

Identification and Classification of New Psychoactive Substances using Raman Spectroscopy and Chemometrics

By

Amira Guirguis

Dissertation

Presented to the University of Hertfordshire in partial fulfilment
of the requirements of the degree of

Doctor of Philosophy

The University of Hertfordshire
School of Life and Medical Science
Department of Pharmacy, Pharmacology and Postgraduate Medicine

January, 2017

Table of Contents

List of Tables	6
List of Figures	8
Abstract	15
Abbreviations	17
Acknowledgements	20
1. Introduction	22
1.1. Background	22
1.2. Chemical analysis of NPS	27
1.3. In-field detection of NPS using Raman instruments	33
1.3.1. In-built algorithms in handheld Raman instruments	33
1.4. The challenge of mixture analysis using Raman spectroscopy	37
1.4.1. Raman applications	37
1.4.2. Chemometric applications	37
1.5. Facing the continuous flood of NPS emergence	41
1.6. Aims of this project	44
2. Evaluation of three handheld Raman instruments for the identification of new psychoactive substances and related compounds	45
2.1. Introduction	45
2.2. Experimental	46
2.2.1. Materials	46
2.2.2. Instrumentation and Methods	48
2.3. Results and discussion	54
2.3.1. Substance selection	54
2.3.2. Comparison of sample spectra to library signatures using the ‘first pass’ in-built identification algorithms	55
2.3.2.1. Truscan™ GP (Thermo-Fisher Scientific)	55
2.3.2.2. FirstGuard™ (Rigaku)	64
2.3.2.3. Progeny™ (Rigaku)	69
2.4. Conclusions	82
3. Comparison of two handheld Raman Spectrometers employing 785 and 1064 nm laser sources for the identification of new psychoactive substances	84
3.1. Introduction	84
3.2. Experimental	85
3.2.1. Chemicals and Reagents	85
3.2.2. Instrumentations and methods	89

3.2.3. Confirmatory analysis of NPS and related substances using gas chromatography – electron ionisation-mass spectrometry (GC-EI-MS)	90
3.3. Results and Discussion	91
3.3.1. Raman analysis of purchased reference standards	91
3.3.2. Raman and GC-MS analysis of NPS internet products	93
3.4. Conclusions	107
4. Development of a pre-processing protocol for Raman spectra of NPS-related powders	109
4.1. Introduction.....	109
4.2. Materials and Methods.....	110
4.2.1. Reagents and chemicals	110
4.2.2. Sample preparation	110
4.2.3. Raman Spectroscopy.....	111
4.2.4. Pre-processing method.....	113
4.2.5. Principal Components Analysis (PCA) (Unscrambler® X 10.3)	114
4.3. Results and Discussion	116
4.3.1. Selection of adulterants and cutting agents.....	116
4.3.2. Step-wise process employed for developing a pre-processing protocol for Raman spectra of NPS-related compounds.....	117
4.3.3. Descriptive statistics (Unscrambler® X 10.3).....	126
4.3.4. Pre-processing of Raman spectra using the Unscrambler® X 10.3.....	127
4.3.5. Combined pre-processing methods.....	154
4.3.6. Variables' reduction.....	162
4.4. Conclusions	167
5. Classification of NPS reference standards using Raman spectroscopy and chemometric approaches.....	168
5.1. Introduction.....	168
5.2. Experimental	169
5.2.1. Materials	169
5.2.2. Sample preparation	170
5.2.3. Raman spectroscopy	174
5.2.4. Raman spectral pre-processing	174
5.2.5. Principal component analysis (PCA).....	175
5.2.6. Prediction via PCA projection	175
5.3. Results and Discussion.....	176
5.3.1. Method development of Raman analysis and chemometrics.....	176
5.3.2. Pre-processing method development	177
5.3.3. Raman and chemometric analysis.....	188

5.4. Conclusions	216
6. Classification of NPS internet samples using NPS, adulterants and cutting agents reference standards <i>via</i> Raman spectroscopy and chemometric approaches – a proof-of-concept.....	219
6.1. Introduction.....	219
6.2. Materials and Methods.....	220
6.2.1. Reagents and chemicals	220
6.2.2. Methods.....	221
6.3. Results and Discussion	226
6.3.1. Selection of cutting agents/ adulterants and NPS internet products	226
6.3.2. Overview of classifications of new datasets using PCA.....	227
6.3.3. Prediction of test NPS internet samples using PCA projection (Unscrambler X 10.4)	243
6.4. Conclusions.....	259
7. Classification of NPS internet samples using handheld Raman spectroscopy, a 1064 nm laser source and chemometric approaches – a proof-of-concept.....	261
7.1. Introduction	261
7.2. Materials and Methods.....	262
7.2.1. Reagents and chemicals	262
7.2.2. Raman analysis	263
7.2.2.1. Sample preparation	263
7.2.2.2. Analysis of NPS reference standards and internet products using handheld Raman Spectroscopy.....	264
7.2.2.3. Raman spectral pre-processing	265
7.2.2.4. Principal Components Analysis	265
7.2.3. Prediction <i>via</i> PCA projection	265
7.3. Results and Discussion.....	266
7.3.1. Handheld Raman method development.....	266
7.3.2. Overview of PCA1 generated from ‘53 representative’ NPS reference standards	268
7.3.3. Prediction of ‘test’ and ‘out-of-model (OOM)’ pure reference samples <i>via</i> PCA projection onto the ‘NPS reference standards/handheld’ model (test sets 1 and 2).....	271
7.3.4. Prediction of test NPS internet samples (test set 3) <i>via</i> PCA projection onto PCA1	276
7.3.5. Prediction of test NPS internet samples (test set 3) <i>via</i> PCA projection onto PCA2	283
7.4. Conclusions	286
8. Conclusions.....	289
Implications for practical use.....	295
Future Work.....	297

9. References.....	299
10. Appendices.....	324
10.1. Appendix A – Chapter 2.....	324
10.2. Appendix B – Chapter 3.....	325
10.3. Appendix C – Chapter 4.....	326
10.4. Appendix D – Chapter 6.....	328
10.5. Appendix E - Published work included in this thesis.....	331

List of Tables

Table 2. 1: Chemical structures of library signatures for NPS, adulterants and cutting agents.	47
Table 2. 2: NPS products purchased from the internet.	48
Table 2. 3: Specifications of handheld Raman instruments evaluated in this study.	49
Table 2. 4: Composition in mg/mg of the binary model mixtures ¹	51
Table 2. 5: Composition of random ternary mixtures.	52
Table 2. 6: Methods used for Raman analysis using handheld Raman instruments (FirstGuard TM and Progeny TM) ¹	52
Table 2. 7: Selectivity of Truscan TM GP ¹ . Table showing the p-values between reference standards and themselves in the library.	56
Table 2. 8: Comparison values of the five model mixtures using three handheld Raman instruments: Progeny TM , FirstGuard TM and Truscan TM GP ¹	60
Table 2. 9: Identification of one or both constituents in model binary mixtures using Truscan TM GP	63
Table 2. 10: Selectivity of FirstGuard TM	66
Table 2. 11: Identification of one or both components in model binary mixtures using the FirstGuard TM	68
Table 2. 12: Selectivity of Progeny TM	70
Table 2. 13: Identification of one or both components in model binary mixtures using Progeny TM	72
Table 2. 14: Identification of ternary mixtures using the in-built algorithm in Truscan TM GP, FirstGuard TM and Progeny TM	74
Table 2. 15: Comparison of the ‘first pass’ matching analysis of NPS products using Raman Truscan TM GP, FirstGuard TM , Progeny TM and corresponding GC-EI-MS results.	76
Table 2. 16: Summary of the performance of three handheld Raman instruments with respect to ‘true’ and ‘false’ judgment results.	78
Table 2. 17: Summary of the handheld Raman instruments’ overall performance for model binary mixtures, random ternary mixtures and NPS internet products.	81
Table 3. 1: Chemical structures of NPS reference standards used in this Chapter	85
Table 3. 2: NPS product specifications.	86
Table 3. 3: Specifications of Xantus TM -1 and FirstGuard TM handheld Raman instruments	89
Table 3. 4: Validation of NPS reference standards using two handheld Raman spectrometers (Xantus TM -1 (785 nm) and FirstGuard TM (1064 nm)) and on board libraries.	92
Table 3. 5: Results from the analysis of aminoindane and arylalkylamine internet products using two handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) and GC-MS ¹	96
Table 3. 6: Results from the analysis of benzodiazepine, piperidine & pyrrolidine, plants & extracts and ‘other’ internet products using two handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) and GC-MS ¹	101
Table 3. 7: Results from the analysis of branded internet products using two handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) and GC-MS ¹	106
Table 3. 8: A summary of the results obtained for NPS identification using the handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) ¹	108
Table 4. 1: Adulterants and cutting agents used in Raman and PCA analysis	110
Table 4. 2: Summary of samples included in the calibration dataset (Set ‘1’)	111
Table 4. 3: Summary of samples included in the new calibration dataset (Set ‘2’)	111
Table 4. 4: Summary of samples included in the validation dataset (Set ‘3’)	111

Table 4. 5: Smoothing parameters.	128
Table 4. 6: Summary of the S/N for all smoothing methods for both C1 and C2 (WiRE 3.4).	129
Table 4. 7: Important peaks for BEN, CAF, CRE and LAC ('partial identifiers') identified by PCA loading plots.	163
Table 5. 1: NPS reference standards used in this study (training, test and out-of-model ('OOM') NPS reference standards).....	171
Table 5. 2: PCA projection results for 21 unknown pure NPS against the PCA model developed using 53 pure NPS	191
Table 6. 1: NPS products purchased from the internet.	221
Table 6. 2: Summary of qualitative and quantitative analysis of NPS internet products using handheld Raman spectroscopy, GC-MS and HPLC	223
Table 6. 3: Analysis of potential overlap between reference spectra for 53 NPS and 22 cutting agent/ adulterants along the PCA scores plot	240
Table 6. 4: PCA projection analysis results	245

List of Figures

Figure 1. 1: The relative prevalence of new psychoactive substances in Europe over time (Figure reproduced with permission from the EMCDDA [9]).	24
Figure 1. 2: The Drug Wheel. The outer ring shows “the controlled NPS under the Misuse of Drugs Act 1971 or The Human Medicines Regulations 2012”, whereas the inner ring shows “the controlled NPS under the Psychoactive Substances Act 2016 (PSA). *Temporary Class Drug Orders (TCDOs)” (Figure reproduced with permission from Dr M. Adley) [32].	25
Figure 1. 3: A simplified Jablonski diagram showing the types of scattering resulting from the interaction between light and matter. The stokes (left), Rayleigh (middle) and anti-stokes (right) scattering (Figure reproduced from Medicines Complete [150]).	32
Figure 1. 4: A screenshot of the ‘actionable’ results obtained with Truscan™ GP using the PVAL algorithm. Sample responses include ‘positive match’ (left), ‘no match found’ (middle) and ‘no positive match’ (right) (© 2015 Thermo-Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo-Fisher Scientific Inc. and its subsidiaries).	34
Figure 1. 5: Raman spectrum of an NPS branded product with no match found (laser excitation wavelength = 785 nm). Reproduced from Analytical Methods [128] with permission from the Royal Society of Chemistry.	38
Figure 1. 6: An example of a two-dimensional plane, which best describes the data (Figure adapted from Camo Software AS, Oslo, Norway) [172].	39
Figure 2. 1: Schematic outlining summary of experiments	53
Figure 2. 2: Normalised Raman spectra of reference standards of a) NPS and adulterants; and b) cutting agents (Truscan™ GP).	57
Figure 2. 3: Normalised Raman spectra for library spectra of DXM25/CAF75, CAF and DXM and the spectrum of the sample 3D4 composed of LID30/CAF70 (Truscan™ GP).	64
Figure 2. 4: Normalised Raman spectra of reference standards of a) NPS and adulterants; and b) cutting agents (FirstGuard™).	65
Figure 2. 5: The Raman spectrum for the MgS reference sample match-tested against the library reference spectra. The chemical structure for bees wax is superimposed on the spectrum (FirstGuard™).	67
Figure 2. 6: Normalised Raman spectra for the samples 1D2 and 4D2 and normalised Raman reference spectra for LID25 CAF75, LID50 CAF50 and CAF (FirstGuard™).	69
Figure 2. 7: Normalised Raman spectra of reference standards of a) NPS and adulterants; and b) cutting agents (Progeny™).	71
Figure 2. 8: Normalised Raman spectra for Pink Champagnes (P2) using Truscan™ GP, FirstGuard™ and Progeny™ and the reference spectra for 2-AI and CAF (Progeny™).	77
Figure 3. 1: Raman spectra of NPS reference standards uploaded in on-board library (FirstGuard™ (1064 nm)).	91
Figure 3. 2: Raman spectra of MPA reference standard using a) Xantus™-1 ($\lambda_{ex} = 785$ nm) and b) FirstGuard™ ($\lambda_{ex} = 1064$ nm).	94
Figure 3. 3: Raman spectra of P11 (MDAI), P23 (MPA), P27 (etizolam), P39 (EPD), P43 (DXM) and P53 (Pink Champagnes) using 785 nm (a) and 1064 nm (b) excitation wavelengths. Product names represent the label claim and not the sample composition that is presented in Tables 3.5 – 3.7. The spectra were normalised to the maximum peak.	95
Figure 3. 4: Raman spectra of selected products and associated reference spectra from the aminoindane (a), arylalkylamine (b), benzodiazepine (c), piperidine & pyrrolidine (d), ‘other’ (e), and branded product (f) EMCDDA categories using a 1064 nm laser excitation wavelength.	

Product numbers and reference names are labelled on the spectra. The spectra were normalised to the maximum peak.....97

Figure 4. 1: A flow chart outlining the initial step-wise pre-processing of Raman spectra ..	113
Figure 4. 2: An example of PCA plots for Raman spectra of cutting agents and adulterants (Unscrambler® X 10.3): a) the explained variance plot, b) the 2D-scores plot, c) the loadings plot; d) the influence plot and e) the leverage plot (outliers are circled).	115
Figure 4. 3: a) Magnified six CAF raw spectra from set '1': C1 (blue), C2 (red), C3 (green), C4 (light blue), C5 (maroon) and C6 (grey); b) Magnified raw spectra from Set '1': C1 to C6 (red), B1 to B3 (blue), CCL 1, 2 and 6 (green) and CCL 3 to 5 (purple) (Unscrambler® X 10.3).	118
Figure 4. 4: The Raman spectrum for C1 displaying a cosmic ray at ca. 3116 cm ⁻¹ (WiRE 3.4).	118
Figure 4. 5: A Section of the Raman spectrum for B1 (detector saturation at ca.1600 cm ⁻¹) (WiRE 3.4).....	119
Figure 4. 6: 'PCA-1' plots derived from the unprocessed Raman spectra for set '1' (Unscrambler® X 10.3). a) the explained variance plot, b) the 2D-scores plot for PC1/PC2, c) the loadings plot; d) the influence plot; e) the leverage plot (PC1) and f) the leverage plot (PC3) (outliers are circled).	122
Figure 4. 7: 'PCA-2' plots of the Raman spectra for Set '2' (Unscrambler® X 10.3): a) the explained variance plot, b) the scores plot and c) the line loadings plot.	124
Figure 4. 8: Six raw (RD) and six interpolated (Int) CAF Raman spectra (Unscrambler® X 10.3).	125
Figure 4. 9: A plot showing multiplicative scatter effects (MSEs) for six CAF spectra (Set '2') (Unscrambler® X 10.3) ¹	126
Figure 4. 10: Summary of pre-processing methods investigated.....	128
Figure 4. 11: Summary of various smoothing of two CAF spectra (C1 and C2 from Set '2'). Spectra were smoothed using the SG algorithm (Unscrambler® X 10.3) using the parameters summarised in Table 4.5.	129
Figure 4. 12: Summary of PCA results for smoothed Set '2' Raman spectra (Unscrambler® X 10.3). Method 1: a) 2D-scores and b) line loading plots. Method 2: c) 2D-scores and d) line loading plots. Method 3: e) 2D-scores and f) line loading plot. Method 4: g) 2D-scores and h) line loading plots.....	131
Figure 4. 13: Summary of baseline correction methods (Unscrambler® X 10.3). Baseline offset method: a) low wavenumber end and b) high wavenumber end. Linear baseline method between the wavenumbers 1800-250 cm ⁻¹ : a) low wavenumber end and d) high wavenumber end. Linear baseline method between the wavenumbers 3000-250 cm ⁻¹ : a) low wavenumber end and f) high wavenumber end. g) The first derivative of two CAF raw spectra (Raw spectra for C1 (green) and C2 (light blue), and first derivative for C1 (blue) and C2 (red)). h) The second derivative of a 100 cm ⁻¹ band (1722-1622 cm ⁻¹) from two CAF raw spectra (Set '2') (Raw spectra for C1 (green) and C2 (light blue) and second derivative for C1 (blue) and C2 (red)).	136
Figure 4. 14: Summary of PCA results for baseline correction methods (Set '2') (Unscrambler® X 10.3). Baseline offset method: a) 2D-scores and b) line loading plots. Linear baseline (1800-250 cm ⁻¹) c) 2D-scores and d) line loading plots. Linear baseline (3000-250 cm ⁻¹) e) 2D-scores and f) line loadings plot. First derivative g) 2D-scores and h) line loading plots. Second derivative i) 2D-scores and j) line loading plots.	137
Figure 4. 15: A CAF peak at 554 cm ⁻¹ (WiRE 3.4)	143
Figure 4. 16: Line plots of normalised and non-normalised CAF spectra (C1 to C6 from Set '2') (Unscrambler® X 10.3). a) Raw spectra, b) Area normalisation, c) Unit vector	

normalisation, d) Mean normalisation, e) Maximum normalisation, f) Range normalisation and g) Peak normalisation. In all Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.....	146
Figure 4. 17: PCA scores and loading plots summarising the normalisation types of six CAF Raman spectra (C1 to C6 from Set ‘2’) in comparison with raw spectra (Unscrambler® X 10.3). (a - b) raw spectra; (c - d) area normalisation; (e - f) unit vector normalisation; (g - h) mean normalisation; (i - j) maximum normalisation; (k - l) range normalisation; (m - n) peak normalisation.....	147
Figure 4. 18: Mean normalisation of Set ‘2’ (Unscrambler® X 10.3). a) line plots of three groupings (CCL (blue), CAF (red) and (BEN (green)) for raw Raman spectra (Set ‘2’); b) line plots of three groupings (CCL (blue), CAF (red) and (BEN (green)) for mean normalised Raman spectra (Set ‘2’); c) The PCA 2D-scores and d) the line loading plots.	149
Figure 4. 19: Raman spectra for C1 to C6 (Set ‘2’) before and after pre-processing using SNV (Unscrambler® X 10.3). In both Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.....	151
Figure 4. 20: SNV of Set ‘2’. a) 2D-scores and b) line loading plots (Unscrambler® X 10.3).	152
Figure 4. 21: Six CAF spectra before and after pre-processing with MSC (Unscrambler® X 10.3). In both Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.	153
Figure 4. 22: MSC of Set ‘2’. a) 2D-scores and b) line loading plots (Unscrambler® X 10.3).	153
Figure 4. 23: Line plots of combined pre-processing methods using six CAF spectra (C1 to C6 from Set ‘2’) (Unscrambler® X 10.3). a) Raw spectra; b) Method 1; c) Method 2; d) Method 3; e) Method 4; f) Method 5; g) Method 6; and h) Method 7. In all Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.	156
Figure 4. 24: Summary of PCA results for the combined pre-processing methods (Unscrambler® X 10.3). a) the scores plot for the raw spectra (Set ‘2’); b) the scores plot for method 1; c) the scores plot for method 2; d) the scores plot for method 3; e) the scores plot for method 4; f) the scores plot for method 5; g) the scores plot for method 6; h) the loadings plot for method 6; i) the scores plot for method 7; j) the loadings plot for method 7.....	159
Figure 4. 25: The 3D-scores plot for the validation samples Set ‘3’ following their pre-processing using method ‘5’ (Unscrambler® X 10.3).....	162
Figure 4. 26: a) The scores plot of the selected-bands spectral data prior to pre-processing, and b) following processing using method ‘5’ (Unscrambler® X 10.3).....	164
Figure 4. 27: A flow chart showing a protocol for the pre-processing of Raman spectra ¹	166
Figure 5. 1: A schematic outlining combined pre-processing steps of spectral Raman data for NPS reference standards.	175
Figure 5. 2: A schematic describing summary of experiments. It includes a summary of both training and ‘query’ sets (i.e. test and ‘OOM’ sets), combined pre-processing steps and details of projection PCA studies.	176
Figure 5. 3: 3D-PCA scores plots of average a) non-smoothed and b) smoothed average spectra of 53 NPS reference standards over the spectral range 3200 – 250 cm ⁻¹ . Average spectra were sample-grouped based on their supercluster membership [204] (Table 5.1).....	179
Figure 5. 4: Raman spectra of selected NPS reference standards (spectral range 3200 – 250 cm ⁻¹). The line plots show different grades of baseline offsets in Figures 5.4a, c and e. Spectra were pre-processed using smoothing and baseline subtraction in Figures 5.4b, d and f respectively (Origin Pro 2016).....	180
Figure 5. 5: PCA explained variance plots over the investigated spectral ranges.	182

Figure 5. 6: PCA a) line loading plot, b) correlation loading plot and c) zoomed-in correlation loading plot over the spectral range 3200 – 250 cm ⁻¹	183
Figure 5. 7: PCA scores plots of 10 replicate spectra of 53 NPS reference standards over the spectral ranges a) 3200 – 250; b) 1750 – 250; c) 1750 – 1000; d) 1750 – 1300 cm ⁻¹ . Replicate spectra were sample-grouped based on their supercluster membership [204].....	184
Figure 5. 8: PCA scores plots of average spectra of 53 NPS reference standards for a) 600 bins, b) 300 bins over the spectral ranges 3200 – 250 and 1750 - 250 cm ⁻¹ respectively. Average spectra were sample-grouped based on their supercluster membership [204].....	186
Figure 5. 9: PCA scores plots of a) 1 single spectrum, b) 3 replicate spectra, c) 5 replicate spectra and d) 10 replicate spectra of 53 NPS reference standards over the spectral range 1750 - 1300 cm ⁻¹ . Spectra were sample-grouped based on their supercluster membership [204]..	187
Figure 5. 10: PCA two-dimensional (2D)-score plots for 530 NPS reference spectra over the spectral range 1750 - 1300 cm ⁻¹ , specifically 2D-scores plot for a) PC1/PC2 and b) PC1/PC3. The scores for 4-HO-DET, GHB and MPA are marked on Figure 5.10a.	188
Figure 5. 11: The chemical structures of α -PVP, α -PBT and MPA. The molecular structure for α -PVP is represented as an example of a cathinone analogue with a typical β k-aromatic group.	190
Figure 5. 12: 3D-scores plot of 530 reference spectra. Projections of supercluster 5 are illustrated in pink simultaneously along the three planes PC1/PC2/PC3 (Origin Pro 2016).	190
Figure 5. 13: PCA projection plot illustrating training, ‘test’ and ‘OOM’ sets. ‘Query’ spectra (i.e. ‘test’ and ‘OOM’ spectra) are projected on the PCA model developed with 53 training NPS (530 NPS reference spectra).	191
Figure 5. 14: a) NPS training members of supercluster 2 with the relevant test and ‘OOM’ samples, cluster membership, TSI and r ² values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructures; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.....	193
Figure 5. 15: a) NPS members of supercluster 3 with relevant test sample, cluster membership, TSI and r ² values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructures; e) Raman spectra and PC line loading plots.....	196
Figure 5. 16: a) NPS training members of supercluster 4 with the relevant test sample, cluster membership, TSI and r ² values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructures; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.	198
Figure 5. 17: NPS training members of supercluster 5 with the relevant test and ‘OOM’ samples, cluster membership, TSI and r ² values.....	200
Figure 5. 18: Plots representing the analysis relevant to supercluster 5. a) PCA projection plots (Training samples (red) versus individual test and ‘OOM’ samples; b) training, test and ‘OOM’ samples classified as per EMCDDA/EDND classification; c) common substructures; d-h) selected Raman spectra and PC line loading plots.....	201
Figure 5. 19: The Raman spectra of zopiclone, THJ-018 and PB-22 and their chemical structures.	203
Figure 5. 20: a) NPS training members of supercluster 6 with the relevant test sample, cluster membership, TSI and r ² values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X10.4); d) common substructures; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.	207
Figure 5. 21: a) NPS training members of supercluster 7 with the relevant ‘OOM’ sample, cluster membership numbers and r ² values and chemical structures; b) PCA projection plot (Unscrambler® X10.4); c) common substructures; d) Raman spectra and PC line loading plots, PCA projection plot for PC1/PC2/PC3.....	209

Figure 5. 22: a) NPS training members of supercluster 8 with the relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4) (Pyrazolam scores are behind flubromazolam scores); d) common substructure; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.....	211
Figure 5. 23: a) NPS training members of supercluster 11 with the relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructure; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.	213
Figure 5. 24: a) NPS training members of supercluster 12 with the relevant test samples, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructure; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.	215
Figure 6. 1: Raman/ chemometrics experimental protocol. Schematic outlining the details of the training and ‘query’ sets, steps of combined spectral pre-processing and chemometrics.	226
Figure 6. 2: Pre-processed average Raman spectra of a) 14 cutting agents and b) eight adulterants over the spectral range 3200 – 250 cm^{-1}	229
Figure 6. 3: a) Three-dimensional (3D)-scores plot of 22 cutting agent/ adulterant reference spectra; b) Two-dimensional (2D)-scores plots for PC1/PC2; c) 2D-scores plots for PC1/PC3; d) Common substructure between PAR and PHE; d) Pre-processed Raman spectra of BEN, PRO, PAR, PHE, DIL, dimethocaine and the line loading plots for PC1/PC2/PC3.	232
Figure 6. 4: a) 3D-scores plot of 23 NPS internet samples; b) 2D-scores plots for PC1/PC2 with an ellipse at 95 % CL; c) 2D-influence plot for PC1/PC2; d) pre-processed replicate Raman spectra for P19 ¹	234
Figure 6. 5: The NPS and non-NPS model. a) 3D-scores plot of 53 NPS and 22 cutting agent/ adulterant reference spectra; b) 2D-scores plots for 13 NPS superclusters and cutting agent/ adulterant classes (PC1/PC2); c) 2D-scores plots for 13 NPS superclusters and cutting agent/ adulterant classes (PC1/PC3).....	238
Figure 6. 6: Pre-processed Raman spectra for: a) 4-MeO-PCP, MXT, N-Me-2C-B and STP; b) BEN, PRO, dimethocaine, 4-MeO- α -PVP and 25H-NBOMe; c) N-Me-2C-B, DIL and mebroqualone); d) 4-AcO-DMT, AB-FUBINACA and GLUT; e) MXT, N-ethylamphetamine, CAF and NIA and line loading plots (PC1/PC2/PC3).....	239
Figure 6. 7: a) 3D-PCA projection scores plot; b) 2D-PCA projection scores plot. a and b are pre-processed Raman spectra of 23 NPS internet products projected onto a PCA model generated from training samples (‘53 representative’ NPS and 22 cutting agents/ adulterants reference standards). The first and second percentage values represent the percentage explained variance by PC1/PC2/PC3 for initial and projected samples respectively; c) PCA projection residual variance plot. This plot shows the calibration and validation plots of the training samples (blue and red respectively) and the projection plot of the projected ‘query’ samples; d) Common substructure between 2-AI and N-Me-2-AI ((2,3-dihydro-1H-inden-2-yl)-12-azane).	244
Figure 6. 8: Overlaid line plots for 2-AI reference standard and replicate spectra for P2. Raman peaks at ca. 1644 and 1374 cm^{-1} are present in P2 R5 (green) but absent in remaining replicate spectra for P2.	246
Figure 6. 9: a) Stacked line plots for average Raman reference spectra for the arylalkylamines 5,6-MDAI, 5-EAPB HCl and MPA, and NPS internet spectra P17 R1 and R10, and P18 R6 and R10; b) 3D-PCA projection scores plot showing the latter reference spectra and samples projected onto the three planes PC1/PC2/PC3 (Origin Pro 2016).....	247

Figure 6. 10: a) The 2D-PCA projection of 23 internet products onto the PCA model generated from the training samples (53 NPS reference standards and 22 cutting agents/ adulterants) (Unscrambler® X10.4). P19 R3, R4 and R5 and the reference tryptamines 4-HO-DET, 5-MeO-DALT, 5-MeO-MiPT and DPT are highlighted with an ellipse; b) Line plots for replicate spectra of P19; c) 3D-PCA projection plot of average reference spectra of tryptamines (i.e. 4-HO-DET, 5-MeO-DALT, 5-MeO-MiPT and DPT) and MPA, and P19 R3 – R5 (Origin Pro 2016).	248
Figure 6. 11: a) Line plots for average reference spectra of 2-AI, 3-MeO-PCE, ketamine, N-Me-2-AI, STP and LID, and replicate spectra of the product P20; b) 3D-PCA projection plot of the latter average reference spectra and samples (Origin Pro 2016).	249
Figure 6. 12: Line plots for average reference spectra of AB-PINACA (red), AB-FUBINACA (light green) and replicate spectra for a) P4; b) P5.	250
Figure 6. 13: a) Line plots for average reference spectra of 5-EAPB, 5,6-MDAI, C8 CP 47-497 and SDB-006, and replicate spectra of P16; b) 3D-PCA projection plot of average reference spectra of 5-EAPB, 5,6-MDAI, C8 CP 47-497 and SDB-006, and replicate spectra of both P16 and P17 (Origin Pro 2016).	251
Figure 6. 14: Pre-processed average Raman spectra for: a) 3-MeO-PCE, 4-MeO-PCP, MXT, etizolam, 2-AI, CAF, LID, THEO, MCC and P6; b) JWH-015, JWH-073, SUC, etizolam, MCC and replicate spectra of P8.	253
Figure 6. 15: a) Pre-processed average Raman spectra for CRE, MAN, TAL, TAU, AM-2201, 5F-APINACA, flubromazepam and replicate spectra of P11; b) 3D-PCA projection scores plot for CRE, MAN, TAL, TAU, AM-2201, 5F-APINACA, flubromazepam and replicate spectra of P11; c) Pre-processed Raman spectra for 5-EAPB, MPA, 4HO-DET, DIL, etizolam, flubromazepam and replicate spectra of P13; 3D-PCA projection scores plot for: d) 5-EAPB, MPA, 4HO-DET, DIL, etizolam, flubromazepam; and e) 25H-NBOMe, DL-4662, flubromazepam, replicate spectra of both P14 and P21.	255
Figure 6. 16: a) Line plots for average reference spectra of 4-HO-DET, 4-AcO-DMT, 5-MeO-DALT, DPT, UR-144, flubromazepam and replicate spectra of P15; b) 3D-PCA projection plot of HO-DET, 4-AcO-DMT, 5-MeO-DALT, DPT, UR-144, flubromazepam and replicate spectra of P15.	257
Figure 6. 17: Line plots for average reference spectra of: a) 2-AI, 5-MeO-DALT, CAF, etizolam, LID and MXT and replicate spectra of P21; b) 2-AI, N-Me-2-AI, N-ethylamphetamine, CAF and replicate spectra of P22; c) 2-AI, CAF, etizolam, ketamine, PAR, THEO and replicate spectra of P23.	259
Figure 7. 1: Sample preparation prior to handheld Raman analysis (Progeny™, Rigaku, USA). a) Sample placed between aluminium plate and cover slip; b) Attaching aluminium plate to cover slip with removable sellotape; c) Final format for sample presentation to the handheld instrument for Raman analysis.	263
Figure 7. 2: Raman and chemometrics experimental protocol: a) three test sets including Raman spectra for 15 NPS ‘test’ reference standards, 4 ‘OOM’ reference standards and 19 NPS internet samples projected onto PCA1 derived from ‘53 representative’ NPS reference standards; b) one test set including Raman spectra for 19 NPS internet samples projected onto PCA2 derived from ‘53 representative’ NPS plus 22 cutting agent/ adulterant reference standards ¹	266
Figure 7. 3: 2D-PCA scores plot (PC1/PC2) illustrating the classification of ‘53 representative’ NPS reference standards classified in previously described superclusters (Chapter 5). Raman spectra were collected using Progeny™ (Rigaku, USA).	270
Figure 7. 4: 2D-PCA projection plot (PC1/PC2) illustrating training, ‘test’ and ‘OOM’ sets. Test sets 1 and 2 (i.e. ‘test’ and ‘OOM’ reference spectra) are projected onto the PCA model	

developed with 53 training NPS, classified in previously described superclusters (Chapter 5). Raman spectra were collected using Progeny™ (Rigaku, USA).	271
Figure 7. 5: PCA projection explained variance plots for: a) ‘test’ (test set 1) and b) ‘OOM’ (test set 2) sets. This is the variance explained by the projection of ‘test’ and ‘OOM’ samples onto PCA1.....	272
Figure 7. 6: 3D-PCA projection plot illustrating training, ‘test’ and ‘OOM’ sets. Test sets 1 and 2 (i.e. ‘test’ (green) and ‘OOM’ (blue) spectra) are projected on the PCA model developed with ‘53 representative’ training NPS (n = 10/ NPS (red)).	273
Figure 7. 7: Average reference Raman spectra for 5-EAPB, MPA, MDMA and line loading plots for PC1 and PC2.....	275
Figure 7. 8: a) 2D-projection PCA scores plot for PC1/PC2; b) 2D-projection PCA scores plot for PC1/PC3. a and b are pre-processed Raman spectra of 19 NPS internet products (test set 3) projected onto PCA1 generated from training samples (‘53 representative’ NPS reference standards).	278
Figure 7. 9: PCA projection explained variance plot. This plot shows the calibration (blue) and validation (red) plots of the training samples (i.e. ‘53 representative’ NPS) and the projection plot (green) of projected ‘query’ samples (i.e. 19 NPS internet products).	279
Figure 7. 10: Normalised replicate Raman spectra for P20 and the average reference Raman spectrum for N-Me-2-AI.....	280
Figure 7. 11: Replicate Raman spectra for P16 and P17 and the average reference spectrum for GHB.	282
Figure 7. 12: a) 2D-projection PCA scores plot for PC1/PC2; b) 2D-projection PCA scores plot for PC1/PC3. a and b are pre-processed Raman spectra of 19 NPS internet products projected onto PCA2 generated from training samples (‘53 representative’ NPS and 22 cutting agents/adulterants reference standards).....	284
Figure 7. 13: PCA projection explained variance plot. This plot shows the calibration (blue) and validation (red) plots of the training samples (i.e. ‘53 representative’ NPS, 14 cutting agents and eight adulterants) and the projection plot (green) of projected ‘query’ samples (i.e. 19 NPS internet products).	285
Figure 7. 14: Replicate Raman spectra for P4 and P5 and the average reference spectrum for SUC.....	286

Abstract

The sheer number, continuous emergence, heterogeneity and wide chemical and structural diversity of New Psychoactive Substance (NPS) products are factors being exploited by illicit drug designers to obscure detection of these compounds. Despite the advances in analytical techniques currently used by forensic and toxicological scientists in order to enable the identification of NPS, the lack of a priori knowledge of sample content makes it very challenging to detect an ‘unknown’ substance. The work presented in this thesis serves as a proof-of-concept by combining similarity studies, Raman spectroscopy and chemometrics, underpinned by robust pre-processing methods for the identification of existing or newly emerging NPS. It demonstrates that the use of Raman spectroscopy, in conjunction with a ‘representative’ NPS Raman database and chemometric techniques, has the potential for rapidly and non-destructively classifying NPS according to their chemical scaffolds. The work also demonstrates the potential of indicating the purity in formulations typical of those purchased by end users of the product i.e. ‘street-like’ mixtures.

Five models were developed, and three of these provided an insight into the identification and classification of NPS depending on their purity. These are: the ‘NPS and non-NPS/benchtop’ model, the ‘NPS reference standards/handheld’ model and the ‘NPS and non-NPS/handheld’ model. In the ‘NPS and non-NPS/benchtop’ model (laser $\lambda_{\text{ex}} = 785 \text{ nm}$), NPS internet samples were projected onto a PCA model derived from a Raman database comprising ‘representative’ NPSs and cutting agent/ adulterant reference standards. This proved the most successful in suggesting the likely chemical scaffolds for NPS present in samples bought from the internet. In the ‘NPS reference standards/handheld’ model (laser $\lambda_{\text{ex}} = 1064 \text{ nm}$), NPS reference standards were projected onto a PCA model derived from a Raman database comprising ‘representative’ NPSs. This was the most successful of the three models with respect to the accurate identification of pure NPS. This model suggested chemical scaffolds in 89% of samples compared to 76% obtained with the benchtop instrument, which generally had higher fluorescent backgrounds. In the ‘NPS and non-NPS/handheld’ model (laser $\lambda_{\text{ex}} = 1064 \text{ nm}$), NPS internet samples were projected onto a PCA model derived from a Raman database comprising ‘representative’ NPSs and cutting agent/ adulterant reference standards. This was the most successful in differentiating between NPS internet samples dependent on their purity. In all models, the main challenges for identification of NPS were spectra displaying high fluorescent backgrounds and low purity profiles.

The ‘first pass’ matching identification of NPS internet samples on a handheld platform was improved to ~50% using a laser source of 1064 nm because of a reduction in fluorescence at this wavelength. We outline limitations in using a handheld platform that may have added to problems with appropriate identification of NPS in complex mixtures. However, the developed models enabled the appropriate selection of Raman signals crucial for identification of NPS *via* data reduction, and the extraction of important patterns from noisy and/or corrupt data.

The models constitute a significant contribution in this field with respect to suggesting the likely chemical scaffold of an ‘unknown’ molecule. This insight may accelerate the screening of newly emerging NPS in complex matrices by assigning them to: a structurally similar known molecule (supercluster/ cluster); or a substance from the same EMCDDA/EDND class of known compounds. Critical challenges in instrumentation, chemometrics, and the complexity of samples have been identified and described. As a result, future work should focus on: optimising the pre-processing of Raman data collected with a handheld platform and a 1064 nm laser λ_{ex} ; and optimising the ‘representative’ database by including other properties and descriptors of existing NPS.

Abbreviations

% HQI	% Hit quality index
2-AI	2-Aminoindane
2D	Two-dimensional
3D	Three-dimensional
4,4'-DMAR	4,4'-Dimethylaminorex
5,6-MDAI	5,6-methylenedioxy-1-aminoindane
5-APB	1-benzofuran-5-ylpropan-2-amine
5-IAI	5-iodo-2-aminoindane
6-APB	1-benzofuran-6-ylpropan-2-amine
ACMD	Advisory Council on the Misuse of Drugs
ACN	Acetonitrile
ANN	Artificial neural networks
AU	Arbitrary units
AUC	Area under the curve
BEN	Benzocaine
CaCO ₃	Calcium carbonate
CAF	Caffeine anhydrous
CC	Correlation coefficients
CCD	Charged coupled device
CCL	Caffeine_Creatine_Lactose mixtures
CL	Confidence limit
CRE	Creatine monohydrate
DART-MS	Direct analysis in real time – mass spectrometry
DEX	Dextrose monohydrate
DIL	Diltiazem hydrochloride
DW	Deionised water
DXM	Dextromethorphan hydrobromide
EDND	European Information System and Database on New Drugs
EI	Electron ionisation
EMCDDA	European Monitoring Centre for Drugs and Drug Addiction
EPS NPS	Enhancing Police Skills on Novel Psychoactive Substances
EU-MADNESS	EUropean-wide, Monitoring, Analysis and knowledge Dissemination on Novel/Emerging pSychoactiveS
EWS	Early Warning System
FEWS	Forensic Early Warning System
FTIR	Fourier transform - infrared
GC-MS	Gas Chromatography Mass spectrometry
GC-MS (EI)	Gas chromatography-mass spectrometry (electron ionisation)
GLU	α -D-glucose anhydrous
GLUT	L(+)-glutamic acid monosodium salt monohydrate
GMP	Good Manufacturing Practices
HPLC	High Performance Liquid Chromatography
ICA	Independent component analysis
i.d.	Internal diameter

InGaAs	Indium Gallium Arsenide
IR	Infrared
IRE	Internal reflection element
LAC	Lactose
LC-MS	Liquid chromatography-mass spectrometry
LID	Lidocaine hydrochloride
LOD	Limit of detection
L-TYR	L-tyrosine
MAN	d-mannitol
MCC	Microcrystalline cellulose
MCS	Minimum common substructure
MDAI	5,6-methylenedioxy-2-aminoindane
MDMA	3,4-methylenedioxy-N-methylamphetamine
MDPV	3,4-Methylenedioxypropylvalerone
MgS	Magnesium stearate
MHRA	Medicines and Healthcare Products Regulatory Agency
ML	Machine learning
MLR	Multiple linear regression
MPA	1-(thiophen-2-yl)-2-methylamino propane
MPD	Methylphenidate
MSC	Multiple scatter correction
MSEs	Multiplicative scatter effects
MVA	Multivariate data analysis
MXT	Methoxetamine
NIA	Niacinamide
NIPALS	Non-linear iterative projections by alternating least squares algorithm
NIR	Near-infrared
NMR	Nuclear magnetic resonance spectroscopy
NPS	New psychoactive substances
NRG	A cathinone brand (Energy)
OOM	Out-of-model
P	Product
PAR	Paracetamol
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PCs	principal components
PHE	Phenacetin
PLS	Partial least squares
PRO	Procaine hydrochloride
PSA	Psychoactive substances Act
PVAL	p-value
R	Reference standard
RMA	Rigaku mixtures algorithm
RSD	Relative standard deviation

S/N	Signal to noise ratio
SERS	Surface enhanced Raman spectroscopy
SIMCA	Soft independent modelling of class analogy
SNV	Standard normal variate
STA	Maize starch
SUC	Sucrose
SVM	Support vector machines
SWGDRUG	Scientific Working Group for the Analysis of Seized Drugs
T ²	Hotelling
TAL	Talc
TAU	Taurine
TCDOs	Temporary class drugs order
Δ ⁹ THC	delta 9-tetrahydrocannabinol
THEO	Theophylline
TSI	<i>Tanimoto</i> similarity index
UNODC	United Nations Office on Drugs and Crime
WCC	Wavelet correlation coefficient
λ _{ex}	Excitation wavelength

Acknowledgements

Foremost, I would like to thank those who have guided and supported me through this insightful learning experience. I am deeply thankful to my supervisors Dr Jacqueline Stair, Dr Stewart Kirton, Dr Suzanne Fergus and Prof. Mire Zloh for their continuous guidance and support throughout my journey, without them the chance to progress in this research would have been impossible. A special thanks to Dr Jacqueline Stair for mentoring me and introducing me to the world of research. I would like to express my sincere gratitude to Dr Stewart Kirton for being my incredibly supportive tutor, mentor, supervisor and line manager! You were always there to answer my never-ending questions! Thank you for giving me perspective, insight and constant motivation into my research and professional development.

I would like to extend my thanks to the Department of Pharmacy, Pharmacology and Postgraduate Medicine, the EU-MADNESS (JUST/2013/DPIP/AG/4823) and the EPSNPS (JUST/2013/ISEC/DRUGS/AG/6429) projects for sponsoring and contributing to the funding of my Ph.D. I would like to thank the Renishaw Support Team, Thermo Scientific Ltd and SciMed Ltd (Rigaku) and particularly Mr Darren Machen for kindly lending us the handheld instruments. I would also like to thank the Joint Pharmaceutical Analysis Group (JPAG) and the Royal Pharmaceutical Society (RPS) for awarding the Geoffrey Phillips Award and the best joint project award for the work presented in this thesis.

I would like to thank Prof. Fabrizio Schifano and Mr John Corkery for introducing me to great opportunities in the research of new psychoactive substances (NPS). Special thank you to Mr John Corkery for his continuous and invaluable support, editing of various pieces of work and disseminating the most up-to-date NPS resources. I would like to thank Prof Gino Martini my professional mentor for his invaluable support and guidance. I would like to thank my colleagues and friends Prof Andrew Hutt and Dr James Collett for their support. A special thanks to Dr Eleftherios Samaras for carrying out the clustering analysis, Dr Jesus Calvo-Castro for introducing me to Origin Pro and for his advice around many issues I encountered in my Ph.D. studies and Dr Katy Kellett for introducing me to mephedrone synthesis. I would like to thank Miss Sara Giroto, Miss Benedetta Berti, Miss Chiara Frisson, Miss Chiara Simionato and Miss Valentina Guarino for their contributions and help in collecting Raman, GC-MS, UV and HPLC data. I would like to thank the laboratory technical staff and in particular Dr Mark Scott for their continuous help with fixing instruments or finding scarce tools to assist me in my research.

