Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Quantitative structure-property relationship approach to predicting xylene separation with diverse exchanged faujasites

Y. Khabzina[a], C. Laroche[a], J. Pérez-Pellitero[a], D. Farrusseng[b*]

[a] IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France

[b] Université de Lyon, Université Claude Bernard Lyon 1, CNRS, IRCELYON – UMR 5256, 2 Avenue Albert Einstein, 69626 Villeurbanne Cedex, France.

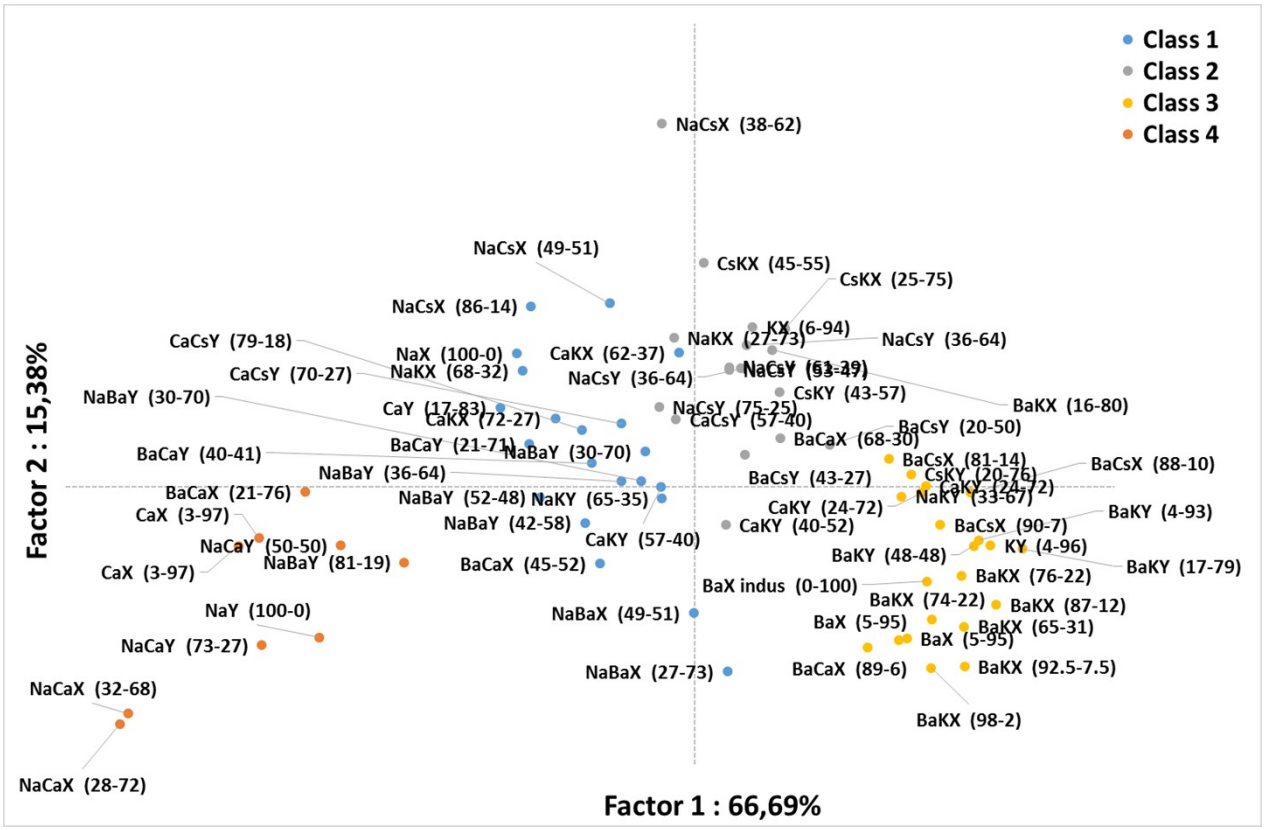

Figure S1. Cluster analysis of 68 exchanged faujasite zeolites based on their xylene separation properties

