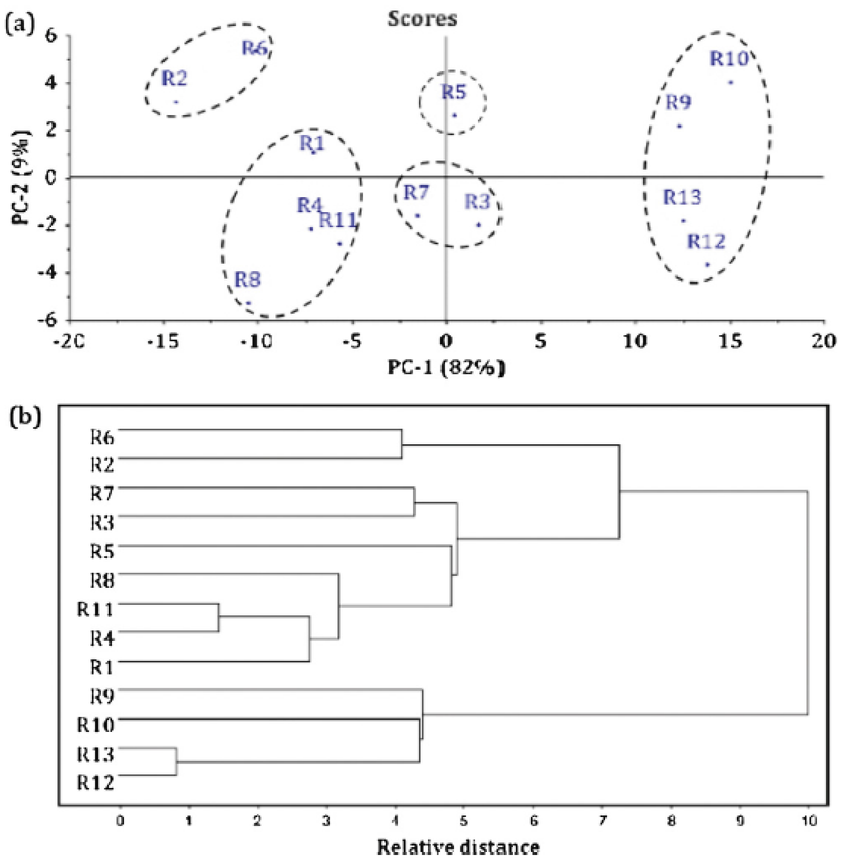


Fig. 3. Score plot from PCA (a) and Dendrogram from AHCA (b) of agglomerated RCD.

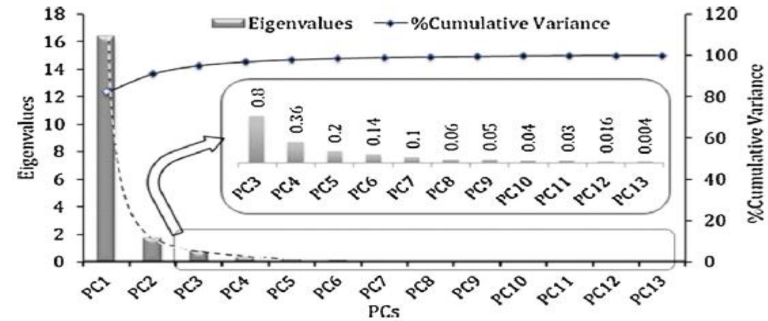
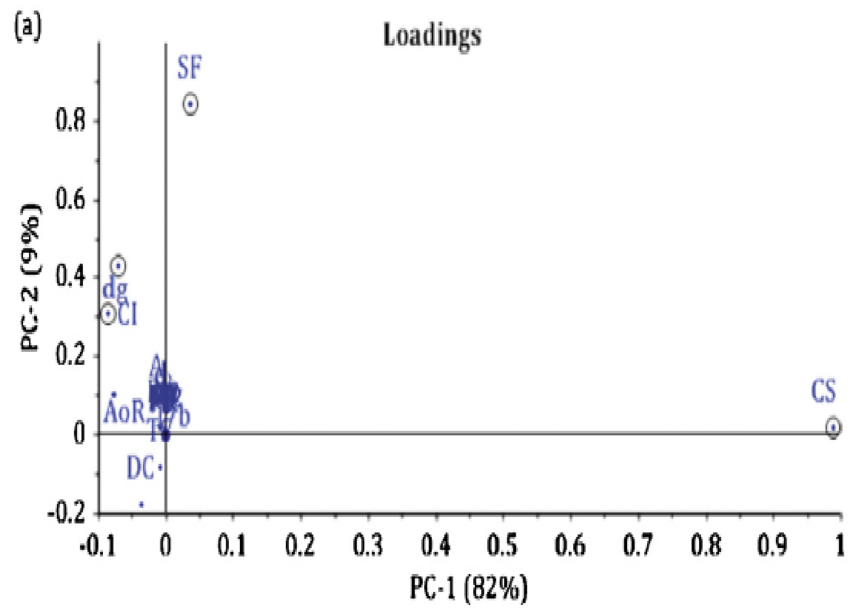
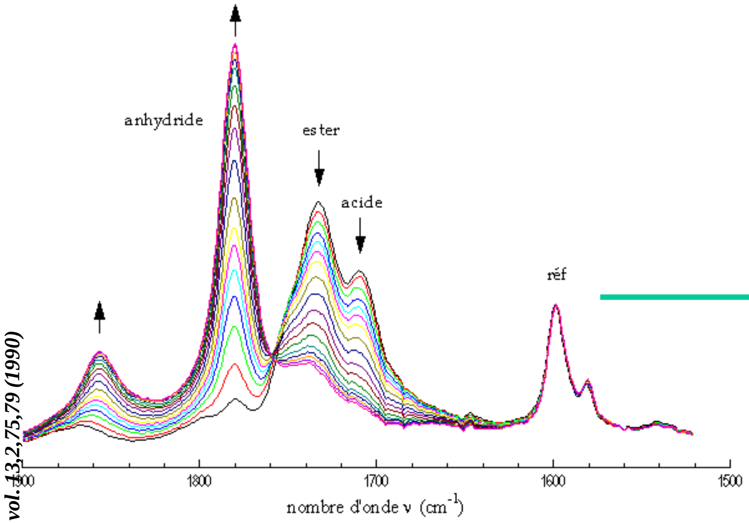


