

# Contents

<b>I. Preface</b>	<b>1</b>
<b>1. Justification, objectives and contributions</b>	<b>3</b>
1.1. Justification	3
1.2. Objectives	4
1.2.1. Traditional and latent variable-based approaches applied to mixture design problems	4
1.2.2. Latent variable-based approaches for efficient processes optimization	5
1.2.3. Latent variable-based approaches applied to the Quality by Design initiative, to increase processes flexibility and guarantee the desired quality	5
1.3. Contributions	6
<b>2. On optimal design of experiments</b>	<b>9</b>
2.1. Introduction	9
2.2. Optimality criteria for the design of experiments	10
2.2.1. D-optimal design of experiments	11
2.2.2. I-optimal and G-optimal design of experiments	13
2.3. Additional considerations for the design of experiments	14
2.3.1. Variance inflation	14
2.3.2. Aliasing	15
2.3.3. Principles of effect-sparsity, hierarchy and heredity	15
<b>3. On latent variable and kernel-based multivariate data analysis</b>	<b>17</b>
3.1. Introduction	17
3.2. Latent variable-based multivariate data analysis techniques	18
3.2.1. Principal Component Analysis (PCA)	18
3.2.2. Partial Least Squares regression (PLS)	19
3.3. Kernel-based techniques	20

3.3.1. Basic principles of kernel-based techniques	21
3.3.2. Pseudo-samples and pseudo-sample projection	22
3.4. Important additional notions: cross-validation and jackknifing	23
<b>4. Materials and methods</b>	<b>25</b>
4.1. Hardware	25
4.2. Software	25
4.3. Datasets and methods	25
<b>II. Mixture design optimization</b>	<b>27</b>
<b>5. Traditional approaches to mixture design</b>	<b>29</b>
5.1. Introduction	29
5.2. The mixture space	31
5.2.1. Assessing the shape of the mixture space	31
5.2.2. Identifying the envelope of the mixture space	34
5.3. Regression model structures in mixture design	37
5.3.1. The Scheffé models	37
5.3.2. The Cox models	44
5.3.3. Mixture-process variable models	46
5.4. Mixture design of experiments	48
5.4.1. Simplex-based DOE	48
5.4.2. DOE in irregular mixture spaces	53
<b>6. Latent variable-based methods for mixture data analysis</b>	<b>57</b>
6.1. Mixture design data analysis with Partial Least Squares	57
6.1.1. Methods and datasets	58
6.1.1.1. <i>Example 1: seven-component octane blending experiment of Cornell</i>	58
6.1.1.2. <i>Example 2: gasoline blending data of Snee</i>	59
6.1.2. Results and discussion	59

---

6.1.2.1. <i>Example 1: seven-component octane blending experiment of Cornell</i>	59
6.1.2.2. <i>Example 2: gasoline blending data of Snee</i>	64
6.1.3. Conclusions and additional considerations	70
6.2. Kernel-PLS and pseudo-sample trajectories	70
6.2.1. Methods	71
6.2.1.1. <i>Pseudo-sample trajectories for mixture data</i>	71
6.2.1.2. <i>Pseudo-sample-based response surface and Scheffé model coefficients</i>	73
6.2.2. Datasets	73
6.2.2.1. <i>Data simulated according to a second-order polynomial model</i>	74
6.2.2.2. <i>Tablet data</i>	74
6.2.2.3. <i>Bubbles data</i>	75
6.2.2.4. <i>Colorant data</i>	75
6.2.2.5. <i>Gasoline data</i>	75
6.2.2.6. <i>Data simulated according to a highly non-linear model</i>	75
6.2.3. Results	76
6.2.3.1. <i>Data simulated according to a second-order polynomial model</i>	76
6.2.3.2. <i>Tablet data</i>	77
6.2.3.3. <i>Bubbles data</i>	77
6.2.3.4. <i>Colorant data</i>	81
6.2.3.5. <i>Gasoline data</i>	81
6.2.3.6. <i>Data simulated according to a highly non-linear model</i>	87
6.2.4. Conclusions	89
<b>7. MiDAs: a software for mixture DOE and data analysis</b>	<b>91</b>
7.1. Introduction	91
7.2. Variable input	93
7.3. Model selection	98

7.4. DOE construction	100
7.5. Data analysis	103
7.5.1. Data analysis with MLR	110
7.5.2. Data analysis with PLS	112
<b>III. Design Space and optimization through the latent space</b>	<b>113</b>
<b>8. Preliminary considerations</b>	<b>115</b>
8.1. Quality by Design and the Design Space	115
8.2. Limitations of the optimization and DOE in the original space	118
8.3. Optimization in the latent space	119
<b>9. Defining the design space in the latent space</b>	<b>123</b>
9.1. Partial Least Squares model fitting and prediction uncertainty	123
9.2. Transferring restrictions to the latent space	125
9.3. Partial Least Squares model inversion	130
9.3.1. The direct inversion	130
9.3.2. Direct inversion-dependant definition of the Null Space	131
9.3.3. Analytical definition of the Null Space	132
9.3.4. Confidence region of the Null Space	134
9.4. Subspace most likely to contain the True Design Space	138
9.4.1. Datasets	139
9.4.1.1. Case study 1: mathematical model	139
9.4.1.2. Case study 2: simulated data with two correlated outputs	140
9.4.1.3. Case study 3: simulated <i>Vinyl-Chloride Monomer manufacturing</i>	141
9.4.2. Results	143
9.4.2.1. Case study 1: mathematical model	143
9.4.2.2. Case study 2: simulated data with two correlated outputs	145