Table S1. Corresponding faujasites

Number	Faujasite	Class
1	NaX (100-0)	1
2	NaKX (68-32)	1
3	NaKY (65-35)	1

4	NaCsX (86-14)	1
5	NaCsX (49-51)	1
6	CaY (17-83)	1
7	NaBaX (49-51)	1
8	NaBaX (27-73)	1
9	NaBaY (52-48)	1
10	NaBaY (42-58)	1
11	NaBaY (36-64)	1
12	NaBaY (30-70)	1
13	NaBaY (30-70)	1
14	CaKX (62-37)	1
15	CaKX (72-27)	1
16	CaKY (57-40)	1
17	CaCsY (70-27)	1
18	CaCsY (79-18)	1
19	BaCaX (45-52)	1
20	BaCaY (21-71)	1
21	BaCaY (40-41)	1
22	NaY (100-0)	4
23	NaCaX (32-68)	4
24	NaCaX (28-72)	4
25	CaX (3-97)	4
26	CaX (3-97)	4
27	NaCaY (73-27)	4
28	NaCaY (50-50)	4
29	NaBaY (81-19)	4
30	BaCaX (21-76)	4
31	NaKX (27-73)	2
32	KX (6-94)	2
33	NaCsY (75-25)	2
34	NaCsX (38-62)	2
35	NaCsY (61-39)	2
36	NaCsY (53-47)	2
37	NaCsY (36-64)	2
38	NaCsY (36-64)	2
39	CsKX (25-75)	2
40	CsKX (45-55)	2
41	CsKY (43-57)	2
42	CaKY (40-52)	2
43	CaCsY (57-40)	2
44	BaKX (16-80)	2
45	BaCsY (20-50)	2
46	BaCsY (43-27)	2
47	BaCaX (68-30)	2
48	NaKY (33-67)	3
49	KY (4-96)	3

50	CsKY (20-76)	3
51	BaX (5-95)	3
52	BaX (5-95)	3
53	BaX indus (0-100)	3
54	CaKY (24-72)	3
55	CaKY (24-72)	3
56	BaKX (76-22)	3
57	BaKX (65-31)	3
58	BaKX (87-12)	3
59	BaKX (74-22)	3
60	BaKX (92.5-7.5)	3
61	BaKX (98-2)	3
62	BaKY (4-93)	3
63	BaKY (17-79)	3
64	BaKY (48-48)	3
65	BaCsX (81-14)	3
66	BaCsX (90-7)	3
67	BaCsX (88-10)	3
68	BaCaX (89-6)	3

Procedure for calculating the accessible volume

1) The positions of the cations within monocationic systems were obtained by geometric optimization performed with the Vienna Ab initio Simulation Package (VASP). This software determines the stable structures of faujasites using DFT (density functional theory) calculations for periodic systems. The PAW method was used to describe the ion-electron interactions, and the PW91 form for the generalized gradient approximation (GGA) was used to determine the exchange energy and correlation energy. The cut-off radius value for the kinetic energy was set to 500 eV, while sampling of the Brillouin zone was limited to the point. Energy convergence was considered to be achieved when all atomic forces were below 0.02 $eVÅ^{-1}$. The calculations were carried out in a cell containing 48 tetrahedral units (i.e., two supercages). Finally, aluminum positions were selected in such a way as to fulfill Loewenstein's law while preventing the presence of Al-O-Al bonds.

2) The cationic distribution was obtained from exchange isotherm models or from analytical results when available.

3) Next, the calculation of the descriptor expressing the void space of the zeolites was carried out using the "Atom Volumes & Surfaces" tool included in the Materials Studio software. To ensure that only the volume of the supercages was considered, the sodalite cages as well as the hexagonal prisms were blocked by "phantom" atoms that prevented any insertion of the probe molecule. The size of the probe molecule (1.4 Å) was chosen because it is close to the distance between a xylene molecule and the crystallographic adsorption sites.

METHODS

1.1. Principal component analysis

Principal component analysis (PCA) is a factorial method that reveals the correlations between variables[i]. This statistical method computes new variables, called principal components (PCs), which are uncorrelated. These PCs are linear combinations of the existing variables, and to each PC corresponds a variance. The first PC has the highest variance. Therefore, only PCs with high variances are retained. This method makes it possible to represent huge amounts of data using a minimum number of variables. In addition, PCA is illustrated by graphics that assist in the visualization of the correlations between the variables.