Fig. 7. Scree plot of all components with cumulative variance.

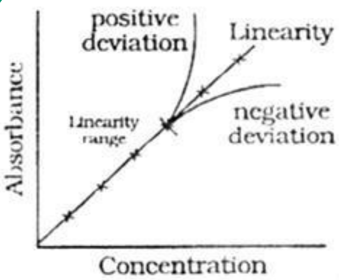


Hard vs soft modeling

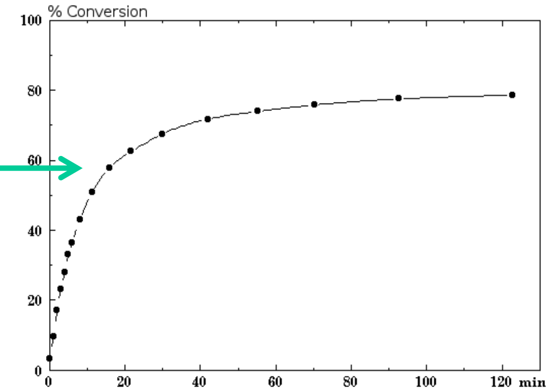


Réactions in the molten state, M. Lamba, J. Druz er F. Mazeres, Plastics and rubber processing and applications, vol. 13,2,75,79 (1990)

- Modélisation « dure »



$$A = \sum_{i=1}^{n} A_i(\varepsilon_{\lambda,i}, l, c_i)$$



Hard vs soft modeling

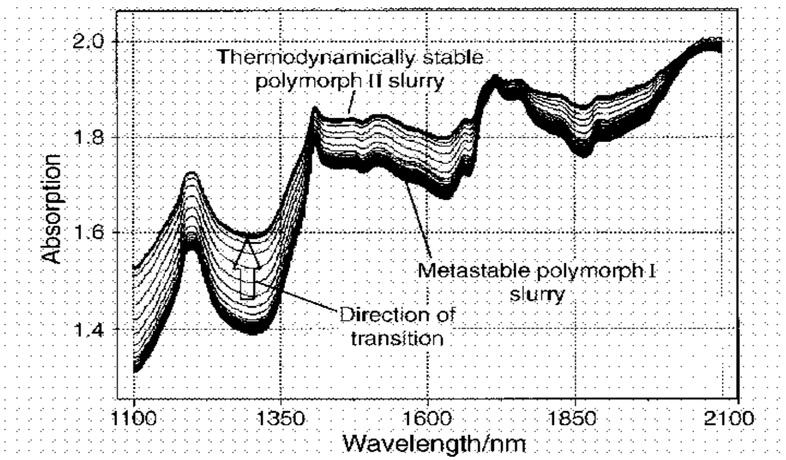
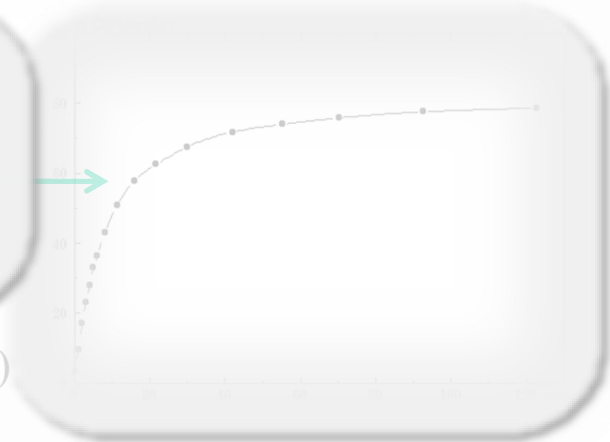