This thesis is dedicated to my daughter Sarah, parents and siblings Nermine, Ray, Josh and Nicky for being so supportive and loving. Last but not least, a special thanks to my chocolate and coffee partner Michelle Botha. Thank you all for your love, support and guidance.

“Anyone who has never made a mistake has never tried anything new.”

Albert Einstein

1. Introduction

1.1. Background

New psychoactive substances (NPS) are ‘novel’ recreational drugs, which people use as intoxicants. Although there is not an internationally agreed definition [1, 2], the Advisory Council on the Misuse of Drugs (ACMD) has proposed the following for describing NPS: “psychoactive drugs, which are not prohibited by the United Nations Single Convention on Narcotic Drugs or by the Misuse of Drugs Act 1971, and which people in the UK are seeking for intoxicant use” [3]. The term ‘novel’ means that these substances not only include newly synthesised, invented or discovered psychoactive substances, but also existing molecules, which are being used recreationally in ‘novel’ ways [4, 5]. The Psychoactive Substances Act (PSA) (TSO, 2016), which came into effect in the UK on May 26, 2016 to control the use of psychoactive intoxicants uses a broader term i.e. “psychoactive substances’ and defines an NPS as a substance ‘(a) capable of producing a psychoactive effect in a person who consumes it, and (b) is not an exempted substance ...’ [6]. The PSA states it is illegal to supply, produce, possess with intention to supply or possess in legal custody (e.g. ‘adult prison or young offender institute’ [7]).

NPS are also called by various other names. One such name, ‘legal highs’ is an inaccurate name marketed to deceive potential users about the ‘legal’ status of these substances and their ‘high’-inducing effect. ‘Smart drugs’ is a term used to describe a subcategory of NPS comprising of over-the-counter drugs, dietary supplements and image enhancers mostly used by students and athletes. ‘Designer drugs’ is another nomenclature used, and is a term to describe a ‘new’ derivative arising from a minor chemical modification to existing drugs of abuse with proven psycho-pharmacological activity, which hence circumvent the law [2, 8]. ‘Research chemical’, ‘bath salts’, ‘fertiliser’, ‘incense’, and ‘plant food’, are other terms used to bypass legislation intended to control the supply and distribution of these substances [2, 9, 10].

In addition, NPS are given different names in different countries. For examples, piperazines were known as ‘party pills’ in New Zealand, cathinones were known as ‘bath salts’ in the US [2] and all ‘legal’ NPS were known as ‘kiken’ in Japan [11]. The majority of NPS are synthetic compounds but the category also includes some herbs, herbal extracts and seeds [12]. NPS are sold in a variety of formulations e.g. tablets, capsules, pellets and powders [12] and recently the NPS market has recently expanded to include diverted medicines (i.e. prescription, pharmacy only or over-the-counter medicines), illegally imported medicines, dietary

supplements, patents including failed patents and drug candidates published in scientific papers [13-16].

The emergence of NPS is not a new phenomenon. The first opioid analogues were identified in 1912 and were subsequently controlled *via* the ‘Hague International Opium Convention’ [17]. The current wave of NPS were initially synthesised or used to evade international conventions and mimic the effects of traditional drugs of abuse such as ecstasy and cocaine [3]. The NPS market has proliferated significantly over recent years thought to be caused, in part, by virtue of the books published by Dr Alexander Shulgin [18, 19] and open access to the internet [10, 20-23]. In collaboration with his wife Ann, Dr Shulgin or ‘Sasha’ the ‘ecstasy Godfather’ published the detailed chemical syntheses, and personal experiences of the effects of, over 230 NPS, mostly from the phenethylamine and tryptamine classes [18, 24]. It is thought that these books have stimulated the global explosion of NPS [2]. In addition, access to NPS has changed dramatically over the past decade from face-to-face sales transactions, between a vendor and a buyer in illicit drug markets, to anonymous, often unmonitored discrete online transactions facilitated by vigorous marketing campaigns [20, 21, 23, 25].

Clandestine chemists are currently outstripping law enforcement and NPS research, with the number of newly occurring NPS being formally notified to the European Early Warning System (EWS) dramatically increasing over the years from 13 [26] to 628 recorded by the EU Information System and Database on New drugs (EDND) by January 2017 [27]. Most NPS have a short market life and the continuous emergence of new forms and analogues of these substances is showing no signs of abating.

Over the past decade, the ban of any particular NPS or class of NPS has led to a rapid replacement in the market, which has been described as a ‘game of cat and mouse’. An example of this goes as far back as 1953, where there were numerous international attempts to control new morphine analogues emerging onto the market [28]. Based on the most recent report by the European Monitoring Centre on Drugs and drug Addiction (EMCDDA), approximately two new NPS per week were formally notified to the EWS in 2015 and over 560 NPS were being monitored [29].

In contrast to the limited numbers and types of traditional drugs of abuse, NPS cover a broad chemical and pharmacological spectrum. To enable, and improve, their identification and classification, the EMCDDA has organised them into several categories [30]. This

classification system has evolved over time to reflect the prevalent European trends but has consistently showed that synthetic cannabinoids, synthetic cathinones and phenethylamines are the most popular classes of NPS since 2010 [29, 30].

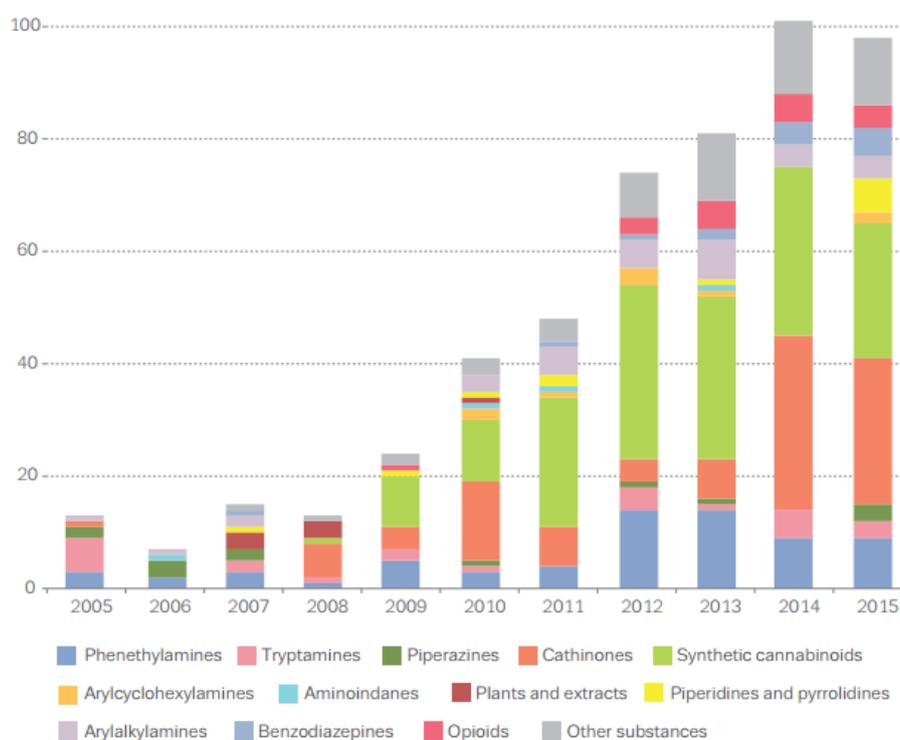


Figure 1. 1: The relative prevalence of new psychoactive substances in Europe over time (Figure reproduced with permission from the EMCDDA [9]).

NPS have been classified in several ways based on the ultimate application. For example, a detailed classification system, which is of interest to psychiatry includes the categorisation of NPS as “synthetic cannabinoids, cathinone derivatives, psychedelic phenethylamines, novel stimulants, synthetic opioids, tryptamine derivatives, phencyclidine-like dissociatives, piperazines, GABA A/B receptor agonists, ...prescribed medications, psychoactive plants/herbs, ...performance and image enhancing drugs” [5].

Another classification system, which can assist decision-making related to treatment in clinical settings includes the categorisation of NPS as four major classes: stimulants, depressants or dissociatives, hallucinogens and synthetic cannabinoids [31]. This classification has recently been expanded on by incorporating opioids, empathogens and psychedelics as new categories. It also includes example NPS membership to these classes based on their legal categorisation (Figure 1.2).

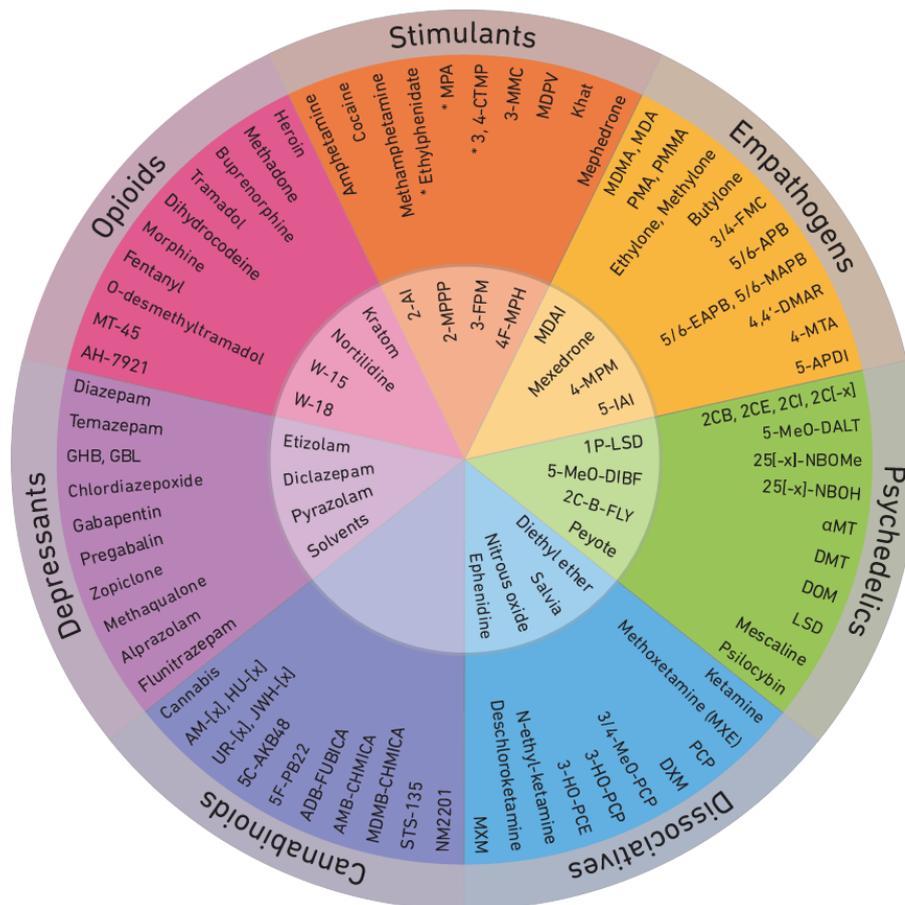


Figure 1. 2: The Drug Wheel. The outer ring shows “the controlled NPS under the Misuse of Drugs Act 1971 or The Human Medicines Regulations 2012”, whereas the inner ring shows “the controlled NPS under the Psychoactive Substances Act 2016 (PSA). *Temporary Class Drug Orders (TCDOs)” (Figure reproduced with permission from Dr M. Adley) [32].

The use of NPS is prevalent amongst clubbers, party goers and those attending music festivals, high-risk drug users e.g. those who inject heroin, abstinence treatment entrants, MSM (i.e. men who have sex with men), people who engage in group sex, and prisoners [33-38]. In addition to the wide prevalence and unsafe practices, the increased diversity of NPS, surge in the number and diversity of users and distributors, the increased availability, rapid speed of emergence, a lack of knowledge on the identity, pharmacology, toxicodynamics and toxicokinetics of the NPS [39, 40], and variation in potency, formulations and routes of administration of these substances are posing significant public health concerns [2, 38, 41]. NPS abuse has been associated with unpredictable toxicities i.e. violence and aggression [42], sympathomimetic, cardiovascular and neurologic symptoms [43-47], major organ failure [48-50], psychosis [43, 46, 51, 52], withdrawal symptoms [53], and fatalities [42, 54-61].

At present, the dearth of pharmacological and toxicological knowledge on NPS increases the potential risks and harms to users and greatly impacts treatment decisions [62]. Furthermore,

the problem of NPS is severely underestimated due to social stigma, self-treatment and lack of awareness of users, which often lead to under-reporting and lack of knowledge of the real size of the problem [63]. Knowledge about NPS is often only captured if a user is being arrested or *via* anecdotes and national surveys [64, 65]. Users usually do not know the actual content in the NPS products they consumed unless a chemical analysis is performed [38]. Moreover, the net impact of NPS abuse on the national health services, crime rates and societal burden is still unknown. This is because NPS fall outside the drug monitoring systems by the public health agencies [66, 67] and are not easily detected using common prison, forensic and toxicology screening tests [68, 69]. Therefore, NPS became preferred among people subject to regular checks such as those working in the military and those under the criminal justice system during their probation period because, again, NPS are not detected using common detection kits [7, 64, 70]. This makes the improved detection of NPS in the field (i.e. hospitals and emergency settings, music festivals, road-side, prisons, border control, etc.) key to optimise and accelerate treatment decisions, for public health purposes and to inform law enforcement bodies.

Unlike pharmaceuticals, NPS are commonly produced in clandestine laboratories and hence, are not produced or controlled under 'Good Manufacturing Practices'. With the reduced availability of drugs of abuse e.g. MDMA in the early 2000s, purity of these compounds declined and illicit drugs were 'cut' to maintain their weight and enhance dealers' profits [71]. As a result, NPS supplemented the repertoire of illicit drugs as adulterants or add-ons. They have also been shown to be produced as, or incorporated into, mixtures to mimic the effect of popular illicit drugs [38, 72]. Compounds sold as pure substances often covertly contained mixtures of one or more active substances of varying purity [38, 73, 74]. For example, seized NPS in 2011/ 2012 have been shown to contain mixtures of up to eight different controlled or uncontrolled NPS or cutting agents [75]. In 2013, the UK 'Forensic Early Warning System (FEWS)' reported that 81 % of seized NPS were mixtures that incorporated more than one active NPS. Of those, 36, 35 and 1 % contained two, three and seven different active NPS, respectively [76]. This trend continued in 2014 with 91 % of seized NPS samples made up of mixtures [77]. Information is still lacking on NPS purity after the UK PSA 2016 came into effect. An extensive review of NPS mixtures particularly the cathinones and the adulterants and cutting agents incorporated in them or co-consumed with them has recently been conducted by Guirguis et al. [38].

The heterogeneity of NPS products in terms of composition, concentrations and number of constituents is one of the tactics used by drug designers to evade detection and circumvent the law [78]. This added complexity makes the identification of NPS and the discrimination between NPS and excipients, using traditional chromatography or spectrometry techniques, limited to analysts with substantial expertise in the interpretation of spectral or chromatographic data.

1.2. Chemical analysis of NPS

Numerous laboratory-based techniques are recognised as validated analytical tools in forensics for the identification and quantification of NPS in a range of sample types such as blood, urine, aqueous solutions, oral fluid, hair, post-mortem matrices and solid samples [79-93]. These techniques include gas chromatography-mass spectrometry (GC-MS), high performance liquid chromatography (HPLC) and liquid chromatography-mass spectrometry (LC-MS).

Gas chromatography-mass spectrometry has been shown to be the predominant chemical technique employed by European Union countries for the analysis of NPS mixtures [94]. Despite being a main confirmatory forensic technique, GC-MS is not available as a handheld device and it is not suitable for high-throughput screening [88]. It is destructive to the samples and involves sample pre-treatment [95], which may filter out the cutting agents in the mixtures, and, hence potentially tamper with a chain of evidence. If the amount of sample seized is sufficient, a subsample is often taken before any pre-treatment is carried out to combat disruption to the legal aspects of a seizure. Nevertheless, sub-sampling may not be representative due to unknown homogeneity in street samples [96]. Reference standards are often required for analysis using GC-MS. However, they may not be available, due to controlling legislation and prohibitive costs, causing unnecessary delays [2, 97]. Analysis using GC-MS may also misidentify thermally unstable NPSs [98].

Unlike GC-MS, other lab-based techniques such as direct analysis in real time – mass spectrometry (DART-MS) has been developed as a rapid, selective, non-contact, high throughput screening tool for the characterisation of cathinone mixtures [86]. This technique does not require sample preparation and provides instantaneous results based on exact masses for samples in their solid form [99]. However, it may be challenging to differentiate between closely related NPS analogues with this technique alone. This is because of extensive

fragmentation and similar fragmentation patterns between positional isomers such as the flephedrone analogues 3-FMC, 4-FMC and 5-FMC [86].

Over the past decade, in-field techniques became available for law enforcement and first responders for the screening and identification of NPS. These techniques have the advantage of transferring the lab to the field. Numerous handheld, portable, remote and disposable kits were developed for the detection of drugs of abuse and NPS [100, 101]. For example, standard presumptive tests such as the Marquis, Mandelin, Scott's and Zimmerman colour tests [102] were applied to the identification of NPS analogues. Although promising, poor selectivity may sometimes require more than one test to be applied for confirmation of identity of the NPS in question [92, 103]. Electrochemistry with mercury dropping electrodes has been evaluated as another presumptive tool for the detection of mephedrone. However, mercury is a known harmful substance and, therefore the method was not encouraged [104]. Conversely, electrochemistry with boron-doped diamond or glassy carbon macro-electrodes was evaluated for the detection of cathinone analogues. Results showed that pH is a major step-limiting key player and, therefore a change in pH may impact the electrochemical selectivity [104]. Very recently a molecular recognition approach has shown promise for the positive identification of mephedrone alone and in mixtures adulterated with benzocaine, caffeine, lidocaine and paracetamol [105]. The main limitation of this sensor molecule is that ion salts of mephedrone may contribute to false positive results and, therefore mephedrone must be in its free base form.

Conventional in-field immunoassay kits are also showing promise for the in-field detection of NPS [106, 107]. However, the time lag required to synthesise them i.e. following the emergence of an NPS, the long analysis time and poor cross-reactivity between closely related NPS, which either lead to false positive or false negative results of NPS new analogues is limiting their use in this field [108-111]. Recently, Randox marketed in-field immunoassay kits for the detection of synthetic cathinones. These kits achieved levels of detection of 0.18 and 9.2 μgL^{-1} . Ellefsen et al. independently validated these tests and demonstrated that false positive results were observed between closely related analogues [107].

Unlike the presumptive tests, portable devices have shown promise in the detection of NPS. Recently, a portable near-infrared (NIR) in conjunction with careful data pre-processing has successfully identified 8/11 NPS forensic samples [112]. Additionally, surface enhanced Raman spectroscopy (SERS), a novel approach, which was applied using a portable Raman device in conjunction with fractional factorial design and achieved a detection limit for

mephedrone of $1.6 \mu\text{g mL}^{-1}$. The method involved enhancing the Raman signal through the galvanisation of the pre-1992 British 2 pence coin with silver [113, 114]. Another SERS attempt successfully discriminated methcathinone from methamphetamine using a gold nanorod and a mini Raman instrument [115]. However, these proof-of-concept methods need to be applied to the vast array of prevalent NPS in order to be credible with respect to a wider application. With the exception of DART-MS and NIR, all the methods described above are destructive to the sample, require a relatively long analysis time, significant instrument expertise, which would preclude use by novice law enforcement and frontline healthcare professionals and the use of chemical reagents, which may be harmful to the operator and environment [101].

Vibrational spectroscopy, available in handheld and portable versions, including infrared (IR) and Raman spectroscopy, have emerged as ‘first pass’ analytical techniques for the screening and identification of seized drugs of abuse, particularly, where a rapid in-field non-destructive, non-invasive identification is required [101, 116-118]. The Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) has classified both IR and Raman as techniques with a higher discriminatory power (i.e. category A) than colour tests and immunoassays (category C) [119]. Infrared has been established since the nineteenth century [120] and has always been a preferred forensic technique due its ability to perform rapid analysis in the solid-state with high selectivity and specificity [121-123]. It has also proven to be a reliable technique for the discrimination of NPS isomers [121, 122], in addition to its availability in handheld and portable versions, which are advantageous for in-field testing for first responders [124]. However, mid-IR and Raman are considered complementary techniques that can be employed together for the full determination of the vibrational modes of a molecule and, hence its molecular structure [125].

Raman spectroscopy offers a number of advantages over IR for the in-field detection of NPS such as non-contact through package analysis [124, 126, 127], low sensitivity to cutting agents [128, 129], water or moisture [116, 124] and physical properties [130], and the possibility of using a smaller sample size (e.g. $1 \times 1 \times 5 \mu\text{m}$) due to high spatial resolution [116]. In contrast, placing the sample in contact with the diamond internal reflectance element (IRE) for IR may tamper with the chain of evidence [131]. Moreover, mid-IR spectrometers cannot characterise fingerprint groups absorbing in the low wavenumber range (i.e. less than 400 cm^{-1}) [120]. For these reasons, Raman spectroscopy has been employed in research and forensic analysis [132,

133] for the characterisation of drugs of abuse such as 3,4-methylenedioxy-N-methylamphetamine (MDMA), cocaine, heroin [134-137] and NPS [128, 138-140].

Despite its low sensitivity to moisture and physical properties compared to IR, Raman spectroscopy has shown success in the identification of polymorphic, hydration and salt forms of samples, thus providing information on their physical state [141]. Various salts form different crystals motifs/ supramolecular structures, which are mainly due to the counter ion of the acid used and based on the conditions (e.g. polymorphic transitions, experimental conditions, etc.) and composition (e.g. presence of solvent molecules) [142]. Hydration states are often distinguishable *via* peak shifting, whereas, salt forms are often distinguishable *via* a change in signal intensity associated with the change in the sample surface position or *via* the presence/ absence of Raman features. For example, characteristic strong Raman bands attributed to the symmetric breathing of the pyrimidine ring in barbiturates and their salts were observed at ca. $629 \pm 8 \text{ cm}^{-1}$ for the free barbituric acids and at ca. $652 \pm 4 \text{ cm}^{-1}$ for the sodium salts, thus, enabling the discrimination between both forms [143]. In the same study, unlike the free acid form, a weak Raman band at ca. $1585 \pm 15 \text{ cm}^{-1}$ was present in the corresponding barbituric sodium salt form only. These effects could be the results of the distortion of the molecules in the crystal lattice and the potential change in site symmetry. Additionally, these effects could be the results of vibration coupling of more than two molecules [142]. In addition, drug substances may undergo polymorphic transitions between various forms. These forms often exist in various crystal states and molecular arrangement within the crystal lattice. Drug polymorphic forms exhibit different physicochemical properties such as particle morphology, bioavailability, solubility, dissolution etc. [144]. Therefore, in this thesis, the different salts and hydration forms were explicitly stated because specific particle morphologies/ polymorphism can exhibit a great impact on the library spectra used to identify unknown NPS.

Our initial published work investigated the identification of NPS in products obtained over the internet, using Raman, NIR and Fourier transform - infrared (FTIR) handheld devices [128]. This study illustrated that handheld Raman outperformed both NIR and FTIR to give a greater discrimination between NPS and excipients i.e. cutting agents and adulterants in internet products. However, the study revealed that the use of on-board libraries and library matching in-built algorithms was limited for identifying NPS in NPS mixtures due to swamping by fluorescence signals; overlapping analyte-excipient bands, where the excipient bands occurred

at the same spectral region as the analyte; sample heterogeneity and variable Raman activity of constituents.

Very recently, in-field analysis of selected NPS using Raman spectroscopy has gained regulatory acceptance in the UK [145] and has been encouraged to be used by law enforcement in the US [146]. In the UK, Trunarc Raman analyser (HazmatLINK Ltd, UK), Alpha IR and Mobile IR (Bruker UK Limited) were approved by the Home Office for the identification of methylmethcathinone and results achieved from these instruments can be used as evidence in court [145].

Raman spectroscopy is a technique that uses inelastic scattering from a sample to produce a fingerprint and enable the identification and quantification of molecules based on their peak positions and relative peak intensities respectively [147, 148]. Raman spectroscopy simply involves the interaction between light (monochromatic laser source) and matter (the vibrating molecules of the sample), which result in the scattering of photons or particles of light. Scattering or other forms of interaction between light and matter such as absorption or emission of light depend on the molecule and its energy states [101] and the wavelength of the incident light [149]. In Raman spectroscopy, two types of scattering occur: Rayleigh (elastic) and Raman (inelastic) [125].

Most of the scattered photons oscillate at the same frequency of the incident electromagnetic field of the laser source. This is termed Rayleigh scattering and involves no exchange of energy between the photon and molecule. In contrast, 1 in a million of the scattered photons exhibits a Raman shift with a frequency greater (stokes) or less (anti-stokes) than the frequency of the incident electromagnetic field of the laser source. This is termed the 'Raman effect' and involves an exchange of energy between the photon and molecule [117]. The Raman shift and the change in energy for inelastic Raman scattering is equivalent to the unique vibrational frequency of the molecule and, hence is more informative about the molecule [101, 120]. Figure 1.3 shows a simplified Jablonski diagram which explains the types of scattering and illustrates the transition of molecules between both the ground and first excited vibrational energy states and the virtual state (it is not a real energy state) [125].

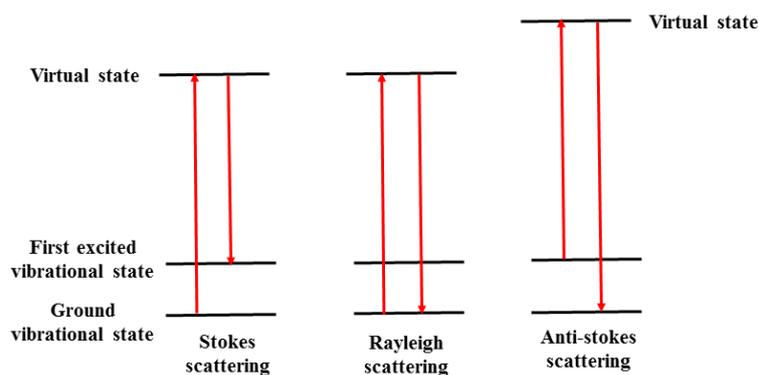


Figure 1. 3: A simplified Jablonski diagram showing the types of scattering resulting from the interaction between light and matter. The stokes (left), Rayleigh (middle) and anti-stokes (right) scattering (Figure reproduced from Medicines Complete [150]).

Raman scattering is associated with the polarisation of the electron clouds on the molecules of the sample and a transient induced dipole. During the normal mode of molecular vibrations, a band arises from Raman active groups because of a change in polarisability i.e. a distortion of the electron clouds [120]. In contrast, IR has a different selection rule, where bands arise as a result of absorbing energy and subsequent change in the dipole moment leading to molecular transitions between vibrational energy levels [125].

In addition, recent studies have evaluated the use of Raman for the analysis of NPS products [113, 114, 132, 140, 151, 152]. Maheux and Copland used combined analytical techniques including Raman spectroscopy for the identification of cathinones in seized NPS samples [151]. The discrimination between cathinone regioisomers [152] and derivatives [132] has also been evaluated using benchtop Raman instruments employing a laser excitation wavelength (λ_{ex}) of 785 nm. Bell et al. recently reported on the use of Raman spectroscopy for the identification of > 200 seized NPS products using an λ_{ex} of 785 nm [140]. From these studies, it is clear that the accurate Raman analysis of NPS products is often challenged by intense fluorescent backgrounds resulting from impurities and/or cutting agents, which masks the weaker Raman signal from the NPS in the mixtures [128, 153, 154].

Numerous approaches were investigated to reduce fluorescent backgrounds in Raman analysis of NPS. One approach is SERS, which enhances the NPS Raman signal while also reducing interference due to fluorescence [113, 114, 155, 156]. However, expertise and sample pre-treatment is often needed. An alternative approach to reduce fluorescent backgrounds and, which requires no sample preparation is the use of a longer NIR λ_{ex} (e.g. between 750 and 1000 nm) [101]. This has been shown to avoid interference from fluorescence and improve identification of traditional drugs of abuse such as cocaine and amphetamine [154, 157]. At

present, a few studies have evaluated the use of Raman spectroscopy with a low energy long laser λ_{ex} (e.g. 1064 nm) to reduce fluorescence in different sample types [117, 158, 159] and investigate the feasibility of improving the identification of drugs of abuse [160]. To our knowledge, this approach has only been evaluated in two recently published studies for a limited range of NPS samples [138, 161].

1.3. In-field detection of NPS using Raman instruments

In-field detection using vibrational spectroscopy based on an instrument's in-built algorithm has been investigated [128]. Despite the dearth of information regarding how proprietary algorithms actually function, published manufacturers' information is presented here. In-built algorithms in handheld instruments based on vibrational spectroscopy are crucial as they produce 'actionable results' for first responders to arrest/ confiscate and/or treat NPS related overdose/ toxicity. In-built algorithms commonly involve mathematical or statistical calculations in conjunction with pre-processing techniques to measure the probability, correlation, consistence or closeness between spectra of the 'unknown' substances against reference spectra stored on-board the instrument. In-built algorithms enable spectral interpretation by non-experts i.e. screening (for example, does it contain mephedrone?) and identification (what is it?) [162].

1.3.1. In-built algorithms in handheld Raman instruments

Algorithms have been developed to enhance the detection of drugs of abuse with Raman spectroscopy using techniques such as discrete peak positions for library searching and multiple-algorithm approaches to improve the probabilistic capability of drug identification [112, 137]. In our initial publication, we assessed the capabilities of a few in-built algorithms in handheld FTIR, NIR and Raman instruments with respect to accurate identification of NPS [128]. In the following Section, in-built algorithms in handheld Raman instruments employed in the work undertaken in this thesis are described.

1.3.1.1. The p-value algorithm

Using handheld Raman instruments such as the TruscanTM GP instrument (Thermo-Fisher Scientific Ltd), Raman responses are measured using the p-value (PVAL) algorithm, where a 'positive match' ($\text{PVAL} > 0.05$) meant that the 'unknown' sample spectrum is consistent with the library spectrum (i.e. > 95 % confidence limit (CL)). A 'no match found' result ($\text{PVAL} \leq 0.001$) means that the 'unknown' sample spectrum was not consistent with any library

reference spectrum. A ‘no positive match’ ($0.001 < PVAL \leq 0.05$) means that the ‘unknown’ sample spectrum was ‘fairly consistent’ with at least one of the library spectra. In such cases, the in-built algorithm is unable to unambiguously correlate the sample spectrum to library spectra but instead, highlights the most similar library spectrum, within the confidence interval described, as the most likely match. Inaccurate suggestions of sample identity may occur if substances in the library or sample mixtures have chemical structures closely related to compounds in the on-board spectral library (Figure 1.4). It is important to note that the PVAL does not measure the correlation coefficient between two spectra (sample and reference spectra), but instead it measures the dissimilarity between them against the uncertainty of the measurement [162].



Figure 1. 4: A screenshot of the ‘actionable’ results obtained with Truscan™ GP using the PVAL algorithm. Sample responses include ‘positive match’ (left), ‘no match found’ (middle) and ‘no positive match’ (right) (© 2015 Thermo-Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo-Fisher Scientific Inc. and its subsidiaries).

The PVAL algorithm is a probability-based correlation between the ‘unknown’ sample spectrum and the library reference spectra. It is calculated using the Bayes’ theorem:

$$P(A|B) = (P(B|A)P(A))/P(B) \quad \text{Equation 1}$$

Where,

$P(A|B)$ = Conditional probability of A such that B exists

$P(B|A)$ = Conditional probability of B such that A exists

$P(A)$ = Marginal probability of A

$P(B)$ = Marginal probability of B

The in-built algorithm, involves an automatic interpolation of the x-axis of the spectrum and the calculation of the 1st and 2nd derivatives to the y-axis prior to PVAL analysis. Time to acquire a spectrum depends on the signal to noise ratio (S/N) of a sample. The S/N cut-off for test samples and library signatures is 5 and > 50 arbitrary units (AU) respectively. The PVAL

algorithm is often a preferred algorithm for the screening of known substances e.g. the screening of raw materials in the field of pharmaceuticals [163] because screening algorithms, such as PVAL, outperform identification algorithms e.g. % HQI (see Section 1.3.1.2) with respect to substance detection due to the ability to tailor the algorithm threshold according to substance [162].

1.3.1.2. The hit quality index algorithm

The hit quality index (HQI) is an alternative to the p-value algorithm. In handheld Raman instruments such as FirstGuard™ (SciMed Ltd (Rigaku)), Raman responses are measured using the HQI (%) algorithm. HQI is a measure of the correlation between the measured spectrum of an unknown material against library reference spectra [164, 165]. The HQI is calculated using the following equation:

$$\text{HQI} = \frac{(\text{library} \cdot \text{unknown})^2 * 100}{(\text{library} \cdot \text{library})(\text{unknown} \cdot \text{unknown})} \quad \text{Equation 2}$$

Where,

‘library’ is the library or reference spectrum

and ‘unknown’ is the unknown spectrum of the sample

An HQI of 100 % means that the correlation between the ‘unknown’ spectrum and the reference spectrum is absolute (i.e. 1). An HQI of 0 % means that the ‘unknown’ spectrum and the reference spectrum are orthogonal to each other [162]. The higher the % HQI, the more consistent the sample spectrum is with the corresponding on-board library spectrum. The HQI algorithm is suitable for the rapid identification of unknown materials as it is capable of searching a large number of spectra in a relatively short time. This method was designed by McCreery et al. for the identification of pharmaceuticals through USP vials [166]. A limitation of the % HQI in-built algorithm is that no threshold is set to determine positive correlations. Therefore, it does not confirm the quality of the identification nor provide the probability of the statistical consistency between the unknown and library spectra [164]. The resulting percentage may actually reflect the spectral contribution of different constituents, and hence be artefactual. Additionally, the % HQI may not be sensitive to changes in the composition of mixtures. Therefore, false positives or misidentification may occur. Another limitation of this correlation-based algorithm involves its inability to discriminate between analogous substances, which share common functional groups [162, 167].

1.3.1.3. The wavelet correlation coefficient and mixture algorithms

A few handheld Raman instruments such as the ProgenyTM (SciMed Ltd (Rigaku)) instrument record Raman responses using multiple in-built algorithms. For example, for the ProgenyTM instrument, Raman responses are measured using % HQI, wavelet correlation coefficient (WCC) and the ‘Rigaku mixtures algorithm’ (RMA). The % HQI or WCC algorithm are often selected for match-testing pure substances against pure reference spectra. Conversely, the RMA is preferred for match-testing mixtures against pure reference spectra because of its claimed ability to identify up to six constituents in mixtures. Using RMA, ProgenyTM is claimed to identify the constituents in mixtures and provide an estimation of the spectral contribution of each of the individual constituents. The higher the value of the match for WCC, % HQI (as described above) and RMA, the higher the correlation between the ‘unknown’ and the library spectra.

The wavelet correlation coefficient (WCC) transform involves a pre-processing step, whereby Raman spectral background and noise (i.e. shot noise, dark-current noise and readout noise) interferences are de-emphasised and Raman signals are accentuated [168, 169]. Shot noise constitutes the “statistical variation in the number of photons” reaching the detector [170]. The dark noise constitutes the “statistical variation in the number of electrons” generated in the absence of the laser light, whereas, the readout noise is the electronic noise in the detector and it plays a key role in determining the limit of detection of the detector [170]. The wavelet transformation improves spectral classification by compressing the data through identifying ‘wavelet functions’ or ‘wavelet coefficients’ that represent specific Raman peaks [168]. ‘Wavelet coefficients’ are calculated through the analysis of the signal, then processed. Subsequently, the signal is re-constructed from the processed ‘wavelet coefficients’ [169]. The WCC algorithm identifies the component with high Raman activity in an unknown sample [171].

The ‘Rigaku mixtures algorithm’ (RMA) is a proprietary detection algorithm and hence, it is not fully understood how it is computed. This algorithm mainly identifies individual components in mixtures, where results are expressed as percentage correlation coefficient (CC) values [171]. The algorithm gives a higher weight to the spectral contribution rather than the actual concentration of constituents in a mixture, and so may be influenced unfavourably by strong Raman active components.

1.4. The challenge of mixture analysis using Raman spectroscopy

To overcome the complexity of Raman analysis of drug mixtures, Raman and chemometric approaches were adopted. The International Chemometrics Society defined chemometrics as “the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods” [172].

1.4.1. Raman applications

To improve the identification of constituents in complex mixtures, several Raman applications can be adopted. These include Raman micro-spectroscopy, enabling the identification of small masses (e.g. down to five pico-grams) [116, 173]; Fourier transform - Raman with beam diameters in the range of 100 - 1000 μm , enabling the examination of large surface areas [174]; transmission Raman spectroscopy, probing and quantifying the bulk content of pharmaceutical tablets or capsules [175], Raman mapping using grid measurements [176] and macro-Raman systems followed by micro-Raman measurements to give information on individual particles [177] and the identification of unique crystal habits using microcrystalline testing followed by Raman micro-spectroscopy [139]. Although promising, these applications are not rapid, are unsuitable for in-field analysis and require substantial scientific expertise to use.

1.4.2. Chemometric applications

In contrast to Raman-derived applications, chemometric approaches to the analysis of complex mixtures, such as multivariate data analysis (MVA) are used. Multivariate data analyses include methods that aim to separate and extract data structure from noise [172, 178]. For example, Raman spectral data include the Raman signal of interest and noise. The application of MVA for mixtures are favoured because they are rapid and practical for forensic applications such as the analysis of ink and paint [179, 180]. Multivariate data analysis demonstrated its suitability for the analysis of spectroscopic data owing to its multivariate nature and large number of highly correlated variables [179]. Furthermore, MVA has the advantage of extracting optimal information from the entire spectrum rather than individual peaks [181]. The use of MVA in conjunction with Raman spectroscopy was revisited by forensic scientists to enable the identification, classification and prediction of amounts of constituents in mixtures of illicit drugs [95, 120, 176, 182-185]. Coupling other vibrational techniques such as FTIR and NIR with chemometrics have also been evaluated for the identification and classification of drugs of abuse and NPS respectively [186, 187]. In the studies described above, chemometric

analysis involved pre-processing of spectral data, exploratory principal component analysis (PCA) followed by classification or prediction methods.

In order to enhance the composition profiling of mixture components from Raman spectra, all factors that are not pertinent to the sample that could dominate the analysis should be removed through spectral pre-processing [120]. Raman spectra may contain noise and fluorescence interferences (Figure 1.5), which could be attributed to: 1) instrument artefacts such as noise (CCD (charged coupled device) noise, background noise), changes in laser power, fluctuations in spectrometer performance and temperature, changes in optics geometry; 2) analysis effects such as different powder opacity between measurements, variations in focal distance etc.; 3) sample effects such as vibrational frequency of scattering molecules, amount of scattering groups, presence of fluorescing species, homogeneity, etc. and 4) environmental effects (e.g. ambient light, cosmic rays etc.) [188]. Variations in baselines and absolute intensities and overlapping Raman bands [176, 184] have also restricted the use of absolute Raman signals in analysis and limited the use of Raman as a forensic tool [188].

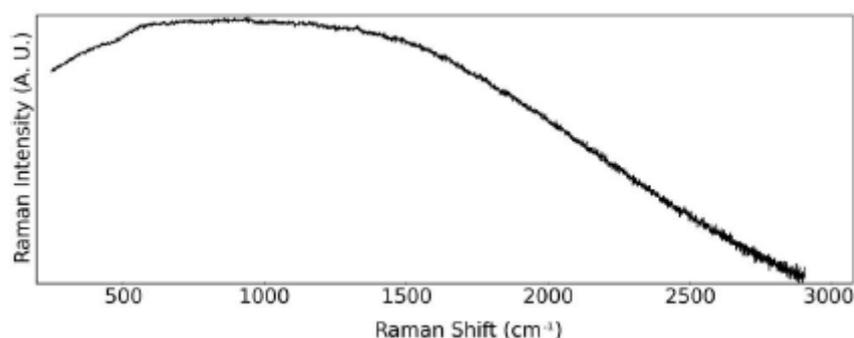


Figure 1. 5: Raman spectrum of an NPS branded product with no match found (laser excitation wavelength = 785 nm). Reproduced from Analytical Methods [128] with permission from the Royal Society of Chemistry.

Statistical pre-processing of Raman spectral data became a preferred non-chemical technique to correct for these variations and is often applied prior to MVA [95, 178]. Collinearity (i.e. assumption that the data points are interrelated) between the data points, variations in absolute intensities, noise and fluorescence have been shown to impact the accuracy of machine learning models [183]. Therefore, pre-processing is indispensable to achieve the optimal extraction of Raman data from noise interferences [176, 189], thus enabling the identification of target compounds in complex mixtures [184, 190].

In general, for the application of MVA in a well-controlled dataset, the number of samples should be much larger than the number of variables. In such cases, the number of degrees of

freedom is small. However, this is not the case with Raman spectral data (i.e. > 500 variables or data points in typical handheld and > 3000 variables in typical benchtop Raman instruments). Therefore, feature selection or variable reduction techniques are often required prior to conducting MVA. Examples of feature selection techniques include the ‘genetic algorithm’ method [191] or increasing the weighting of significant spectral features of specific substances [192]. The latter method is more suitable for a small dataset including a limited number of samples or a homogeneous dataset with a narrow range of analytes.

Following pre-processing and/ or feature selection/ variable reduction, exploratory PCA is usually conducted. PCA is known as “the mother of all methods in multivariate data analysis” and it forms the basis for classification of complex data matrices [178]. Exploratory PCA reduces the dimensionality of the dataset by transforming the original data matrix such that the largest variance between samples is explained by discrete orthogonal principal components (PCs) [120]. This concept is very useful in spectral data because the number of variables or data points is largely reduced, maximum information is extracted and influence of noise is reduced [193]. Figure 1.6 shows a descriptive schematic of data reduction by PCA.

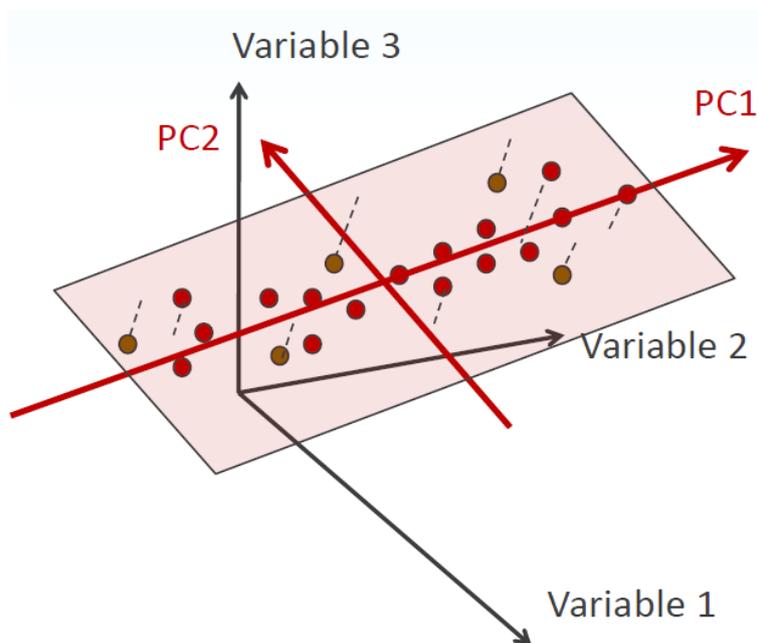


Figure 1. 6: An example of a two-dimensional plane, which best describes the data (Figure adapted from Camo Software AS, Oslo, Norway) [172].

The first PC represents the eigen vector with the highest eigen value (variance), whereas the second PC, which is orthogonal to PC1, has the second highest eigen value (Figure 1.6) [178, 193]. Exploratory PCA is a useful unsupervised method, which can be used to visualize and find patterns, similarities and differences in multivariate data and identify outliers [178].