1.2. Optimal regression – Furnival and Wilson's algorithm

This algorithm is specifically employed to sort through predictive variables using a stepwise method. For a number of variables that does not exceed 40, the procedure consists in adding successive variables to the best regression with one variable, in order to obtain the highest possible correlation coefficient (R^2)[i]. This procedure supplies several model scenarios. The

selections of variables that appear in the models are based on the results of Student and Fisher significance tests[i]. This method provides a graphical tool that shows the evolution of R^2 as a function of the number of variables and that designates the relevant variables that should be included in the model.

1.3. Multiple linear regression

Multiple linear regression is an algorithm that provides a linear model between a predictable variable (y) and a set of predictive variables (x_k) as follows:

$$y^{pred} = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$$

y and x are continuous variables, b_0 is a constant, and b_k are the coefficients of the regression calculated by the algorithm. Estimation of these coefficients is carried out using the method of least squares by minimizing $\sum_{i=1}^{n}(y_i^{exp} - y_i^{pred})^2$.

This statistical approach illustrates the degree of dependence of the predictive variables (x) on the predictable variable (y) and thereby evaluates the goodness of fit by measuring the correlation coefficient R^2 between the experimental values (y_i^{exp}) and the predicted values (y_i^{pred}), according to the definition:

$$R^2 = \frac{\sum_i (y_i^{pred})^2}{\sum_i (y_i^{exp})^2}$$

Moreover, the standard deviation (s.d.) criterion, which corresponds to the degree of error in the predicted model, is calculated as follows:

$$s.d = \sqrt{\frac{1}{i\_tot}\sum_{i}(y_i^{exp} - y_i^{pred})^2}$$

In addition to the overall measurement of the quality of the prediction through the calculation of R^2, multiple linear regression provides another criterion that tests the statistical significance of the regression coefficients obtained, through a Student's t-test. The purpose is to assess the influence of the predictive variables (x) on the predictable variable (y) by calculating a probability (p). The probability threshold is usually set to 0.05. If the calculated probability (p) is below this value of 0.05, then the hypothesis that the variable (x) has no influence on the variable (y) is rejected.

1.4. Discriminant analysis

Discriminant analysis is a predictive technique that consists in allocating individuals into identified classes. This attribution is based on a linear model that correlates the distribution of the classes (qualitative variables) with the predictive variables (quantitative/qualitative).

We note that the algorithms for all of the above statistical methods were implemented in SPAD, a commercial software program.