Fig. 2 Typical crystal slurry spectra collected during conversion of trovafloxacin mesylate polymorph I into polymorph II in butanol at 95 °C.

- Modélisation « dure »

$$A = \sum_{i=1}^{n} A_i(\varepsilon_{\lambda,i}, l, c_i)$$



Modèle chimiométrique multivarié

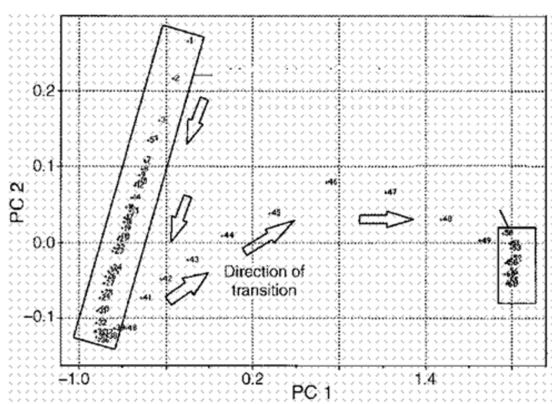


Fig. 3 Scores plot of principal component 1 (PC 1) *versus* principal component 2 (PC 2) for typical conversion.

- Modélisation « comportementale » ou « soft modeling ».

Timothy Norris, Paul K Aldridge and S. Sonja Sekulic.
Analyst june 1997 vol 122 (549-552)

Hea-Eun Lee, Min-Jeong Lee, Woo-Sik Kim, Myung-Yung Jeong, Young-Sang Cho, Guang Jin Choi,
In-line monitoring and interpretation of an indomethacin anti-solvent crystallization process by near-infrared spectroscopy (NIRS),
International Journal of Pharmaceutics, Volume 420, Issue 2, 28 November 2011, Pages 274-281

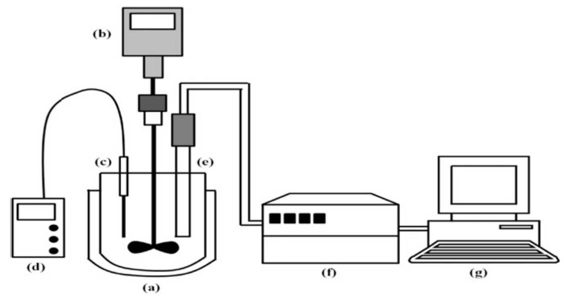


Fig. 1. A schematic diagram of experimental apparatus: (a) glass jacketed re
(b) agitator, (c) temperature probe, (d) temperature indicator, (e) NIR fiber-
probe, (f) NIR main body and (g) personal computer.