Coupling of PCA with Raman spectroscopy has strengthened Raman analysis by enabling the exploration of the main structure of the data, identification of outliers [179], exploration of latent variables in complex Raman data [184] e.g. Raman spectral data of illicit drug mixtures [136, 154, 192] and resolution of spectral data to discriminate between different components in mixtures [95]. In these studies, Ryder et al. analysed simple binary mixtures incorporating one drug of abuse (i.e. cocaine, heroin or MDMA) and one cutting agent [192]. However, they could not design a universal model for the classification of all three drugs. They developed multiple models instead [192]. Subsequently, they developed baseline subtraction methods with the use of support vector machine (SVM) to improve the robustness of their methods [176, 190]. Noonan et al. developed a universal PCA method for simulated quaternary mixtures, each incorporating a drug and three excipients [95]. In their study, spectra of closely-related substances were differentiated as a result of robust pre-processing [95].

The use of chemometrics has recently been employed in the determination of NPS classes in NPS seized samples by comparing the pre-processed NIR spectra to reference spectra using PCA [187]. PCA can also be employed to build calibration models and then test ‘unknown’ samples against pre-defined classes in a calibration set [184]. However, as mixtures increase in complexity (i.e. low concentration of NPS in the mixture, colourful constituents, excipients with greater Raman activity than the target analyte), the ability of MVA to separate the different components is reduced [184]. This was observed with cocaine mixtures [191], but has been shown to improve by incorporation of spectral pre-processing [183, 192].

Classification methods are usually applied using the calculated PCs. Classification methods include soft methods such as ‘Soft Independent Modelling of Class Analogy (SIMCA)’ and regression methods such as ‘principal component regression (PCR)’. SIMCA usually measures the ‘Mahalanobis distance’ from the ‘unknown’ sample and the model developed using the scores of the retained PCs and assign ‘unknown’ samples to pre-defined classes [184, 194, 195]. However, SIMCA has been shown to result in a high number of inaccurate classifications as compared to regression-based classification of complex mixtures [184]. This is because SIMCA performs best when sample classes are well defined, in such cases, the distance of ‘unknown’ samples to the model is small and the leverage value is low owing to the low variance within individual classes. However, when samples are adulterated with excipients assigned to multiple classes, the misclassification rate increases [184]. Alternatively, PCR is used for the identification of an ‘unknown’ substance by assigning it to a pre-defined class

based on set threshold values [184]. PCR works best for simulated mixtures as prior knowledge of the mixture will allow the optimisation of these thresholds. Other regression methods such as ‘multiple linear regression (MLR)’ have shown to be limited in efficiently identifying the components in complex NPS data matrices [120].

Data reduction *via* PCA has also been shown to improve Machine learning (ML) models [183]. Machine learning comprises computer algorithms that are trained to recognise latent patterns in complex data [176]. ML techniques such as ‘support vector machines (SVM)’ and ‘artificial neural networks (ANN)’ have been shown to outperform MVA techniques such as PCR in their ability to identify patterns in corrupt data. Both SVM and ANN have been evaluated in the qualitative and quantitative Raman analysis of complex drug mixtures e.g. illicit drug mixtures [176, 191]. The simplest SVM method is a linear SVM used for binary classifications by increasing the margin, which may separate between two classes e.g. NPS or non-NPS [176]. In contrast, ANN is used for the prediction of non-linear translation between the input (layers of neurons) e.g. training set of Raman spectral data and the output resulting from each layer of neurons. The cumulative outputs from all layers represent the predicted outcome e.g. amounts of NPS in various mixtures. Training of the model may not be straight forward as the weight of neurons’ inputs, the number of hidden layers and the number of neuron in each hidden layer have to be adjusted for each dataset [191]. In the studies contained in this thesis, chemometric analysis involved pre-processing of spectral data, exploratory PCA followed by prediction *via* PCA projection methods.

1.5. Facing the continuous flood of NPS emergence

Previously unseen NPS continue to emerge at a rapid rate. This sheer number of NPS and the speed with which they are emerging hamper their accurate detection and classification. This issue is compounded, as challenges associated with obtaining commercially sold NPS reference standards (i.e. cost, time and licensing requirements) hinder the building of extended instrumental libraries, thus raising analytical and forensic obstacles [94]. As a result, newly emerging NPS are often undetected using library-based matching algorithms because of sample complexity, absence from libraries, limitations of in-built algorithms [128] or high selectivity of in-field techniques such as commercial immunoassays [104].

If the spectral data for a compound is not stored in on-board libraries, then an expert interpretation of the spectral data is often required for the full or partial identification of that

compound. It is very challenging to fully elucidate the chemical structure of an unknown substance using spectral data such as IR, Raman or mass spectra, particularly if no reference standard is available for this substance. This is due to unknown correlations between structural and spectral information e.g. IR ‘structure-specific coupled vibrations’ [196, 197], which requires, a high structural similarity between the ‘unknown’ and library substance in order to identify the ‘unknown’ compound and give confidence to the match (although false positive results may still occur [197]).

Despite the fact that spectral similarity approaches are known to be ‘relative’ because they may incorporate other analogy principles such as complementarity and equivalence, they are essential in exploring relationships between spectral and structural information [197, 198].

As a result of the vast number of NPS and their continuous emergence, the ACMD has suggested that in an effort to gauge the potential effects of a never-before-seen substance, the psychoactivity of a new analogue could be translated or extrapolated from previous neurochemical knowledge on existing NPS [4]. Although not fully proven owing to the impact slight molecular changes can have on receptor binding affinities (and subsequently on psychoactive effect), structural similarity with existing NPS may assist in categorising according to the class of substance, and hence predicting potential pharmacological activity for newly emerging NPS. In fact, numerous methods have been used to measure structural similarity. These include computing the *Tanimoto* similarity index (TSI), and Euclidean distances between molecules and molecular vectors respectively [197, 199].

In contrast, methods used to measure spectral similarity include in-built library-matching algorithms (as described above) such as the P-value, % HQI, WCC or simply correlation coefficients between spectra [199]. However, the increase in complexity of datasets makes interpretation of measures of similarity/ dissimilarity between ‘unknown’ and library substances challenging [199]. Seized street samples of NPS are often complex mixtures [75] and PCA has been shown to extract hidden relationships between spectral data and chemical structures for the classification of diverse chemical structures [200]. However, these studies postulated a positive and perhaps definitive correlations between both structural and spectral similarities [200, 201], which may not be possible for complex NPS mixtures.

As previously mentioned, clandestine chemists have always been one step ahead of law enforcement, there has been an interest in developing predictive models that may enable law

enforcement agents to identify an unknown NPS using the wide chemical space of NPS identified to date. However, designing a ‘representative’ library or an extended database of NPS reference spectra is not an easy task and library design should be fit for purpose. Ideally, the library should contain a limited number of NPS that adequately represent all the NPS available in the entire NPS chemical space. For example, ‘primary screening libraries’ are commonly designed in drug discovery processes for the purpose of identifying lead compounds [202, 203].

Due to the analytical, instrumental, forensic and law enforcement challenges, initial work was developed by Zloh et al. [204] to investigate ‘principles of structural similarity *in-silico*’ of 478 NPS recorded by the EMCDDA/EDND (up to January 2015). Similarity methods coupled with chemical spectral data (e.g. MS and FTIR), pre-processing and chemometrics have been shown to group or classify highly similar or dissimilar substances, respectively [200, 202, 205]. An *in-silico* ‘hierarchical cluster analysis’ was conducted to select ‘representative’ NPS from the known 478 NPS, which have maximum diversity in chemical structure across the chemical space defined by the dataset. Clustering involved dividing the dataset into chemically diverse superclusters in which members have some similarities to each other but maximum dissimilarity to members belonging to different superclusters [199]. This was achieved using a TSI of 50 % and a minimum common substructure of nine atoms *via* the ChemAxon software (ChemAxon KFT, Hungary). Fifty-three disparate NPS were selected to represent the 21 superclusters identified. The ‘hierarchical clustering method’ and spectral pre-processing efficiency have been shown to influence the classification of chemically diverse substances [200, 202, 205].

Hitherto, the analytical, forensic, instrumental and sample complexity challenges for the identification of NPS and NPS in mixtures have been presented and discussed. The synthetic analogues of traditional drugs of abuse, NPS are proliferating on a global scale. The chemical complexity of NPS formulations/ mixtures and heterogeneity regarding their composition, concentrations, number and variable Raman activity of constituents, presence of unknown constituents, overlapping NPS-exciipient bands, swamping by fluorescence signals and limitations of in-field detection tools hinder their identification and classification. In this thesis, it is hypothesised that the use of handheld Raman spectroscopy coupled with chemometrics may improve the in-field identification and classification of NPS products.

1.6. Aims of this project

To develop Raman spectroscopic methods coupled with chemometric techniques and propose the likely identity of chemical scaffolds of newly emerging NPS with a means to improve the in-field identification and classification of NPS.

Objectives:

1. To evaluate the strengths, limitations and analytical challenges of handheld Raman spectroscopy using NPS-related substances and recommend a suitable handheld Raman instrument.
2. To develop a pre-processing protocol and a variable reduction method for Raman spectral data.
3. To build a chemically and structurally diverse spectral database using benchtop and handheld Raman spectroscopy.
4. To build a PCA model for the 'representative' NPS Raman database and NPS-related substances (i.e. adulterants and cutting agents).
5. Validate and challenge the developed models using NPS-related test samples.

2. Evaluation of three handheld Raman instruments for the identification of new psychoactive substances and related compounds

2.1. Introduction

This Chapter reports on a preliminary study to evaluate the performance of three handheld Raman instruments (i.e. Truscan™ GP (Thermo-Fisher Scientific Ltd), FirstGuard™ and Progeny™ (SciMed Ltd - Rigaku)) for the identification of new psychoactive substances (NPS) in solid powder mixtures. This preliminary study was necessary to give an overview of available handheld Raman instruments, and subsequently recommend a Raman instrument for developing Raman and chemometric methods that may assist first responders in the accurate screening of NPS in the field.

Initial published work has justified the use of handheld Raman spectroscopy over handheld infrared (IR) spectroscopy for the detection of NPS in internet products [12, 128]. It has also highlighted the analytical challenges encountered with the detection of NPS in solid mixtures. Very few studies have investigated the detection of NPS mixtures using handheld and portable Raman spectroscopy [128, 138, 206]. In fact, the study by Brewster et al. evaluates the use of a portable Raman instrument in the identification of NPS in NPS liquid mixtures [206].

This Chapter evaluates three handheld Raman instruments for the identification of pure substances, model binary mixtures, random ternary mixtures and NPS internet products in their solid form. In this respect, a qualitative analysis was performed in order to evaluate: 1) the ability of the in-built algorithm (i.e. p-values (PVAL), percentage hit quality index (% HQI), wavelet correlation coefficient (WCC) and ‘Rigaku mixtures algorithm (RMA)’) to accurately screen the analyte of interest. 2) mixture sensitivity i.e. the ability of the in-built algorithm to identify low amounts of analyte with various Raman scattering activity in powder mixtures; and 3) specificity i.e. the ability of the in-built algorithm to discriminate between substances with similar chemical structure or class [207]. This work also evaluates the false positive and false negative rates using different mixtures of increasing complexity. This is because NPS are structurally similar to traditional drugs of abuse. Thus false positive results may lead to unlawful arrests, whereas false negative results may prevent the detection of a new NPS and, hence increase public health risks.

The aims of this study were to recommend a Raman instrument by evaluating the strengths and limitations of each instrument included in the study, and the effectiveness of both in-built algorithms and on-board libraries for the ‘first pass’ screening and identification of the drug constituent in solid mixtures. In addition, the challenges that may be encountered in NPS internet samples and the practical considerations for the in-field detection of NPS using handheld Raman spectroscopy were also evaluated.

2.2. Experimental

2.2.1. Materials

Powder reference standards of two NPS, i.e. 2-aminoindane hydrochloride (2-AI) and dextromethorphan hydrobromide (DXM) were obtained from Sigma Aldrich (Dorset, UK). Powder reference standards of common [208, 209] adulterants, i.e. benzocaine (BEN), caffeine (anhydrous) (CAF), lidocaine hydrochloride (LID), paracetamol (PAR), phenacetin (PHE), procaine hydrochloride (PRO) and theophylline (THEO) were also purchased from Sigma Aldrich (Dorset, UK). Diltiazem hydrochloride (DIL) was obtained from Medicines Testing lab (UK). Powder reference standards of common [208, 209] cutting agents, i.e. calcium carbonate (CaCO_3), creatine monohydrate (CRE), dextrose monohydrate (DEX), α -D-glucose anhydrous (GLU), magnesium stearate (MgS), microcrystalline cellulose (MCC), sucrose (SUC) and taurine (TAU) were purchased from Sigma Aldrich (Dorset, UK). Lactose (LAC) was obtained from Thermo-Fisher Scientific (Loughborough, UK), whereas talc (TAL) was obtained from BDH (UK). In this Chapter, reference standards were analysed in their pure form or in simulated mixtures. Justification for the selection of these substances is presented in Section 3.1 of this Chapter. In addition to the reference standards (Table 2.1 and Appendix A Table A2.1), six NPS products were purchased from the internet (Table 2.2).

Table 2. 1: Chemical structures of library signatures for NPS, adulterants and cutting agents.

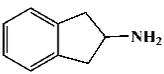
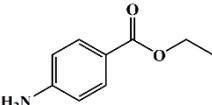
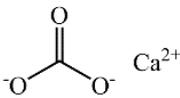
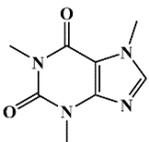
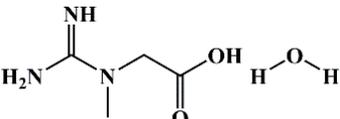
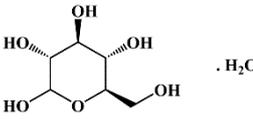
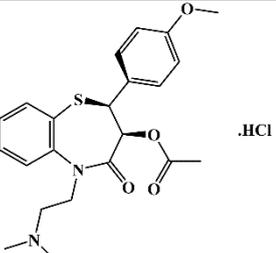
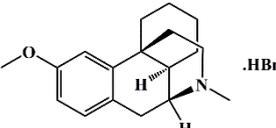
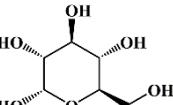
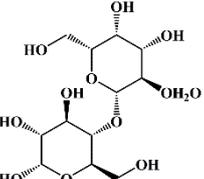
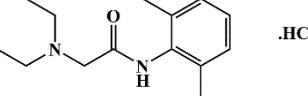
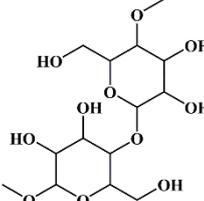
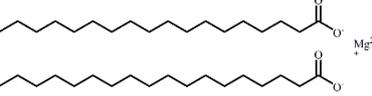
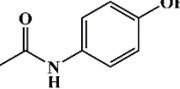
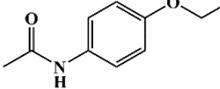
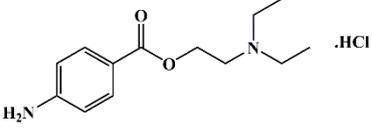
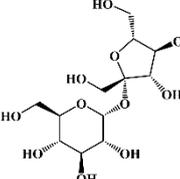
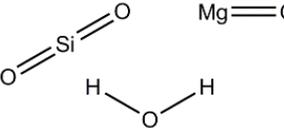
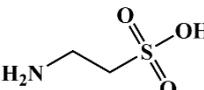
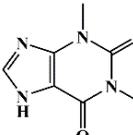
2-AI	BEN	CaCO ₃
		
CAF	CRE	DEX
		
DIL	DXM	GLU
		
LAC	LID	MCC
		
MgS	PAR	PHE
		
PRO	SUC	TAL
		
TAU	THEO	
		

Table 2. 2: NPS products purchased from the internet.

Product No. ¹	Label claim (ingredients)	Colour/ Form	Supplier (website)
P1	2-AI	White powder	www.benzo-fury.me.uk
P2	Pink Champagnes (aminoindane, caffeine, cola vera, aminoacid complex)	Brown powder in capsules	www.herbalhighlights.co.uk
P3	DXM	White powder	www.benzo_fury.me.uk
P4	DXM	White powder	www.highstore.net
P5	DXM	White powder	www.highstore.net
P6	DXM	White powder	www.highstore.net

¹P is the product number

2.2.1.1. Sample preparation

Library signatures (ca. 100 mg) were analysed through clear glass vials (Kimble Chase vial screw thread with PTFE cap, China). Due to limitations in the amounts of DIL reference standard procured, library signatures were measured through the sample (ca. 2 mg) placed on an aluminium plate (HSA14521A - Weight dish alum 43 mm, Fisherbrand). NPS internet products were obtained in the form of either powders or capsules. Powders and capsules were emptied into glass vials for analysis.

The powders were weighed using a Mettler Toledo balance capable of measurements from 0.01 mg up to 220 g. The balance was enclosed in a safety cabinet (BIGNEAT F3-XIT). To ensure homogeneity of mixtures, the powders in vials were mixed using a VORTEX-GENIE2 (Scientific industries, Inc.) for 30 sec., then tapped, shaken, then re-vortex mixed for 30 sec. immediately before each analysis.

2.2.2. Instrumentation and Methods

2.2.2.1. Handheld Raman instruments

Three handheld Raman instruments TruscanTM GP (ThermoFisher Scientific Inc., Wilmington, MA), FirstGuardTM and ProgenyTM (SciMed (Scientific & Medical Products Ltd), Rigaku, USA) were evaluated for the identification of reference standards (i.e. two NPS, eight adulterants and ten cutting agents), NPS and selected adulterants in model mixtures (Tables 2.4 and 2.5) and NPS internet mixtures (Table 2.2). Specifications of the three instruments are described in Table 2.3.

Table 2. 3: Specifications of handheld Raman instruments evaluated in this study

Raman instruments Specifications	Truscan™ GP (Thermo-Fisher Scientific)	FirstGuard™ (Rigaku)	Progeny™ (Rigaku)
Instrument image ¹			
Laser wavelength (nm)	785 +/-0.5	1064	1064
Laser output power (mW)	250 +/-25	30 – 490	30 – 490
Laser spot diameter (µm)	25	20	20
Spectral resolution (cm ⁻¹)	8 - 10.5	15 - 18	8 - 11
Spectral range (cm ⁻¹)	250 - 2875	200 - 2000	200 - 2500
Numerical aperture	0.33	0.25	0.25
Grating	Transmission volume phase (VPG)	Transmission volume phase (VPG)	Transmission volume phase (VPG)
Detector	TE Cooled CCD	TE Cooled InGaAs	TE Cooled InGaAs
Working distance (mm)	18	20	20
Calibration reference standard	Polystyrene	Benzonitrile	Benzonitrile
Weight (kg)	0.9	2.7	1.6
Dimensions (LxWxH) (cm)	20.8 x 10.7 x 4.3	12.2 x 31.1 x 31.4	29.9 x 8.1 x 7.4
Library	No standard library	Standard library (266 items) and user library	Standard library (12290 items) and user library
Operation and analysis software	Raman, WindowsXP/Vista/Win7	Micro2020, WindowsXP/Vista/Win7	RRT Progeny software version 0.001-26 140521
Data analysis	Eigen software	Grams, Unscrambler	Grams, Unscrambler
Data export formats	PDF, csv and txt	SPC and csv	PDF, .XML and .TXT
Algorithms	p-value (PVAL)	Percentage hit quality index (% HQI)	Percentage wavelet correlation coefficient (WCC), HQI and ‘Rigaku mixtures algorithm’
Operating temperature (°C)	-10 to +40	-10 to +30	-20 to +50
Battery	Rechargeable internal lithium ion battery (> 3 hours operation)	Switchable lithium ion battery (4 hours operation)	Switchable lithium ion battery (4 - 5 hours operation)
Other	no adjustable probe tip	Focus adjustable probe tip, a vial holder	Focus adjustable probe tip, a vial holder, adjustable exposure time 5 ms to 30 sec.

¹Instrument images were reproduced with courtesy of ThermoFisher Scientific Inc., Wilmington, MA and SciMed, Rigaku, USA.

For Truscan™ GP, vials were held in close contact with the nozzle piece. In contrast, for FirstGuard™, vials were placed in a vial holder and were shielded manually using a black

rubberised fabric (0.12 mm thick, ThorLabs, USA). For Progeny™, vials were placed in enclosed vial holders.

The instruments were calibrated daily, immediately prior to analysis, (i.e. using a polystyrene rod standard for Truscan™ GP, or benzonitrile reference standard for both FirstGuard™ and Progeny™) in order to verify the performance of the instruments [210]. For both FirstGuard™ and Progeny™, library signatures were measured (scan time was ca. 1 min) after optimising the focal distance using CAF (a relatively strong Raman scatterer) [211]. Using Truscan™ GP, signature scan times ranged from 1 min. e.g. for CAF and up to 80 min. e.g. for TAL (a relatively poor Raman scatterer). The scan time was automatically adjusted to optimise the S/N ratio (> 50 AU).

For the three instruments, reference standards were used as library signatures and test samples. For both Truscan™ GP and FirstGuard™, library spectra were built for reference standards (i.e. 2-AI, BEN, CaCO₃, CAF, CRE, DEX, DIL, DXM, GLU, LAC, LID, MCC, MgS, PAR, PHE, PRO, SUC, TAL, TAU and THEO) and binary mixtures (25:75, 50:50, and 75:25 for 2-AI/CAF, DXM/CAF, LID/CAF, PRO/CAF and PHE/CAF). In contrast, for Progeny™, libraries were built for pure reference standards only. This is because the instrument is equipped with an in-built ‘Rigaku mixtures algorithm (RMA)’, claimed capable of matching mixtures to pure substances in the library [212].

In order to evaluate the instruments’ in-built identification algorithms, model binary dilution mixtures containing various % m/m of mixture 1 (2-AI/CAF), 2 (DXM/CAF), 3 (LID/CAF), 4 (PRO/CAF) and 5 (PHE/CAF) were prepared and analysed against pure and discrete mixtures in the library (see above). The model mixtures were prepared by adding different amounts of pure CAF powder to each of the drugs’ reference standards: 2-AI, DXM, LID, PRO and PHE to get variable percentages in the range of 5 - 95 % m/m (Table 2.4).

Table 2. 4: Composition in mg/mg of the binary model mixtures¹

Mixture number	Mixture type	Drug/CAF (mg/mg)	Mixture number	Mixture type	Drug/CAF (mg/mg)
1D1	2-AI/CAF	4.9/95.1	4D1	PRO/CAF	5.0/95.0
1D2	2-AI/CAF	10.0/90.0	4D2	PRO/CAF	10.0/90.0
1D3	2-AI/CAF	20.0/80.0	4D3	PRO/CAF	23.8/76.2
1D4	2-AI/CAF	29.9/70.1	4D4	PRO/CAF	30.7/69.3
1D5	2-AI/CAF	40.3/59.7	4D5	PRO/CAF	40.1/59.9
1D6	2-AI/CAF	59.9/40.1	4D6	PRO/CAF	63.4/36.6
1D7	2-AI/CAF	69.9/30.1	4D7	PRO/CAF	70.0/30.0
1D8	2-AI/CAF	80.1/19.9	4D8	PRO/CAF	83.0/17.0
1D9	2-AI/CAF	90.0/10.0	4D9	PRO/CAF	90.2/9.8
1D10	2-AI/CAF	95.1/4.9	4D10	PRO/CAF	94.7/5.3
2D1	DXM/CAF	5.5/94.5	5D1	PHE/CAF	5.0/95.0
2D2	DXM/CAF	9.8/90.2	5D2	PHE/CAF	10.0/90.0
2D3	DXM/CAF	20.0/80.0	5D3	PHE/CAF	20.0/80.0
2D4	DXM/CAF	30.6/69.4	5D4	PHE/CAF	30.0/70.0
2D5	DXM/CAF	39.9/60.1	5D5	PHE/CAF	40.0/60.0
2D6	DXM/CAF	60.1/39.9	5D6	PHE/CAF	60.0/40.0
2D7	DXM/CAF	70.0/30.0	5D7	PHE/CAF	70.0/30.0
2D8	DXM/CAF	80.1/19.9	5D8	PHE/CAF	80.0/20.0
2D9	DXM/CAF	89.8/10.2	5D9	PHE/CAF	90.0/10.0
2D10	DXM/CAF	95.2/4.8	5D10	PHE/CAF	95.0/5.0
3D1	LID/CAF	5.6/94.4			
3D2	LID/CAF	10.2/89.8			
3D3	LID/CAF	21.2/78.8			
3D4	LID/CAF	29.6/70.4			
3D5	LID/CAF	39.6/60.4			
3D6	LID/CAF	59.6/40.4			
3D7	LID/CAF	70.0/30.0			
3D8	LID/CAF	80.5/19.5			
3D9	LID/CAF	89.2/10.8			
3D10	LID/CAF	95.0/5.0			

¹D = dilution number, 1D = mixture 1, 2D = mixture 2, 3D = mixture 3, 4D = mixture 4 and 5D = mixture 5

The method was validated using six random ternary mixtures prepared with common and largely available adulterants and cutting agents (i.e. BEN, CAF, GLU, LAC, LID, MCC, MgS, PAR, PHE, PRO and TAL) (Table 2.5) [208, 209]. The mixtures were made by randomly mixing different amounts of cutting agents and adulterants such that each mixture contained one, two or three adulterants weighing ≥ 25 % of the total mixture weight [136].

Table 2. 5: Composition of random ternary mixtures

Mixture number	Constituent A	Constituent B	Constituent C	Weight of constituent A (mg)	Weight of constituent B (mg)	Weight of constituent C (mg)
1	PRO	PAR	MgS	45	31	24
2	PAR	BEN	CAF	28	53	19
3	BEN	LAC	MgS	35	51	14
4	MCC	TAL	PAR	26	44	30
5	BEN	PRO	MCC	25	32	43
6	PAR	BEN	PRO	36	33	31

The method was also validated using six NPS internet products (Table 2.2). Parameters and settings were not adjustable by the operator but were automatically optimised for Truscan™ GP. Conversely, using the FirstGuard™ and Progeny™ instruments, parameters were adjustable and a few methods were developed. Using the FirstGuard™, methods A, B or C were employed, whereas methods D and E were used with Progeny™ as described in Table 2.6. Using the FirstGuard™, all samples were initially run using method A. Method B was used for samples, which displayed a poor S/N. Method C was used for samples, which were burnt from high exposure time and laser power (i.e. coloured samples and samples containing fluorescing impurities). Using Progeny™, all samples were initially run using method D. Method E was used for samples, which were burnt from high exposure time and laser power. Using both FirstGuard™ and Progeny™, the background noise was subtracted from each measurement.

Table 2. 6: Methods used for Raman analysis using handheld Raman instruments (FirstGuard™ and Progeny™)¹

Parameters	Method name				
	A	B	C	D	E
Exposure time (ms)	1000	5000	500	2000	2000/1000
Averaging of spectra	2	2	2	10	10
Laser power (mW)	300	490	200	490	300/ 200
Baseline correction	Yes	Yes	Yes	Yes	Yes
Subtract dark background	Yes	Yes	Yes	Yes	Yes

¹For Truscan™ GP, parameters were not adjustable by the operator but were automatically optimised using an in-built proprietary method.

A schematic of summary of experiments is summarised in Figure 2.1.

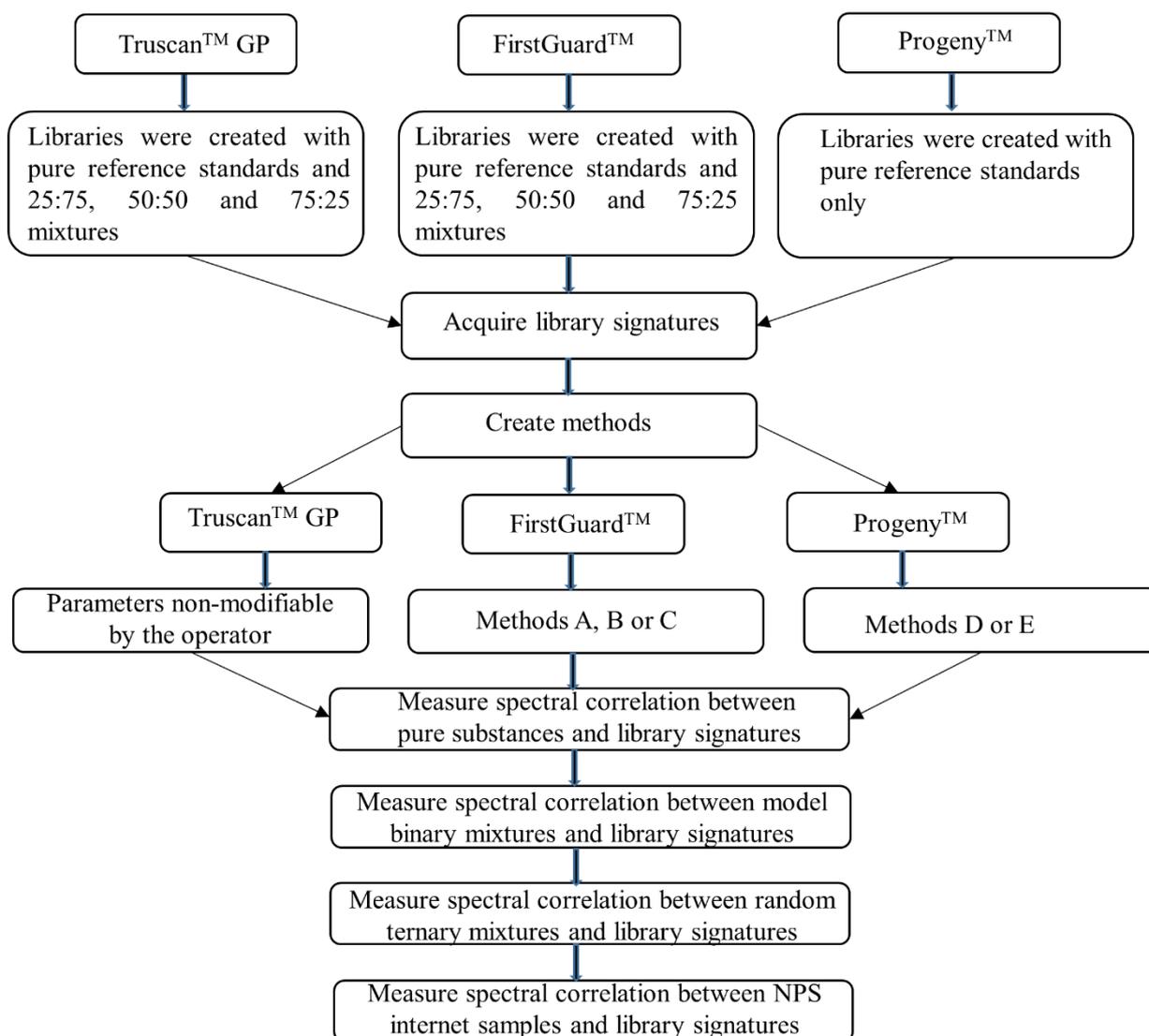


Figure 2. 1: Schematic outlining summary of experiments

2.2.2.2. Confirmatory analysis of NPS internet products using GC-MS

Gas chromatography-electron ionisation-mass spectrometry (GC-EI-MS) analysis was used to confirm the identity of compounds present in the purchased internet products (Table 2.2). The method was adapted from Assi et al. [128]. The analysis was performed using a Varian 240-MS ion trap MS equipped with a Varian 450-GC gas chromatography instrument and a Varian 8400 auto-sampler from Agilent Technologies (Berkshire, UK). Samples were analysed using electron ionisation (EI) with a scan range from m/z 40 – m/z 1000. An Agilent Technologies column (30 m x 0.25 mm x 0.25 μ m) coated with a 0.50 mm film of 50 % phenyl – 50 % methyl polysiloxane was used with helium gas as the mobile phase at a flow rate of 1 mL min^{-1} . A CP-1177 injector was held at 275 °C and was used in split mode (10:1) for most samples, but in the splitless mode when low signals were observed. An injection volume of 1

μL was used for all samples. The column temperature was programmed as follows: 50 °C for 2 min, ramped to 300 °C at 15 °C min^{-1} , held for 5 min, then cooled to 50 °C; the total run time was 28.67 min. The mass spectra obtained were compared to the purchased reference standards and the following EI spectral libraries: NIST (Version 1.0.2.2), SWGDRUG MS (Version 2.1 (2014)) and Cayman (Version 04292014).

2.3. Results and discussion

2.3.1. Substance selection

A pragmatic approach was adopted for the selection of substances in this pilot study. When this study was conducted, ‘legal’ NPS, commercially available in large quantities were selected to enable the preparation of a variety of mixtures. Selected NPS and adulterants were members of different classes such as aminoindanes, anaesthetics, analgesics and methyl-xanthines. Substances selected to build the libraries and evaluate the performance of the instruments were all common active substances, adulterants or cutting agents found in NPS products, with various Raman scattering activity [38, 213].

The NPS and adulterants selected were 2-AI, BEN, CAF, DIL, DXM, LID, PAR, PHE, PRO and THEO (Table 2.1). Prior to the UK blanket ban in 2016, both 2-AI and DXM were readily available for sale on the internet at low cost [6]. Aminoindanes were ‘legal’ analogues of amphetamine. Their popularity was linked to their pharmacological effects as potent serotonin-releasing substances [84]. 2-aminoindane is one of the most popular aminoindanes [214]. In contrast, DXM is a non-opioid cough suppressant sold over-the-counter, which may induce hallucinations at high doses [215]. It is the d-isomer of the codeine analogue levorphanol. Dextromethorphan gained its popularity from its psychostimulant effect manifested by inhibiting the re-uptake of the neurotransmitter serotonin at therapeutic doses [216]. Caffeine, on the other hand, is one of the most common adulterants found in NPS products due to its stimulant effect [213]. Benzocaine, LID and PRO are used as local anaesthetics, together with the analgesic PHE, are routinely used to cut cocaine and NPS products [79, 213, 217, 218].

In this study six products were purchased from the internet such that the claimed NPS on the label was either 2-AI or DXM, the NPSs studied in this Chapter.

2.3.2. Comparison of sample spectra to library signatures using the ‘first pass’ in-built identification algorithms

In this Section, the performance of three handheld Raman instruments (i.e. Truscan™ GP, FirstGuard™ and Progeny™) was evaluated for the identification of pure substances, model binary mixtures, random ternary mixtures and NPS internet products in their solid form.

2.3.2.1. Truscan™ GP (Thermo-Fisher Scientific)

Using the Truscan™ GP instrument, Raman responses were measured using the p-value (PVAL) algorithm, where a ‘positive match’ ($PVAL > 0.05$) meant that the sample spectrum is consistent with the library spectrum (i.e. $> 95\%$ CL). A ‘no match found’ ($PVAL \leq 0.001$) meant that the sample spectrum was not consistent with library spectra. A ‘no positive match’ ($0.001 < PVAL \leq 0.05$) meant that the sample spectrum was partially consistent with the library spectrum (see Chapter 1).

2.3.2.1.1. Analysis of pure substances

The accuracy and selectivity of the Truscan™ GP instrument were measured by testing reference standards against themselves in the library, a method known as ‘negative testing’ [219]. This included substances from a similar class and/ or structure such as [DEX, GLU, LAC, SUC]; [BEN, LID, PRO] and [THEO, CAF] (Table 2.1). ‘Negative testing’ is important to evaluate the specificity of the instrument and to verify the level of false-positives which may result [219]. In this Chapter, the top correlation was the match reported (Table 2.7).

Using the Truscan™ GP instrument, analysing reference standards against themselves in the library resulted in the successful identification of 19/20 substances including those from a similar class and/ or structure. The PVAL showed the statistical consistence between sample and library spectra with PVAL greater than 0.05 (see blue boxes in Table 2.7). The higher the PVAL, the stronger the consistence of the sample spectrum to the corresponding library reference spectrum.

Table 2. 7: Selectivity of Truscan™ GP¹. Table showing the p-values between reference standards and themselves in the library.

	NPS		Adulterants								Cutting agents									
	2-AI	DXM	BEN	CAF	DIL	LID	PAR	PHE	PRO	THEO	CaCO ₃	CRE	DEX	GLU	LAC	MCC	MgS	SUC	TAL	TAU
2-AI	0.5050																			
DXM		0.4729																		
BEN			0.6636																	
CAF				0.6606																
DIL					0.4121															
LID						0.5345														
PAR							0.6432													
PHE								0.6938												
PRO									0.5092											
THEO										0.5179										
CaCO ₃											0.5251									
CRE												0.5175								
DEX													0.6595							
GLU														0.3333						
LAC															0.4115					
MCC																0.5422				
MgS																	0.0845			
SUC																		0.4992		
TAL																			0.0091	
TAU																				0.5743

¹Single measurements were taken for testing reference standards against themselves in the library. The yellow box was a 'no positive match' response, whereas all the blue boxes represent positive correlations.

Substances identified included two NPS, eight adulterants and ten cutting agents. These substances were consistent with library spectra with PVAL ranging from 0.0845 – 0.6938. The wide range of PVAL, possibly indicates the difference between strong and weak Raman scatterers (Figure 2.2).

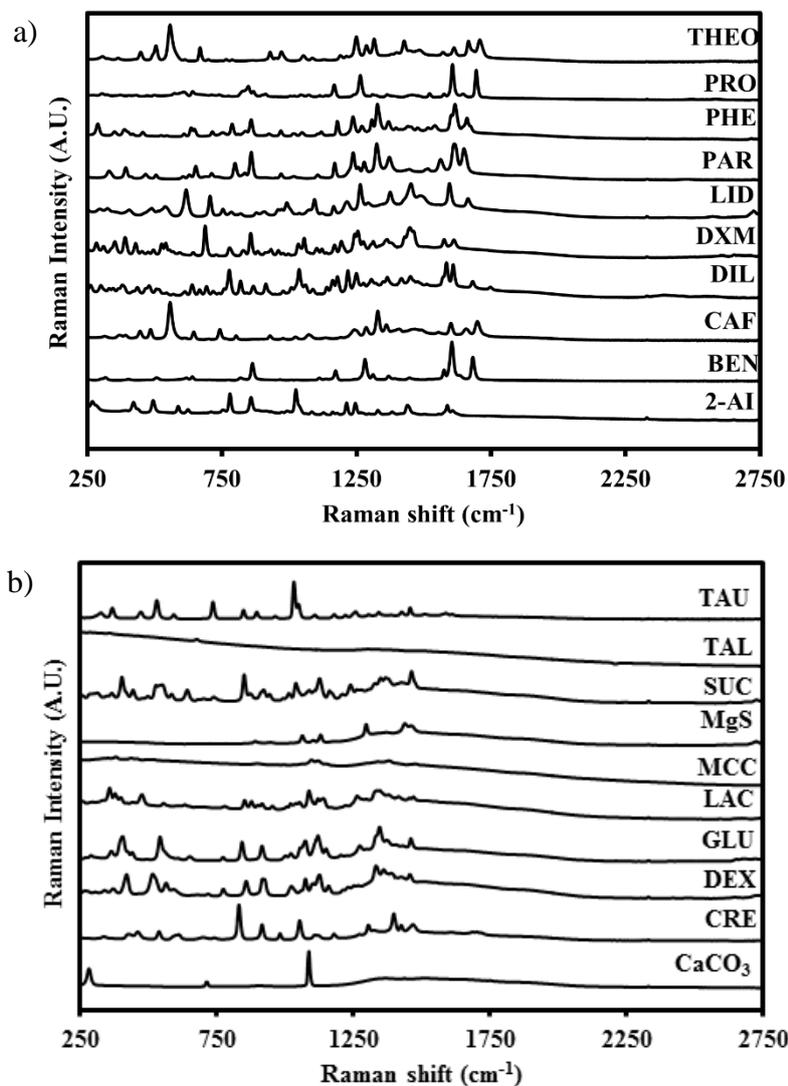


Figure 2. 2: Normalised Raman spectra of reference standards of a) NPS and adulterants; and b) cutting agents (TruscanTM GP).

The PVAL for NPS and adulterants ranged from 0.4121 – 0.6938, whereas the PVAL for cutting agents ranged from 0.0845 – 0.6595. In general, substances such as NPS and adulterants with unsaturated bonds and aromatic rings are known to induce a greater change in polarizability, and hence are generally strong Raman scatterers [125]. This explains the consistent PVAL results obtained for NPS and adulterants as opposed to most cutting agents.

Using the PVAL probabilistic algorithm, prominent Raman bands in spectra for NPS and adulterants has enabled the determination of the statistical consistence between the sample and library spectra. This was demonstrated through the analysis of TAL, where the instrument response was ‘no positive match’ with a poor PVAL of 0.0091. By examining Figure 2.2b, the Raman spectrum for TAL showed a raised baseline and no Raman bands, which resulted in negative identification of the substance. Talc is a known weak Raman scatterer which intrinsically exhibits high fluorescent background with a 785 nm laser λ_{ex} , distorting the Raman features and masking the weaker Raman signal in mixtures [220, 221].

Similarly for MgS, the Raman spectrum displayed very weak bands and the instrument responded with an accurate ‘positive match’ to MgS but with a poor PVAL of 0.0845 compared to other reference standards. An exception to this was the result obtained for MCC, where the spectrum displayed very weak Raman bands and the instrument responded with an accurate ‘positive match’ to MCC and with a relatively high PVAL of 0.5422. In such a case, a high consistence between the sample and library spectrum could be due to a poor quality library spectrum. In other words, both library (Figure 2.2b) and sample spectra for MCC have poor S/N and, therefore it is not surprising that the instrument responded with a relatively high PVAL to indicate high consistence between both spectra.

Truscan™ GP has shown to be highly selective and accurate with its ability to discriminate between substances belonging to the same class of materials and/ or with similar chemical structure. Truscan™ GP has correctly discriminated between the anaesthetics (i.e. BEN, LID and PRO) and the sugars (i.e. DEX, GLU, LAC and SUC) despite the great structural and spectral similarity between both BEN and PRO (Table 2.1 and Figure 2.2a) and DEX and GLU (Table 2.1 and Figure 2.2b) respectively. Similarly, Truscan™ GP has correctly discriminated between CAF and THEO. Theophylline is one of CAF metabolites, where a methyl group at position 7 in the five-membered ring is replaced by a hydrogen atom during metabolism [222]. Hitherto, results obtained show the potential of Truscan™ GP in identifying NPS, adulterants and cutting agents reference standards when tested against themselves in the library. However, defined Raman features for a few cutting agents were difficult to obtain due to swamping by intense fluorescent background when using a laser of λ_{ex} 785 nm. The next Section evaluates the Truscan™ GP instrument for the identification of a drug of interest in model binary mixtures.

2.3.2.1.2. Analysis of model binary dilution mixtures

Following on from our initial published work, where library spectra were created using pure reference standards and 50:50 mixtures [128], in this study, we widened the scope of the library by adding 25:75 and 75:25 mixtures. In the initial study, adding 50:50 mixtures to the library has been shown to improve the accuracy of the ‘first pass’ matching algorithm [128]. In this Chapter, widening the scope of the library was performed to evaluate the ability of the instrument to identify individual components in simple mixtures, the impact of the Raman strength of the ‘second’ component in the mixture on the identification of the NPS or drug of interest, and the range within which the NPS or drug of interest is identified (Table 2.8). Target % m/m is quoted in the discussion instead of the actual masses in mg/mg to enable comparison between the different mixtures. The target % m/m was deviated from the actual masses by a maximum amount of less than 2 %. For the exact masses, see Table 2.4.

All model binary mixtures positively correlated with the nearest mixture composition in the library except for LID/CAF. For example, in mixtures 1, 2, 4 and 5, 20/80 and 30/70 % m/m mixtures positively correlated with 25/75 % m/m library mixtures, 60/40 % m/m mixtures positively correlated with 50/50 % m/m library mixtures, and 70/30 and 80/20 % m/m mixtures positively correlated with 75/25 % m/m library mixtures. The LID powders were challenging during analysis and exhibited electrostatic interactions with electric charges on glass. Lidocaine spectra also exhibited raised baselines, which may have influenced the in-built decision tool (Figure 2.2a). Based on the Raman activity of the drug mixed with CAF, results varied at high concentrations (i.e. drug/ CAF 90/10 % m/m). For example, for 2-AI/CAF and DXM/CAF model mixtures, at 90/10 % m/m concentrations, 2-AI and DXM only were identified respectively. In contrast, PRO/CAF and PHE/CAF mixtures, at 90/10 % m/m concentrations, 75/25 % m/m of PRO/CAF and PHE/CAF were identified respectively. This could possibly be due to sample inhomogeneity, given that powders were not homogenised by grinding and homogeneity was not confirmed with wet laboratory techniques such as high performance liquid chromatography (HPLC). Samples were intentionally not ground to mimic street samples.