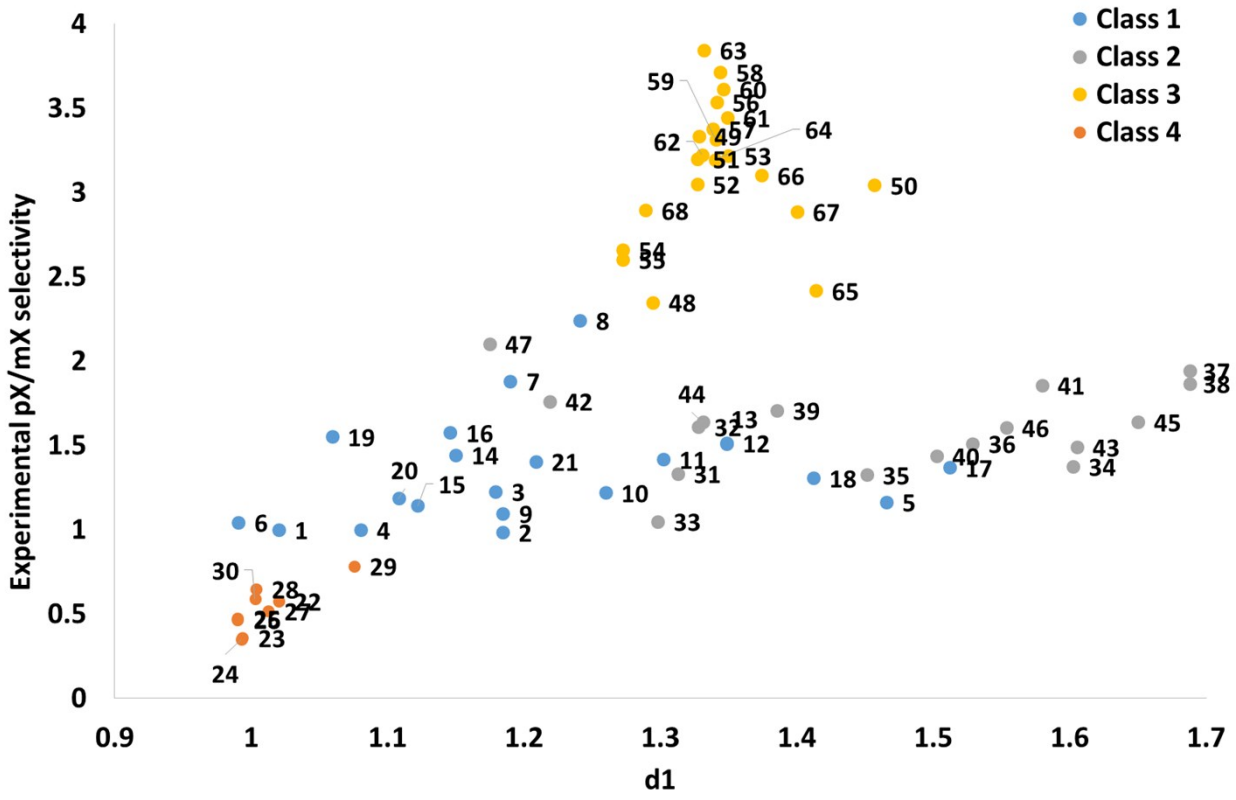

Figure S2. Evolution of experimental pX/mX as a function of d1

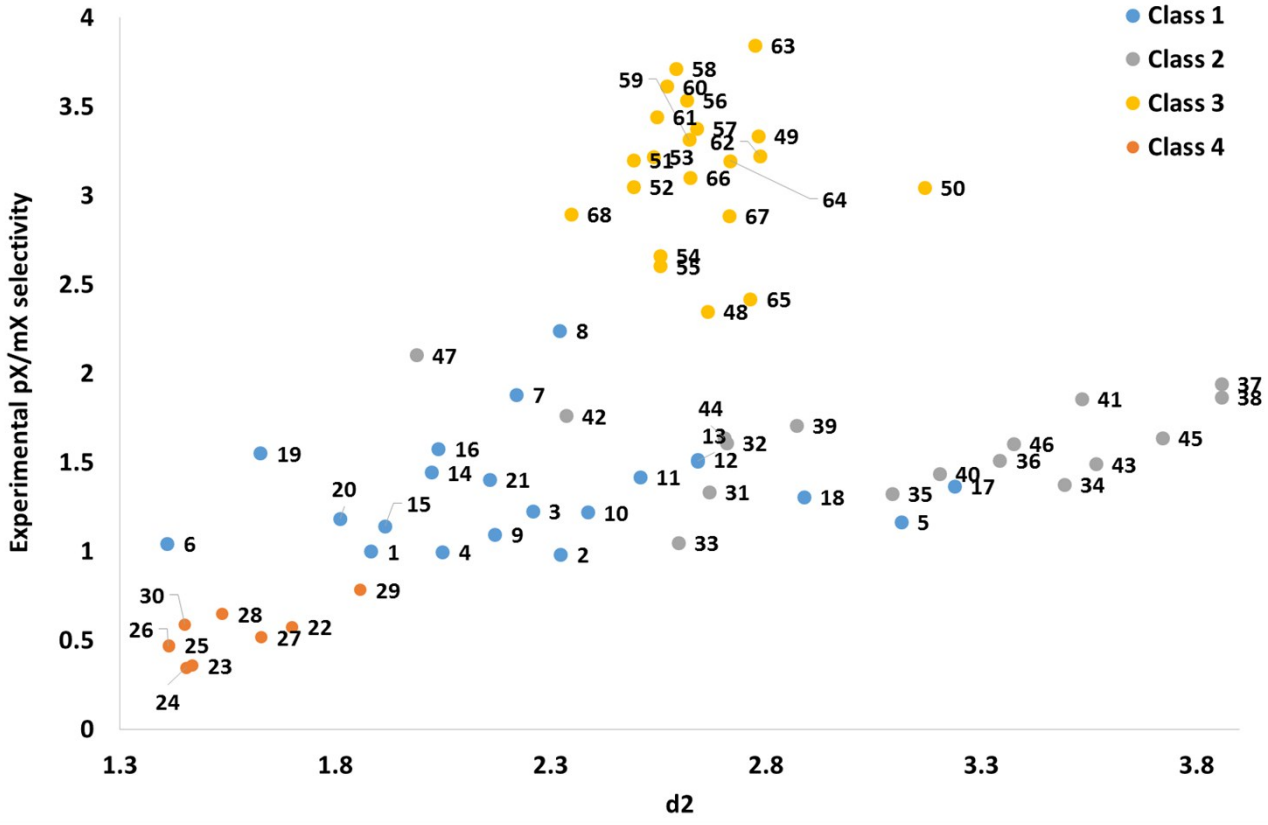

Figure S3. Evolution of experimental pX/mX as a function of d_2

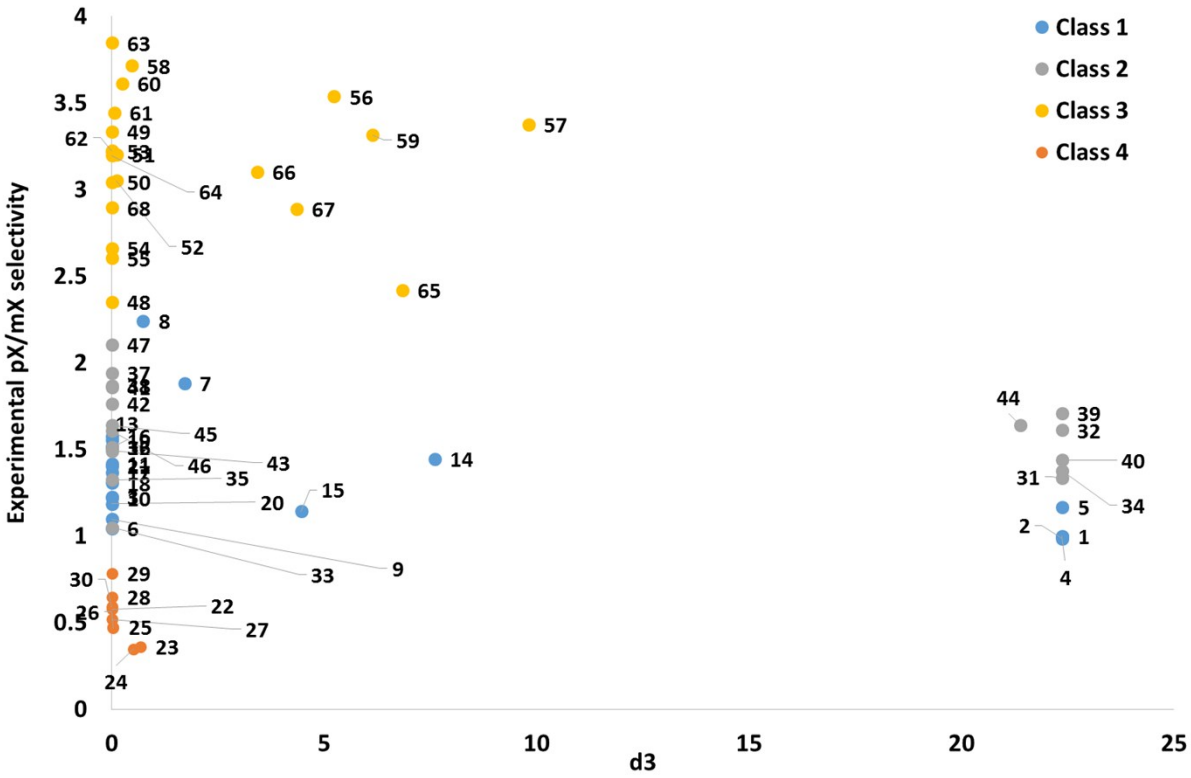