---

9.4.2.3. Case study 3: simulated Vinyl-Chloride Monomer manufacturing	147
9.4.3. Conclusions	149
9.5. Subspace least likely to fall outside of the True Design Space	150
9.6. Assessing the adequacy of a PLS-regression model for inversion	153
9.6.1. Assessment via direct inversion	154
9.6.2. Assessment by comparison with the closest solution in the null-space	154
9.7. Additional considerations	156
<b>10. Optimization problem formulation in Quality by Design</b>	<b>157</b>
10.1. Introduction	157
10.1.1. Optimization in the original space through the latent space	157
10.1.2. Optimization in the latent space	159
10.2. Quadratic optimization formulation	163
10.2.1. Optimization of a linear combination of outputs	163
10.2.2. Optimization for exploration and DOE in the latent space	165
10.2.3. Tackling the maximization/minimization problem	168
10.2.3.1. Defining feasible minimum/maximum values as the desired ones	168
10.2.3.2. Changing the sign of the weight given in the objective function	169
10.2.3.3. Finding extreme achievable values below the Hotelling $T^2$ limit	170
10.3. Linear optimization formulation	172
10.4. A sequential optimization approach	174
<b>11. Two real case studies of optimization in the latent space</b>	<b>177</b>
11.1. Introduction	177
11.2. Methods	177
11.3. Datasets	177
11.3.1. Case study 1: minimizing two output variables simultaneously	178

11.3.2. Case study 2: maximizing a linear combination of outputs	178
11.4. Results and discussion	178
11.4.1. Case study 1: minimizing two output variables simultaneously	178
11.4.2. Case study 2: maximizing a linear combination of outputs	180
11.5. Conclusions	183
<b>12. On experimentation to improve the design space estimation</b>	<b>185</b>
12.1. Introduction	185
12.2. Methods	186
12.2.1. Method 1: classical DOE and OLS model inversion approach	186
12.2.2. Method 2: DOE in the latent space and LVRMI approach	189
12.2.3. Method 3: Optimization in the latent space and LVRMI approach	191
12.3. Datasets	192
12.3.1. Case study 1: mathematical mod	192
12.3.2. Case study 2: simulated Vinyl-Chloride Monomer manufacturing	192
12.4. Results and discussion	193
12.4.1. Case study 1: mathematical mod	193
12.4.1.1. Detailed procedure for the application of all three methods	193
12.4.1.2. Assessing the performance of each of the three methods	195
12.4.1.3. Additional considerations	200
12.4.2. Case study 2: simulated Vinyl-Chloride Monomer manufacturing	201
12.4.2.1. Detailed procedure for the application of all three methods	201
12.4.2.2. Assessing the performance of each of the three methods	203
12.5. Conclusions	205

<b>IV. Epilogue</b>	<b>207</b>
<b>13. Conclusions and perspectives</b>	<b>209</b>
13.1. Accomplishment of the objectives	209
13.1.1. Objective I - Traditional and latent variable-based approaches applied to mixture design problems	209
13.1.2. Objective II - Latent variable-based approaches for efficient processes optimization	210
13.1.3. Objective III - Latent variable-based approaches applied to the Quality by Design initiative, to increase processes flexibility and guarantee the desired quality	211
13.2. Future research lines	212
<b>14. Appendices</b>	<b>213</b>
14.1. Annex to Part I	213
14.1.1. Relationship between the Euclidean distance matrix, $\mathbf{D}$ , and the inner product matrix, $\mathbf{X} \cdot \mathbf{X}^T$	213
14.1.2. Practical meaning of the pseudo-samples in the feature space	214
14.2. Annex to Part II	215
14.2.1. Relationship between the Scheffé and Cox models coefficients	215
14.2.2. Projection of a point/vector onto the intersection of a group of hyperplanes	216
14.3. Annex to Part III	218
14.3.1. Relationship between the result of the PLS-regression direct inversion and the point in the combined null space closest to the centre of projection/with lowest leverage	218
14.3.2. Analytical expression for the confidence region of the null space for a linear combination of outputs using OLS-type expression for the prediction's confidence interval	219
14.3.3. Confidence interval for the prediction of a linear combination of outputs in PLS using OLS type expression	222
<b>Bibliography</b>	<b>227</b>