Table 2. 8: Comparison values of the five model mixtures using three handheld Raman instruments: Progeny™, FirstGuard™ and Truscan™ GP¹

Library			Truscan-GP ($\lambda_{\text{ex}} = 785 \text{ nm}$, PVAL ²)		FirstGuard™ ($\lambda_{\text{ex}} = 1064 \text{ nm}$, % HQI ³)		Progeny™ ($\lambda_{\text{ex}} = 1064 \text{ nm}$, % RMA ⁴)	
			Pure and mixtures (25:75; 50:50; 75:25)		Pure and mixtures (25:75; 50:50; 75:25)		Pure substances	
Code ⁵	Test set	Target % m/m ⁶	Library signature X_n/Y_m ⁷	PVAL	Library signature X_n/Y_m ⁷	% HQI	Library signature X_v/Y_w ⁸	RMA (% CC) ⁹
1D1	2AI/CAF	5/95	CAF	0.3266	LID25/CAF75	98.0	CAF	0.99
1D2	2AI/CAF	10/90	CAF	0.3214	LID25/CAF75	96.9	CAF	0.99
1D3	2AI/CAF	20/80	2AI25/CAF75	0.6316	2AI25/CAF75	95.0	2-AI36/CAF64	1.00
1D4	2AI/CAF	30/70	2AI25/CAF75	0.5697	2AI50/CAF50	96.6	2-AI62/CAF38	0.99
1D5	2AI/CAF	40/60	2AI50/CAF50	0.7290	2AI50/CAF50	92.9	2-AI37/CAF63	1.00
1D6	2AI/CAF	60/40	2AI50/CAF50	0.2280	2AI50/CAF50	93.6	2-AI60/CAF40	1.00
1D7	2AI/CAF	70/30	2AI75/CAF25	0.7360	2AI75/CAF25	93.4	2-AI46/CAF54	1.00
1D8	2AI/CAF	80/20	2AI75/CAF25	0.3069	2AI75/CAF25	92.5	2-AI	0.99
1D9	2AI/CAF	90/10	2AI	0.2710	2AI75/CAF25	90.7	2-AI	1.00
1D10	2AI/CAF	95/5	2AI	0.2409	2AI75/CAF25	90.8	2-AI	1.00
2D1	DXM/CAF	5/95	CAF	0.4399	CAF	96.6	CAF	1.00
2D2	DXM/CAF	10/90	CAF	0.5378	LID25/CAF75	97.6	CAF	1.00
2D3	DXM/CAF	20/80	DXM25/CAF75	0.5512	DXM50/CAF50	66.8	CAF	1.00
2D4	DXM/CAF	30/70	DXM25/CAF75	0.1124	DXM50/CAF50	95.3	DXM45/CAF55	0.99
2D5	DXM/CAF	40/60	DXM50/CAF50	0.3038	DXM50/CAF50	95.7	DXM73/CAF27	1.00
2D6	DXM/CAF	60/40	DXM75/CAF25	0.1286	DXM50/CAF50	96.9	DXM33/CAF67	1.00
2D7	DXM/CAF	70/30	DXM75/CAF25	0.6594	DXM	82.4	DXM56/CAF44	1.00
2D8	DXM/CAF	80/20	DXM75/CAF25	0.1068	DXM	88.1	DXM26/CAF74	1.00
2D9	DXM/CAF	90/10	DXM	0.0995	DXM	81.9	DXM	0.99
2D10	DXM/CAF	95/5	DXM	0.4461	DXM	94.4	DXM	0.99
3D1	LID/CAF	5/95	DXM25/CAF75	0.1857	CAF	95.5	CAF	0.99
3D2	LID/CAF	10/90	CAF	0.125	CAF	92.1	CAF	1.00

3D3	LID/CAF	20/80	CAF	0.0789	CAF	90.4	LID15/CAF85	0.96
3D4	LID/CAF	30/70	No positive match to DXM25/CAF75	0.0003	LID50/CAF50	92.6	LID19/CAF81	0.94
3D5	LID/CAF	40/60	No match found	-	LID50/CAF50	84.0	LID38/CAF62	0.87
3D6	LID/CAF	60/40	No match found	-	LID50/CAF50	75.2	LID25/CAF75	0.80
3D7	LID/CAF	70/30	No match found	-	Not match found	-	LID65/CAF35	0.80
3D8	LID/CAF	80/20	No match found	-	Not match found	-	LID81/CAF19	0.75
3D9	LID/CAF	90/10	No match found	-	Not match found	-	LID86/CAF14	0.76
3D10	LID/CAF	95/5	LID	0.3860	LID	90.5	LID	0.99
4D1	PRO/CAF	5/95	CAF	0.5130	LID25/CAF75	97.1	CAF	1.00
4D2	PRO/CAF	10/90	PRO25/CAF75	0.0801	LID50/CAF50	92.6	PRO58/CAF42	1.00
4D3	PRO/CAF	20/80	PRO25/CAF75	0.480	PRO25/CAF75	96.7	PRO43/CAF57	1.00
4D4	PRO/CAF	30/70	PRO50/CAF50	0.5013	PRO25/CAF75	96.1	PRO49/CAF51	1.00
4D5	PRO/CAF	40/60	PRO50/CAF50	0.4977	PRO25/CAF75	94.1	PRO68/CAF32	1.00
4D6	PRO/CAF	60/40	PRO50/CAF50	0.2852	PRO75/CAF25	99.2	PRO21/CAF79	1.00
4D7	PRO/CAF	70/30	PRO75/CAF25	0.4336	PRO75/CAF25	99.0	PRO63/CAF37	1.00
4D8	PRO/CAF	80/20	PRO75/CAF25	0.2289	PRO75/CAF25	98.2	PRO	1.00
4D9	PRO/CAF	90/10	PRO75/CAF25	0.2213	PRO75/CAF25	99.0	PRO	1.00
4D10	PRO/CAF	95/5	PRO75/CAF25	0.0801	PRO75/CAF25	98.5	PRO	1.00
5D1	PHE/CAF	5/95	CAF	0.3403	CAF	72.0	CAF	1.00
5D2	PHE/CAF	10/90	PHE25/CAF75	0.2329	PHE	97.4	PHE20/CAF80	1.00
5D3	PHE/CAF	20/80	PHE25/CAF75	0.499	PHE25/CAF75	88.9	PHE26/CAF74	1.00
5D4	PHE/CAF	30/70	PHE25/CAF75	0.5917	PHE25/CAF75	90.7	PHE47/CAF53	1.00
5D5	PHE/CAF	40/60	PHE50/CAF50	0.5393	PHE25/CAF75	88.3	PHE55/CAF45	1.00
5D6	PHE/CAF	60/40	PHE75/CAF25	0.7219	PHE75/CAF25	91.4	PHE42/CAF56	1.00
5D7	PHE/CAF	70/30	PHE75/CAF25	0.6671	PHE	95.3	PHE53/CAF47	1.00
5D8	PHE/CAF	80/20	PHE75/CAF25	0.5368	PHE	96.8	PHE26/CAF74	1.00
5D9	PHE/CAF	90/10	PHE75/CAF25	0.2434	PHE	97.4	PHE	1.00
5D10	PHE/CAF	95/5	PHE75/CAF25	0.1426	PHE	95.9	PHE	1.00

¹Single measurements were taken for testing pure reference standards against themselves in the library.

²PVAL: p-value.

³HQI: Hit Quality Index.

⁴RMA: 'Rigaku mixtures algorithm'.

⁵D: dilution number.

⁶Target % m/m (mass/mass), refer to Table 2.4.

⁷X_n/Y_m: X and Y constitute both constituents in the mixtures. 'n' is the percentage mass of constituent X. 'm' is the percentage mass of constituent Y.

⁸X_v/Y_w: X and Y constitute both constituents in the mixtures. 'v' and 'w' do not constitute the % m/m but the spectral contribution of both constituents as calculated by the in-built algorithm 'RMA'.

⁹RMA results were reported as percentage correlation coefficient.

Conversely, mixtures containing LID/CAF showed inconsistent results. LID was not identified at amounts < 95 %. LID/CAF mixtures positively correlated to CAF at CAF concentration > 80 % m/m, possibly because CAF is a relatively stronger Raman scatterer than LID. In addition, inconsistent results were obtained from LID/CAF mixtures, possibly because of significant differences in the observed physical properties of both powders, which may have resulted in non-homogeneity. Particle shape and powder surface charge lead to observed powder segregation and electrostatic interactions with glass. No correlation was found for LID/CAF mixtures at concentrations ranging from 40/60 to 90/10 % m/m (Table 2.8).

In our initial published work [128], binary mixtures of 2-AI, DXM and LID with CAF were tested against a spectral library built with pure reference standards and 50:50 binary mixtures using Truscan™ RM ($\lambda_{\text{ex}} = 785 \text{ nm}$), another handheld Raman instrument from Thermo-Fisher Scientific. In this study on mixtures, where the PVAL obtained was less than 0.05, mixture components were identified in discovery mode, where the sample spectrum was compared to all library spectra. In this initial work, 2-AI, DXM and LID were identified at amount ranges of 20.0 – 74.6, 20.0 – 80.1 and 20.0 – 89.7 % m/m respectively [128]. In this Chapter, testing model binary mixtures against spectral libraries built with pure reference standards and 25:75, 50:50, 75:25 mixtures has shown a significant improvement in positive correlation results. The addition of a discrete number of equally spaced concentrations of binary mixtures to the library has improved the accuracy of correlations. In this respect, NPS constituents (i.e. 2-AI and DXM) in the model mixtures were identified in the range 20 – 95 % m/m. The adulterant constituents (LID, PRO and PHE) in model binary mixtures were identified in the range ≥ 95 % m/m, 10 – 95 % m/m and 10 – 95 % m/m respectively (Table 2.8).

Based on the in-built algorithm, Raman activity of mixture components, mixture homogeneity, the amount of each constituent, the quality of the library and sample spectra and powder properties, Truscan™ GP identified either one or two components in model binary mixtures.

As NPS ‘street-like’ samples are made of complex mixtures [76], it is important to evaluate the feasibility of identifying the NPS in the mixture as well as the ‘other’ component(s) in the mixture and to develop an understanding on how the ‘other’ component(s) may hamper the identification of the target drug [221]. For instance, NPS or adulterant identification has been influenced with the Raman activity of the drug mixed with CAF relative to CAF and its composition in the mixture. In this respect, in both PRO/CAF and PHE/CAF, false negative results were obtained for the drug of interest (i.e. PRO and PHE respectively) with CAF only being identified at CAF concentration > 95 % m/m, whereas both components were identified in the range 20 – 80 % m/m. In LID/CAF mixtures, false negative results of both components in the mixtures occurred at LID concentrations 5 – 90 % m/m (Table 2.9).

Table 2. 9: Identification of one or both constituents in model binary mixtures using Truscan™ GP

Mixture type	Drug/CAF mixture composition (% m/m)	Drug only identified ¹ (%)	Drug/CAF mixture composition (% m/m)	Drug and CAF identified ² (%)
2-AI/CAF	≥90	20	20-80	60
DXM/CAF	≥90	20	20-80	60
LID/CAF	≥95	10	0	0
PRO/CAF	- ³	0	10-95	90
PHE/CAF	- ³	0	10-95	90

¹is the number of mixtures (expressed as a %), where the drug mixed with CAF was solely identified. ²is the number of mixtures (expressed as a %), where both constituents in binary mixtures were identified (i.e. drug and CAF). ³means none of the constituents were identified in all mixtures.

Based on the in-built algorithm, the Truscan™ GP instrument has not produced any false positive results except for LID/CAF mixtures, where LID/CAF mixtures correlated to DXM25/CAF75 mixtures with a ‘positive match’ and a ‘no positive match’ at LID concentration of 5 and 30 % m/m respectively. By examining the Raman spectra (Figure 2.3), it was observed that the library spectrum for DXM25/CAF75 mainly showed prominent peaks for CAF at 555, 1329, 1600 and 1699 cm⁻¹, possibly attributed to C-N-CH₃, C-N, C=C and C=N vibrations respectively. The sample Raman spectrum (3D4) showed an intense fluorescent background possibly attributed to glass in the region 1500 - 1200 cm⁻¹, and the only Raman bands observed over the region 740 – 440 cm⁻¹ are thought to be attributed to CAF. This explains why these mixtures correlated with false positive matches to DXM25/CAF75 mixtures.

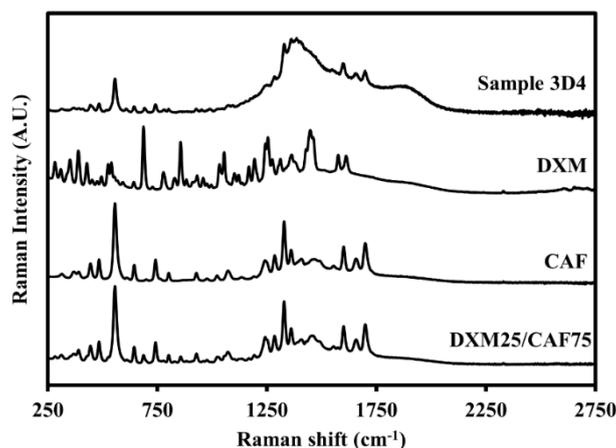


Figure 2. 3: Normalised Raman spectra for library spectra of DXM25/CAF75, CAF and DXM and the spectrum of the sample 3D4 composed of LID30/CAF70 (Truscan™ GP).

In summary, the Truscan™ GP instrument ($\lambda_{\text{ex}} = 785 \text{ nm}$) has identified 95 % of pure substances when they were match-tested against themselves in the library, but fluorescence has clearly distorted the Raman features in the cutting agents spectra. Poor correlations were observed for poor Raman scatterers e.g. TAL and for mixtures, which exhibited fluorescent background e.g. LID/CAF mixtures. In model binary mixtures, the NPS constituents (i.e. 2-AI and DXM) were identified in the range 20 – 95 % m/m, whereas the adulterant constituents (LID, PRO and PHE) were identified in the range ≥ 95 % m/m, 10 – 95 % m/m and 10 – 95 % m/m respectively.

2.3.2.2. FirstGuard™ (Rigaku)

Using the FirstGuard™ instrument, Raman responses were measured using the % HQI algorithm. An HQI of 100 % means that the correlation between the ‘unknown’ spectrum and the reference spectrum is absolute (i.e. 1). An HQI of 0 % means that the ‘unknown’ spectrum and the reference spectrum are orthogonal to each other (see Chapter 1). Positive correlations were assigned to a % HQI \geq a threshold of 85 %, as defaulted by the manufacturer.

2.3.2.2.1. Analysis of pure substances

In order to enable a direct comparison between the results obtained with Truscan™ GP and the results obtained with FirstGuard™, all conditions must be the same. However, this is not the case. For example, peak relative intensities may change slightly by using a different laser λ_{ex} (e.g. 1064 instead of 785 nm) [221]. In addition, the absolute intensity may decrease by increasing the laser λ_{ex} . This is because Raman scattering is proportional to the fourth power of the frequency of the laser source [120]. Moreover, the quality of the spectral library may change, not only because of the change of the laser λ_{ex} but also because the powder area being

irradiated is different, which may have a great impact on the correlation results for mixtures using the in-built % HQI algorithm. This was observed with the quality of the reference spectra for NPS, adulterants and cutting agents (Figure 2.4) as compared to Figure 2.2.

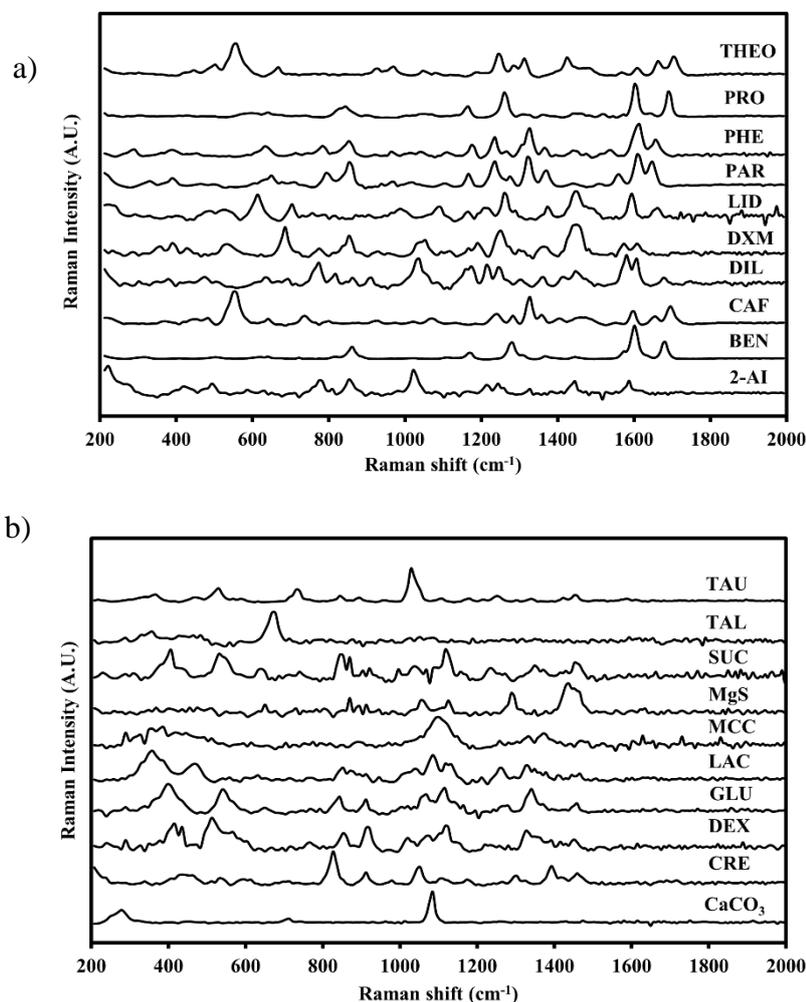


Figure 2. 4: Normalised Raman spectra of reference standards of a) NPS and adulterants; and b) cutting agents (FirstGuard™).

As with the Truscan™ GP instrument, a similar analysis was performed with FirstGuard™ using the % HQI algorithm to evaluate the accuracy and selectivity of the instrument. The % HQI values describe the correlation between sample and library spectra. The higher the % HQI, the stronger the correlation with library reference spectra. The HQI algorithm is highly dependent on the S/N in the sample spectrum [212]. In this Section, the top match was the match reported. Using the in-built algorithm, 19/20 samples (i.e. two NPS, eight adulterants and nine cutting agents) were identified (Table 2.10).

Table 2. 10: Selectivity of FirstGuard^{TM1}

	NPS		Adulte rants								Cutting agents										
	2-AI	DXM	BEN	CAF	DIL	LID	PAR	PHE	PRO	THEO	CaCO ₃	CRE	DEX	GLU	LAC	MCC	MgS	SUC	TAL	TAU	
2-AI	71.7 ± 1.2																				
DXM		92.3 ± 1.0																			
BEN			99.2 ± 0.2																		
CAF				98.8 ± 0.2																	
DIL					96.1 ± 3.0																
LID						92.1 ± 0.9															
PAR							96.8 ± 0.9														
PHE								97.1 ± 0.7													
PRO									98.6 ± 0.1												
THEO										97.7 ± 0.3											
CaCO ₃											95.6 ± 0.9										
CRE												95.4 ± 1.6									
DEX													96.3 ± 1.0								
GLU														92.5 ± 2.2							
LAC															94.1 ± 2.3						
MCC																87.3 ± 4.2					
MgS																	*				
SUC																		96.0 ± 0.38			
TAL																			97.0 ± 0.5		
TAU																					98.4 ± 0.7

¹Triplicate measurements were taken for testing pure reference standards against themselves in the library; *MgS correlated to bees wax with a % HQI of 90 ± 4; % HQI ± SD is presented in blue boxes.

The % HQI is presented \pm standard deviation (SD) between triplicate measurements. Results showed that 2-AI had the lowest % HQI (71.7 ± 1.2), whereas BEN had the highest % HQI (99.2 ± 0.2). Both MCC and DIL had the largest SD between three replicate measurements. The library spectrum for MCC showed poor S/N with poorly defined Raman features, possibly because it is a poor Raman scatterer (Figure 2.4b). Conversely, the library spectrum for DIL showed adequate S/N (Figure 2.4a). However, Raman analysis was challenging because of the small sample size, the challenge to pack the sample, to present it to the instrument, to optimise the focal distance and place the sample at the focal point. In addition, the weight and geometry of the instrument made it very challenging to collect high quality spectra under the previously described conditions (Section 2.2.1.1.).

A false positive correlation to bees wax was obtained for MgS. This was not surprising as bees wax has a similar chemical structure to MgS (see Table 2.1 for MgS chemical structure). Figure 2.5 shows the spectrum obtained and the chemical structure for bees wax.

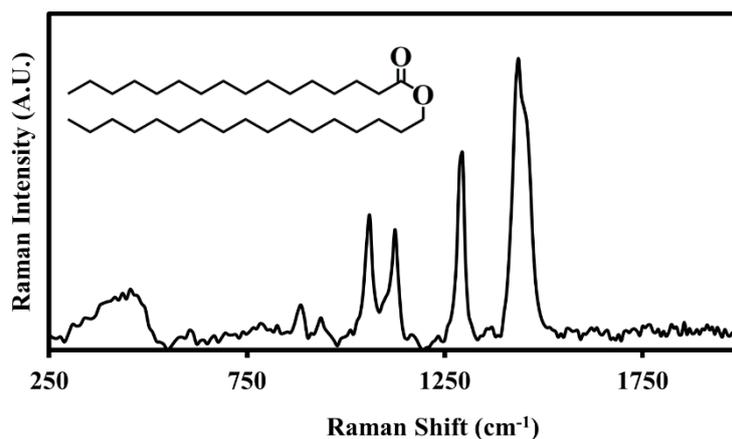


Figure 2. 5: The Raman spectrum for the MgS reference sample match-tested against the library reference spectra. The chemical structure for bees wax is superimposed on the spectrum (FirstGuardTM).

2.3.2.2.2. Analysis of model binary dilution mixtures

Similar to Section 2.3.2.1.2., the model binary mixtures were match-tested against pure and discrete mixtures in the library (Table 2.8).

Model binary mixtures were match-tested against all libraries (pure reference standards and 25:75, 50:50, 75:25 mixtures). The NPS constituents (2-AI and DXM) in the model mixtures were identified in the range 20 – 95 % m/m using the in-built algorithm in the FirstGuardTM instrument (i.e. % HQI). The adulterants constituents (LID, PRO and PHE) in the model binary

mixtures were identified in the range 30 – 60 % m/m and ≥ 95 % m/m for LID, 20 – 95 % m/m for PRO, and 10 – 95 % m/m for PHE (Table 2.8).

Based on the ‘first pass’ in-built algorithm, the FirstGuard™ instrument identified either one or two components in model binary mixtures. In this respect, the ‘drug’ in ‘Drug/CAF’ mixtures was identified in DXM/CAF, LID/CAF and PHE/CAF at concentrations of 70 - 90, ≥ 95 , and ≥ 70 % m/m respectively. Both components were identified at different concentrations for all mixtures (Table 2.11).

Table 2. 11: Identification of one or both components in model binary mixtures using the FirstGuard™

Mixture type	Drug/CAF mixture composition (% m/m)	Drug only identified ¹ (%)	Drug/CAF mixture composition (% m/m)	Drug and CAF ² (%)
2-AI/CAF	- ³	0	20-95	8
DXM/CAF	70-95	4	20-60	4
LID/CAF	≥ 95	4	30-60	3
PRO/CAF	- ³	0	20-95	8
PHE/CAF	≥ 70	5	20-60	4

¹is the number of mixtures (expressed as a %), where the drug mixed with CAF was solely identified. ²is the number of mixtures (expressed as a %), where both constituents were identified (i.e. drug and CAF). ³means none of the constituents were identified in all mixtures.

Based on the in-built algorithm, the FirstGuard™ instrument generated a number of false positive results in all mixtures except for PHE/CAF mixtures. False positive results to LID/CAF mixtures in the library mainly occurred at drug concentrations of 10 % m/m or less in mixtures with CAF except for LID/CAF mixtures where false positive results occurred at drug concentrations of 30 – 60 % m/m. By examining the Raman spectra for these samples and comparing them to the library spectra, it was observed that CAF has again largely contributed to the composition of the library spectra and, hence resulted in false positives in mixtures with high CAF composition (Figure 2.6). Figure 2.6 shows selected examples, where samples falsely correlated with LID/CAF mixtures from the library. This demonstrated the complexity of building libraries with mixtures and the feasibility of identifying mixtures’ components. This also demonstrated the impact of different Raman activity, powder characteristics (i.e. powder surface charge and particle shape leading to observed powder segregation and electrostatic interactions with glass) and amounts of components in a mixture as well as mixture homogeneity.

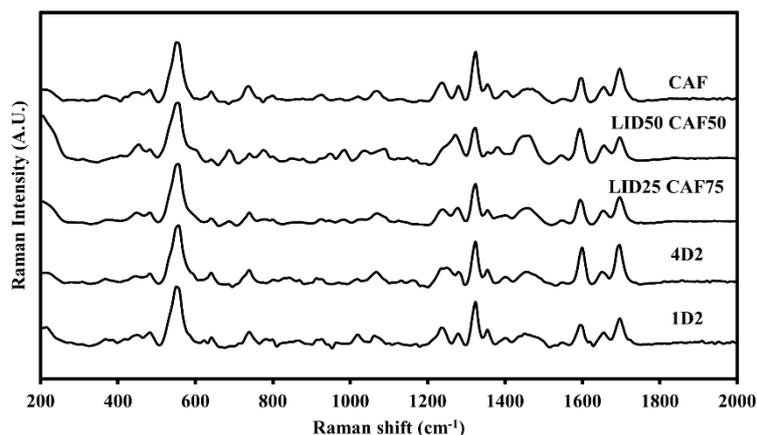


Figure 2. 6: Normalised Raman spectra for the samples 1D2 and 4D2 and normalised Raman reference spectra for LID25 CAF75, LID50 CAF50 and CAF (FirstGuard™).

In summary, the FirstGuard™ has identified 95 % of pure substances when they were match-tested against themselves in the library. However, the examination of the spectra clearly shows reduced fluorescence signals as compared to Truscan™ GP (laser $\lambda_{\text{ex}} = 785 \text{ nm}$), but it also shows the reduced S/N as a result of using a longer laser λ_{ex} of 1064 nm. In model binary mixtures, the NPS constituents (i.e. 2-AI and DXM) were identified in the range 20 – 95 % m/m, whereas the adulterant constituents (LID, PRO and PHE) were identified in the range 30 – 60 % m/m and $\geq 95 \%$ m/m for LID, 20 – 95 % m/m for PRO, and 10 – 95 % m/m for PHE respectively. The FirstGuard™ has generated a greater number of false positive results in all mixtures except for PHE/CAF mixtures, which is thought to be because of the individual contribution of drug constituents in library spectra.

2.3.2.3. Progeny™ (Rigaku)

The in-built algorithm wavelet correlation coefficient (WCC) was employed to evaluate the accuracy and selectivity of the instrument by match-testing pure solid powders against pure substances in the library including similar class of materials and/ or structure such as [DEX, GLU, LAC, SUC]; [BEN, LID, PRO] and [THEO, CAF]. WCC is claimed to be the best search algorithm for noisy spectra compared to both PVAL and HQI [212]. The closer the WCC value to 1.00, the stronger the correlation with the library reference spectra. In this Section, only the top match was reported. The WCC values \pm SD between triplicate measurements are shown in the blue boxes (Table 2.12). Results showed that 100 % of NPS, adulterants and cutting agents' reference standards were differentiated from each other.

Table 2. 12: Selectivity of Progeny™

	NPS		Adulterants								Cutting agents									
	2-AI	DXM	BEN	CAF	DIL	LID	PAR	PHE	PRO	THEO	CaCO ₃	CRE	DEX	GLU	LAC	MCC	MGS	SUC	TAL	TAL
2-AI	1±0																			
DXM		1±0																		
BEN			1±0																	
CAF				1±0																
DIL					1±0															
LID						0.99 ± 1.36E-16														
PAR							1±0													
PHE								1±0												
PRO									1±0											
THEO										1±0										
CaCO ₃											1±0									
CRE												1±0								
DEX													0.99 ± 1.36E-16							
GLU														1±0						
LAC															0.99 ± 1.36E-16					
MCC																1±0				
MGS																	0.99 ± 1.36E-16			
SUC																		0.97 ± 1.36E-16		
TAL																			0.97 ± 1.36E-16	
TAU																				1±0

¹Triplicate measurements were taken for testing pure reference standards against themselves in the library.

Based on the in-built algorithm, ProgenyTM identified 20 pure substances (i.e. 2 NPS, 8 adulterants and 10 cutting agents) with a WCC values ranging from 0.97 for SUC and TAL (relatively weak Raman scatterers) to 1.00 for all the remaining substances tested. The WCC values between triplicate measurements were highly reproducible for the analysis of the reference standards and resulted in a maximum percentage relative standard deviation (RSD) of 1.4×10^{-16} . The Raman spectra of the reference standards were also examined (Figure 2.7).

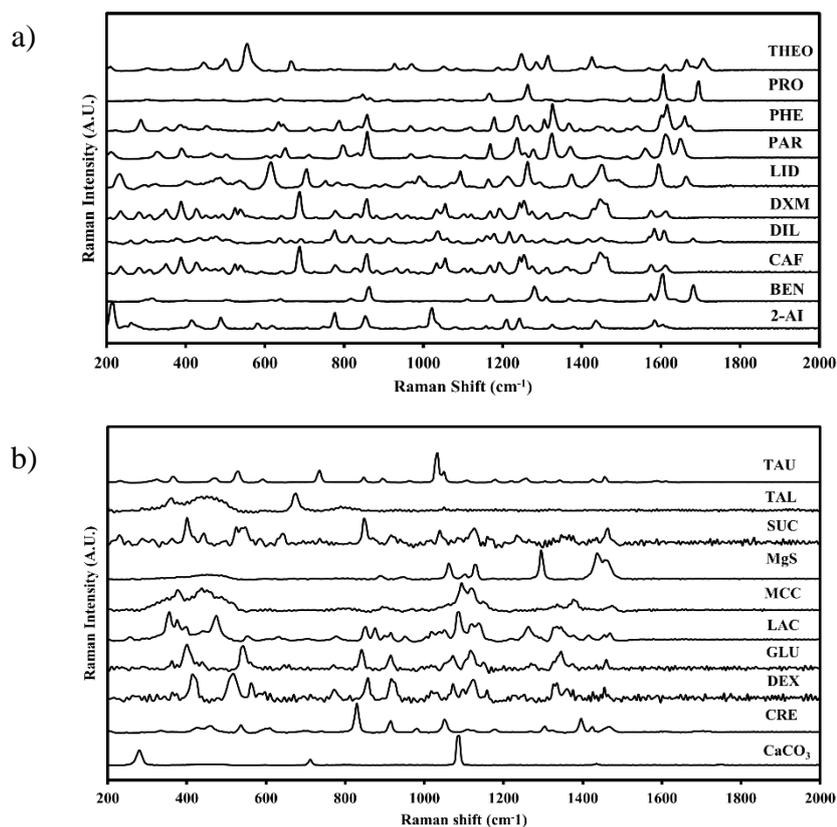


Figure 2. 7: Normalised Raman spectra of reference standards of a) NPS and adulterants; and b) cutting agents (ProgenyTM).

By comparing Figure 2.7b to Figure 2.4b, visual inspection of the spectra showed the ProgenyTM instrument, Raman features were better defined and this is possibly, owing to the software used and proprietary pre-processing performed by both instrument.

In contrast to match-testing pure reference standards against the library, the in-built RMA is claimed capable of correlating the spectra of drugs mixtures to reference spectra in the library by comparing each mixture spectrum to all library signatures [212]. Therefore, using ProgenyTM, libraries were created with pure reference standards only and RMA. Results were reported as percentage correlation coefficient (% CC). Results of match-testing binary mixtures against the library are reported as X_v/Y_w , where X and Y are suggested constituents in the test

mixtures and ‘v’ and ‘w’ represent the suggested composition of each constituent in the test mixture. ‘v’ and ‘w’ do not represent the % m/m of both constituents. ‘v’ and ‘w’ did not follow a trend when the percentage mass of drug mixed with CAF changed (i.e. increased or decreased) in the mixtures.

Results showed that using RMA, mixtures 1 (2-AI/CAF), 2 (DXM/CAF), 3 (LID/CAF), 4 (PRO/CAF) and 5 (PHE/CAF) were consistent with mixtures of both components in the library in the range of 20 - 70 % m/m for 2-AI, 30 – 80 % m/m for DXM, 20 – 90 % m/m for LID, 10 - 70 % m/m for PRO, and 10 – 80 % m/m for PHE respectively. Above and below these ranges, only the component in higher amount in the mixture was detected as anticipated (Table 2.8).

Using the ‘first pass’ in-built algorithm, Progeny™ identified either one or two components in model binary mixtures. The ‘drug’ in ‘Drug/CAF’ mixtures and both components in the mixtures were identified at various concentrations (Table 2.13).

Table 2. 13: Identification of one or both components in model binary mixtures using Progeny™

Mixture type	Drug/CAF mixture composition (% m/m)	Drug only identified ¹ (%)	Drug/CAF mixture composition (% m/m)	Drug and CAF identified ² (%)
2-AI/CAF	≥80	3	20-70	5
DXM/CAF	≥90	2	30-80	5
LID/CAF	≥95	1	30-90	7
PRO/CAF	>80	3	20-90	6
PHE/CAF	>70	2	10-80	7

¹is the number of mixtures (expressed as a %), where the drug mixed with CAF was solely identified.

²is the number of mixtures (expressed as a %), where both constituents were identified (i.e. drug and CAF).

In summary, Progeny™ has identified 100 % of pure substances when they were match-tested against themselves in the library. Visual inspection of Raman spectra showed better definition of Raman features even for poor Raman scatterers as compared to the FirstGuard™. In model binary mixtures, the NPS constituents (i.e. 2-AI and DXM) were identified in the range 20 – 70 % m/m for 2-AI, 30 – 80 % m/m for DXM, 20 – 90 % m/m for LID, 10 - 70 % m/m for PRO, and 10 – 80 % m/m for PHE respectively. Based on the in-built algorithm, the Progeny™ instrument has not generated any false positive results. In the next Section 2.3.2.4. and 2.3.2.5., the three instruments are challenged with random ternary mixtures and NPS internet samples.

2.3.2.4. Testing the complexity of random ternary mixtures

In this Section, the aim of this analysis was to evaluate the feasibility of identifying one adulterant as a single reference or in a binary mixture using the instruments Truscan™ GP and FirstGuard™. For the Progeny™ instrument, the aim of the analysis was to evaluate the performance of the in-built algorithm ‘RMA’ in identifying all the adulterants in the mixtures. The mixtures were match-tested in triplicate against the libraries developed using the three handheld Raman instruments (Table 2.14).

Using the in-built algorithm in Truscan™ GP (i.e. PVAL), triplicate measurements of random ternary mixtures yielded reproducible results in all six mixtures despite mixing the sample between measurements and irradiating the samples at different powder areas (Table 2.14). The analysis of 3/6 mixtures (i.e. mixtures 1, 5 and 6) resulted in false negatives, whereas the analysis of mixtures 2 – 4 resulted in the reproducible identification of a single adulterant as expected. In mixtures 2 – 4, the adulterant identified was the strongest Raman scatterer, which was in equal or in a greater amount than other adulterant(s) in the mixture. When the adulterant identified was mixed with cutting agents only, it was identified even if its amount was less than that of the cutting agents. In mixtures 1, 5 and 6, no adulterant was identified, possibly because the constituents had relatively similar Raman strength and amounts to each other in the mixture and this is one of the challenges faced in the analysis of mixtures.

For the FirstGuard™, one adulterant was identified in at least two out of the three measurements in all six mixtures. However, in mixtures 2 and 3, false positive correlations to p-hydroxycocaine and a false negative in mixture 4 occurred in one measurement. The adulterant identified was as a single reference or in a binary mixture. The amount of the adulterant identified in the mixtures did not reflect its true mass. The presence of mixtures in the library may lead to false positive results i.e. identifying a component that is not truly in the mixture. For example, for mixture 5, BEN50/CAF50 was identified in all three measurements, which means one true positive component (i.e. BEN), one false positive component (i.e. CAF) and one false negative component (i.e. PRO) (Table 2.14).

Table 2. 14: Identification of ternary mixtures using the in-built algorithm in Truscan™ GP, FirstGuard™ and Progeny™

Truscan™ GP								
Mixture number	Constituents in the mixtures	% mg/mg/mg	1st Measurement		2nd Measurement		3rd Measurement	
			Library signature	PVAL	Library signature	PVAL	Library signature	PVAL
1	PRO/PAR/MgS	45/31/23	No match found		No match found		No match found	
2	PAR/BEN/CAF	28/53/18	BEN	0.1084	BEN	0.1067	BEN	0.1077
3	BEN/LAC/MgS	35/51/12	BEN	0.355	BEN	0.3362	BEN	0.3358
4	MCC/TAL/PAR	26/44/30	PAR	0.0067	PAR	0.0058	PAR	0.0061
5	BEN/PRO/MCC	25/32/43	No positive match to BEN	0.0003	No positive match to BEN	0.0003	No positive match to BEN	0.0003
6	PAR/BEN/PRO	36/33/31	No match found		No match found		No match found	
FirstGuard™								
			Library signature	% HQI	Library signature	% HQI	Library signature	% HQI
1	PRO/PAR/MgS	45/31/23	PRO	84.1	PRO75/CAF25	83.7	PRO75/CAF25	82.3
2	PAR/BEN/CAF	28/53/18	BEN	98.5	BEN50CAF50	93.3	p-hydroxycocaine	59.7
3	BEN/LAC/MgS	35/51/12	BEN50CAF50	97	BEN	94.3	p-hydroxycocaine	62.6
4	MCC/TAL/PAR	26/44/30	Unknown		PAR	92.4	PAR	70.7
5	BEN/PRO/MCC	25/32/43	BEN50CAF50	95.9	BEN50CAF50	92.6	BEN50CAF50	96.3
6	PAR/BEN/PRO	36/33/31	BEN	91.2	BEN50CAF50	90.3	PRO	68.2
Progeny™								
			Library signature	% CC	Library signature	% CC	Library signature	% CC
1	PRO/PAR/MgS	45/31/23	PRO	0.99	PRO	0.99	PRO	0.98
2	PAR/BEN/CAF	28/53/18	BEN	0.99	BEN	0.99	BEN	0.99
3	BEN/LAC/MgS	35/51/12	BEN	0.99	BEN	0.99	BEN	0.99
4	MCC/TAL/PAR	26/44/30	MCC/TAL/PAR	0.99	MCC/TAL/PAR	0.99	MCC/TAL/PAR	0.99
5	BEN/PRO/MCC	25/32/43	BEN/PRO	0.99	BEN/PRO	0.99	BEN/PRO	0.99
6	PAR/BEN/PRO	36/33/31	BEN/PRO	0.99	BEN/PRO	0.99	BEN/PRO	0.99

For Progeny™, at least one adulterant was reproducibly identified for each measurement in all six mixtures ($WCC \geq 0.98$) with a maximum of two adulterants being identified per mixture (Table 2.14). Adulterants were identified in mixtures 3, 4 and 5 despite the presence of one or two cutting agents in the mixtures such as LAC, MCC, MgS and TAL. It was noted that PAR was not identified in ternary mixtures containing other adulterants, yielding false negatives in mixtures 1 and 6.

In conclusion, using the Truscan™ GP and FirstGuard™ instruments, one adulterant was identified in 50 and 100 % of mixtures. Match-testing against mixtures in the library gave rise to false positives, possibly due to non-homogeneity of the mixtures. Using Progeny™, at least one adulterant was identified in 100 % of mixtures, two adulterants were identified in two out of four mixtures. With all instruments, the adulterant identified was the strongest Raman scatterer, which was in equal or greater amount than other adulterant(s) in the mixture. This finding corroborated with published literature on pharmaceutical materials [223]. When the adulterant identified was mixed with cutting agents only, it was identified even if its amount was less than that of the cutting agents.

2.3.2.5. Testing the complexity of NPS internet products

The three instruments were evaluated for their ability to identify the NPS and adulterant constituents in NPS products as characterised by GC-EI-MS qualitative analysis. The constituents identified, their major ion peaks (m/z) and retention time (min) are shown in Table 2.15.

Product 1 (i.e. P1) was labelled as 2-aminoindane (2-AI) but was confirmed to contain CAF [224, 225]. P2 was labelled as 'Pink Champagnes', claimed to contain aminoindane, CAF, cola vera and amino acid complex. The GC-EI-MS analysis of P2 confirmed the presence of 2-AI [226] and CAF [224, 225]. The products P3 - P7 claimed to contain DXM, which was confirmed with the GC-EI-MS [227] except for P3, which was confirmed to contain CAF and methoxetamine (MXT) (Table 2.15).

Table 2. 15: Comparison of the ‘first pass’ matching analysis of NPS products using Raman Truscan™ GP, FirstGuard™, Progeny™ and corresponding GC-EI-MS results.

Product No.	Label claim	Truscan™ GP		FirstGuard™		Progeny™		GC-EI-MS		
		Library signature	PVAL	Library signature	% HQI	Library signature	RMA (% CC)	Library signature	Major ion peak (m/z) ¹	Retention time (min)
P1	2AI	No match found	-	No match found	-	CAF	1.00	CAF	194	14.4
P2	Pink Champagnes	No match found	-	CAF	75.9	2AI/CAF	0.98	CAF	194	14.4
								2AI	133	9.8
P3	DXM	No positive match to CAF	0.0018	CAF	80.0	CAF	0.96	CAF	194	14.4
								MXT	190	14.8
P4	DXM	DXM	0.1293	DXM	60.0	DXM/Starch	1.00	DXM	271	16.6
P5	DXM	DXM	0.1311	DXM	60.4	DXM/Starch	1.00	DXM	271	16.6
P6	DXM	DXM	0.1154	No match found	-	DXM/Starch	0.99	DXM	271	16.6

¹Corresponds to the major ion peaks for the main constituents in each NPS product.

Using the in-built algorithms (i.e. PVAL for Truscan™ GP, % HQI for FirstGuard™ and RMA for Progeny™), the three handheld Raman instruments provided different ‘actionable results’ for the NPS internet products. For Truscan™ GP, the NPS content was confirmed in 3/6 NPS internet products (i.e. P4 - P6). For FirstGuard™, the NPS content was confirmed in 2/6 NPS internet products (i.e. P4 and P5). However, the adulterant content was confirmed in 2/6 products (i.e. P2 and P3), yielding two false negatives for the NPS in the mixtures. This could possibly be because of a larger composition of the adulterant in the mixtures. For Progeny™, the NPS and adulterant content in all six mixtures were identified except one false negative in 1/6 mixtures. Mixtures of DXM with starch were identified with Progeny™ for P4 - P6. It is not surprising that starch has not been identified with GC-EI-MS as it is not readily soluble in methanol and has low volatility.

In order to further explore these results, the products’ Raman spectra were examined. Figure 2.8 is an example of a selected product (P2), where no correlation to an NPS was identified by Truscan™ GP, the adulterant only was identified with FirstGuard™ and both constituents (i.e. NPS and adulterant) were identified with Progeny™. The Raman spectra showed that for Truscan™ GP, an intense fluorescent background has completely masked the Raman signals and, hence no correlation was made to an NPS. Coloured species in NPS branded products are known to absorb monochromatic radiation, which may lead to fluorescence emission with a laser λ_{ex} of 785 nm [117]. The effect of the laser λ_{ex} on the identification of NPS is investigated in Chapter 3. For FirstGuard™, it was difficult to distinguish Raman features for CAF from the spectrum due to poor S/N. For Progeny™, although the spectra showed a relatively adequate S/N and an accurate ‘first pass’ correlation to both 2-AI and CAF, prominent peaks were observed at ca. 222, 782, 863, and 1020 cm^{-1} and were consistent with 2-AI only.

