Contents

I Prologue **1**

1 Justification, objectives and contributions **3**
1.1 Justification 4
1.2 Objectives 7
1.3 Contributions 9
1.3.1 Papers in peer-reviewed journals 9
1.3.2 Poster conference contributions 10
1.3.3 Oral conference contributions 10
1.3.4 Chapters of books 12
1.3.5 Software 12
1.3.6 Awards 12

2 On latent variable-based regression models **13**
2.1 Introduction 14
2.2 Partial Least Square (PLS) Regression 14
2.2.1 Prediction uncertainty 16
2.2.2 Model inversion 17
2.2.3 Optimization problem formulation 21
2.3 Sequential Multi-block (SMB) PLS regression 22
Appendices 25
2.A SMB PLS regression 25

3 Materials **27**

3.1 Hardware . . . 28
3.2 Software . . . 28
3.3 Datasets . . . 28

II Novel methodological contributions 29

4 On the properties of PLS for analyzing Design of Experiments 31
4.1 Introduction . . . 32
4.2 Two-level factorial designs . . . 34
4.2.1 Full factorial designs: 2^k . . . 34
4.2.2 Fractional factorial designs: 2^{k-p} . . . 39
4.3 Traditional approaches applied to two-level factorial designs with missing runs 40
4.4 PLS applied to two-level factorial designs with missing runs . . . 41
4.4.1 Lack of resources to execute a factorial design (scenario i)) . . . 41
4.4.2 Unexpected problems in the execution of some runs (scenario ii)) . . . 44
4.5 Illustrative examples . . . 45
4.5.1 First illustrative example: 2^4 . . . 45
4.5.2 Second illustrative example: 2^{6-2} . . . 51
4.6 Discussion and conclusions . . . 54
Appendices . . . 56
4.A Definition of PLS coefficients from NIPALS algorithm in a full factorial design 56
4.B Definition of PLS coefficients from the criterion of maximum variance in a full factorial design . . . 58
4.C The second PLS component lacks predictive ability in a full factorial design 59
4.D Illustrating the latent space in a full factorial design . . . 60
4.E Partitioning of the sum of squares for PLS in a full factorial design . . . 61
4.F Lists of recommended combinations of missing runs for the most popular designs . . . 63

5 Defining multivariate raw materials specifications 67
5.1 Introduction . . . 68
5.2 Data requirements . . . 72
5.3 Defining the design space in the latent space by means of PLS . . . 72
5.3.1 Design space with no uncertainty . . . 72
5.3.2 High confidence design space . . . 74
5.4 Exploiting the model . . . 80
5.5 Industrial case studies . . . 81
5.5.1 First industrial case study: cereal extraction process . . . 81
5.5.2 Second industrial case study: blown film process . . . 89
5.6 Conclusion . . . 91

Appendices . . . 93
5.A Specification confidence limits for the l-th critical quality attributes . . . 93

6 Defining multivariate raw material specifications via SMB-PLS 97
6.1 Introduction . . . 98
6.2 Data requirements . . . 99
6.3 The SMB-PLS model in the raw material paradigm . . . 100
6.4 Defining the design space in the latent space by means of SMB-PLS . . . 101
6.5 Multivariate raw material specification region . . . 102
6.5.1 Without improved control . . . 102
6.5.2 Under improved control . . . 104
6.6 Presence of known disturbances affecting control actions . . . 105
6.7 Industrial case study . . . 106
6.8 Conclusions . . . 117
Appendices . . . 118
6.A SMB-PLS weights transformed to be independent between components . . . 118

7 Latent space-based multivariate capability index 121
7.1 Introduction . . . 122
7.2 Data requirements . . . 123
7.3 Supplier's raw material operating space (RMOS) . . . 123
7.4 Latent space-based multivariate capability index . . . 124
7.5 Diagnosing assignable causes . . . 126
7.6 Proposed methodology . . . 127
7.7 Industrial case study . . . 128
7.8 Conclusions . . . 132

III Novel applications 135

8 Health application: COVID-19 Pandemic 137
8.1 Introduction . . . 138
8.2 Machine learning models for early estimation of COVID-19 mortality risk in hospitalized patients . . . 138
8.3 Methods: Reformulation of the optimization problem . . . 139
8.3.1 PLS customized optimization problem formulation . . . 139
8.3.2 Nonlinear PLS customized optimization problem formulation . . . 140
8.3.3 SMB-PLS customized optimization problem formulation . . . 143
8.4 A Latent variable-based alternative to clinical trials upon new diseases . . . 144
8.4.1 Simulated case study . . . 146
8.4.2 Spanish society of hospital pharmacy case study . . . 157

8.5 Conclusion . . . 163
Appendices . . . 164
8.A Function: H . . . 164

9 Industrial application: Multivariate Six Sigma 165
9.1 Introduction . . . 166
9.2 Methods . . . 167
9.2.1 Six Sigma's DMAIC methodology . . . 167
9.2.2 Optimization of the latent space-based multivariate capability index 167
9.3 Results . . . 168
9.3.1 Define . . . 168
9.3.2 Measure . . . 169
9.3.3 Analyze . . . 170
9.3.4 Improve . . . 173
9.3.5 Control . . . 174
9.4 Conclusion . . . 176

IV Graphical user interface 177

10 Dragonet: a software for data analysis and process optimization 179
10.1 Introduction . . . 180
10.2 Importing data . . . 180
10.3 Building a model . . . 182
10.4 Data analysis . . . 184
10.5 Process optimization . . . 187
10.6 Defining the high-confidence design space . . . 190
10.7 Conclusions . . . 192

V Epilogue 193

11 Conclusions 195
11.1 Meeting the objectives . . . 196
11.2 Future research lines and transfer activities . . . 199

Bibliography 201