Figure S4. Effect of the occupancy of sites III on the pX/mX selectivity

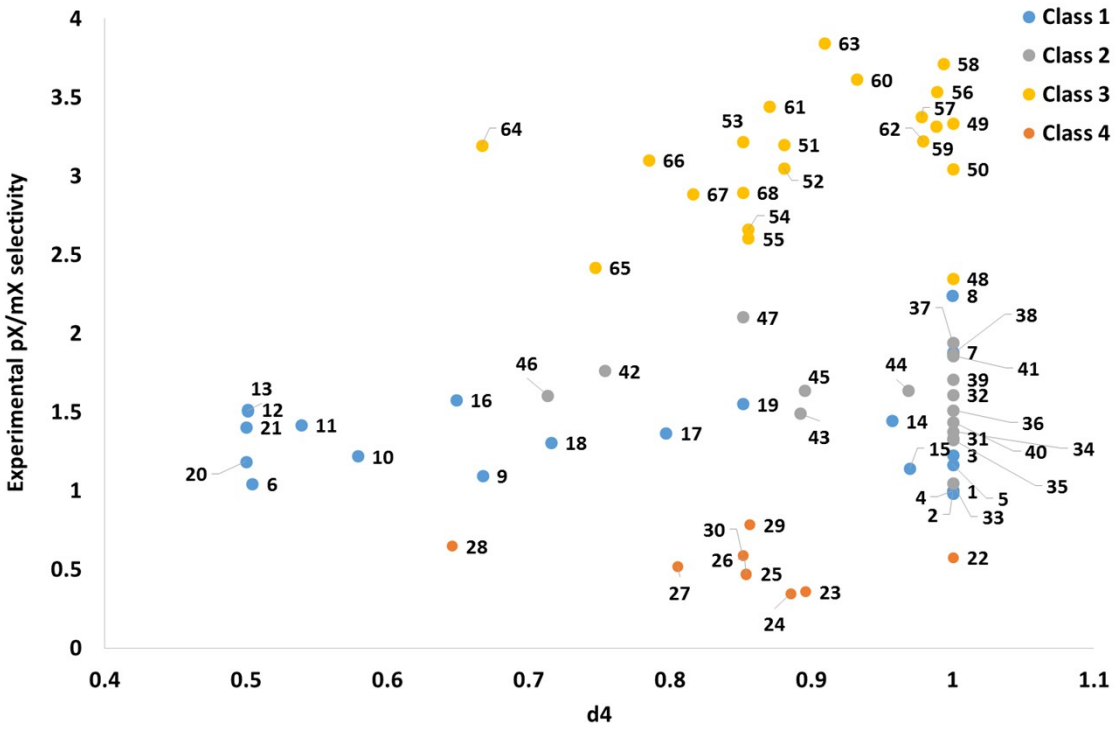

Figure S5. Evolution of pX/mX selectivity as a function of SII saturation

Table S2. Cationic distribution established using the exchange isotherm models

Crystallographic sites	SI,I'	SII	SIII	pX/mX

KX (6-94)	4.8 Na 27 K	31.4 K	22 K	1.61
KY	1.8 Na 18 K	32 K		3.34
NaKX (27-73)	19.5 Na 12.3 K	3.7 Na 28.1 K	21.8 K	1.33
NaKY (33-67)	13.1 Na 6.7 K	3.7 Na 28.6 K		2.35
BaKX (16-80)	5.7 Ba 17 K	1.3 Ba 28.9 K	22.7 K	1.64
BaKX (87-12)	15.9 Ba	21.6 Ba 10.9 K		3.61