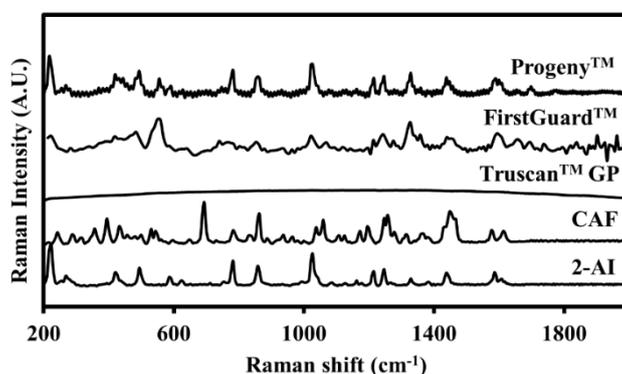


Figure 2. 8: Normalised Raman spectra for Pink Champagnes (P2) using Truscan™ GP, FirstGuard™ and Progeny™ and the reference spectra for 2-AI and CAF (Progeny™).

The following Section summarises the performance of the three instruments in terms of the generation of ‘true judgments’ or ‘false judgments’ in Table 2.16 and summarises the overall results for the three instruments in Table 2.17. A ‘true judgment’ is explained as true positive and true negative results, whereas a ‘false judgment’ is explained as ‘false positive’ and ‘false negative’ results [131]. A ‘true positive’ result is a constituent that is truly included in the mixture. For example, 2-AI/CAF correlating to 2-AI/CAF constitutes two true positives. A ‘true negative’ result is a constituent that is not included in the mixture. This type of information cannot be obtained using the in-built algorithms of the three instruments i.e. does it contain cocaine/ and the answer is ‘yes’ or no’ and, hence is not included in the Table. A ‘false positive’ result is a constituent that was identified but in fact, it is not truly included in the mixture. For example, 2-AI/CAF correlating to DXM/CAF constitutes one false positive for DXM. A ‘false negative’ result is the number of constituents unidentified and/ or displayed as a ‘no match found’ or ‘unknown’. Ideally, a greater number of true positives and less numbers of true negatives, false positives and false negatives are preferred.

In Table 2.16, a ‘no positive match to compound x’ result for Truscan™ GP is considered as a positive correlation to compound ‘x’. This is because it is still a suggestion made by the decision tool of the instrument, but the correlation did not exceed the threshold. The numbers presented in Table 2.16 are total numbers of NPS and adulterants identified in all types of mixtures investigated in this preliminary study.

Table 2. 16: Summary of the performance of three handheld Raman instruments with respect to ‘true’ and ‘false’ judgment results.

Handheld Raman Instrument	Mixtures	True judgements	False judgements	
			True positives	False positives
Truscan™ GP	Binary Mixture	75	2	35
FirstGuard™		73	6	27
Progeny™		80	0	20
Truscan™ GP	Ternary Mixture	4	0	13
FirstGuard™		8	6	5
Progeny™		8	0	5
Truscan™ GP	NPS products	4	0	8
FirstGuard™		4	0	6
Progeny™		7	0	1

In model binary mixtures, Truscan™ GP generated a greater number of false negatives as compared to the other instruments, whereas FirstGuard™ generated a greater number of false positives than the other instruments. In contrast, Progeny™ generated a greater number of true positives and least number of false negatives. In addition, Progeny™ did not generate any false positives.

For ternary mixtures, Truscan™ GP generated a greater number of false negatives and least number of true positives compared to the other instruments. Again, FirstGuard™ generated the greatest number of false positives. In contrast, Progeny™ generated a similar number of true positives as FirstGuard™ and least number of false negatives. In addition, Progeny™ did not generate any false positives.

For NPS products, all instruments did not generate any false positive results. Both Truscan™ GP and the FirstGuard™ generated true negative and false negative results for both NPS and adulterants. In contrast, Progeny™ generated a greater number of true positives and least number of false negatives compared to the other instruments. In addition, Progeny™ did not generate any false positives. It is remarkable that Progeny™ was consistent with the gold standard GC-EI-MS results for NPS internet products except for P3, where MXT was not identified. Progeny™ also proved its ability to identify cutting agents in P4, P5 and P6. However, further studies to confirm the cutting agents' content in these products are outside the scope and purpose of this study.

In summary, different algorithms have different advantages and disadvantages based on the application [164]. For example, setting a threshold to confirm screening/ identification may lead to false positive or negative results [162]. In contrast, not setting a threshold enables the identification of NPS present in low amounts in mixtures but may also lead to false positives. In-built algorithms may impact selectivity of the instrument if the NPS in the unknown sample is absent from the library or if an analogous NPS is present instead. In the latter case, analogous substances with high structural similarity display minor differences in their Raman fingerprints [152], which may lead to false positive results.

Furthermore, some algorithms, such as the 'hit quality index (HQI)' has been shown to depict slope and offset as differences from library spectra [228] and may erroneously consider fluorescent background as a sample identifier [212]. The use of in-built algorithms for the screening of complex NPS mixtures is dependent on a number of factors such as the number

of constituents in the mixtures, diversity of Raman scattering activity of the constituents, percentage composition of the drug of interest in the mixture, the content and size of the library and the ability of in-built algorithms to identify mixture constituents with the least false positives and false negatives [118, 229]. A summary of the overall performance of three handheld Raman instruments for model binary mixtures, random ternary mixtures and NPS products purchased from the internet is shown in Table 2.17.

Table 2. 17: Summary of the handheld Raman instruments' overall performance for model binary mixtures, random ternary mixtures and NPS internet products.

Handheld Raman instrument	Analysis of pure reference standards	Analysis of model binary mixtures	Range within which all constituents were identified in model binary mixtures	Analysis of random ternary mixtures	Analysis of NPS internet products
Truscan™ GP	19/20 reference standards identified with a PVAL \geq 0.0845	Library correlations were found for all the mixtures with a PVAL \geq 0.0789	Positives correlations were achieved within the following ranges: 1D 20 - 95 % m/m 2D 20 - 95 % m/m 3D > 95 % m/m 4D 10 - 95 % m/m 5D 10 - 95 % m/m	At least one adulterant was identified in 3/6 (50 %) mixtures.	Positive library correlations to an NPS were found in 3/6 products (i.e. P4 - P6).
FirstGuard™	19/20 reference standards identified with a % HQI \geq 71.7 \pm 1.2	Library correlations were found for all the mixtures with a % HQI \geq 66.8	Positives correlations were achieved within the following ranges: 1D 20 - 95 % m/m 2D 20 - 95 % m/m 3D 30 - 60 and > 95 % m/m 4D 20 - 95 % m/m 5D 10 - 95 % m/m	At least one adulterant was identified twice in 6/6 (100 %) mixtures and two adulterants were identified at least once in 1/4 (only four samples had more than one adulterant).	Positive library correlations to an NPS were found in 2/6 products (i.e. P4 and P5). In P2 and P3, the adulterant only was identified.
Progeny™	20/20 reference standards identified with a % CC \geq 97.0 \pm 0.4 $\times 10^{-4}$	Library correlations were found for all the mixtures with WCC values \geq 0.75	Positives correlations were achieved within the following ranges: 1D 20 - 95 % m/m 2D 30 - 95 % m/m 3D 20 - 95% m/m 4D 10 - 95 % m/m 5D 10 - 95 % m/m	At least one adulterant was identified in 6/6 (100 %) mixtures in all measurements. Two adulterants were identified in 2/4 (only four samples had more than one adulterant). Cutting agents were correctly identified in 1/6 mixtures.	Positive library correlations to an NPS were found in 5/6 products (P1, 2, 4-6). In P3, the adulterant only was identified.

2.4. Conclusions

In this preliminary study, Progeny™ operating with a 1064 nm laser source and multiple algorithms is the recommended instrument since it has demonstrated to be more suitable for the identification of NPS in ‘street-like’ NPS products. In this respect, Progeny™ has shown better selectivity to the same class of material and/ or similar chemical structure than both Truscan™ GP and FirstGuard™ with a 100 % identification of pure substances. The ‘drug’ constituent in model binary mixtures was identified within a minimum range of 20 – 95 % m/m for Truscan™ GP and 20 – 95 % m/m for FirstGuard™ and 10 – 95 % m/m for Progeny™. In contrast to both Truscan™ GP and FirstGuard™, using RMA, Progeny™ did not generate any false positive correlations for model binary mixtures, random ternary mixtures and NPS internet products. The Progeny™ instrument identified the NPS/ adulterant content in all NPS internet products in consistence with GC-EI-MS confirmatory analysis, except for one product, where only one constituent was identified.

The presence of equally-spaced discrete binary mixtures (25/75, 50/50 and 75/25) in the library has improved the accuracy of positive correlations of model binary mixtures. However, it had a great impact on the qualitative identification using both Truscan™ GP and FirstGuard™ because of inhomogeneity of samples used in creating the spectral libraries and because the spectral libraries made of discrete mixtures did not reflect the true composition of the model mixtures. Conversely, Progeny™, using RMA and pure reference standards only in the library outperformed the other instruments in the identification of the drug(s) of interest.

This pilot study has highlighted important factors that influence the identification of NPS mixtures using handheld Raman spectroscopy. First, instrument factors such as the laser λ_{ex} (i.e. 785 versus 1064 nm), where an adequate Raman spectrum could not be obtained for a few cutting agents and NPS products using a high energy short laser λ_{ex} of 785 nm. However, identification of constituents in NPS internet products was improved with a longer laser λ_{ex} of 1064 nm, possibly because of reduced fluorescence. The design and weight of the instruments influenced the stable samples presentation for analysis may have had a great impact on collecting Raman spectra on DIL placed on aluminium plate rather than in a glass vial. The content and quality of library spectra had an impact on the number of positive correlations using the instruments’ in-built algorithms. Fixed non-adjustable parameters by the operator and the use of proprietary methods may not be suitable for the identification of NPS mixtures and

may cause sample damage. Second, sample factors such as sample heterogeneity, colour, number of adulterants in a single mixture, Raman scattering activity of constituents in mixtures, presence of fluorescing impurities. Third, in-built algorithm factors such as the presence of a threshold for positive correlations and impact of poor S/N on correlation values and algorithm calculation.

This study highlighted the strengths and limitations of each instrument in the identification of different NPS, adulterants and cutting agents routinely encountered in seized NPS street samples. This work illustrated that the effectiveness of both in-built algorithms and on-board libraries in improving the identification of solid model mixtures depends on the quality of the spectra. This study illustrated the issues that may be encountered in street samples such as the presence of multiple constituents, constituents with various Raman scattering activity and the presence of fluorescing excipients/ impurities such as cutting agents. Results revealed the practical considerations for the in-field detection of NPS using handheld Raman instruments such as the ease of presenting the sample to the instrument and the feasibility of obviating sample damage with the laser.

Handheld Raman spectroscopy has shown promise as a non-contact non-destructive technique that could be employed for the identification of NPS mixtures with unknown composition in the field. However, in-built algorithms can lead to misidentification or failure of identification of NPS products based on instrument, algorithm, library and sample factors. Further work is needed to evaluate the precision of the instruments by means of measuring a greater number of replicates. Future work may include the evaluation of 'first pass' matching by altering the threshold of the % HQI algorithm. The next Chapter will evaluate the impact of the laser λ_{ex} on the possibility of improving the identification of NPS in a wider range of NPS products.

3. Comparison of two handheld Raman Spectrometers employing 785 and 1064 nm laser sources for the identification of new psychoactive substances

3.1. Introduction

This Chapter evaluates two Raman laser excitation wavelengths (λ_{ex}) 785 and 1064 nm for the identification of new psychoactive substances (NPS) purchased from the internet. This work is significant, as there remains a need for improved selective and rapid in-field detection of NPS by law enforcement and healthcare professionals.

Due to the intrinsic nature of NPS samples in terms of the presence of impurities such as fluorescing chemicals and coloured ingredients, identification of NPS content can still be challenging using standard handheld Raman instruments (λ_{ex} of 785 nm) (see Chapter 2 and Assi et al. [128]). Numerous methods were employed to reduce fluorescent background. For example, surface enhanced Raman spectroscopy (SERS) has shown promising improvement in enhancing the Raman signal while also reducing fluorescent background [113, 114]. However, invasive sample preparation is often needed. An alternative approach to reduce fluorescence and, which requires no sample preparation is the use of a longer λ_{ex} (e.g. between 750 and 1000 nm) [101]. This approach has been shown to evade interference from fluorescence and improve identification of traditional drugs of abuse such as cocaine and amphetamine [154, 157]. At present, a few studies have evaluated the use of Raman spectroscopy with a long λ_{ex} (e.g. 1064 nm) to reduce fluorescence [117, 158, 159] and investigate the feasibility of improving the identification of drugs of abuse [160] and a limited range of NPS [138, 161]. Thus, there is still limited information on the use of Raman spectroscopy for the wide range of chemically diverse NPS products and the feasibility to improve NPS identification by using a longer λ_{ex} (i.e. 1064 instead of 785 nm).

Previous work has shown preliminary results for improved identification of NPS in model NPS mixtures and a limited range of NPS internet samples with the ProgenyTM instrument (λ_{ex} = 1064 nm) (see Chapter 2). However, with the FirstGuardTM instrument (λ_{ex} = 1064 nm), there was a compromise between reduced fluorescent background and reduced S/N. Thus there was a need to further investigate the impact of a longer laser wavelength. The aim of this Chapter was to evaluate the feasibility of improving the identification of a wide range of NPS products

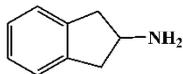
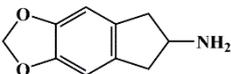
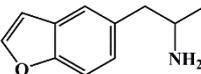
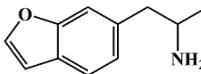
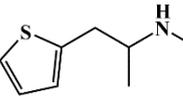
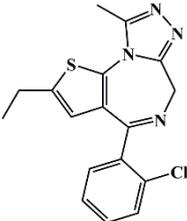
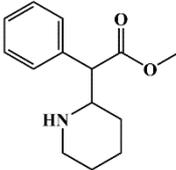
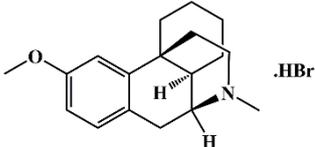
purchased from the internet using two handheld Raman instruments i.e. XantusTM-1 ($\lambda_{\text{ex}} = 785$ nm) and FirstGuardTM ($\lambda_{\text{ex}} = 1064$ nm) (SciMed Ltd). The instruments were selected because they operate with the same software, settings and algorithms (i.e. % HQI) but differ with respect to the laser λ_{ex} .

3.2. Experimental

3.2.1. Chemicals and Reagents

The reference standards of eight NPS drugs, eight adulterants and twelve cutting agents were used for the study. The NPS reference standards 2-aminoindane (2-AI), 5,6-methylenedioxy-2-aminoindane (5,6-MDAI), 1-benzofuran-5-ylpropan-2-amine (5-APB), 1-benzofuran-6-ylpropan-2-amine (6-APB), 1-(thiophen-2-yl)-2-methylamino propane (MPA), etizolam and methylphenidate (MPD) were purchased from LGC standards (Teddington, UK); dextromethorphan hydrobromide (DXM) was purchased from Sigma Aldrich (Dorset, UK) (Table 3.1).

Table 3. 1: Chemical structures of NPS reference standards used in this Chapter

2-AI	5,6-MDAI	5-APB	6-APB
			
MPA	Etizolam	MPD	DXM
			

The adulterants benzocaine (BEN), caffeine (anhydrous) (CAF), lidocaine hydrochloride (LID), paracetamol (PAR), phenacetin (PHE) and theophylline (THEO) were purchased from Sigma Aldrich (Dorset, UK); diltiazem hydrochloride (DIL) was obtained from the Medicines Testing Lab (UK); and procaine hydrochloride (PRO) was obtained from British Drug Houses (London, UK). The cutting agents calcium carbonate (CaCO_3), creatine monohydrate (CRE), dextrose monohydrate (DEX), α -D-glucose anhydrous (GLU), lactose (LAC), L-tyrosine (L-TYR), magnesium stearate (MgS), microcrystalline cellulose (MCC), niacinamide (NIA), sucrose (SUC), talc (TAL) and taurine (TAU) were purchased from Sigma Aldrich (Dorset,

UK). Sixty NPS products were purchased from the internet, under a Home Office licence. In order to ensure a wide chemical diversity, the products were selected according to their label claim, such that they represented most of the NPS classes as per the EMCDDA (European Monitoring centre for Drugs and Drug Addiction) classification [30]. In addition, a variety of formulations such as powders, capsules, pellets, tablets and seeds were purchased. Most of the products were ‘legal’ NPS at the time of purchase and, hence were widely available over the internet. A few NPS were under temporary class drugs orders (TCDOs) such as 5-APB, 6-APB and MPA and these were more challenging to obtain. Most of these NPS were associated with formal notifications to the EU Early Warning System (EWS) and high toxicity [214, 230-234]. Hence, they pose serious risks to public health. Additional details related to the 60 NPS are provided in Table 3.2.

Table 3. 2: NPS product specifications

Product No.	Product name	Label Claim	Dosage Form	Colour of the powder	Purchase Date	Website Code ¹
Aminoindanes						
1	2-AI	2-aminoindane	Powder	Dark beige	28/01/2013	1
2	2-AI	2,3-dehydro-1H-inden-2-amine	Powder	Beige	16/01/2013	2
3	5-IAI	5-iodo-2-aminoindane	Powder	White	29/02/2012	3
4	5-IAI	5-IAI	Powder	Beige	29/02/2012	3
5	MDAI	5,6-methylenedioxy-2-aminoindane	Powder	Dark white	14/02/2012	4
6	MDAI	6,7-dihydro-5H-cyclopenta[F][1,3]benzodioxol-6-amine	Powder	Dirty white	28/01/2013	5
7	MDAI	6,7-dihydro-5H-cyclopenta[F][1,3]benzodioxol-6-amine	Powder	Dirty white	28/01/2013	1
8	MDAI	6,7-dihydro-5H-cyclopenta[F][1,3]benzodioxol-6-amine	Powder	Dirty white	28/01/2013	1
9	MDAI	5,6-MDAI	Powder	Dirty white	16/01/2013	6
10	MDAI	MDAI	Powder	Beige	28/01/2013	7
11	MDAI	None	Tablets	Antique pink	28/01/2013	7
12	MDAI	MDAI	Powder	Dirty white	16/01/2013	2
Arylalkylamines						
13	APB	APB	Tablets	Pink/ fuchsia	28/01/2013	5
14	5-APB	1-benzofuran-5-yl-propan-2-amine	Powder	Beige	16/01/2013	6
15	5-APB	5-(2-aminopropyl)benzofuran	Powder	Brown/ dark yellow	28/01/2013	3
16	5-APB	5-APB	Powder	Brown/ dark yellow	28/01/2013	8
17	5-APB	5-(2-aminopropyl)benzofuran	Powder	Reddish brown	28/01/2013	2
18	6-APB	benzofuran-6-yl-propane-2-amine	Powder	Beige	16/01/2013	6
19	6-APB	6-(2-aminopropyl)benzofuran	Powder	Dirty white	16/01/2013	2

20	5-MAPB	5-MAPB (part missing for label overlapping) benzofuran-5-yl)-N-methylpropane-2-amine	Tablets	Salmon pink	28/01/2013	5
21	MPA	1-(thiophen-2-yl)-2-methylaminopropamine	Powder	Dirty white	28/01/2013	7
22	MPA	N-methyl-1-(thiophen-2-yl)propane-2-amine	Powder	Beige	16/01/2013	6
23	MPA	methiopropamine	Powder	Beige	28/01/2013	3
24	MPA	MPA	Powder	Dirty white	28/01/2013	8
25	MPA	methiopropamine; N-methyl-1-(thiophen-2-yl)propan-2-amine	Powder	Dirty white	16/01/2013	2
26	MPA	MPA	Powder	Dirty white	28/01/2013	5
Benzodiazepines						
27	Etizolam	6-(2-chlorophenyl)-4-ethyl-13-methyl-3-thia-1,8,11,12-terazatricyclo[8.3.0.0.2,6]trideca-2(6),4,7,10,12-pentaene	Tablets	White and turquoise coating	28/01/2013	5
28	Etizolam	6-(2-chlorophenyl)-4-ethyl-13-methyl-3-thia-1,8,11,12-terazatricyclo[8.3.0.0.2,6]trideca-2(6),4,7,10,12-pentaene	Tablets	White and turquoise coating	16/01/2013	6
29	Etizolam	Etizolam	Tablets	White and turquoise coating	28/01/2013	3
30	Etizolam	Etizolam	Tablets	White powder with pink coating	16/01/2013	2
31	Etizolam	Etizolam	Tablets	White powder with turquoise coating	28/01/2013	5
32	Etizolam	Etizolam	Pellets	White powder with turquoise coating	28/01/2013	8
33	Pyrazolam	1-methyl[1,2,4]triazol-6-(2-pyridinyl)-8-bromo-1,4-benzodiazepine	Pellets	White and pale yellow	28/01/2013	8
34	Pyrazolam	1-methyl[1,2,4]triazol-6-(2-pyridinyl)-8-bromo-1,4-benzodiazepine	Tablets	White and pale yellow	28/01/2013	5
35	Pyrazolam	1-methyl[1,2,4]triazol-6-(2-pyridinyl)-8-bromo-1,4-benzodiazepine	Tablets	White and pale yellow	28/01/2013	5
Piperidines & Pyrrolidines						
36	Ethyl phenidate	(RS)ethyl-2-phenyl-2-piperidin-2-ylacetate	Powder	White	28/01/2013	5
37	Ethyl phenidate	(RS)ethyl-2-phenyl-2-piperidin-2-ylacetate	Powder	Dirty white	28/01/2013	7
38	Ethyl phenidate	Ethylphenidate	Powder	Dirty white	28/01/2013	3
39	Ethyl phenidate	Ethylphenidate	Powder	Dirty white	28/01/2013	8
40	Ethyl phenidate	Ethylphenidate; ethyl-2-phenyl-2-piperidin-2-ylacetate	Powder	Dirty white	16/01/2013	2
41	Ethyl phenidate	Ethylphenidate	Powder	Dirty white	28/01/2013	3
Plants & Extracts						
42	LSA Morning Glory Seeds	None	Seeds	Cappuccino with black pieces from seed shells	14/02/2014	9

Other						
43	DXM	DXM; dextromethorphan; (+)-3-methoxy-17-methyl-(9 α ,13 α ,14 α)-mor	Powder	Dirty white	28/01/2013	3
44	DXM	DXM	Powder	White	28/01/2013	10
45	DXM	DXM	Powder	Dirty white	28/01/2013	10
46	DXM	DXM	Powder	Dirty white	28/01/2013	10
47	DXM	DXM	Powder	White	14/02/2012	11
Branded products						
48	Blow	Unknown	Powder	Dirty white	28/01/2013	9
49	Blurberry	Aminoindan, caffeine, cola vera, amino acid complex.	Capsules content	Brown	01/03/2012	9
50	High beams	Ipomoea convolvulaceae, amino acid complex, cola vera, zingiber officinale, ginseng, caffeine.	Capsules content	Cinnamon brown	14/02/2014	9
51	Magic Beans	Caffeine, cola vera, amino acid complex.	Capsules content	Beige	14/02/2014	9
52	Pink Champagnes	Aminoindan, caffeine, cola vera, amino acid complex.	Capsules content	Cinnamon brown	14/02/2012	9
53	Pink Champagnes	Aminoindan, caffeine, cola vera, amino acid complex.	Capsules content	Cinnamon brown	14/02/2012	9
54	Pink Champagnes	Aminoindan, caffeine, cola vera, amino acid complex.	Capsules content	Cinnamon brown	14/02/2012	9
55	Pink Champagnes	Aminoindan, caffeine, cola vera, amino acid complex.	Capsules content	Cinnamon brown	14/02/2012	9
56	Pink Panthers	1-(thiophen-2-yl)-2-methyl aminopropane, 5,6-methylenedioxy-2-aminoindan	Capsules content	Dirty white	14/02/2014	12
57	Punk plus	L-tyrosine, caffeine, gelatin, niacinamide, green tea extracts, salix alba, eleuterococcus senticosus, taurine, poveira cupono, colo vera, magnesium stearate	Capsules content	Light brown	14/02/2014	9
58	Recovery	Ascorbic acid, niacinamide, ground rice powder, D-alpha, tocopheryl acetate, niacinamide, ferrous fumarate, retinyl acetate, calciun pantothenate, chromium chloride, folic acid, cyanocobalamin, cholecalciferol, pyridoxine HCl, thiamine concentrate, riboflavin, magnesium stearate	Capsules content	Beige	14/02/2014	9
59	Route 56	Route 56	Pellets	Dark pink	28/01/2013	3
60	White Pearls	Ketones, caffeine, zingiber officinale, amino acid complex.	Capsules content	Pale yellow	14/02/2012	9

¹For the website details, see Appendix B Table A3.1.

Powders and capsules were emptied into clear glass vials (Kimble Chase vial screw thread with PTFE cap, China) for Raman analysis, while the pellets and tablets were crushed using an agate mortar and pestle before transferring into glass vials. The glass vials were vortex-mixed before collection of each spectrum using a VORTEX-GENIE2 (Scientific industries, Inc., USA) mixed for 30 sec, shaken, then the process repeated. For GC-EI-MS analysis, solutions (1 mg mL⁻¹) of each standard and product was prepared in HPLC grade methanol from Fisher

Scientific (Loughborough, UK) except for the benzodiazepine tablets, which were concentrated to ca. 45 mg mL⁻¹ and filtered through 0.2 µm PTFE membrane filters (National Scientific Company, USA) prior to analysis.

3.2.2. Instrumentations and methods

Two handheld Raman instruments with different laser sources, XantusTM-1 and FirstGuardTM (SciMed Ltd, Rigaku, USA), were employed for the analysis of NPS products. Specifications of both instruments as shown in Table 3.3 are very similar. One of the main differences in specifications is the laser λ_{ex} .

Table 3. 3: Specifications of XantusTM-1 and FirstGuardTM handheld Raman instruments

Raman Instrument Specifications	Xantus TM -1	FirstGuard TM
Instrument image ¹		
Laser λ_{ex} (nm)	785	1064
Laser output power (mW)	30 - 490	30 - 490
Laser spot diameter (µm)	200	100
Spectral resolution (cm ⁻¹)	7 - 10	15 - 18
Spectral range (cm ⁻¹)	200 - 2000	200 - 2000
Numerical aperture	0.3	0.3
Grating	Transmission volume phase (VPG TM)	Transmission volume phase (VPG TM)
Detector	TE Cooled CCD	TE Cooled InGaAs
Working distance (mm)	20	20
Weight (kg)	2.2	2.7
Dimensions (LxWxH) (cm)	12.5 x 23.3 x 8.5	12.2 x 31.1 x 31.4
Library	Standard library (266 items) and user library	Standard library (266 items) and user library
Operation and analysis software	Micro2020, WindowsXP/Vista/Win7	Micro2020, WindowsXP/Vista/Win7
Data analysis	Grams, Unscrambler	Grams, Unscrambler
Data export formats	SPC and csv	SPC and csv
Algorithm	Percentage hit quality index (% HQI)	Percentage hit quality index (% HQI)
Operating temperature (°C)	-10 to +30	-10 to +30
Battery	Switchable lithium ion battery (4 hours operation)	Switchable lithium ion battery (4 hours operation)
Other	no adjustable probe tip	Focus adjustable probe tip.

¹Instrument images reproduced with courtesy of SciMed, Rigaku, USA.

Four methods were developed to collect the Raman spectra depending on the nature of the substance and included method A (1000 ms exposure time; 300 mW laser power; 2 accumulations), method B (5000 ms exposure time; 490 mW laser power; 2 accumulations),

method C (500 ms exposure time; 200 mW laser power; 2 accumulations), and method D (1000 ms exposure time; 100 mW - 1 mW laser power; 2 – 25 accumulations). All samples were initially run using method A, but method B was used for samples that displayed a poor Raman signal. Method C was used for samples that were burned from long exposure time and/or high laser power (i.e. coloured samples or samples containing fluorescing chemicals). Method D was developed in an attempt to collect Raman signals from challenging samples, which exhibited intense fluorescent background and/ or burned with method C. This was performed by adopting an iterative approach to reducing the laser power and increasing the number of accumulations. All methods used baseline correction, and the dark background was corrected for every 15 minutes. The instruments were calibrated each day immediately before analysis using a benzonitrile reference standard (Rigaku, US). Most reference standards and products were analysed directly through glass vials after optimisation of the vial holder attachment with respect to the focal point. For NPS standards that were limited in quantity, approximately 2 mg were placed on aluminium plates (HSA14521A - Weight dish alum 43 mm, Fisherbrand) and the signal was optimised using the adjustable probe tip. All substances were analysed in triplicate. Raman spectra of the reference substances (n = 28) were added to the on-board factory spectral library, which was composed of 260+ spectra of common chemical substances.

For a ‘first pass’ identification, the spectra from the NPS products were automatically compared to the on-board reference library and reported as a percentage hit quality index (% HQI) correlation. The mean \pm the standard deviation (SD) of the top hit was calculated from the triplicate measurements and reported. If the correlations between the triplicate analyses were inconsistent, this was reported instead of a mean value. The findings of the matching algorithms were evaluated by a comparison to the GC-MS results and also by comparing the spectra of the products to the reference spectra. Miss S. Giroto and Miss B. Berti, visiting Erasmus students, contributed to data collection under my supervision.

3.2.3. Confirmatory analysis of NPS and related substances using gas chromatography – electron ionisation-mass spectrometry (GC-EI-MS)

GC-MS analysis was used to confirm the identity of compounds present in the purchased internet products. The method was adapted from Kellett [235]. The analysis was performed using electron ionisation (EI) only. The mass spectra obtained were compared to the purchased reference standards and the following EI spectral libraries: NIST (Version 1.0.2.2),

SWGDRUG MS (Version 2.1 (2014)) and Cayman (Version 04292014). Miss S. Girotto, a visiting Erasmus student, contributed to data collection under my supervision.

3.3. Results and Discussion

3.3.1. Raman analysis of purchased reference standards

Eight NPS, eight adulterants and 12 cutting agents were purchased as reference standards from the internet. A spectrum was collected for each substance and stored in the on-board library of both handheld instruments. Figure 3.1 shows the Raman spectra of NPS reference standards uploaded in the on-board library.

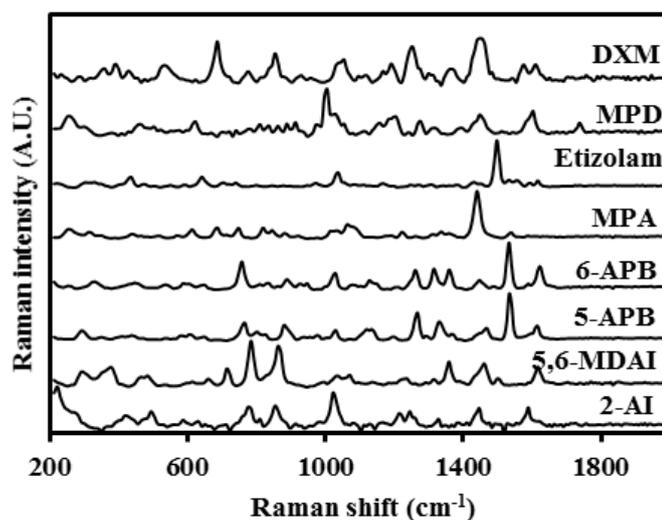


Figure 3. 1: Raman spectra of NPS reference standards uploaded in on-board library (FirstGuard™ (1064 nm)).

The examination of the reference spectra for NPS showed that the aminoindane analogues are distinguishable by characteristic Raman peaks at 713 and 1350 cm^{-1} , possibly attributed to $\text{C}=\text{C}$ cis-di-substituted deformation vibrations and methoxy stretching vibration of the methylenedioxy moiety in 5,6-MDAI [138]. The chemical structures of the arylalkylamine 5-APB and its positional isomer 6-APB are very similar as they only differ in the substitution position of the alkylamine on the 5- and 6- positions of the benzofuran group respectively. However, their Raman spectra can be distinguished from each other by a key discriminating peak at 1350 cm^{-1} for 6-APB.

The spectrum for MPA, another arylalkylamine, has shown a prominent peak at 1438 cm^{-1} . Similarly, the spectrum for etizolam showed a characteristic peak at 1496 cm^{-1} . The piperidine MPD showed numerous characteristic peaks at 996, 1189, 1431, 1587 and 1723 cm^{-1} . DXM

from the ‘other’ category also showed numerous characteristic peaks at 686, 853, 1245 and 1439 cm^{-1} .

To test the accuracy and selectivity of each instrument, the standards were match-tested against the library spectra as test samples. Table 3.4 shows the % HQI results of both handheld Raman instruments. Characteristic Raman peaks were identified using a laser λ_{ex} of 785 nm.

Table 3. 4: Validation of NPS reference standards using two handheld Raman spectrometers (XantusTM-1 (785 nm) and FirstGuardTM (1064 nm)) and on board libraries.

	Xantus TM -1 (785 nm) % HQI \pm SD ¹	FirstGuard TM (1064 nm) % HQI \pm SD	Characteristic Raman peaks at λ_{ex} 785 nm (cm^{-1})
Aminoindanes			
2-AI	99.51 \pm 0.02	72 \pm 1	427, 496, 779, 855, 1021
5,6-MDAI	93 \pm 2	78 \pm 2	715, 782, 855, 1351
Arylalkylamines			
5-APB	97 \pm 2	94.0 \pm 0.2	763, 878, 1026, 1266, 1324, 1535
6-APB	99.6 \pm 0.2	74.3 \pm 0.1	756, 1026, 1527, 1621, 1258, 1310, 1356
MPA	90 \pm 10	85 \pm 1	565, 610, 684, 743, 823, 850, 1068, 1085, 1442
Benzodiazepines			
Etizolam	97.9 \pm 0.4	96 \pm 3	437, 642, 1032, 1495
Piperidines & Pyrrolidines			
MPD	99.0 \pm 0.2	95 \pm 2	999, 622, 1199, 1272, 1445, 1594
Other			
DXM	99.93 \pm 0.06	96 \pm 1	686, 853, 1051, 1250, 1446

¹SD = standard deviation

Using the 785 nm source, 27 out of 28 standards were consistent with their library signature with % HQI values ranging from 90 \pm 10 to 100 %. Microcrystalline cellulose resulted in inconsistent correlations to LAC and amylose from potato. The spectra indicated that the mismatch was likely the result of a high fluorescent background with poorly defined Raman features seen for all three replicates. When using the longer 1064 nm wavelength, 27 out of 28 standards correlated to their library signature. Of those, 23 standards were consistent with their library signature with % HQIs ranging from 90.0 \pm 0.9 to 100.0 \pm 0.1 %. Four samples correlated with % HQIs < 90 %. These were 2-AI, 5,6-MDAI, 6-APB and MPA with % HQI ranging from 72 \pm 1 to 85 \pm 1. The only mismatch for the 1064 nm instrument was for MgS. The MgS spectrum correlated to the signature of bees wax, a structurally similar compound, as the first hit (90 \pm 4 % HQI), but correlated to MgS in all measurements as the second hit (80 \pm

2 % HQI). In summary, both instruments showed selectivity for the majority of standards run as test samples, but fluorescence affected one sample using the 785 nm and poor Raman scatterers gave slightly reduced % HQIs (i.e. 70 – 90 %) for the 1064 nm. Slight reductions in % HQI for these particular standards were perhaps also a consequence of needing to run these standards on aluminium plates with a small sample size. Given the reliable performance of both instruments on known reference standards, the same protocols were applied to complex NPS internet samples. Although an optimisation protocol was followed, variations in the beamwidth and distance to the target substance can influence spectral quality and, hence affecting the measurement reproducibility and % HQI value during validation with pure and NPS products.

3.3.2. Raman and GC-MS analysis of NPS internet products

The effect of using different λ_{ex} (i.e. 785 and 1064 nm) for the identification of NPS internet products was assessed. In this study, 60 NPS products were analysed using two handheld Raman instruments using a ‘first pass’ in-built matching algorithm and evaluation of Raman spectra. The NPS products analysed covered a wide range of categories according to the EMCDDA classification [30] (Table 3.2). GC-MS was employed to confirm the identity of volatile/ semi-volatile compounds present in the NPS products.

3.3.2.1. Aminoindanes

Twelve aminoindane samples, purchased from the internet, were analysed using both Raman instruments and GC-MS (Table 3.5). Internet products included three of the most popular aminoindane substances, 2-AI, 5-IAI and MDAI [214, 236]. Aminoindanes are amphetamine analogues [237] and have been shown to be potent serotonin-releasing substances [214]. The GC-MS results indicated that 11 out of the 12 products did contain an aminoindane, while only nine products contained the aminoindane reported on the label claim. Two aminoindane substances only (i.e. 2-AI and 5,6-MDAI) were identified [226]. The adulterants found included CAF and MPA [238]. Two samples, P4 and P10, showed peaks in the chromatogram that were not identified *via* the MS libraries. The GC-MS results were then used to compare with the Raman results. When using the standard 785 nm laser, only four of the NPS products (i.e. P6, 8, 9 and 12) correlated to an aminoindane substance using the in-built algorithm, even though most samples were confirmed to contain an aminoindane *via* GC-MS. These four products correlated to 5,6-MDAI, as confirmed by GC-MS, with % HQIs ranging from 60 ± 8 to 84 ± 10 . A high fluorescent background and low Raman signal were observed for P1, 2, 4, 10 and 11. These products correlated to MPA at % HQIs ranging from 96.3 ± 0.8 to 97.0 ± 0.3 ,

but MPA was not confirmed using GC-MS except in P11. This was likely the result of the MPA standard spectrum (i.e. the spectral library signature) displaying a high fluorescent background (Figure 3.2). Consequently, the MPA signature correlated highly to NPS product spectra with similar backgrounds and little to no Raman bands, resulting in the false positives.

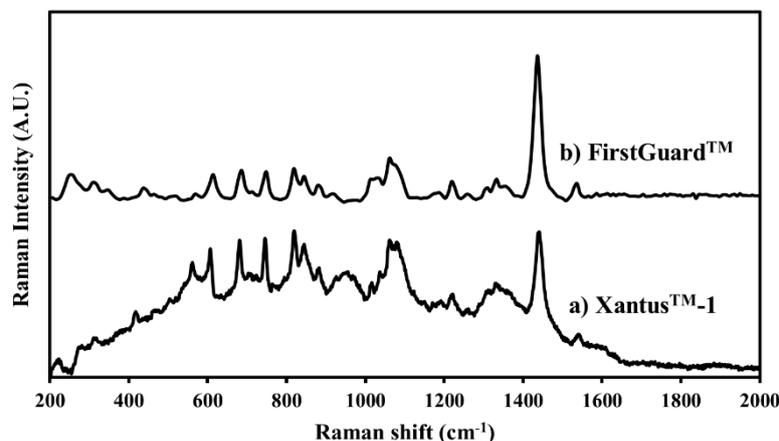


Figure 3. 2: Raman spectra of MPA reference standard using a) Xantus™-1 ($\lambda_{\text{ex}} = 785 \text{ nm}$) and b) FirstGuard™ ($\lambda_{\text{ex}} = 1064 \text{ nm}$).

For example, the Raman spectrum of P11 (Figure 3.3a) showed small peaks at ca. 714, 782 and 864 cm^{-1} , which match Raman bands for 5,6-MDAI, but due to the large fluorescent background, this sample correlated to MPA at $96.5 \pm 0.2 \%$. As the Raman band for MPA at 1438 cm^{-1} was not visible, this correlation was also likely a false positive. Products 5 and 7 showed a few higher intensity Raman peaks on a fluorescent background but resulted in no correlations (‘no match’). This may have occurred as the HQI algorithm considers fluorescence signals as additional characteristics of the unknown sample [228]. Interestingly, whilst P3 was confirmed to contain both 5,6-MDAI and CAF, the spectra collected using the 785 nm source correlated to CAF ($87 \pm 2 \%$ HQI) demonstrating the challenge of identifying NPS in a complex mixture adulterated with a relatively strong Raman scatterer.

The 12 samples were then analysed using the 1064 nm source, where 9 of the 12 samples (P1, 2, 5 - 9, 11 and 12) correlated to the NPS present in the sample with HQIs ranging from 60 ± 6 to $91.3 \pm 0.4 \%$. Again, P3 correlated to the adulterant CAF ($80 \pm 10 \%$), which was confirmed from the GC-MS analysis, while P4 correlated to benzyl alcohol ($80 \pm 2 \%$), which was not confirmed. Three samples (i.e. P3, 4, and 11) had mixtures of active ingredients indicated by GC-MS, where the highest Raman correlation presumably indicated the active ingredient either in highest concentration or with the strongest Raman signal(s). When using the 1064 nm instrument, fluorescence was significantly reduced for many of the products as shown for P11

which correlated to MDAI at 64 ± 2 % HQI (Figure 3.3b). This resulted in improved spectral definition and subsequent identification.

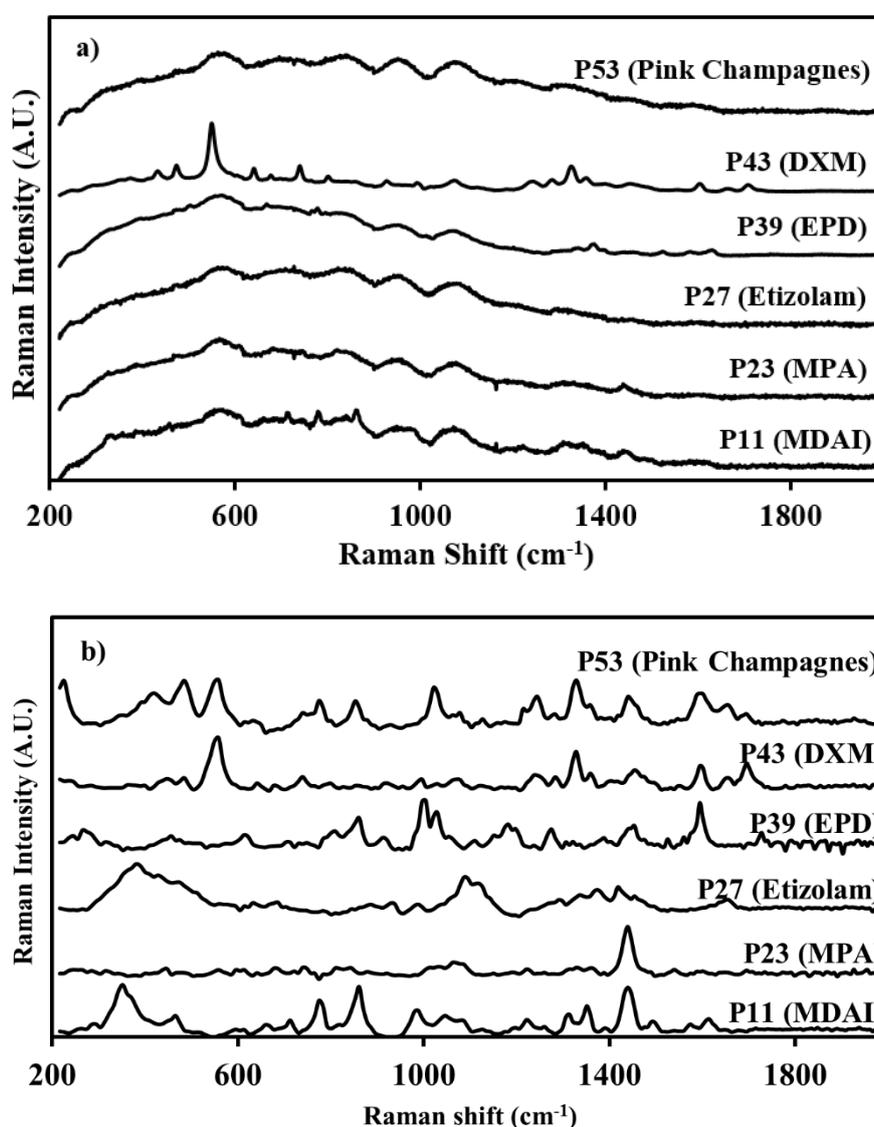


Figure 3. 3: Raman spectra of P11 (MDAI), P23 (MPA), P27 (etizolam), P39 (EPD), P43 (DXM) and P53 (Pink Champagnes) using 785 nm (a) and 1064 nm (b) excitation wavelengths. Product names represent the label claim and not the sample composition that is presented in Tables 3.5 – 3.7. The spectra were normalised to the maximum peak.