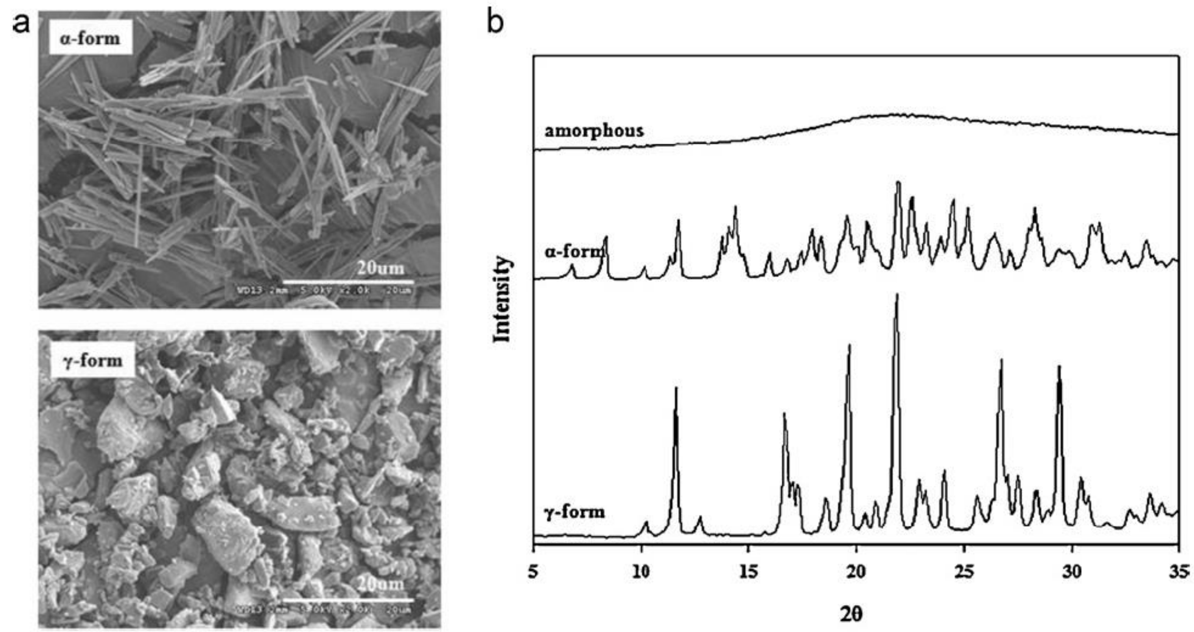


Fig. 2. SEM photographs and XRD patterns of standard IMC powders.

Hea-Eun Lee, Min-Jeong Lee, Woo-Sik Kim, Myung-Yung Jeong, Young-Sang Cho, Guang Jin Choi,
In-line monitoring and interpretation of an indomethacin anti-solvent crystallization process by near-infrared spectroscopy (NIRS),
International Journal of Pharmaceutics, Volume 420, Issue 2, 28 November 2011, Pages 274-281

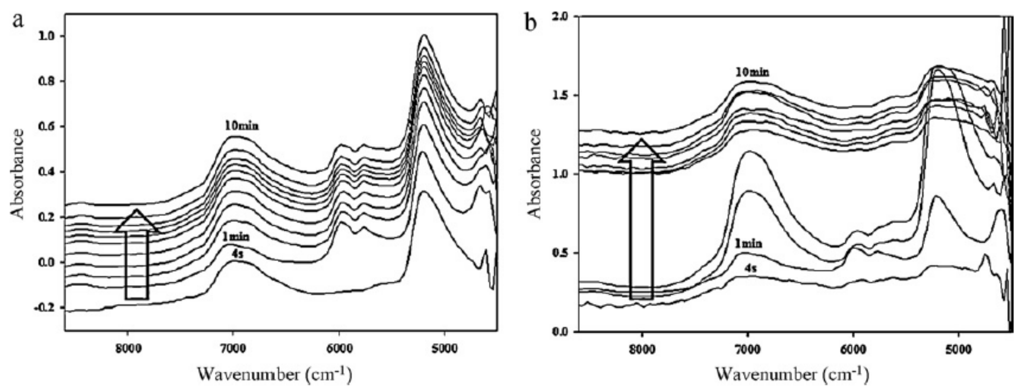


Fig. 3. NIR spectra acquired during IMC crystallization process (the first spectrum at 4 s and the rest 10 spectra at every minute from 1 to 10): (a) S-to-A scheme and (b) A-to-S scheme.

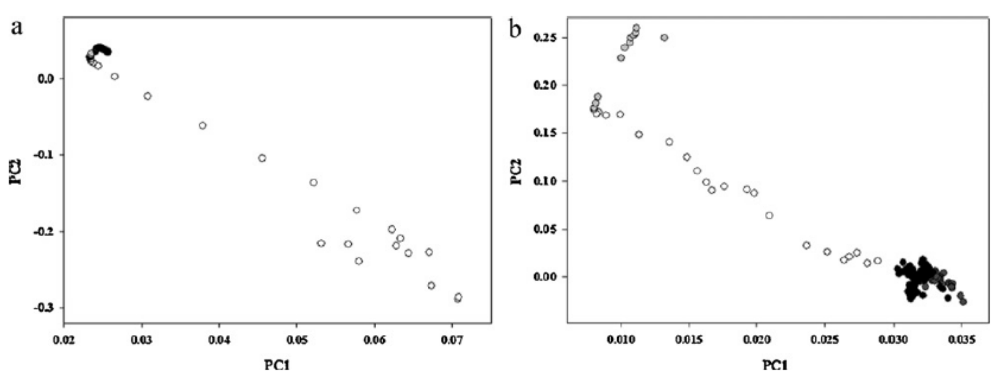


Fig. 4. PC score plots of IMC crystallization process: (a) S-to-A scheme and (b) A-to-S scheme.

"Although pure α-form IMC powder was resulted under A-to-S scheme, a mixture of the α-form and γ-form was produced for S-to-A case."