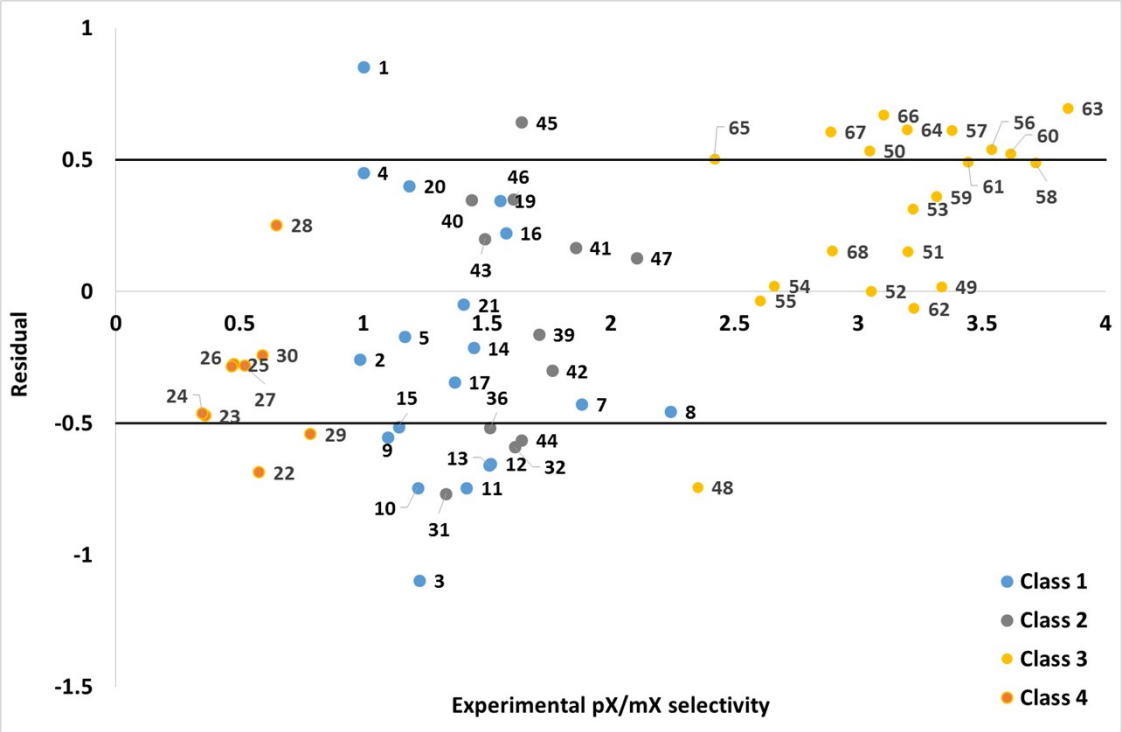

Figure S6. Residual plot

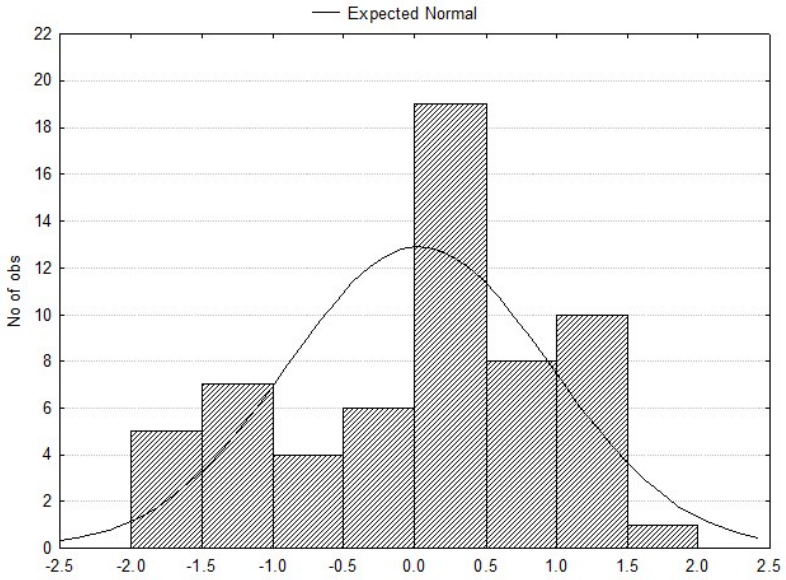

Figure S7. Distribution of standard residuals

---

i L. Lebart, M. Piron, A. Morineau, Statistique exploratoire multidimensionnelle, 4ème édition DUNOD