Table 3. 5: Results from the analysis of aminoindane and arylalkylamine internet products using two handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) and GC-MS¹

Product No.	Product name	Handheld Raman 785 nm		Handheld Raman 1064 nm		GC-MS		
		ID	% HQI	ID	% HQI	RT (min)	Base Peak (m/z)	MS ID
Aminoindanes								
1	2-AI	MPA ³	96.6 ± 0.4	2-AI	80 ± 4	9.9	133	2-AI
						12.6	160	5,6-MDAI
2	2-AI	MPA ³	96.3 ± 0.8	2-AI	91.3 ± 0.4	9.8	133	2-AI
3	5-IAI	CAF	87 ± 2	CAF	80 ± 10	12.6	160	5,6-MDAI
						14.4	194	CAF
4	5-IAI	MPA ³	96.0 ± 0.7	Benzyl Alcohol	80 ± 2	9.0	58	MPA
						9.4	133	2-AI
						12.6	160	5,6-MDAI
						14.64	86	NC
5	MDAI	No match		5,6-MDAI	60 ± 6	12.8	160	5,6-MDAI
6	MDAI	5,6-MDAI	84 ± 10	5,6-MDAI	80.3 ± 0.2	12.8	160	5,6-MDAI
7	MDAI	No match		5,6-MDAI	80 ± 2	12.7	160	5,6-MDAI
8	MDAI	5,6-MDAI	60 ± 8	5,6-MDAI	80.7 ± 0.1	12.8	160	5,6-MDAI
9	MDAI	5,6-MDAI	75 ± 6	5,6-MDAI	80.1 ± 0.5	12.9	160	5,6-MDAI
10	MDAI	MPA ³	97.0 ± 0.3	No match		14.0	192	NC
						14.4	191	NC
						15.2	206	NC
						15.5	177	NC
11	MDAI	MPA ³	96.5 ± 0.2	5,6-MDAI	64.0 ± 1.8	8.7	58	MPA
						12.6	160	5,6-MDAI
12	MDAI	5,6-MDAI	67 ± 4	5,6-MDAI	80 ± 2	12.8	160	5,6-MDAI
Arylalkylamines								
13	APB	MPA ³	95.2 ± 0.4	5-APB	80 ± 2	11.6	44	5-APB
						16.4	126	Pyrovalerone ²
14	5-APB	MPA ³	97.2 ± 0.6	5-APB	55.0 ± 0.6	11.6	44	5-APB
15	5-APB	MPA ³	94 ± 1	5-APB	60 ± 1	11.4	44	5-APB
16	5-APB	MPA ³	95.8 ± 0.7	No match		11.7	44	5-APB
17	5-APB	MPA ³	95.1 ± 0.1	Data acquisition failed		11.8	44	5-APB
						12.3	134	5-APDB ²
18	6-APB	Inconsistent correlations		MCC	77.4 ± 0.2	11.6	44	5-APB
19	6-APB	Inconsistent correlations		6-APB	50 ± 3	11.4	44	5-APB
						11.7	44	6-APB
						12.9	160	5,6-MDAI
20	5-MAPB	Inconsistent correlations		5-APB	60 ± 4	12.1	58	5-MAPB ²
21	MPA	MPA ³	94 ± 2	MPA	82.9 ± 0.7	9.1	58	MPA
						16.6	271	DXM
22	MPA	MPA ³	98.1 ± 0.2	MPA	60 ± 2	9.2	58	MPA
23	MPA	MPA ³	97.2 ± 0.1	MPA	78.2 ± 0.8	8.8	58	MPA
						11.9	58	NC
						12.0	44	6-APB
						14.4	194	CAF
						17.4	110	5-MeO-DALT ²
24	MPA	MPA ³	98.1 ± 0.2	MPA	81.2 ± 0.8	8.8	58	MPA
25	MPA	MPA ³	96 ± 1	MPA	80 ± 2	8.7	58	MPA
26	MPA	MPA ³	96 ± 3	MPA	80 ± 2	8.8	58	MPA

¹ID: identification; RT: retention time; NC: not confirmed.

²Raman spectrum of substance not present in both Raman libraries.

³Suspected false positive.

By examining the product spectra and comparing them to reference spectra (Figure 3.4), P1 showed Raman peaks, which were consistent with 2-AI (i.e. at 775, 844, 1020, 1205, and 1236 cm^{-1}) (Figure 3.4a). Although GC-MS analysis confirmed that both 2-AI and 5,6-MDAI were

present in this product. The product P11 was also confirmed to contain two NPS, MPA and 5,6-MDAI. Most peaks were consistent with 5,6-MDAI (i.e. at 713 and 1355 cm^{-1}). However, characteristic peaks for MPA were also visible at 1038 and 1433 cm^{-1} (Figure 3.4a). Interestingly, two products, P4 and P11, contained a combination of MDAI and MPA, which has been reported to have synergistic and/ or additive psychoactive effects [239].

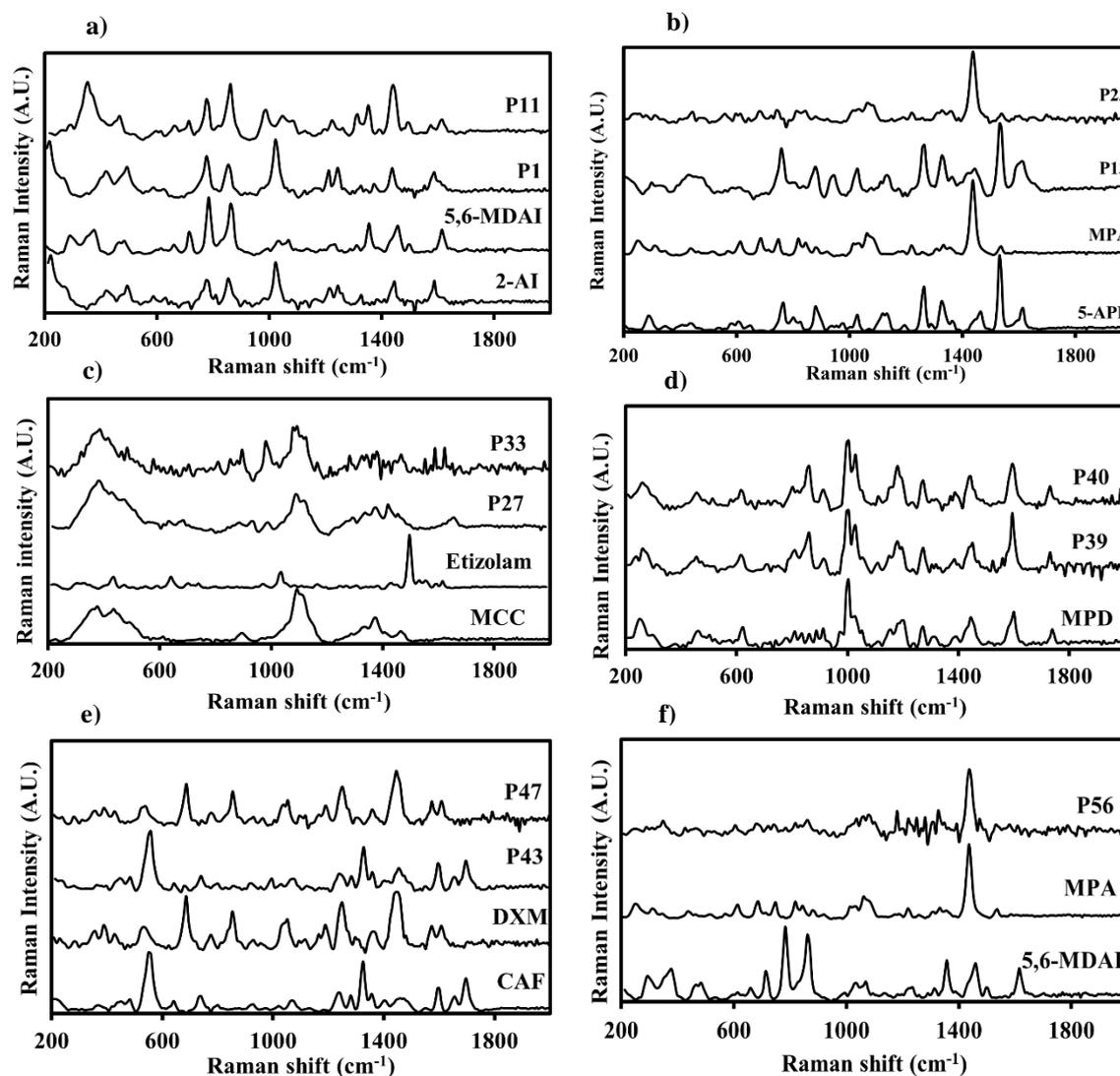


Figure 3. 4: Raman spectra of selected products and associated reference spectra from the aminoindane (a), arylalkylamine (b), benzodiazepine (c), piperidine & pyrrolidine (d), ‘other’ (e), and branded product (f) EMCDDA categories using a 1064 nm laser excitation wavelength. Product numbers and reference names are labelled on the spectra. The spectra were normalised to the maximum peak.

In summary, the use of a lower energy wavelength (i.e. 1064 nm) reduced fluorescence, which improved the S/N and, subsequently the matching identification of the NPS for 5 of the 12 aminoindane internet products, resulting in a total of 9 products with a correctly identified NPS.

3.3.2.2. Arylalkylamines

Fourteen arylalkylamines samples, purchased from the internet, were analysed using both Raman instruments and GC-MS (Table 3.5). These included the aminopropylbenzofurans (APB) isomer/analogues 5-APB, 6-APB and 1-(Benzofuran-5-yl)-*N*-methylpropan-2-amine (5-MAPB). Methyl aminopropamine (MPA) products, reported to exert stimulant effects [239], were also investigated. The GC-MS results indicated that all 14 products contained an arylalkylamine, while 13 products contained the arylalkylamine reported on the label claim. The arylalkylamine substances identified *via* GC-MS were 5-APB [240], 6-APB [240], 5-MAPB, MPA [238] and 5-APDB (1-(2,3-dihydro-1-benzofuran-5-yl)propan-2-amine). Other adulterants found included pyrovalerone [241], MDAI [226], DXM [227], CAF [224, 225] and 5-MeO-DALT (5-methoxy-*N,N*-diallyltryptamine) [219]. One sample, P23, showed one peak in the chromatogram that was not confirmed using MS. The GC-MS results were then used to compare with the Raman results.

When using the standard 785 nm laser, most samples (i.e. P13 - P17 and P21 - P26) correlated to MPA (94 ± 2 to 98.1 ± 0.2 % HQI) while three (i.e. P18 - P20) resulted in inconsistent correlations. Upon investigation of the spectra (see Figure 3.3a for the spectrum of P23), all samples showed intense fluorescent background with little to no distinct Raman bands indicating false positive correlations to the MPA signature. Therefore, after visual inspection it was suggested that no true positive correlations to an arylalkylamine were found in any of the 14 products using the 785 nm source.

The 14 products were then analysed using the 1064 nm source, where 10 products (i.e. P13 - 15, 19, 21 - 26) correlated to the NPS present in the sample with HQIs ranging from 50 ± 3 to 82.9 ± 0.7 %. Again, a reduction in fluorescence improved S/N in spectra and, thus matching identification of the NPS products. The spectra from the NPS products that correlated to MPA using an λ_{ex} of 1064 nm showed a distinct peak around 1442 cm^{-1} with no fluorescence indicating true positive correlations (see Figure 3.3b for the spectrum of P23). Figure 3.4b compares the spectra for P23 (MPA) and P13 (5-APB) and their associated reference spectra. By comparing the products and reference spectra, MPA was identified by a unique strong dominant peak, as previously described. For P13, the Raman spectrum showed peaks at 758, 1258, 1326, and 1530 cm^{-1} , which were consistent with 5-APB; the spectrum correlated to 5-APB with a % HQI of 80 ± 2 . There was a correlation to MCC (77.4 ± 0.2 %) for P18 even though 5-APB was present in the sample. This suggests that the cutting agent concentration

may potentially be incorporated in the mixture in a higher proportion compared to the active ingredient [128].

In the case of P20, a 60 ± 4 % correlation to 5-APB was found as no 5-MAPB signature was present in the Raman library. This demonstrates that a substance may be correlated to a similar structural analogue, which can assist with identifying suspect previously unseen NPS. Samples P16 and P17 resulted in ‘no match’ and failed data acquisition due to sample burning, even when using low power, as they were both of a dark colour [120].

Five products (i.e. P13, 17, 19, 21, and 23) were shown by GC-MS to contain a mixture of active ingredients, where the highest Raman correlation indicated the active ingredient either found in highest concentration or with the strongest Raman signal (Figure 3.3). For example, P23 was shown to contain MPA, 6-APB, 5-MeO-DALT and CAF by GC-MS, yet correlated to MPA at 78.2 ± 0.8 % using the 1064 nm Raman instrument. Figure 3.3b presents the Raman spectrum for this sample clearly showing a Raman band at ca. 1438 cm^{-1} for MPA using the 1064 nm source. The use of a lower energy laser wavelength reduced fluorescence from both the cutting agents incorporated in the mixture and coloured samples and, subsequently improved identification of the NPS ingredient for 11 of 14 arylalkylamines products.

3.2.2.3. Benzodiazepines

Nine benzodiazepine (BZD) samples, purchased from the internet, were analysed using both Raman instruments and GC-MS (Table 3.6). The products included two different BZDs, etizolam and pyrazolam. Benzodiazepines are pharmaceuticals, which exert depressant, anxiolytic, hypnotic and muscle relaxant effects. Due to the high level of excipients present in the tablet/ pellet formulations, it was necessary to remove the coating and increase the concentration of the analysis solution (ca. 45 mg mL^{-1}) to enable detection of the NPS via GC-MS. Both etizolam and pyrazolam were then identified in all nine samples (i.e. P27 – P35) using GC-MS. Etizolam was confirmed with a base peak of m/z 342 and a retention time of 23 min, whereas pyrazolam was confirmed with a base peak of m/z 353 and a retention time of 13.5 min.

Both P30 and P32 also contained unclaimed cannabinoids, which included 1-naphthalenyl[1-(4-penten-1-yl)-1H-indol-3-yl]-methanone (JWH-022), 1-naphthalenyl[1-(4-penten-1-yl)-1H-indol-3-yl]-methanone (AM-2201) and 1-Naphthoyl indole. The combination of cannabinoids with BZDs could be intentional as synthetic cannabinoids, such as the receptor agonist AM-

2201 (aka 'Black Mamba'), have been reported to induce tonic-clonic seizures [242], while BZDs have known anti-convulsant effects. Interestingly, P30 and P32 also shared three active ingredients despite being purchased from different websites and having a different appearance, suggesting a similar supply chain.

When using the standard 785 nm laser, no correlations to a BZD were found in any of the nine samples even those samples, which were confirmed to contain a BZD *via* GC-MS. Pyrazolam was not present in the Raman libraries but the products were included to investigate analogue selectivity. However, P30, P32 and P33 - P35 correlated to the cutting agents LAC (87 ± 5 to 91 ± 5 %) and MCC (83 ± 3 to 87 ± 2 %) respectively. High fluorescent background occurred in four samples (i.e. P27 – P29 and P31), which again resulted in false positive correlations to MPA (96.9 ± 0.4 to 98.5 ± 0.7 %) as MPA's characteristic peak at 1438 cm^{-1} was not visible in the spectra.

Using the 1064 nm source, no correlations to a BZD resulted for any of the nine samples. However, P30, P32 and P33 - P35 correlated to the same cutting agents as found with the 785 nm source with similar % HQIs. Due to reduced fluorescence, P27 – P29 correlated to MCC (70 ± 1 to 70 ± 9 %). For example, Figure 3.4c demonstrates this reduction in fluorescence and subsequent matching to MCC when comparing the two wavelength sources for P27. The subsequent identification of MCC is better illustrated in Figure 3.4c using P27 and P33 where the MCC signature is clearly visible when comparing it with an MCC reference (c.a., 397, 1094, and 1355 cm^{-1}). For P27, the strong signature peak of etizolam at 1496 cm^{-1} (Figure 3.1) is clearly not visible. As mentioned above both MCC and LAC do not readily dissolve in methanol and have low volatility; hence, it is often not detected with GC-MS analysis, but may still be present in the sample. Microcrystalline cellulose is a common diluent used in pharmaceutical tablets and LAC is commonly used in direct compression tableting applications and is also used as a tablet filler and binder. As most of the BZD products were in tablet or pellet forms, the presence of these excipients in high concentrations is likely. One sample P31 correlated to a factory library signature labelled as phosphorus at 60 ± 3 %.

Products 27, 28 and 29 were purchased from three different websites (Table 3.2); however, the three batches, all turquoise in colour, have been shown to contain MCC and etizolam using the 1064 nm Raman spectrometer and GC-MS, respectively. This may indicate that despite being sold on different websites and compressed with different tablet dies, the powder mix could have originated from the same supplier.

Table 3. 6: Results from the analysis of benzodiazepine, piperidine & pyrrolidine, plants & extracts and ‘other’ internet products using two handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) and GC-MS¹

Product No.	Product name	Handheld Raman 785 nm		Handheld Raman 1064 nm		GC-MS		
		ID	% HQI	ID	% HQI	RT (min)	Base Peak (m/z)	MS ID
Benzodiazepines								
27	Etizolam	MPA ³	96.9 ± 0.4	MCC	70 ± 1	23.1	342	Etizolam
28	Etizolam	MPA ³	97.49 ± 0.05	MCC	70 ± 4	23.2	342	Etizolam
29	Etizolam	MPA ³	96.9 ± 0.7	MCC	70 ± 9	23.2	342	Etizolam
30	Etizolam	LAC	87 ± 5	LAC	80 ± 7	16.9	339	JWH-022 ²
						19.3	359	AM-2201 ²
						22.9	270	1-NI ^{2,4}
						23.2	342	Etizolam
31	Etizolam	MPA ³	98.5 ± 0.7	Phosphorous	60 ± 3	23.1	342	Etizolam
32	Etizolam	LAC	91 ± 5	LAC	88.6 ± 2.3	16.7	339	JWH-022 ²
						21.8	268	1-N-2-MI ^{2,4}
						22.9	270	1-NI ^{2,4}
						23.2	342	Etizolam
33	Pyrazolam	MCC	86 ± 2	MCC	80 ± 6	13.5	353	Pyrazolam ²
34	Pyrazolam	MCC	87 ± 2	MCC	90 ± 2	13.5	353	Pyrazolam ²
35	Pyrazolam	MCC	83 ± 3	MCC	81 ± 3	13.5	353	Pyrazolam ²
Piperidines & pyrrolidines								
36	Ethyl phenidate	MPD	64 ± 4	MPD	76.6 ± 0.4	14.0	84	MPD
37	Ethyl phenidate	MPD	65 ± 2	MPD	76.8 ± 0.4	14.0	84	MPD
38	Ethyl phenidate	Inconsistent correlations		No match		19.4	359	AM-2201 ²
39	Ethyl phenidate	Cetyl-pyridinium chloride	61 ± 20	MPD	80 ± 2	14.0	84	MPD
40	Ethyl phenidate	MPD	63 ± 5	MPD	76.6 ± 0.2	14.0	84	MPD
41	Ethyl phenidate	Inconsistent correlations		No match		19.4	359	AM-2201 ²
Plants & Extracts								
42	LSA Morning Glory Seeds	MPA ³	96 ± 2	No match		No match		
‘Other’								
43	DXM	Talc	69 ± 3	CAF	80 ± 10	14.4	194	CAF
						14.8	190	MXE ²
44	DXM	Inconsistent correlations		DXM	60 ± 9	16.6	271	DXM
45	DXM	DXM	57 ± 1	DXM	64.0 ± 0.4	16.6	271	DXM
46	DXM	DXM	59 ± 6	DXM	63.8 ± 0.4	16.6	271	DXM
47	DXM	DXM	84.1 ± 0.2	DXM	90 ± 4	16.6	271	DXM

¹ID: identification; RT: retention time; NC: not confirmed.

²Raman spectrum of substance not present in both Raman libraries.

³Suspected false positive.

⁴1-NI: 1-Naphthoyl indole; 1-N-2-MI: 1-Naphthoyl-2-methyl indole.

In summary, no NPS ingredients were identified in the BZD products using the instruments matching algorithms and the visual examination of Raman spectra. The products were largely composed of excipients used for tablet and pellet manufacture (i.e. MCC and LAC), which was also observed in a recent study using a benchtop Raman spectrometer [140]. A comparison of

the spectra from the 785 and 1064 nm showed a reduction in fluorescence, which did improve identification of the cutting agents in six of the nine products. This indicates that for these BZDs and perhaps other NPS in tablet form, identification of the active ingredient can be challenging (i.e. resulting in false negatives), as the amount of active ingredient may significantly be lower relative to the excipients. The use of spectral subtraction is a possible tool for mixtures that may reduce these types of false negatives; it could be used to improve the identification of NPS with low content in the presence of larger amounts of cutting agents that results in a larger Raman signal than the NPS [140]. However, careful spectral pre-processing should be made in order to prevent the loss of Raman bands. An alternative method is the use of ‘polynomial curve fitting’ [243]. Nevertheless, it largely depends on users’ expertise.

3.3.2.4. Piperidines & pyrrolidines

Six ethylphenidate (piperidine) samples, purchased from the internet, were analysed using both Raman instruments and GC-MS (Table 3.6). Ethylphenidate (EPD) is a synthetic analogue of MPD and was encountered in the UK and Europe for the first time in 2011 [75]. The GC-MS analysis confirmed the presence of MPD [13] in four products (i.e. P36, 37, 39 and 40) with a base peak indicating the piperidinium ion at m/z 84 and the tropylium ion at m/z 91. However, P38 and P41 correlated to unclaimed AM-2201 using GC-MS. Until recently, EPD was uncontrolled in the UK, perhaps a reason why the controlled substance MPD was substituted for EPD in these samples. These products are an example of how NPS may be marketed as ‘legal’ products, where in fact they may contain a controlled drug. Thus, only MPD was added to the Raman libraries as EPD was not identified *via* GC-MS. Products 38 and 41 were purchased from the same website and shared a similar active adulterant (i.e. AM-2201).

When using the standard 785 nm laser, P36, 37 and 40 correlated to MPD (63 ± 5 to 65 ± 2 %). Products 38 and 41 showed inconsistent Raman responses due to high fluorescent backgrounds. The Raman spectra of P39 correlated to cetylpyridinium chloride (61 ± 20 %), where very weak Raman signals were observed on top of a high fluorescent background (Figure 3.4a), again suggesting a false positive result.

Using the 1064 nm source, four of the six samples (i.e. P36, 37, 39 and 40) correlated to MPD, the analogue of EPD, with a % HQI of 76.6 ± 0.2 to 80 ± 2 . Figure 3.3b displays the Raman spectra of P39 showing distinct peaks at 784, 1029 and 1588 cm^{-1} which correspond to Raman bands for the MPD signature, not cetylpyridinium chloride. Figure 3.4d confirms this for P39

and P40 as an example. Products 38 and 41 resulted in ‘no match’ despite showing Raman bands at 511, 668, 775, 1012, 1370, 1516 and 1622 cm^{-1} as both Raman instruments did not have a library signature for AM-2201, but these peaks do correspond to AM-2201 reference spectra (see Chapter 5).

In summary, the 785 nm source successfully identified the NPS ingredient in three products, while use of the 1064 nm source reduced fluorescence, improved algorithm matching and subsequent NPS identification for one of the three remaining products in addition to evidence for the presence of AM-2201 in two products. Interestingly, no products contained EPD, but four were adulterated with the schedule II substance MPD.

3.3.2.5. Plants & Extracts

LSA morning glory seeds (P42) were purchased from the internet and analysed using both Raman instruments and GC-MS. Morning glory seeds are known to be sacred seeds, originally used by some Mexican Indian tribes. They contain lysergic acid amide (LSA), which is the non-alkylated amide analogue of the schedule I controlled lysergic acid diethyl amide (LSD) [244]. As these samples were seeds, before analysis they were ground using an agate mortar and pestle. LSA morning glory seeds were characterised with the black shell and grey content. Using GC-MS, the content of the seeds was not identified as no chromatographic peaks were observed. When using the standard 785 nm laser, P42 correlated to MPA ($96 \pm 2\%$) as seen with other products using the 785 nm laser. However, this is again a potential false positive result as the content was not confirmed using GC-MS. Using the 1064 nm source, no match was found and the sample was burned upon analysis using all methods. Seed samples such as these, which are dark in colour, are particularly problematic when analysing *via* Raman even with the 1064 nm wavelength.

3.3.2.6. ‘Other’

Five DXM samples, purchased from the internet, were analysed using both Raman instruments and GC-MS (Table 3.6). In addition to its classification by the EMCDDA, DXM has also been classified as ‘other’ in the UNODC (2014) report [245]. DXM is a non-opioid anti-tussive drug. It is the d-isomer of the codeine analogue levorphanol, which inhibits the re-uptake of serotonin at therapeutic doses [216]. The GC-MS analysis identified DXM [227] in four of the five products (i.e. P44 – P47).

However, the GC-MS analysis of P43 showed the presence of an unclaimed NPS i.e. methoxetamine (MXE) and CAF. This means product 43 contained two stimulants (i.e. CAF and MXE); stimulant cocktails such as P43 have been shown to cause cardiotoxic symptoms in previous studies [246].

When using the standard 785 nm laser, correlations to DXM were found in three of the five samples (P45 – P47) with % HQIs ranging from 57 ± 1 to 84.1 ± 0.2 %. Product 43 correlated to TAL (69 ± 3 %) (Figure 3.3a) and P44 resulted in inconsistent correlations; these spectra displayed Raman signals on top of slightly fluorescent backgrounds.

Using the 1064 nm source, four of the five samples (i.e. P44 - P47) correlated to the NPS on the label (DXM) with an HQI of 60 ± 9 to 90 ± 4 %, while P43 correlated to the unclaimed adulterant CAF at a % HQI of 80 ± 10 %. A reduction in fluorescence was most noticeable for P43 (Figure 3.3b), which correlated to the unclaimed adulterant CAF at a % HQI of 80 ± 10 %. Figure 3.4e shows an example of two products' spectra, P43 and P47, and their similarity to the highest correlation signature spectrum. The spectra for P43 and CAF are very similar with characteristic Raman peaks at 549, 1325, 1600 and 1690 cm^{-1} even though MXE was also identified using GC-MS; and the spectra for P47 and DXM are also very similar with characteristic peaks at 686, 852, 1242, and 1436 cm^{-1} .

In summary, the 785 nm source successfully identified the NPS ingredient in three products, while use of the 1064 nm source reduced fluorescence, improved NPS identification for one of the two remaining products using the matching algorithm and visual inspection of the spectra.

3.3.2.7. Branded products

Thirteen branded products were analysed using both Raman instruments and GC-MS (Table 3.7). Internet products are often branded with names such as blast, bliss, bloom and blow [38, 152]. Although these products are marketed with brand names, all branded products in this study did have a label claim stating ingredients except for P48 and P59. Six products (i.e. P49, P52 – P56) named an NPS on the label such as AI, 5,6-MDAI and MPA, nine products (i.e. P49 – 55, 57 and 60) named CAF on the label, and ten products (i.e. P49 – 55, 57, 58 and 60) named herbal extracts and dietary supplements on the label.

The analysis using GC-MS identified seven different NPS substances. These were MPD in P48 and P59 (base peak m/z 84, RT 14.1 min); 2-AI in P49 - P55 (base peak m/z 133, RT 9.4 min); MDAI in P56 (base peak m/z 160, RT 12.7 min); MPA in P56 and P59 (base peak m/z 59, RT

8.8 min); 6-APB in P59 (base peak m/z 44, RT 11.7 min); 5-MeO-DALT in P59 (base peak m/z 110, RT 17.3 min); and 5-APDB in P60 (base peak m/z 44, RT 12.3 min) (Table 3.7). Interestingly, both P23 labelled as MPA and P59 labelled as 'Route 56' were purchased from the same website and shared four active ingredients (i.e. MPA, 6-APB, CAF and 5-MeO-DALT), again suggesting a similar supply chain.

When using the standard 785 nm laser, correlations to a NPS substance were found in 8 out of 13 samples, even though all but one sample was confirmed to contain an NPS *via* GC-MS. However, for seven of these products (i.e. P49 - 55 and 57 - 60) the spectra showed very high fluorescent backgrounds resulting in false positive correlations to MPA (92.5 ± 0.4 to 99.1 ± 0.1 %) (see Figure 3.3a for the spectrum of P53). Products 57 - 60 also showed a high level of fluorescence using the 785 nm source and resulted in inconsistent correlations. The calculation of the HQI algorithm has been shown to be affected by background fluorescence in unknown spectra, depicting slope and offset as differences from library spectra [228]. Pink panthers (P56) correlated to 5,6-MDAI with a % HQI of 80 ± 7 , whereas the product Blow (P48) correlated to BEN with a % HQI of 76 ± 10 .

Using the 1064 nm source, two additional NPS, MPA (80 ± 2 %) and delta 9-tetrahydrocannabinol (THC) (59.3 ± 0.8 %), were identified in P56 and P60. The presence of MPA was confirmed using GC-MS, however THC could potentially be a false positive result, since THC was not identified by the MS libraries. The presence of the adulterants BEN (% HQI 90 ± 7) and CAF (% HQI 54.2 ± 0.2) was confirmed in P48 and P54 using GC-MS. Products 55, 57 and 59 correlated to phosphorous (70 ± 3 %), L-TYR (54.7 ± 0.7 %) and MCC (60 ± 6 %), respectively.

Figure 3.4f shows the improved Raman spectra of P53 after using the 1064 nm source, where clear distinct Raman bands are visible. The improved Raman spectra still resulted in 'no match' using the algorithm, however characteristic Raman peaks for CAF (e.g. 549 and 1322 cm^{-1}) and 2-AI (e.g. 775, 844, and 1030 cm^{-1}) were clearly visible. The improved Raman spectra still resulted in 'no match', but further use of a 'mixtures algorithm' may improve identification. Jones et al. evaluated Raman spectra of NPS mixtures by subtracting the spectra of pure substances sequentially after identifying the substances using a Raman microscope [140].

Interestingly, the product Pink Panthers (P56), which was confirmed to contain both 5,6-MDAI and MPA with GC-MS analysis correlated to 5,6-MDAI using the 785 nm source and to MPA using the 1064 nm source. This may be the result of mixture heterogeneity inherent to branded

products, despite considerable efforts to vortex mix. Figure 3.4f shows peaks in the Raman spectrum for P56 corresponding to MPA (i.e. 595, 677, 810, 1052, and 1436 cm^{-1}). Peaks corresponding to 5,6-MDAI were not observed.

In summary, the identification of branded NPS products was very challenging using both handheld Raman instruments resulting in many inconsistent correlations, ‘no match’ finds, and false positive matches to MPA. This was mainly due to high fluorescence signals using the 785 nm, which was significantly reduced using the 1064 nm source, and the chemical complexity of the samples. Since peaks were observed with the 1064 nm source and were consistent with NPS reference spectra, ‘mixtures algorithm’ or spectral subtraction would be alternative methods to improve the identification of samples such as NPS internet samples.

Table 3. 7: Results from the analysis of branded internet products using two handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm) and GC-MS¹

Product No.	Product name	Handheld Raman 785 nm		Handheld Raman 1064 nm		GC-MS		
		ID	% HQI	ID	% HQI	RT (min)	Base Peak (m/z)	MS ID
Branded products								
48	Blow	BEN	76 ± 10	BEN	90 ± 7	12.4 14.1	120 84	BEN MPD
49	Blurberry	MPA ³	99.1 ± 0.1	No match		9.5 14.4	133 194	2-AI CAF
50	High beams	MPA ³	97.7 ± 0.1	No match		9.4 14.4	133 194	2-AI CAF
51	Magic Beans	MPA ³	92.5 ± 0.4	Inconsistent correlations		9.4 14.4	133 194	2-AI CAF
52	Pink Champagnes	MPA ³	98.0 ± 0.4	Inconsistent correlations		9.4 14.4	133 194	2-AI CAF
53	Pink Champagnes	MPA ³	98.4 ± 0.2	No match		9.2 14.4	133 194	2-AI CAF
54	Pink Champagnes	MPA ³	98.4 ± 0.1	CAF	54.2 ± 0.2	9.4 14.4	133 194	2-AI CAF
55	Pink Champagnes	MPA ³	98.3 ± 0.2	Phosphorous	70 ± 3	9.3 14.5	133 194	2-AI CAF
56	Pink panthers	5,6-MDAI	80 ± 7	MPA	80 ± 2	8.9 12.7	58 160	MPA 5,6-MDAI
57	Punk plus	Inconsistent correlations		L-TYR	54.7 ± 0.7	10.7 14.4	106 194	NIA CAF
58	Recovery	Inconsistent correlations		Data acquisition failed		7.1	71	No match
59	Route 56	Inconsistent correlations		MCC	60 ± 6	8.8 11.6 14.0 14.4 17.3	58 44 84 194 110	MPA 6-APB MPD CAF 5-MeO-DALT ²
60	White Pearls	Inconsistent correlations		Delta 9-THC	59.3 ± 0.8	12.3 14.4	44 194	5-APDB ² CAF

¹ID: identification; RT: retention time; NC: not confirmed.

²Raman spectrum of substance not present in both Raman libraries.

³Suspected false positive.

3.4. Conclusions

In this Chapter, the use of handheld Raman spectroscopy showed promise for the in-field identification of NPS in a wide range of NPS internet samples. Results demonstrated that a 1064 nm laser source significantly reduced background fluorescence of NPS products and subsequently improved identification using an in-built ‘first pass’ matching algorithm. In this study, handheld Raman spectroscopy with two laser excitation sources was used to identify NPS in internet products using a ‘first pass’ matching algorithm (% HQI) as well as visual inspection of Raman spectra.

The matching algorithm approach successfully identified an NPS in 29 out of 60 (48 %) diverse and chemically complex internet products using a 1064 nm laser source. An overview of the results is presented in Table 3.8. An increase in the laser excitation wavelength from 785 to 1064 nm improved positive NPS identification (i.e. from 11 to 29 substances). These improvements were mainly the result of reduced fluorescence, most likely originating from cutting agents and coloured constituents in the products. Correlations between the internet products with the NPS signatures, that were confirmed with GC-MS, ranged from 57.0 to 84.1 % with RSDs < 10 % using the 785 nm source and from 60.0 to 91.3 % with RSDs < 7 % using the 1064 nm source. Thus, reduced % HQI thresholds may be required when monitoring NPS products in the field.

A higher number of false positives and false negative were observed when using the 785 nm source, again resulting mainly from the fluorescent background produced by these samples. False negatives observed for both wavelength sources were also due to low NPS concentration, absence from the instrument’s library and/ or high chemical complexity of the product. For example, no NPS was identified and no etizolam Raman bands were observed for the etizolam tablets and pellets as they were largely composed of common excipients with a relatively low etizolam concentration. Chemically complex samples, such as some of the ‘branded products’, did not correlate to an NPS signature but did show marked improvement in the Raman spectra and characteristic Raman bands upon using the 1064 nm source. In these cases, spectral subtraction could be useful to further assist identification. Further work should focus on optimisation of instrument parameters, spectral pre-processing and investigation of ‘mixtures algorithms’ with improved NPS libraries.

Table 3. 8: A summary of the results obtained for NPS identification using the handheld Raman spectrometers ($\lambda_{\text{ex}} = 785$ and 1064 nm)¹

Category	λ_{ex} 785 nm	λ_{ex} 1064 nm
Identification of NPS	11	29
Identification of adulterant	2	4
Identification of cutting agent ²	7	14
Fluorescence	38	0
Inconsistent correlation	10	2
No match	2	7
False positive for an NPS	28	1
False negative for an NPS	46	28

¹Raman spectrometers were used with set parameters as stated in the method Section.

²Cutting agents were not confirmed via GC-MS.

4. Development of a pre-processing protocol for Raman spectra of NPS-related powders

4.1. Introduction

This Chapter outlines a step-by-step protocol for the pre-processing of Raman spectra of mixtures containing adulterants and cutting agents commonly incorporated in mixtures of new psychoactive substances (NPS). Variations in Raman spectra resulting from instrument artefacts, sample, environment and/ or analysis effects are unavoidable, which may hamper the identification of NPS using Raman spectroscopy. Therefore, a need remains for a universal protocol for the pre-processing of Raman spectra of NPS-related substances that could assist in extracting the maximum chemical information from spectra and, subsequently improve the in-field detection of NPS by first-responders.

For frontline emergency staff and law enforcement agents, detection of NPS using a ‘first pass’ library-matching algorithm is critical. However, in-built algorithms have been shown to be limited in their classifications of the NPS internet products due to sample complexity. Swamping by fluorescence signals, overlapping analyte-excipient bands, sample heterogeneity and variable Raman activity of constituents are potential challenges in NPS products (see Chapters 2 and 3 and Assi et al. [128]).

Therefore, in this study, exploratory multivariate analysis (MVA) such as principal components analysis (PCA) was evaluated for the feasibility of classifying simple solid-form mixtures containing common adulterants and cutting agents. Coupling of PCA with Raman spectroscopy has been shown to strengthen Raman analysis by enabling the exploration of latent variables in complex Raman data [184] e.g. Raman spectral data of illicit drug mixtures [136, 154, 192]. However, pre-processing of Raman spectral data is a prerequisite to the application of PCA.

In this Chapter, the aims are to develop a step-by-step protocol for the pre-processing of Raman spectral data for solid drug powders and evaluate the effect of pre-processing of Raman data on the classification of known pure substances and mixtures of known composition using PCA. Pre-processing of Raman spectra has been shown to improve the classification of drugs of abuse using MVA [95, 176, 184]. A few studies have addressed the challenges encountered in Raman analysis of NPS [138, 140]. In this Chapter, the optimal combined pre-processing

sequence was performed on a selected Raman dataset for the application of PCA. This work also investigated the feasibility of using PCA of pre-processed Raman data as a tool to explore substance identifiers, which may, in turn improve the classification of drug mixtures.

4.2. Materials and Methods

4.2.1. Reagents and chemicals

In this initial work, a pre-processing method was developed using selected adulterants (i.e. benzocaine and caffeine) and cutting agents (i.e. creatine and lactose) that are commonly incorporated in NPS mixtures [208, 209]. Five reference standards were commercially obtained and were used for analysis as supplied (Table 4.1).

Table 4. 1: Adulterants and cutting agents used in Raman and PCA analysis

Materials	Batch Number	Supplier	Abbreviations
Benzocaine	100M0213V	Sigma Aldrich, Dorset, UK	BEN
Benzocaine	SLBB1067V	Sigma Aldrich (Dorset, UK)	BEN
Caffeine (anhydrous)	1428211V	Fluka Analytical, Sigma Aldrich, Dorset, UK	CAF
Creatine monohydrate	SLBH1411V	Sigma Aldrich, Dorset, UK	CRE
Lactose	893920	Fisher Scientific, Loughborough, UK	LAC

4.2.2. Sample preparation

Samples were analysed in their pure powder forms (white powders) and simulated ternary mixtures (Tables 4.2 – 4.5). Pure reference samples (i.e. CAF and BEN) and ternary mixtures (CAF_CRE_LAC or CCL) were weighed, ground using an agar mortar and pestle, then placed on microscope glass slides (75 x 25 x 1 mm) wrapped with aluminium foil. The microscope stage was moved in the x, y and z positions, then the focal distance was re-adjusted between measurements. Samples were mixed on the microscope slides using a stainless steel micro spatula for one minute, tapped and flattened prior to Raman analysis. In this preliminary study, four datasets are discussed in Section 4.2.2.1. to 4.2.2.4.

4.2.2.1. Calibration dataset (Set ‘1’)

‘Set 1’ consists of the samples employed to develop the pre-processing method and to calculate the PCA model’s principal components (PCs) (Table 4.2).

Table 4. 2: Summary of samples included in the calibration dataset (Set ‘1’)

Sample name	No. of samples	No. of analysis measurements/ sample	Sample weight (mg)	Sample labels	Batch No.
BEN	1	3	30	B1 to B3	100M0213V
CAF	6	1	30	C1 to C6	1428211
CCL	6	1	10:10:10	CCL1 to CCL6	1428211 SLBH1411V 893920

4.2.2.2. New calibration dataset (Set ‘2’)

Set ‘2’ consists of the original calibration set (Set ‘1’) after the rejection of spectra with errors (i.e. B1) and ‘zapping’ of cosmic rays from C1 spectrum, which is described in Section 4.3.2.1. (Table 4.3).

Table 4. 3: Summary of samples included in the new calibration dataset (Set ‘2’)

Sample name	No. of samples	No. of analysis measurements/ sample	Sample weight (mg)	Sample labels	Batch No.
BEN	1	2	30	B2 and B3	100M0213V
CAF	5	1	30	C2 to C6	1428211
CCL	6	1	10:10:10	CCL1 to CCL6	1428211 SLBH1411V 893920

4.2.2.3. Validation dataset (Set ‘3’)

Set ‘3’ includes a subset that was not employed to develop the pre-processing method or to calculate the model’s PCs (Table 4.4).

Table 4. 4: Summary of samples included in the validation dataset (Set ‘3’)

Sample name	No. of samples	No. of analysis measurements/ sample	Sample weight (mg)	Sample labels	Batch No.
BEN	1	3	30	B4 to B6	SLBB1067V
BEN	1	3	30	B7 to B9	100M0213V
CAF	3	1	30	C7 to C9	1428211

4.2.2.4. Selected-bands calibration dataset (Set ‘4’)

Set ‘4’ consists of selected-bands from the new calibration Set ‘2’. The data matrix contained the following bands of wavenumbers (cm^{-1}): [2965-2975]; [2950-2960]; [1690-1705]; [1590-1605]; [1320-1335]; [735-745] and [550-560]. This is described in details in Section 4.3.6.

4.2.3. Raman Spectroscopy

Spectra were acquired using a benchtop Renishaw inViaTM Raman microscope (Serial No. 29U091) employing a laser with excitation wavelength (λ_{ex}) of 785 nm and a 10 % laser output

power of ca. 20 mW (10.6 mW at sample using a x50 objective lens), a high sensitivity ultra-low noise RenCam CCD detector, held in a vacuum of -70 °C, an ultra-high precision diffraction grating of 1200 lines/ mm. Parameters employed for both calibration of the Raman instrument and measurements of all samples were a 10 sec. exposure time, 1 accumulation, 10 % laser power, x50 objective lens (spot radius = 638 nm). The instrument was calibrated regularly prior to analysis using a static silicon wafer reference. Calibration was carried out in order to ensure wavenumber accuracy and adjusting the wavenumber (cm^{-1}) to $520 \pm 0.6 \text{ cm}^{-1}$. One spectrum was acquired for each pure reference standard or mixture over a spectral range of 3200 - 100 cm^{-1} . Spectral data were exported as WiRE (Windows-based Raman Environment) version 3.4 (.wxd) files, which were converted to text file format (.txt), then saved as Microsoft Excel 2013 files (.xlsx), imported into the Unscrambler® X 10.3 software for pre-processing and chemometric analysis.

4.2.4. Pre-processing method

In this study, the step-wise process employed for developing a pre-processing method for Raman spectra is summarised in Figure 4.1.

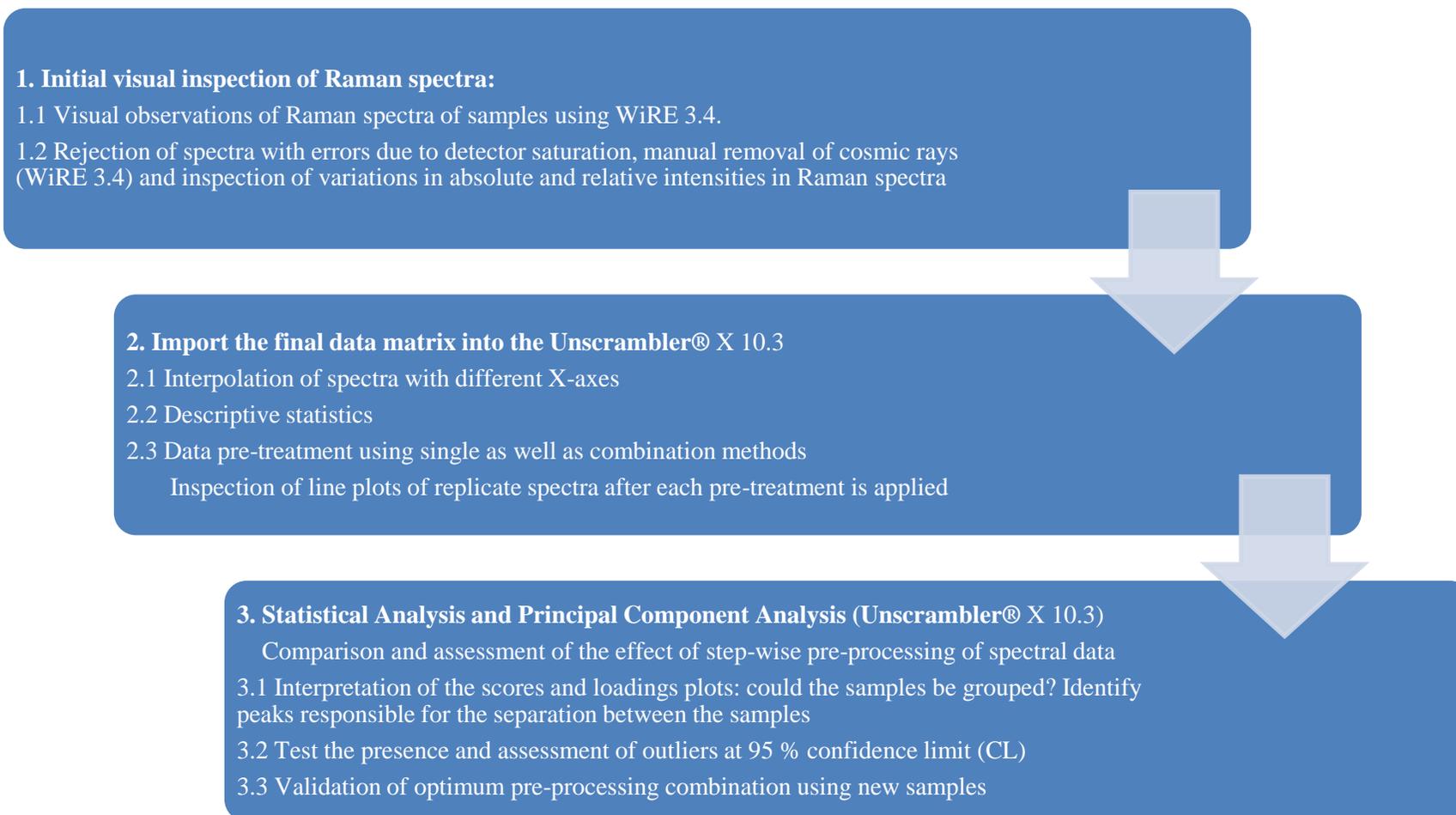


Figure 4. 1: A flow chart outlining the initial step-wise pre-processing of Raman spectra

4.2.5. Principal Components Analysis (PCA) (Unscrambler® X 10.3)

PCA was performed using the Unscrambler® X 10.3 (CAMO PROCESS AS, Oslo, Norway) software. PCA was calculated using the NIPALS (Non-linear iterative projections by alternating least squares) algorithm because the Raman data matrix may contain missing values as well as large amount of variables (i.e. 3777 wavenumbers or data points) [172]. In this study, the plots generated by PCA models (i.e. explained variance, scores, loadings, influence and leverage plots (Figure 4.2)) were examined and compared based on their performance (e.g. correct classification, reduction of noise and redundant variables) and assessed in terms of the problem in question (i.e. efficient classification of known pure substances/ mixtures and the presence of outliers) [172]. Figure 4.2 shows an example of PCA plots for Raman spectra of adulterants and cutting agents.

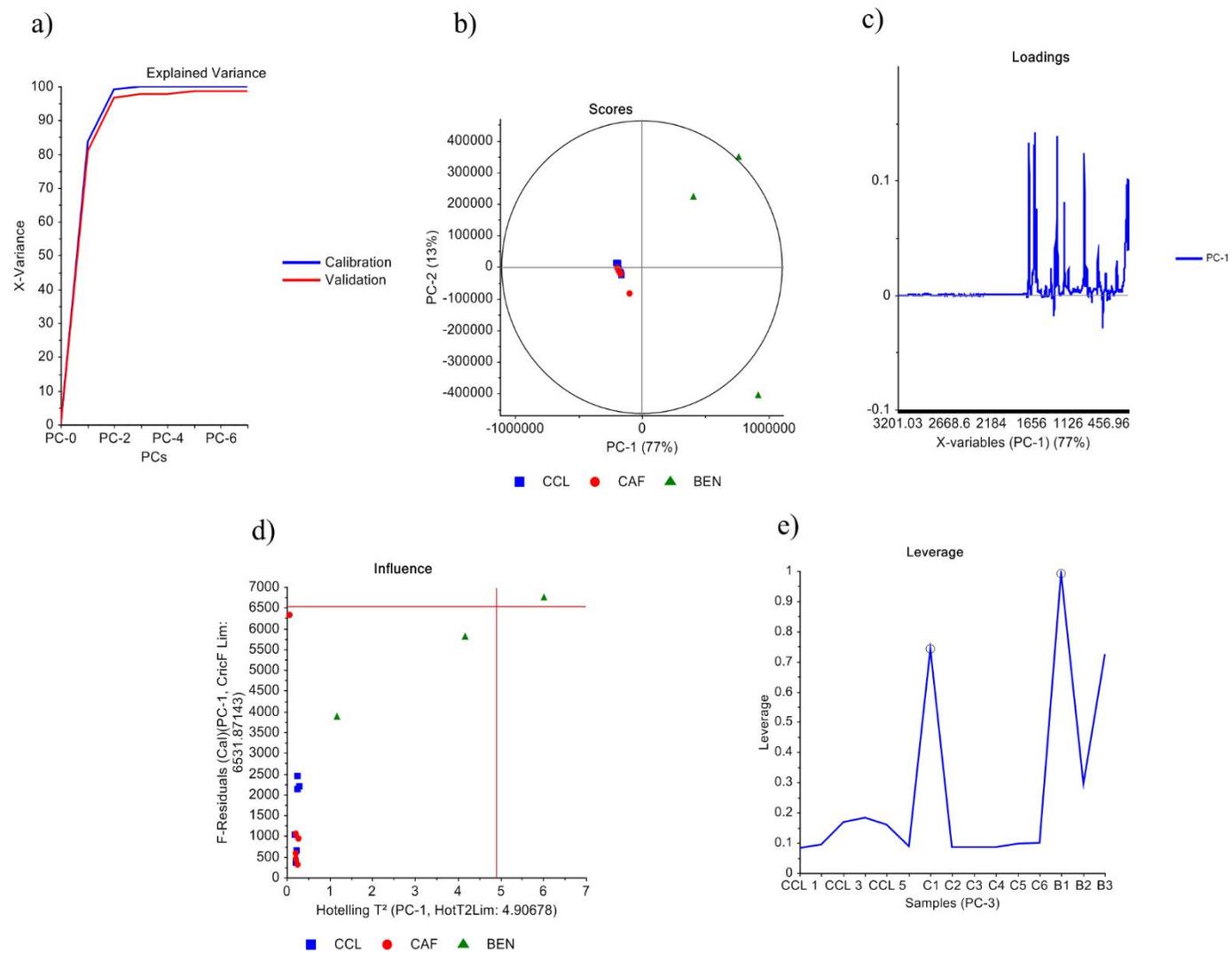


Figure 4. 2: An example of PCA plots for Raman spectra of cutting agents and adulterants (Unscrambler® X 10.3): a) the explained variance plot, b) the 2D-scores plot, c) the loadings plot; d) the influence plot and e) the leverage plot (outliers are circled).

In general, the PCA explained variance plot (Figure 4.2a) describes both calibrated and full cross-validated variance. The former is the variance in the calibrated dataset as explained by the PCs and the latter iteratively takes a sample out of the calibration set and tests remaining samples against the calibration model. Close values between calibrated and validated variance, in the explained variance plot, indicates that the model is well explained by the PCs.

In the two-dimensional (2D) scores plot for PC1/PC2 (Figure 4.2b), each score represents a spectrum. The scores plot profiles the differences between spectra according to their patterns along the PCs with respect to the variables (i.e. data points) with high loadings (i.e. contribution to the PC). The scores plot describes the relative significance of the PCs with respect to the variance they explain. PC1 explains the largest variance in the model, PC2 explains the second largest variance, etc. The Hotelling T^2 ellipse in the scores plot represents the samples within a designated critical limit. In this study, the 95% confidence limit (CL) is employed as the statistical critical limit. The loadings plot (Figure 4.2c) describes the correlation between the variables (i.e. data points) and the PCs, such that it describes the important Raman peaks that explain the variance by a specified PC. The scores and loadings plots are often interpreted simultaneously such that the relationship between the samples and the loadings is dependent on the location of the scores (spectra) in the scores plot.

The influence plot with Hotelling (T^2) (Figure 4.2d) describes whether samples/ spectra fit the model, whether they are influential on the model or whether they could potentially be outliers. It compares the distance of the spectra from the mean along the PCs against critical limits using the F-test. The leverage plot (Figure 4.2e) confirms whether potential outliers previously identified in both scores (outside of the 95 % CL) and influence plots (high leverage/ Hotelling T^2 and high F-residual limits) are true outliers [172].

4.3. Results and Discussion

4.3.1. Selection of adulterants and cutting agents

A review was conducted to select adulterants and cutting agents associated with NPS from the literature, forensic and government reports (2008 - 2014). NPS products including ‘NRG’ and ‘bath salt’ products seized from ‘headshops’, purchased over the internet for research purposes or reported by the EMCDDA have been shown to contain BEN, CAF, CRE and LAC *inter alia* adulterants and cutting agents [73, 79, 132, 217, 247-253]. In this pilot study, BEN, CAF, CRE and LAC (Table 2.1 and Appendix C Figure A4.1) were selected because they are commonly

found in NPS products and because they have different Raman activities. For example, BEN is a very strong Raman scatterer, whereas LAC is a relatively weak Raman scatterer. A small number of samples was employed to enable the examination of the effects of the step-wise pre-processing of Raman data on the classification of pure substances and mixtures. Our previous work has shown that CAF usually has the largest Raman contribution in the spectra of simulated binary and ternary mixtures (Assi et al. [128] and Chapter 2). Therefore CAF was selected as the ‘analyte’ in the mixtures to test its effect on the classification of pure CAF versus CAF mixtures (i.e. CCL mixtures). The Raman spectra of the high energy amorphous LAC have been reported to have broad peaks with some baseline offsets [254, 255]. Therefore LAC was selected as the cutting agent in this study for its potential challenging Raman spectra.

4.3.2. Step-wise process employed for developing a pre-processing protocol for Raman spectra of NPS-related compounds

4.3.2.1. Visual inspection of Raman spectra

Initial visual inspection of Raman spectra was conducted to evaluate the extent by which instrument artefacts influence the quality of replicate spectra and to establish the necessity and types of possible pre-processing methods that would need to be undertaken (Figure 4.1). As discussed above, Raman spectra often show variations in baselines, absolute intensities and noise interferences that could be attributed to 1) instrument artefacts such as noise (CCD, background noise, etc.), changes in laser power and optics geometry, fluctuations in spectrometer performance and temperature; 2) analysis effects such as variations in powder opacity and focal distance etc.; 3) sample effects such as amount and vibrational frequency of Raman active molecules, presence of fluorescing impurities etc. and 4) environmental effects (e.g. ambient light, cosmic rays etc.) [188]. Therefore prior to analysis, pre-processing of Raman data is often indispensable to remove variations that are not pertinent to the samples and to isolate the Raman data from noise.

Visual inspection of set ‘1’ showed baseline offsets (i.e. potential additive scatter effects), a significant variation in the overall peak absolute intensities and amplification of intensities over certain bands (i.e. potential multiplicative scatter effects (MSEs)). It also showed the regions with important chemical information. Spectra were examined individually and in groups to see whether there is a single spectrum issue or a sample trend (Figure 4.3). CAF (Figure 4.3a) and BEN replicate spectra showed variations in absolute intensities and variations in regions, where a cosmic ray was present (e.g. for C1) or where the detector was saturated (e.g. for B1).

Otherwise, all replicate spectra for pure substances were overlaid with most peaks present in all replicates. Overlaid spectra for CCL mixtures showed different visible peaks between CCL 1, 2 and 6 against CCL 3 to 5 (Figure 4.3b).

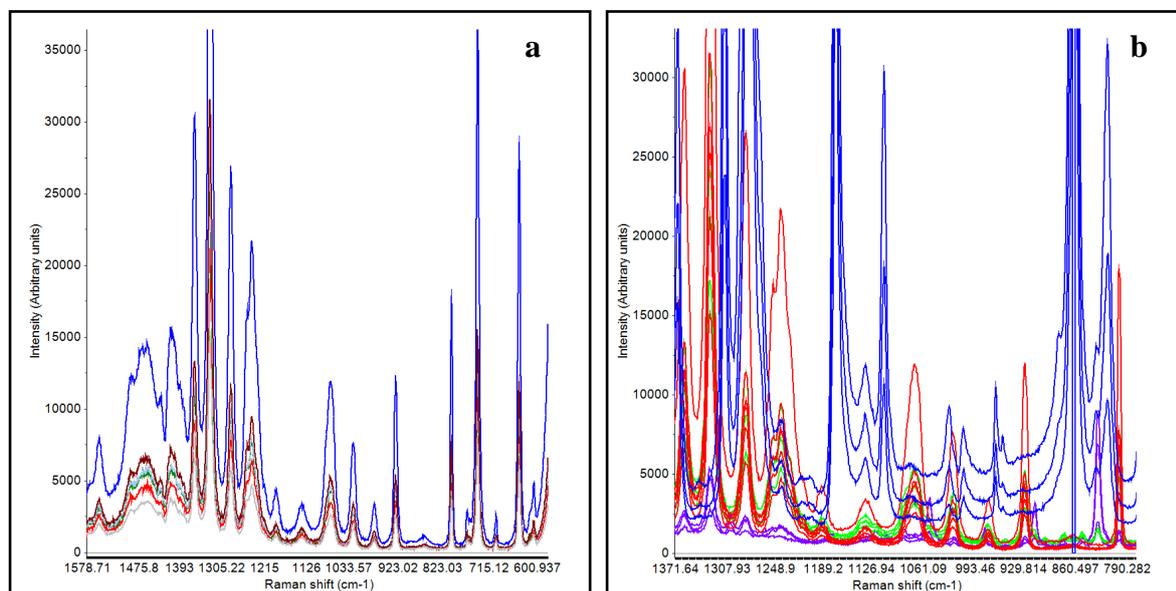


Figure 4. 3: a) Magnified six CAF raw spectra from set ‘1’: C1 (blue), C2 (red), C3 (green), C4 (light blue), C5 (maroon) and C6 (grey); b) Magnified raw spectra from Set ‘1’: C1 to C6 (red), B1 to B3 (blue), CCL 1, 2 and 6 (green) and CCL 3 to 5 (purple) (Unscrambler® X 10.3).

Visual examination of Raman spectra (set ‘1’) showed that some spectra featured cosmic rays (e.g. for C1 at 3116 cm^{-1}) (Figure 4.4). Cosmic rays are known to skew analysis results and should be ‘zapped’ prior to the application of MVA [256]. In this study, cosmic rays were manually removed from C1 spectrum with the ‘zapping’ function using the software WiRE 3.4 [116].

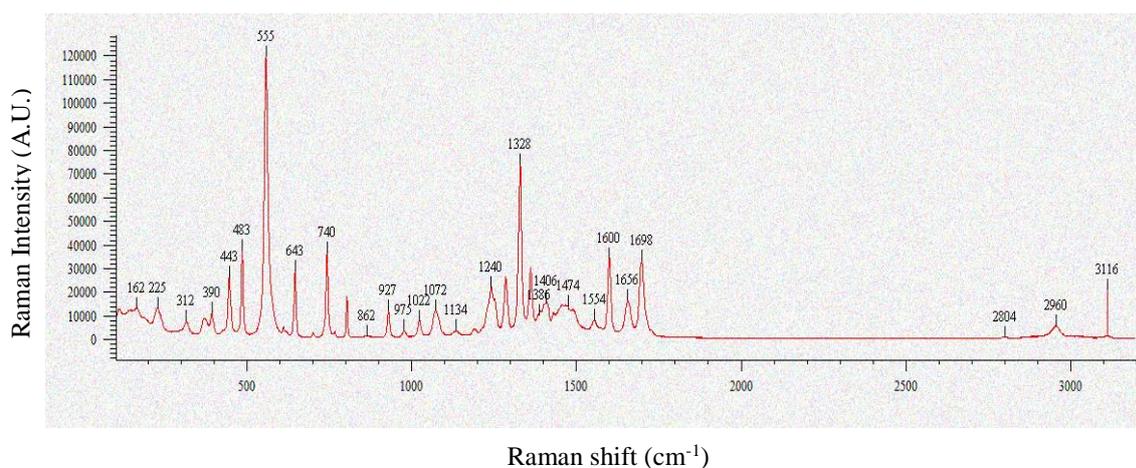


Figure 4. 4: The Raman spectrum for C1 displaying a cosmic ray at ca. 3116 cm^{-1} (WiRE 3.4).

Visual examination of the spectra also showed that the detector was saturated at 860 and 1600 cm^{-1} during the Raman measurement for B1 (set '1') (Figure 4.5). This was expected since BEN is a very strong Raman scatterer. This was potentiated by using a laser power of 10 %.

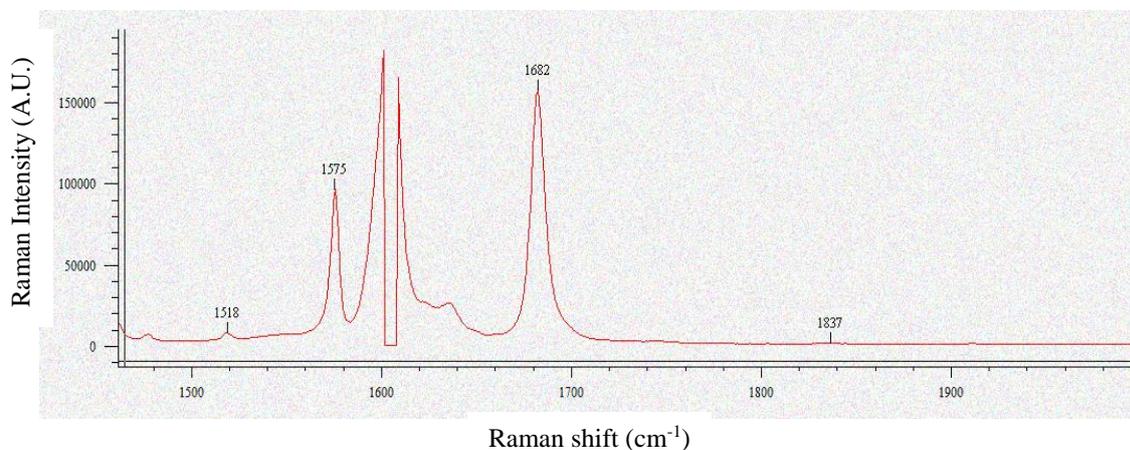


Figure 4. 5: A Section of the Raman spectrum for B1 (detector saturation at ca.1600 cm^{-1}) (WiRE 3.4).

Variation in absolute intensities was observed between replicate Raman spectra (Figure 4.3a). Therefore the signal to noise ratio (S/N) in all spectra (set '1') was calculated. The signal is sample specific related to the molecular Raman activity [257]. This was observed in this dataset such that the S/N for a BEN sample (3113) was 33 times greater than that of a CCL mixture (93). Therefore in this Chapter, no S/N cut-off was designated since the smallest S/N was 93 times greater than the noise. However, the effect of the wide range in S/N (93 – 3113) would be examined in the analysis. Heraud et al. [258] rejected spectra with maximum absolute signal less than 5,000 counts, whereas Ryder et al. [259] rejected spectra when absolute intensities was 70 % lower than average. The latter method was not fully explained in their publication. In this Chapter, the lowest maximum absolute signal was about 6,000 counts (S/N = 93).

Variations in relative intensities between the highest two peaks in each spectrum were also investigated. Relative peak intensities ranged from 0.61 to 0.65 for CAF; from 0.58 to 0.75 for BEN and from 0.65 to 1.20 for CCL mixtures. Greater variations in peak relative intensities were observed for the mixtures as compared to pure substances. According to the British Pharmacopoeia, absolute and relative intensities are influenced by the 'state of polarisation' (i.e. orientation of oscillations of light waves) of the excitation laser, orientation and efficiency of the collecting optics, laser wavelength and laser power at both the source and the sample, focus on the sample, the working distance, the type of sample and its geometry, the packing density and opacity of powdered samples, inter- and intra-day variations and instrument and

analyst variations [260]. A day-to-day variation of $\pm 10\%$ in relative Raman band intensities is common [260]. Set '1', including Set '2' (Section 4.2.2.2.), were measured on the same day.

Further investigations were conducted using exploratory PCA to determine whether the issues identified *via* visual inspection of Raman spectra i.e. baseline offsets, scatter effects, presence of cosmic rays, saturation of the detector and a wide range in S/N in Raman spectra could potentially skew the analysis. The plots generated by 'PCA-1' (set '1') were examined. The explained variance plot (Figure 4.6a) showed that cross-validation was skewed with an abnormal drop along PC2 with a large difference between both calibrated and validated explained variance. This indicated that the model may not be representative for new samples. In this model, PC1 explained 77 and 57 % calibrated and validated variance respectively. The variance in the dataset was explained with three PCs.

The scores and loadings plots (Figure 4.6b and 4.6c) showed that the variance in BEN samples (i.e. B2 and B3) was explained by both PC1 and PC2 with high positive scores and higher than average values for variables (i.e. data points) with positive interpretable loadings respectively. In the scores plot B1 falls outside the ellipse at 95% CL, whereas B3 is at the borders of the ellipse. Therefore, both samples could potentially be outliers and were further investigated. Positive correlation between CCL mixtures and pure CAF samples (both CCL and CAF grouped together), which indicated that these samples are not well described by the model as no variation/ separation was achieved. Linkage between CCL and CAF was expected since CCL mixtures contain CAF (a stronger Raman scatterer than both CRE and LAC). On the other hand, C1 was separated from remaining CAF samples and this could possibly be because it contained a cosmic ray (identified through visual inspection) and had greater S/N compared to remaining CAF and CCL samples (set '1'). C1 was delineated from B2 and B3. This is because they are different chemicals, which gives rise to different Raman spectra and that is why the algorithm is discriminating between them. The line loadings plot (Figure 4.6c) showed important spectral regions for PC1 which mainly correlate to BEN samples based on their positions in the scores plot (Figure 4.6b).

To determine whether both B1 and B3 are true outliers, the influence and leverage plots were examined. The influence plot with T^2 -PC1 (Figure 4.6d) showed that B1 had high T^2 and high F-residuals. B1 did not fit the model (far top right quadrant) and is likely a true outlier, whereas B3 had high T^2 and high F-residuals but fits the model. Both samples are influential on the model. In contrast, C1 had high F-residuals but fits the model. The leverage plots (PC1 and

PC3) (Figures 4.6e and 4.6f) showed that B1 had the highest leverage, and, hence largely contribute to the variance in the model compared to average samples. Both B1 and C1 were identified as outliers. The difference between the first three PCs have been shown to be due to errors in the Raman spectra (i.e. saturation of the detector (PC1), intense samples (PC1/PC2/PC3) and sample featuring a cosmic ray (PC3).

In summary, the plots generated by 'PCA-1' for set '1' demonstrated that samples with spectral errors such as saturation of the detector are outliers that should be rejected because they do not fit the model but largely contribute to the variance. It also demonstrated that samples, which feature cosmic rays are potentially outliers; however, they fit the model and, hence could be retained. The effect of 'zapping' (i.e. removing) cosmic rays was investigated. The wide-range S/N influence the model and samples with very high S/N largely contribute to the variance, despite fitting the model (S/N = 3113 for B3), and, hence normalisation methods of spectra are assessed. Baseline offsets and scatter effects in set '1' potentially caused insignificant variance in the model, possibly because spectra of reference standards were collected with a high performance benchtop Raman instrument. However, their effect was further investigated in this Chapter.

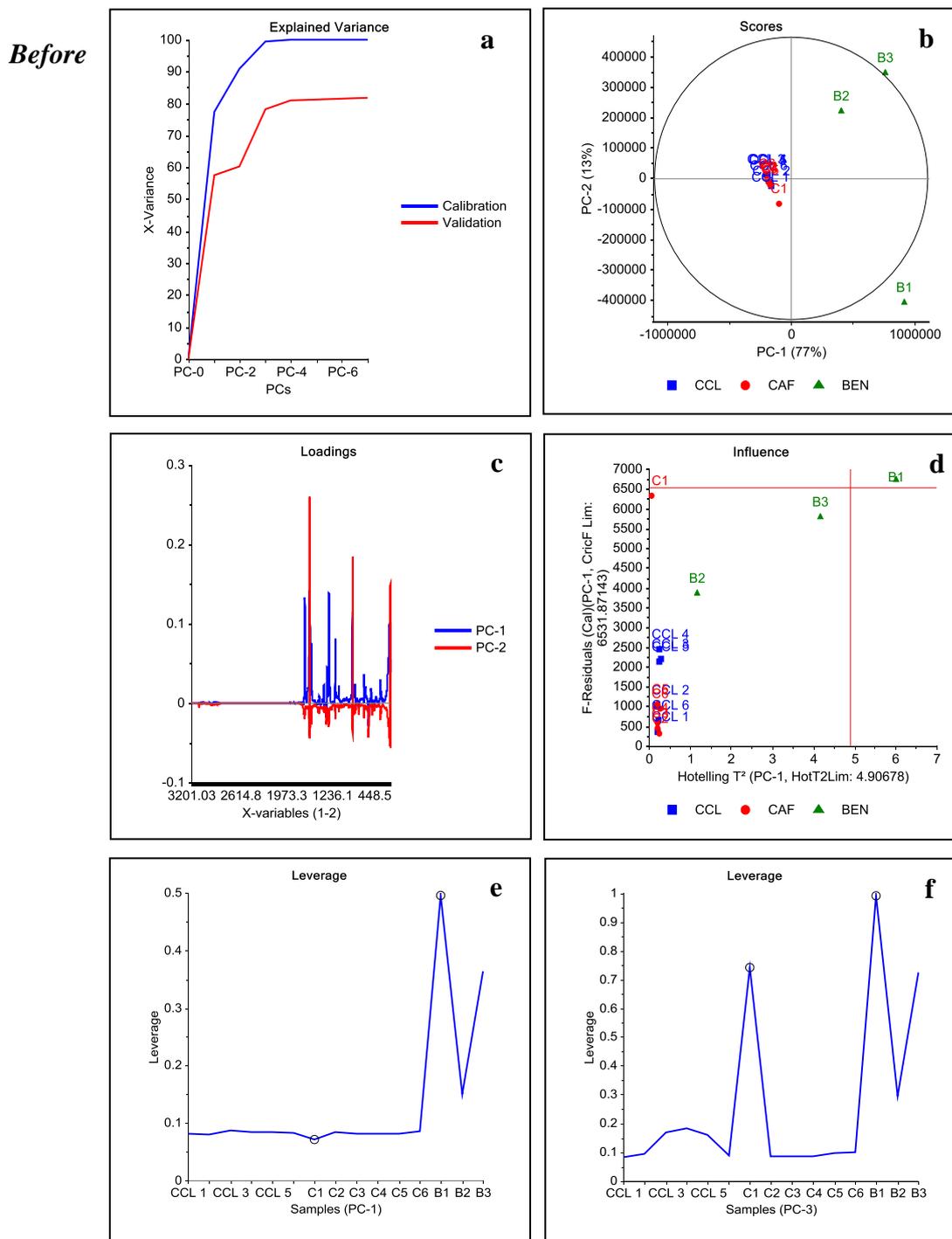


Figure 4. 6: ‘PCA-1’ plots derived from the unprocessed Raman spectra for set ‘1’ (Unscrambler® X 10.3). a) the explained variance plot, b) the 2D-scores plot for PC1/PC2, c) the loadings plot; d) the influence plot; e) the leverage plot (PC1) and f) the leverage plot (PC3) (outliers are circled).

Cosmic rays were ‘zapped’ from C1 spectrum using the software WiRE 3.4 and the spectrum was retained for analysis. In contrast, B1 spectrum, which featured detector saturation was discarded from the dataset. The new dataset is named Set ‘2’. ‘PCA-2’ was then generated from

Set '2' to test the effects of 'zapping' cosmic rays and discarding spectra with errors on the model performance.

In 'PCA-2', the explained variance plot (Figure 4.7a) showed that cross-validation has improved. This was observed with the closeness between the values of both calibrated and validated explained variance, such that one PC explained 84 and 81 % calibrated and validated variance respectively and two PCs explained 99.8 and 97 % cumulative calibrated and validated variance respectively. In 'PCA-2', the largest variance between the spectra was explained with reduced number of PCs (two PCs) and increased percentage of explained variance. This equally means reduced percentage of residual variance or errors in modelling.

The 2D-scores plot for PC1/PC2 (Figure 4.7b) showed that all spectra lie in the ellipse at 95 % CL. BEN spectra (i.e. B2 and B3) are both explained by PC1 with high positive scores and higher than average values for variables with positive loadings (Figure 4.7c). Both B2 and B3 largely contribute to the variance in the model possibly because of their high S/N compared to remaining samples in the dataset (Set '2'). Both CCL mixtures and pure CAF samples showed a slight separation. CCL mixtures positively correlated with CAF samples. C1 is still separated from the remaining CAF samples, potentially because of its high S/N as compared to the remaining CAF samples. C1 has high positive score for PC2. In the loadings plot (Figure 4.7c), large loadings, accounting for BEN samples, and, ideally should be used for the interpretation of the model. Yet, this is not desired as BEN samples have the greatest S/N in the dataset and, hence may be dominating the model, which may not be a true explanation of the variance in the dataset. Therefore normalisation of spectra is required.

After

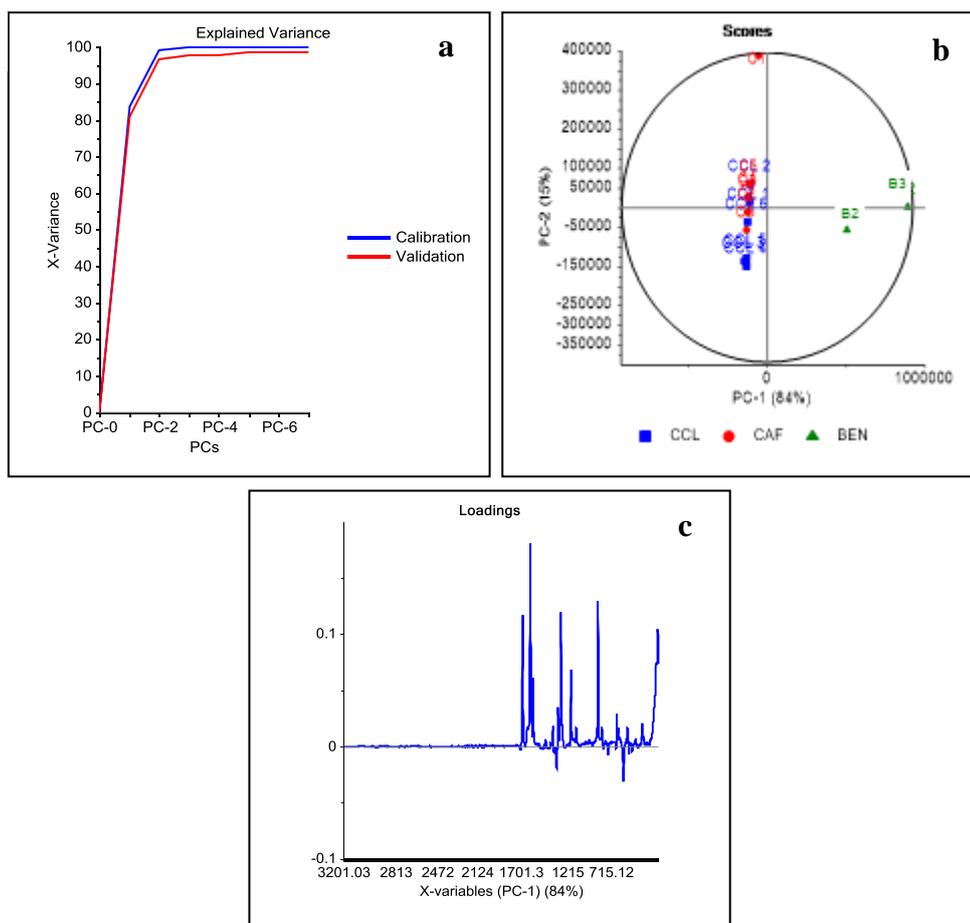


Figure 4. 7: ‘PCA-2’ plots of the Raman spectra for Set ‘2’ (Unscrambler® X 10.3): a) the explained variance plot, b) the scores plot and c) the line loadings plot.

Results for both ‘PCA-1’ and ‘PCA-2’ have demonstrated the effect of instrumental, experimental and environmental artefacts on PCA performance in terms of substance classification, noise reduction and presence of outliers. Results have also highlighted the importance of both initial visual inspection of Raman spectra and the application of PCA for the initial assessment of the quality of Raman spectra prior to analysis. In conclusion, it is recommended to reject Raman spectra featuring detector saturation and to ‘zap’ cosmic rays, whilst retaining these spectra. It is also advised to normalise spectral data because a wide-range in S/N largely contribute to the variance in the model, which may not be a true explanation of the variance that is based on samples’ properties.

4.3.2.2. Importing the final dataset (Set ‘2’) into the Unscrambler® X 10.3

Following the initial visual inspection of Raman spectra and the application of PCA for the initial assessment of the quality of spectra, the final dataset was imported into the Unscrambler® X 10.3 as Excel 2010 files (Microsoft) (Figure 4.1). Using the Unscrambler®

X 10.3, individual datasets ought to have the same X-axis to enable the comparison between different spectra. Although not visually observed, Raman spectra measured on different days had slightly misaligned data points due to different starting wavenumbers (cm^{-1}) i.e. 3200.95 cm^{-1} (Set '3') and 3201.03 cm^{-1} (sets '1' and '2'), a difference of 0.08 cm^{-1} . However, the total number of data points remains the same (i.e. 3777 data points). Therefore interpolation of spectra was essential before combining spectra into one analysis set for processing using the Unscrambler® X 10.3. Descriptive statistics were also performed, after the data was imported into the Unscrambler® X 10.3, to evaluate scatter effects in Raman spectra.

4.3.2.3. Interpolation of Raman spectra

Interpolation of Raman spectra was undertaken using the Unscrambler® X 10.3. Six pure CAF samples (C1 to C6 from Set '2') were employed to test the impact of interpolation on Raman spectra. Spectra were interpolated from $3201.03 - 100.55 \text{ cm}^{-1}$ to $3200.95 - 100.41 \text{ cm}^{-1}$. Upon visual inspection, the effect of interpolation could not be visually observed because interpolated and non-interpolated spectra were fully overlaid (Figure 4.8). Therefore a correlation matrix Table (not shown) was calculated using Excel 2010 (Microsoft). Results demonstrated that there was a positive correlation between interpolated and non-interpolated CAF and between all CAF spectra themselves ($R^2 = 0.9928 - 0.9999$). A high positive correlation demonstrated that there was no significant difference between interpolated and non-interpolated spectra. In this study, spectra were only be interpolated when both sets '2' and '3' were combined for pre-processing method validation in Section 4.3.5.1.

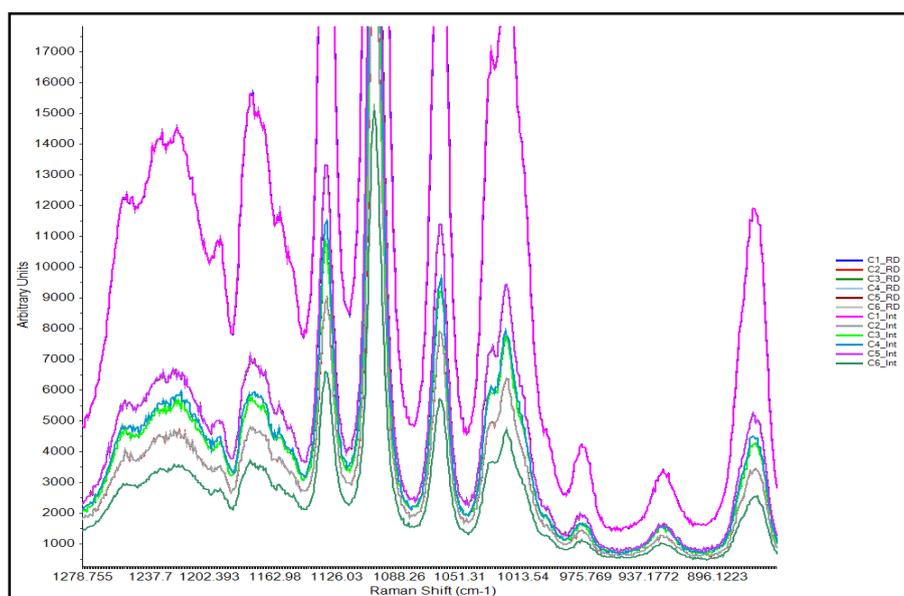


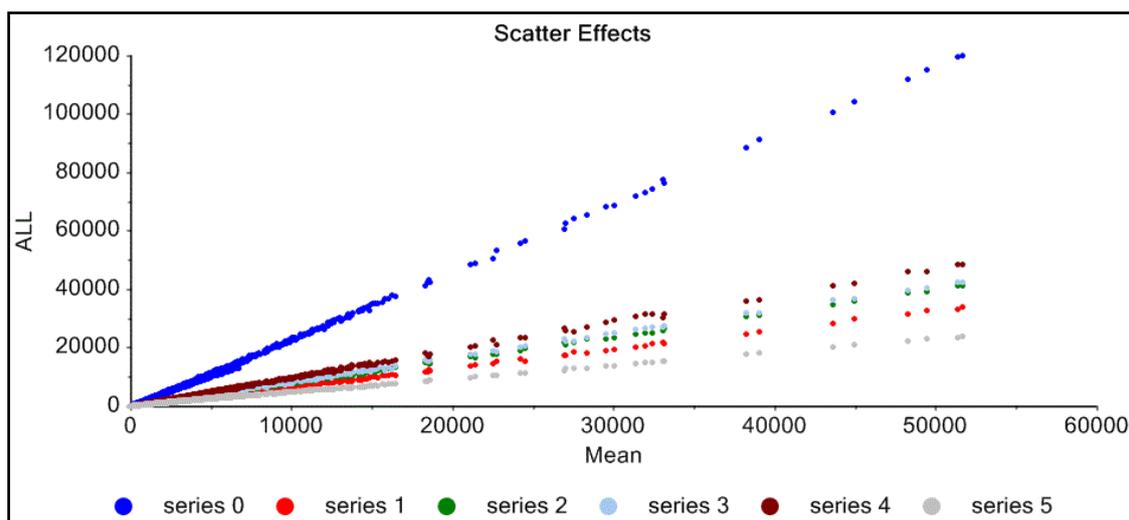
Figure 4. 8: Six raw (RD) and six interpolated (Int) CAF Raman spectra (Unscrambler® X 10.3).

4.3.3. Descriptive statistics (Unscrambler® X 10.3)

In general, after importing any dataset into the Unscrambler® X 10.3, an overview of Raman spectra could alternatively be achieved using numerous descriptive statistical methods such as multiple scatter plots, mean and standard deviation plots. In this Section, descriptive statistics were performed to investigate the presence of possible scatter effects in Set ‘2’, initially identified through visual inspection of Raman spectra (Section 4.3.2.1.).

4.3.3.4. Investigation of the presence of scatter effects in Raman spectra

Descriptive statistics were employed to establish the presence and type of scatter effects (i.e. multiplicative and/ or additive). The presence of scatter effects are common in Raman spectra and may result from instrument artefacts such as fixed pattern noise inducing a change in photon response by the detector [261]. The scatter effects testing algorithm (Unscrambler® X 10.3) compares individual spectra to a mean spectrum, which is automatically selected from the dataset. Figure 4.9 represents a 2-D scatter plot for six CAF spectra (Set ‘2’) (labelled ‘ALL’ on the Y-axis) plotted against the mean spectrum (X-axis). The software’s algorithm regresses the spectra (represented by row vectors) against the offset and slope of the mean spectrum, hence correcting the spectra to the ‘same scatter level’ as the mean spectrum [172, 178]. Results showed that there is a difference in slope, thus MSEs were confirmed (amplification of some peak intensities). MSEs were also confirmed for CCL and BEN spectra (Figures not shown). Therefore spectral pre-processing is required to remove MSEs prior to MVA.



¹(Series 0: C1; series 1: C2; series 2: C3; series 3: C4; series 4: C5; and series 5: C6)

Figure 4. 9: A plot showing multiplicative scatter effects (MSEs) for six CAF spectra (Set ‘2’) (Unscrambler® X 10.3)¹.

4.3.4. Pre-processing of Raman spectra using the Unscrambler® X 10.3

This Section provides a detailed description of the modus operandi of various spectral pre-processing techniques. Hitherto, it has been established that instrument, sample, analysis and environment artefacts (i.e. saturation of the detector, baselines offsets, multiplicative scatter effects (MSEs), variable S/N and cosmic rays) have contributed to the Raman signals in the measured spectra. It is concluded that spectra with detector saturation should be discarded, whereas spectra featuring cosmic rays could be retained after being ‘zapped’. The final dataset (Set ‘2’), imported into the Unscrambler® X 10.3, was not interpolated because all the samples had the same x-axis. Imported spectra were pre-processed to reduce baselines offsets, MSEs and variations in absolute intensities (Figure 4.1). Removal of these interferences has been shown to help in extracting the Raman signal of samples and in quantifying the different constituents [262] (e.g. amount of scattering species or constituents in mixtures). Over pre-processing may, however, lead to over-fitting the data to chemometric models (e.g. PCA) which could be misleading [172]. In several studies a maximum of three pre-processing methods were applied to Raman data in order to prevent over-fitting the model to the data [132, 148, 172, 176, 184, 258, 263]. However, some datasets may require more than three pre-processing methods [95]. Pre-processing may remove important chemical information through data reduction e.g. with the use of PCA, multiplicative scatter correction (MSC) or other pre-processing methods [172].

In this study, selected pre-processing methods were evaluated for their applicability to PCA (Figure 4.10). Smoothing methods were studied as they have been shown to improve the performance of chemometric models [182]. Baseline correction, normalisation and multiplicative scatter correction pre-processing methods were evaluated because baseline offsets, variable absolute intensities and MSEs were initially identified through visual inspection and descriptive statistics (Set ‘1’, including Set ‘2’ (Section 4.2.2.2.)). The order in which combined pre-processing methods should be applied was also evaluated. The effects of each pre-processing method was examined using pure substances, then evaluated using mixtures.

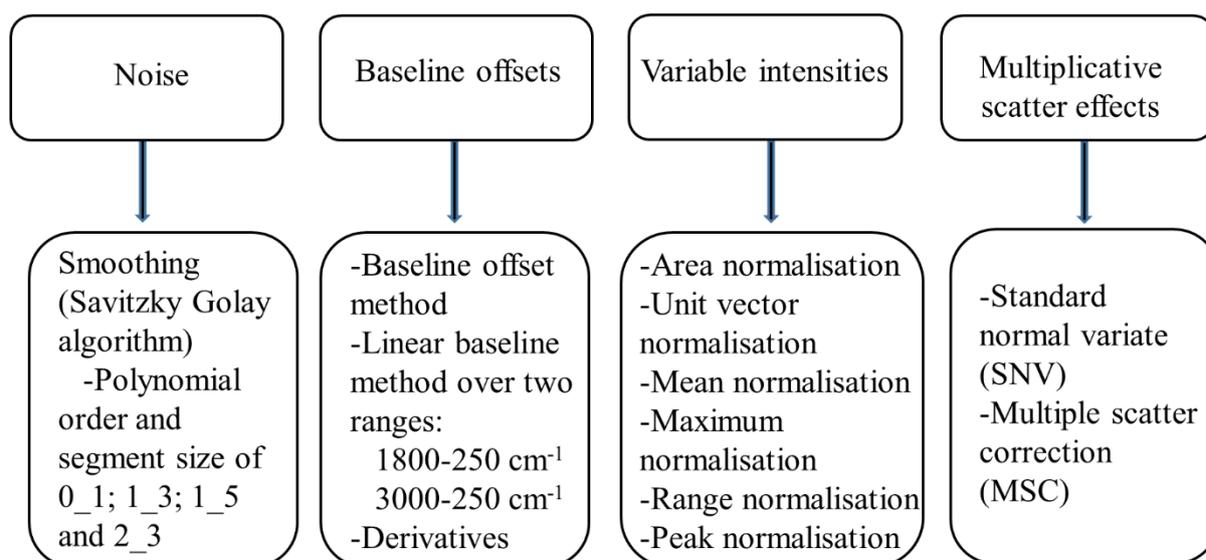


Figure 4. 10: Summary of pre-processing methods investigated.

4.3.4.1. Smoothing

Smoothing using the Savitzky Golay (SG) algorithm (Unscrambler® X 10.3) is a common pre-processing method employed to reduce/ filter high frequency noise in spectral data by fitting a polynomial to a segment of data points. Larger segments and lower polynomial orders lead to more smoothing being generated. The spectra for two pure CAF (C1 and C2) were employed to enable the visualisation of the effect of smoothing on the spectra. Four different sets of parameters were evaluated starting from default parameters, then working up (Table 4.5).

Table 4. 5: Smoothing parameters.

Method number	Polynomial Order	Size of smoothing segment	Abbreviations
1	0	3	SG_0_3
2	1	3	SG_1_3
3	1	5	SG_1_5
4	2	3	SG_2_3

Small segment sizes were selected in order to ensure the preservation of small Raman peaks [95]. Pre-processed spectra were visually compared to the raw spectra (RD). The effect of each smoothing method for both C1 and C2 could not visually be observed and, hence smoothed spectra could not be distinguished from non-smoothed spectra (Figure 4.11).

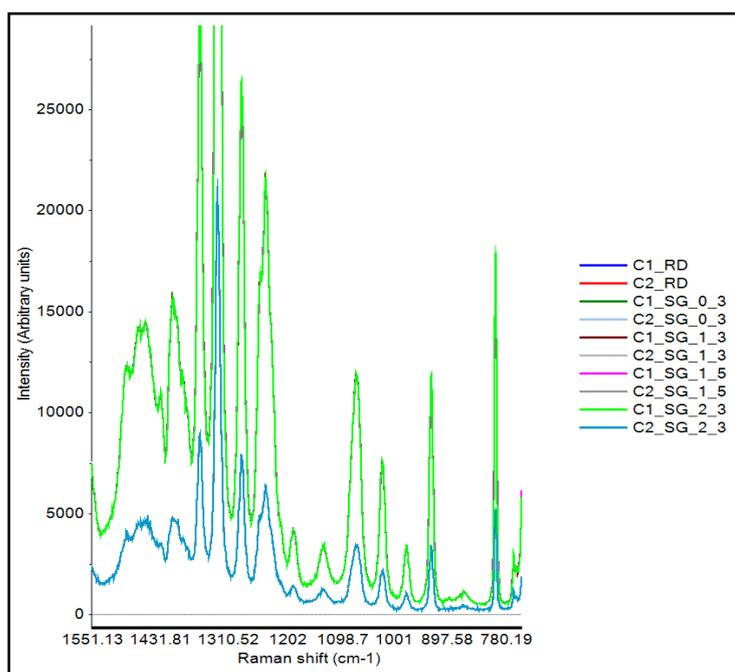


Figure 4. 11: Summary of various smoothing of two CAF spectra (C1 and C2 from Set ‘2’). Spectra were smoothed using the SG algorithm (Unscrambler® X 10.3) using the parameters summarised in Table 4.5.

The S/N of smoothed spectra was then compared to that of the raw spectra (Set ‘2’). Despite the lack of visual analysis of the Raman spectra, results in Table 4.6 demonstrated that the S/N ratios have almost doubled with both smoothing methods 2 and 3 for C1, and have more than doubled using smoothing methods 1, 2 and 3 for C2. This could be due to a reduction in noise as compared to raw spectra. Noise levels were not reduced using smoothing method 4.

Table 4. 6: Summary of the S/N for all smoothing methods for both C1 and C2 (WiRE 3.4).

Smoothing parameters	Signal (A.U.)	Noise (A.U.)	S/N
C1_Raw spectrum	118345	70	1691
C1_SG_0_3	116412	55	2117
C1_SG_1_3	116412	36	3234
C1_SG_1_5	112522	36	3126
C1_SG_2_3	118145	90	1313
C2_Raw spectrum	33536	74	453
C2_SG_0_3	32700	26	1258
C2_SG_1_3	33800	32	1025
C2_SG_1_5	33300	20	1665
C2_SG_2_3	33250	80	416

Each smoothing method was then evaluated using Set ‘2’. For each smoothing method, a PCA was generated and compared to ‘PCA-2’ (Figure 4.7). The PCA generated for each smoothing method (Figure 4.12) was almost identical to that of the raw spectra (Set ‘2’) demonstrating

that the effect of the above smoothing parameters was insignificant with respect to the classification of the tested substances using PCA. In this study, smoothing (SG) with a polynomial order of 1 and segments of 3 and 5 smoothing points were considered and evaluated in the final combined pre-processing methods because both methods consistently reduced noise and enhanced the S/N for both C1 and C2.

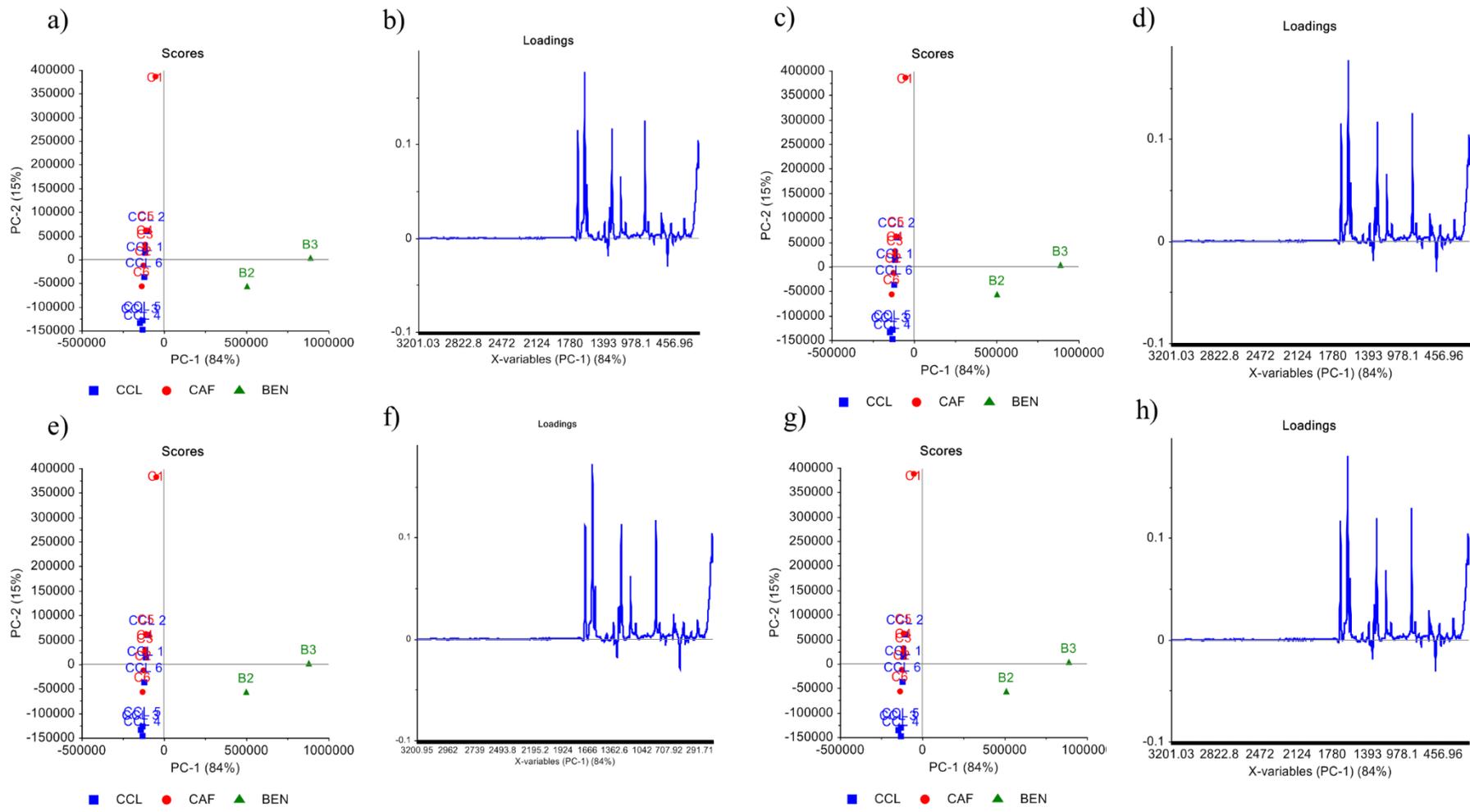


Figure 4.12: Summary of PCA results for smoothed Set ‘2’ Raman spectra (Unscrambler® X 10.3). Method 1: a) 2D-scores and b) line loading plots. Method 2: c) 2D-scores and d) line loading plots. Method 3: e) 2D-scores and f) line loading plot. Method 4: g) 2D-scores and h) line loading plots.

4.3.4.2. Correction of baseline offsets

Baseline offsets are common instrumental artefacts, which could be corrected or reduced using various pre-processing methods and are often corrected prior to spectral normalisation [261]. In this study, baseline offsets were initially identified through visual examination of the Raman spectra (Set '1', including Set '2') (Figure 4.3). In this Section, two pure CAF spectra (C1 and C2 from Set '2') were employed to enable the visualisation of the effect of baseline correction methods (i.e. baseline offset, linear baseline and derivatives) (Figure 4.10). Then the methods investigated were used to pre-process Set '2'. A PCA was then generated for each method and compared to that of the raw spectra ('PCA-2', Figure 4.7).

4.3.4.2.1. The baseline offset method

The baseline offset correction algorithm corrects the dataset by subtracting the minimum absolute intensity value from all intensity values across the spectrum [229]. The line plots for pre-processed CAF spectra (C1 and C2 from Set '2') using the baseline offset method have shown that there was no significant improvement in reducing the offset observed over the region 1850 - 100 cm^{-1} (Figure 4.13a). The latter region has been shown to include more information on functional group frequencies and the molecular fingerprint, hence more informative about CAF. The high wavenumber end of the spectra has not shown a significant change from the raw spectra (Figure 4.13b). A PCA was generated from the pre-processed spectra for Set '2' and compared to that of the raw spectra to test the impact of this baseline offset method (Figures 4.14a and 4.14b). The PCA showed no change to the model or improvement to the samples' classification after the removal of the baseline offsets from the Raman spectra as compared to PCA-2 for unprocessed Set '2' (Section 4.3.2.1 and Figures 4.7a to 4.7c). In conclusion, the baseline offset correction method did not improve the PCA sample classification but transformed the values in the data matrix to values with positive signs. The significance of a single sign in the data matrix is further investigated with respect to combined pre-processing methods (see Section 4.3.5.).

4.3.4.2.2. Linear baseline method

Linear baseline correction method is an alternative baseline correction method which involves designing a new baseline for the entire dataset between two selected wavenumbers. The selected wavenumbers become zeroed and the remaining data points are transformed or pre-

processed (interpolated or extrapolated) accordingly [172]. One disadvantage of this method is that it assumes that the offset is the same between all samples.

Linear baseline correction between the wavenumbers 1800 and 250 cm^{-1} (region of chemical information for CAF spectra) (Appendix C Figure A4.2) for both C1 and C2 (Set '2') has slightly reduced the offset over the region 1800 - 250 cm^{-1} (Figure 4.13c). However, a drop of the baseline below zero was observed over the uncorrected region 3201 - 1800 cm^{-1} (Figure 4.13d). Set '2' was then pre-processed using the linear baseline correction method over the range 1800 - 250 cm^{-1} . The PCA results showed that substance classification has not improved as compared to the raw spectra (Figures 4.14c and 4.14d). A slight change to the explained variance occurred such that PC1 explained 82 % calibrated variance down from 84 % (PCA-2, Section 4.3.2.1 and Figures 4.7a to 4.7c) and PC2 explained 17 % calibrated variance up from 15 %. The overall explained variance remained unchanged.

Linear baseline correction over the region 3000 - 250 cm^{-1} (across most of the spectral range) (Figure 4.1) has resulted in the drop of the baseline over the region 3201 - 1800 cm^{-1} below zero but has significantly reduced the offset over the region 1800 - 100 cm^{-1} (Figures 4.12e and 4.12f). Set '2' was then pre-processed using the linear baseline correction method over the range 3000 - 250 cm^{-1} . A PCA was then generated and compared to that of the raw spectra (Set '2') to test the effect of this pre-processing method (Figures 4.14e and 4.14f). The PCA results were very similar to that of the raw spectra, with a slight reduction in the percentage of explained calibrated and validated variance (83 and 80 % for PC1, and 16 and 16 % for PC2 respectively) and increased residuals for some samples (e.g. B2) in the influence plot (not shown). In conclusion, the linear baseline offset correction method has not improved the PCA sample classification. Corrected regions showed slight improvement in the line plots at the expense of the uncorrected regions and, therefore the linear baseline correction method was not considered in the final combined pre-processing methods.

4.3.4.3. Derivatives

Unlike smoothing, derivatives filter low-frequency baselines and enhance high-frequency signals [264]. As the band width increases (low-frequency) the derivative value increases and the peak maxima decreases and *vice versa* [265]. Pre-processing using derivatives is dependent on both the frequency and the collinearity between the data points [264]. Derivatives measure the slope of spectral peaks by estimating the differences between discrete adjacent points [265] and, hence correct for baseline offsets and resolve overlapping peaks. In this study, the first

and second derivatives were calculated using the SG algorithm to evaluate if they would correct for baseline offsets and improve the classification of samples *via* PCA.

4.3.4.3.1. The first derivative

The first derivative measures the slopes of the curve at every data point. Therefore it would theoretically reduce offsets without affecting slopes and correct for fluorescence variation [192, 266]. Since the difference between additive baseline offsets is a constant and since the derivative of a constant is zero, the spectral mean becomes zero and the spectra become replaced with the slopes of the curves. The original peaks maxima also become zeroed since there is no change in the slope [172]. In this Section, two CAF spectra (C1 and C2 from Set '2') were employed for the calculation of the first derivative to enable the visualisation of the pre-processed spectra. The first derivative was calculated using the SG algorithm with a polynomial order of 2 and a smoothing segment of 3 points (default parameters) (Figure 4.13g). The visual inspection of the first derivative line plots showed a better resolution of the major peaks but with observed loss of chemical information (absence of small Raman peaks after the pre-processing was conducted).

Set '2' was then pre-processed using the first derivative (SG) using default parameters (a polynomial order of 2 and a segment size of three points). A PCA was then generated and compared to that of the raw data (i.e. PCA-2, Section 4.3.2.1.). The PCA showed a slight improvement from that of the raw spectra with two PCs explaining 99 % of the variance. The PCA scores plot (Figure 4.14g) showed that both CCL mixtures and pure CAF samples are not well described by the model. Despite positive correlations between them, they have shown slight ungrouping. C1 was still separated from remaining CAF spectra. The loadings plot (Figure 4.14h) showed important variables for PC1 (peaks at ca. 860, 1170, 1279, 1574, 1603 and 1679 cm^{-1} (positive loadings), 643, 863, 1149, 1285, 1608, 1686 and 1579 cm^{-1} (negative loadings). The peaks at ca. 643, 1170, 1279, 1285, 1603 and 1608 cm^{-1} are all consistent with CAF, whereas 860, 863 and 1149 cm^{-1} are consistent with LAC and 1574, 1579, 1679 and 1686 cm^{-1} were consistent with BEN (Appendix C Figures A4.1 – A4.4). The application of the first derivative alone has been shown to slightly improve classification models such as soft independent modelling of class analogy (SIMCA) [184] and to produce satisfactory PCA clusters [95]. This demonstrated that pre-processing is sample and application specific.

4.3.4.3.2. The second derivative

The second derivative measures the derivative of the first derivative, hence filtering additive baselines as well as ‘linear tilts’ [264]. The second derivative accentuates high frequency signals leading to simultaneous increase in noise [264]. However, the filter segment size could be enlarged to reduce noise [265]. Furthermore, the second derivative helps in resolving and sharpening neighbouring peaks and in reducing fluorescence effects [116]. It has been shown to improve quantitative analysis [172] and enhance robustness of measurements against fluorescence or container variation [192, 266]. In this Section, two CAF spectra (C1 and C2 from Set ‘2’) were employed in the calculation of the second derivative to enable the visualisation of the pre-processed spectra.

The second derivative was calculated using the SG algorithm with a polynomial order of 2 and a smoothing segment of 3 points (default parameters). The visual inspection of the second derivative line plots was difficult to interpret due to the large number of variables. Therefore the second derivative of bands (about 100 cm^{-1} range) was performed. By examining the pre-processed CAF spectra, regions with chemical information were: [813-714], [713-613], [612-509] and [508-403] cm^{-1} . In contrast, regions, which showed loss of information (disappearance of peaks) included: [1722-1639], [1638-1553], [1377-1288], [1287-1197], [1196-1104], [1103-1109], [1108-913] and [402-296] cm^{-1} . Figure 4.13h showed relative loss of chemical information. The performance of the second derivative was not improved by increasing the size of the segment up to nine points (estimated width at half the height of the maximum CAF peak at 555 cm^{-1}).

Set ‘2’ was then pre-processed using the second derivative (SG). A PCA was then generated and compared to that of the raw data (i.e. PCA-2, Section 4.3.2.1) (Figures 4.14i and 4.14j). The PCA showed reduced explained variance, reduced cross-validated variance, loss of orthogonality and clustering of the variables around the mean in the 2-D scatter loadings plot (not shown), which corroborated with published literature [95]. As previously stated, the second derivative measures the changes in the slope of the curve. Consequently, the loss of large amount of chemical information may have resulted in the loss of orthogonality.

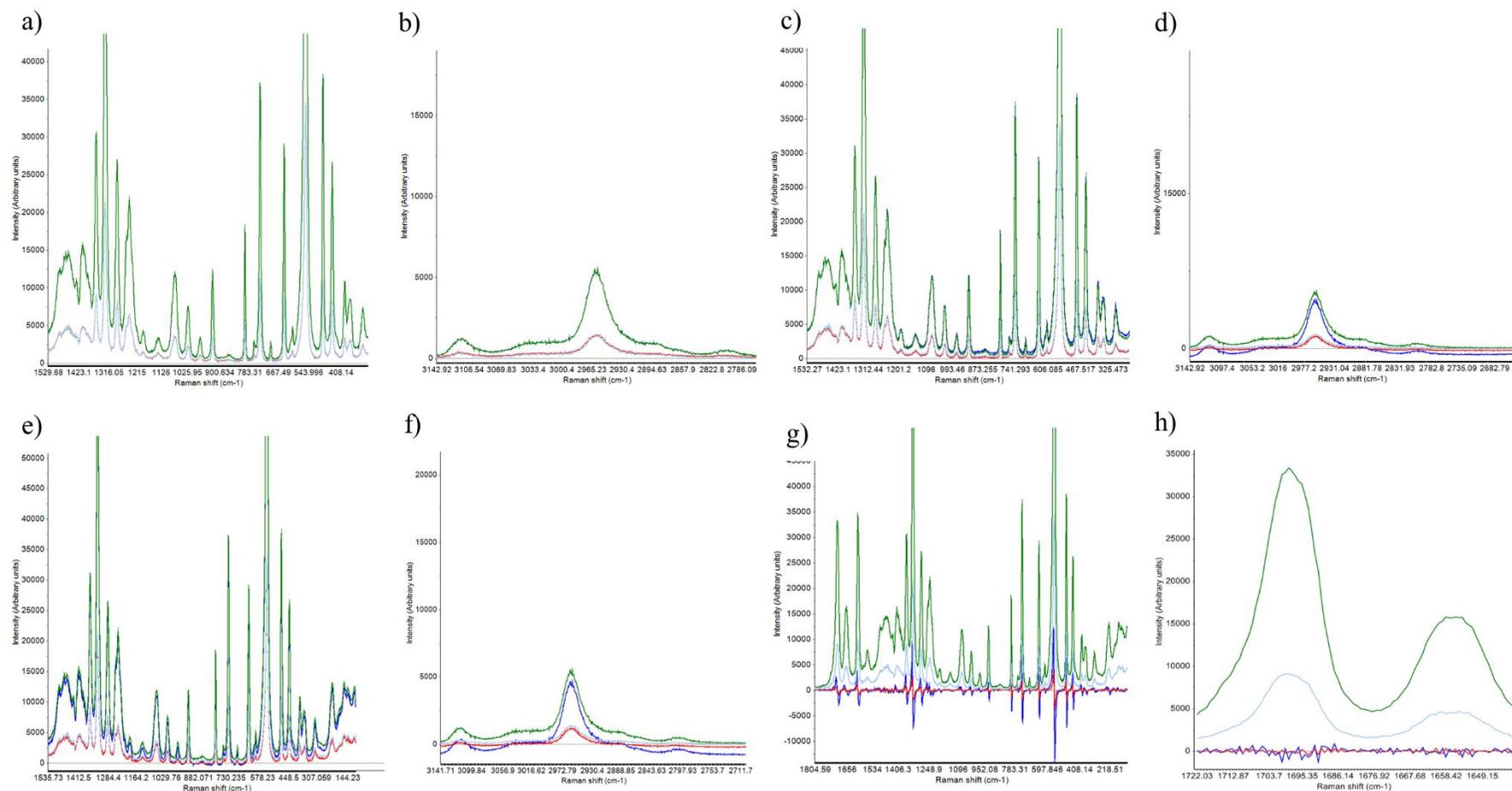


Figure 4. 13: Summary of baseline correction methods (Unscrambler® X 10.3). Baseline offset method: a) low wavenumber end and b) high wavenumber end. Linear baseline method between the wavenumbers 1800-250 cm^{-1} : a) low wavenumber end and d) high wavenumber end. Linear baseline method between the wavenumbers 3000-250 cm^{-1} : a) low wavenumber end and f) high wavenumber end. g) The first derivative of two CAF raw spectra (Raw spectra for C1 (green) and C2 (light blue), and first derivative for C1 (blue) and C2 (red)). h) The second derivative of a 100 cm^{-1} band (1722-1622 cm^{-1}) from two CAF raw spectra (Set '2') (Raw spectra for C1 (green) and C2 (light blue) and second derivative for C1 (blue) and C2 (red)).

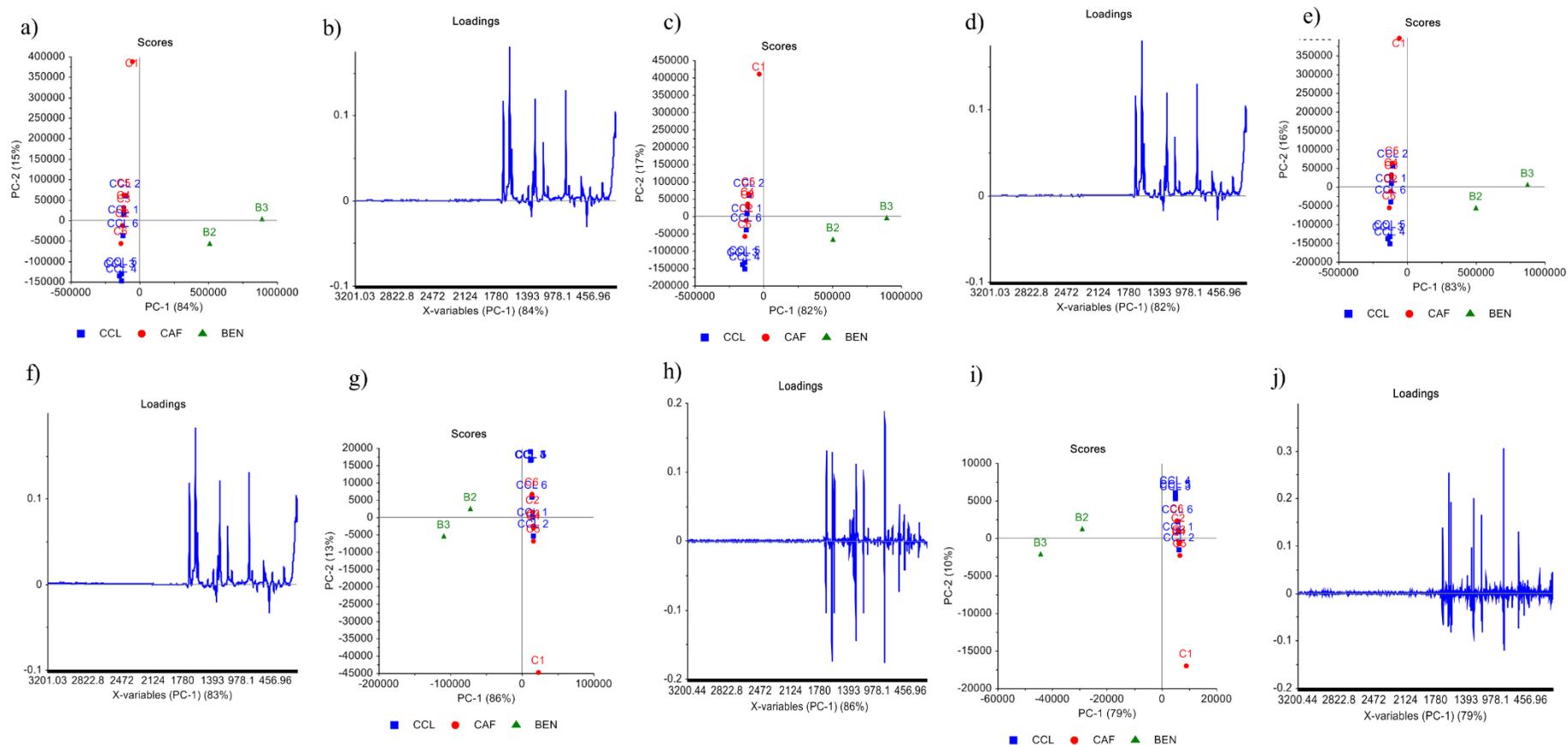


Figure 4. 14: Summary of PCA results for baseline correction methods (Set '2') (Unscrambler® X 10.3). Baseline offset method: a) 2D-scores and b) line loading plots. Linear baseline (1800-250 cm^{-1}) c) 2D-scores and d) line loading plots. Linear baseline (3000-250 cm^{-1}) e) 2D-scores and f) line loadings plot. First derivative g) 2D-scores and h) line loading plots. Second derivative i) 2D-scores and j) line loading plots.

In conclusion, the baseline offset method has not improved substance classification using PCA in this dataset. However, the effect of baseline offset correction is further investigated with combined pre-processing methods (see Section 4.3.5.). The linear baseline correction slightly reduced the baseline offset over particular spectral regions based on the selected wavenumber range. However, it increased the baseline offset over remaining regions. Using the linear baseline method, the PCA model did not improve as compared to the raw data, and, therefore in this study the linear baseline correction method is not recommended. For this dataset, the first derivative of the Raman spectra for Set '2' showed better PCA classification results than the second derivative. The first derivative was able to explain a greater percentage of both calibrated and validated variances than the second derivative. Derivatives have been shown to increase noise in PCA and partial least squares (PLS) models [182]. Unlike the first derivative in this study, the second derivative only has been shown to increase the noise in the PCA model. This was demonstrated by the reduced percentage of explained calibrated and validated variance, the skewness of the explained validated variance and the increased percentage of calibrated (7 % up from 0.6 %) variance explained by the later PCs (noise). The effect of both the first and second derivatives is further investigated in combined pre-processing methods (see Section 4.3.5.).

4.3.4.4. Normalisation methods

Variation in absolute intensities is common in Raman spectra [184]. In this study, variations in absolute intensities and wide-range S/N were initially identified and have been shown to impact sample classification using PCA (PCA-2 Section 4.3.2.1). Therefore normalisation of the spectra in this dataset was recommended because it scales the spectral data and ensures an even distribution of the variance between samples. In other words, normalisation enhances the variance between samples by correcting for interfering scaling variations. Normalisation methods evaluated in this study include area, peak, maximum, unit vector, range and mean normalisation (Figure 4.10).

Six CAF samples (C1 to C6 from Set '2') were employed to enable the visualisation of the effect of normalisation methods on Raman spectra. Subsequently, for each normalisation method, a PCA was generated for C1 to C6 to investigate their effects on reducing the separation between replicate samples as compared to the PCA for raw spectra (C1 to C6, Section 4.3.4.4.1. below). Finally, one normalisation method was selected and employed to normalise Set '2' to evaluate its effect on the entire dataset.

4.3.4.4.1. PCA of the raw spectra of six pure CAF samples

The line plots for the CAF raw spectra displayed variable absolute intensities, which is common in Raman spectroscopy (Figure 4.16a). The PCA results for C1 to C6 (Set '2') showed that all spectra were distinguished/ separated from their raw Raman data (i.e. CAF spectra did not form a cluster). The scores plot (Figure 4.17a) explained about 100 % of the variance in the data, where one PC explained 99.89 % calibrated variance and 99.50 % validated variance. Since all spectra were for pure CAF, the main variance between the spectra was due to instrumental artefacts (i.e. differences in absolute intensity, S/N and/ or slight differences in peak shifting). In the scores plot, C1 (upper right quadrant) was well explained by PC1. C1 had a long distance from the mean along PC1 (ca. 312000) and largely contributed to the variance in the PCA model, potentially due to its highest S/N among the six CAF spectra. Based on their position in the scores plot, C1 has been shown to be located in the opposite quadrant to C6. This could be because C1 had the highest S/N, whereas C6 had the lowest S/N among the six spectra, which explain the wide separation. The variance in C2, C3 and C6 (upper left quadrant) are explained by PC2. They are not well described since the difference in S/N (453, 549 and 442 respectively) and in peak shifting are insignificant. Both C2 and C6 are positively correlated. Both samples have been shown to have the smallest S/N ratio in the dataset.

The scores plot showed that all CAF samples lie in the model with 95 % CL. The loadings plot (Figure 4.17b) showed that C1 had high positive scores and high values for variables with positive loadings. In conclusion, most of the variance in the PCA model for pure CAF was explained by the spectrum with the highest S/N. Correlations between samples were influenced by their S/N. These results demonstrated that normalisation was essential to scale the raw spectra.

In this study, six types of normalisation methods (area, peak, maximum, unit vector, range and mean normalisation) were evaluated and compared (Figure 4.10).

4.3.4.4.2. Area normalisation

Area normalisation sums up the area under the curve to a constant, which is the same for all spectra [172]. In other words, area normalisation makes the area under all six spectra equal. This algorithm transforms the data matrix such that all the values in the data matrix become positive. The line plots for the CAF area normalised spectra (C1 to C6 from Set '2') are shown

in Figure 4.16b). Magnified peaks (not shown) displayed overall absolute intensities in the descending order of the S/N of the CAF spectra ($C1 > C5 > C4 > C2$ and $C3 > C6$).

The PCA scores plot (Figure 4.17c) explained about 91 % of the variance in the data, where one PC explained 83.42 % calibrated variance and 72.45 % validated variance. The explained calibrated variance for PC1 was reduced from 99.89 % (raw spectra, Section 4.3.4.4.1) to 83.42 %, possibly due to reduced variation in the area under the curve among normalised replicate spectra. Compared to Figure 4.17a, it was observed that the coordinates of the scores plot were reduced, showing that area normalisation reduced the distance between the samples (distance between C1 and C6 is 0.0057 scores along PC1). Reduction in the distance between samples illustrates the improvement achieved by area normalisation in grouping replicate spectra of a pure substance. The scores plot showed that the variance in both C1 and C6 (highest and lowest S/N) is explained by PC1, whereas the variance in C5 (second highest S/N) is explained by PC2 (minimal variance explained (8 %)). C1 is widely separated from C6 due to extreme overall absolute intensities in both spectra. Both C2 and C3 are positively correlated, with low impact on the model.

The scores plot showed that all CAF samples lie in the model with 95 % CL. The loadings plot (Figure 4.17d) showed important spectral variables ((peaks at ca. 554, 1331, 739, 1604 and 1694 cm^{-1} (positive loading for C1)) along PC1 and ((peaks at ca. 552 cm^{-1} (positive loading for C1), 1328, 1703 and 2961 cm^{-1} (negative loadings for C6)) along PC2, which correlate to C1 and C6 respectively based on their positions in the scores plot. In the loadings plot, for all the variables, which have positive loadings for PC1, C1 had higher than average values for these variables. Despite area normalisation, C1 is relatively dominating the model showing that the effect of this algorithm on the large variation in S/N was not very significant.

In conclusion, the separation between replicate CAF Raman spectra was reduced by area normalisation, despite being influenced by wide-range S/N. Area normalisation has produced the smallest variance between replicate CAF spectra across PC1 as compared to other normalisation methods as described below.

4.3.4.4.3. Unit vector normalisation

Unit vector normalisation involves transforming the spectra to vectors of length 1 or unity. The line plots for the CAF unit vector normalised spectra (C1 to C6 from Set '2') are shown in

Figure 4.16c. Magnified peaks (not shown) displayed variable overall absolute intensities between spectra but not in a specific order.

The scores plot (Figure 4.17e) explained about 86 % of the variance in the data, where one PC explained 74 % calibrated variance and 57 % validated variance. The large difference between both calibrated and validated variance showed that the model would not be valid for new samples. The variance in the model was explained with four PCs. Compared to Figure 4.17a, it was observed that the coordinates of the scores plot were reduced (distance between C1 and C6 is 0.11 scores along PC1). In other words, the distance between the spectra was reduced and the variance between the samples was reduced. The scores plot showed that the variance in both C1 and C5 was well explained by PC1. C1 had a slightly longer distance from the mean and a higher score (i.e. 0.06 scores) along PC1 than C5 (i.e. 0.05 scores). Despite unit vector normalisation of the Raman data, the locations of both C1 and C6 in the scores plot demonstrates that C1 was still widely separated from C6 due to differences in overall absolute intensities in both spectra. The variance in C2, C3, C4 and C6 was explained by PC2 (12 %). Both C2 and C3 were positively correlated, with a low impact on the model.

The scores plot showed that all CAF samples lie in the model with 95 % CL. The loadings plot (Figure 4.17f), showed the important variables for PC1 (peaks at ca. 554, 741, 1330, 1599 and 1696 cm^{-1}), which correlate to C1 and C5 based on their positions in the scores plot. The loadings plot also showed the important variables for PC2 (peaks at ca. 552.3 cm^{-1} (positive loadings for C2 and C3), 1328, 1399 and 2956 cm^{-1} (negative loadings for C4 and C6). C1 has high positive scores, hence has high values for variables with positive loadings. In the loadings plot, for all the variables, which have positive loadings, C1 and C5 had higher than average values for these variables. In conclusion, the variance between replicate CAF Raman spectra was reduced by unit vector normalisation, despite being influenced by wide-range S/N. The unit vector normalised model was not valid for cross-validated samples.

4.3.4.4.4. Mean Normalisation

Unlike area and unit vector methods, where normalisation algorithms were applied to the entire dataset, mean normalisation normalises individual spectra by dividing all intensities in the spectrum by the average of intensities, thus profiling the relative intensities around 1 [172]. The line plots of the mean normalised spectra for C1 to C6 (Set '2') are shown in Figure 4.16d.

Magnified peaks (not shown) displayed absolute intensities in the descending order of the S/N of the CAF spectra (C1 > C5 > C4 > C3 and C2 > C6).

The scores plot (Figure 4.17g) explained about 91 % of the variance in the data, where one PC explained 83 % calibrated variance and 72 % validated variance. These results were similar to area normalisation of CAF spectra (Section 4.3.4.4.2). The model was explained by three PCs before plateauing. Compared to Figure 4.17a, it was observed that the coordinates of the scores plot were reduced by mean normalisation, showing that mean normalisation reduced the distance between the samples (distance between C1 and C6 is 22 scores along PC1). C1 has a longer distance from the mean along PC1 and a higher score (12 scores) than C6 (-10 scores). The scores plot showed that the variance in both C1 and C6 was explained by PC1. With mean normalisation of the Raman data, the locations of both C1 and C6 demonstrates that C1 is negatively correlated to C6 and is separated from C6 by the longest distance along PC1 among all normalisation methods. This is probably due to differences in intensities in both spectra. C3 and C5 are explained by PC2. Both C2 and C3 are positively correlated.

The scores plot showed that all CAF samples lie in the model with 95 % CL. The loadings plot (Figure 4.17h), showed the important spectral variables for PC1 (peaks at ca. 554, 742, 1329 and 1597 cm^{-1}), which correlate to C1 based on its position in the scores plot. The loadings plot also showed the important variables for PC2 (peaks at ca. 552 cm^{-1} (positive loadings for C3), 1328, 1700 and 2969 cm^{-1} (negative loadings for C5)). C1 has high positive scores, hence has high values for variables with positive loadings. In the loadings plot for all the variables which have positive loadings for PC1, C1 has higher than average values for these variables. In conclusion, the variance between replicate CAF Raman spectra was reduced by mean normalisation, despite being influenced by wide-range S/N. However, mean normalisation has produced the largest variance between replicate CAF spectra across PC1 in comparison to other normalisation methods.

4.3.4.4.5. Maximum Normalisation

Similar to mean and range normalisation, the maximum normalisation algorithm pre-processes individual spectra by normalising the Raman response to a chosen peak (the maximum peak) at a fixed Raman position and laser wavelength. The peak remained unchanged in all spectra. This is performed by dividing each spectrum by the maximum absolute value in the spectrum. The maximum peak in all spectra becomes equivalent to 1 or unity and, hence was not included

in the PCA model. In this dataset (C1 to C6 from Set '2'), the maximum peak was at 554.4 cm^{-1} for all spectra except for C5, where the maximum peak was at 555.4 cm^{-1} . This difference in peak position is insignificant in comparison to band widths for major peaks. For example, the data points 554.0 cm^{-1} and 555.0 cm^{-1} in a CAF spectrum are estimated to be very close at the top of the peak since the band width is about 60 cm^{-1} (Figure 4.15).

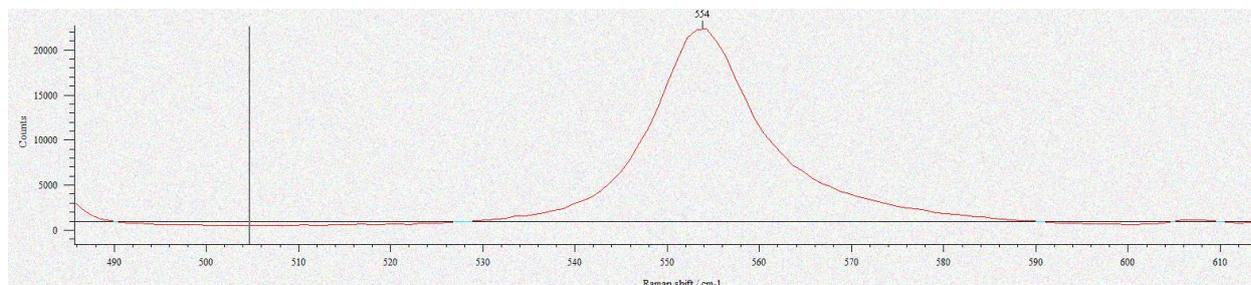


Figure 4.15: A CAF peak at 554 cm^{-1} (WiRE 3.4)

Changes in peak positions could possibly be due to changes in sample morphology resulting from mixing and grinding of the powders [135, 136]. Crystallinity of powders, lattice stress factors, bond contracting or interactions between mixture constituents may also induce band shifting [267].

The line plots of the maximum normalised spectra are shown in Figure 4.16d. Magnified peaks (not shown) displayed variable absolute intensities between spectra, but not in a specific order. Different slopes were observed between peaks in the regions 1492 - 1323 cm^{-1} and 300 - 100 cm^{-1} . This could be due to multiplicative scatter effects. The PCA results of the maximum peak normalised data showed that the variance between the samples was reduced.

The scores plot (Figure 4.17i) explained about 88 % of the variance in the data, where one PC explained 75 % calibrated variance and 61 % validated variance. The model was explained by three PCs before plateauing. Compared to Figure 4.17a, it was observed that the coordinates of the scores plot were reduced by maximum normalisation, showing that maximum normalisation reduced the distance between the samples (distance between C1 and C6 is 0.69 scores along PC1). The scores plot showed that the variance in both C1 and C6 was explained by PC1. C1 has a slightly longer distance from the mean along PC1 (scores of -0.37 scores) than C6 (0.33 scores). With maximum normalisation of the Raman data, the locations of both C1 and C6 demonstrates that C1 is widely separated from C6. C5 is explained by PC2. Both C2 and C3 are positively correlated.

The scores plot showed that all CAF samples lie in the model with 95 % CL. The loadings plot (Figure 4.17j) showed the important spectral variables for PC1 (peaks at 162.1 cm^{-1} (positive loadings for C6), 557, 1330, 1600 and 1701 cm^{-1} (negative loadings for C1)). The loadings plot also showed the important variables for PC2 (peaks at 560, 1331, 1600 and 1701 cm^{-1} (positive loadings for C5). C1 has high negative scores, hence has high values for variables with negative loadings. In the loadings plot for all the variables which have negative loadings, C1 has higher than average values for these variables. In conclusion, the variance between replicate CAF Raman spectra was reduced by maximum normalisation, despite being influenced by wide-range S/N.

4.3.4.4.6. Range normalisation

Similar to mean normalisation, the range normalisation algorithm pre-processes individual spectra. Range normalisation normalises spectra by dividing all intensities in individual spectra by the range of intensities i.e. maximum value minus minimum value [172]. The line plots of the range normalised CAF spectra (C1 to C6 from Set '2') are shown in Figure 4.16f. Slopes were observed over the ranges 1492 - 1323 cm^{-1} and 300 - 100 cm^{-1} . This could be due to multiplicative scatter effects. The absolute intensity values at the maximum peak (at 554.4 cm^{-1}) in all CAF spectra ranged from 0.9986 to 1.000. C5 again was an exception, where the maximum range normalised intensity value (0.9995) was at 555.4 cm^{-1} (see Section 4.3.4.4.5.). Magnified peaks (not shown) displayed differences in intensities among remaining peaks. The PCA results of the range normalised data was very similar to that of maximum normalisation, and hence, are not discussed in details (Figures 4.17k and 4.17l). The PCA results demonstrated that the difference in peak position for C5 was insignificant.

4.3.4.4.7. Peak normalisation (peak no. 3357 at 554.4 cm^{-1})

Similar to area and unit vector normalisation, the peak normalisation algorithm normalises the entire dataset according to one reference peak (a single wavenumber), which must be present in all spectra [172]. Peak normalisation sets the maximum response or absolute intensity of the highest peak to 1 [172, 229]. CAF spectra (C1 to C6 from Set '2') were peak normalised to the peak no. 3357 at 554.4 cm^{-1} . After the application of peak normalisation to C1 to C6 (Set '2'), slopes were observed over the ranges 1492 - 1323 cm^{-1} and 300 - 100 cm^{-1} , which could be due to multiplicative scatter effects. The absolute intensity values at the selected peak (at 554.4 cm^{-1}) in all CAF spectra became 1. Therefore, this column was excluded from the PCA analysis as it contained a constant number (i.e. no variance). The line plots of the peak normalised CAF

spectra are shown in Figure 4.16g. Magnified peaks (not shown) showed differences in intensities among remaining peaks. The PCA results for the peak normalised data was very similar to that of both the maximum and range normalisation because the same peak was employed for this pre-processing. Therefore, the PCA results of the peak normalised spectra is not discussed in details (Figures 4.17m and 4.17n).

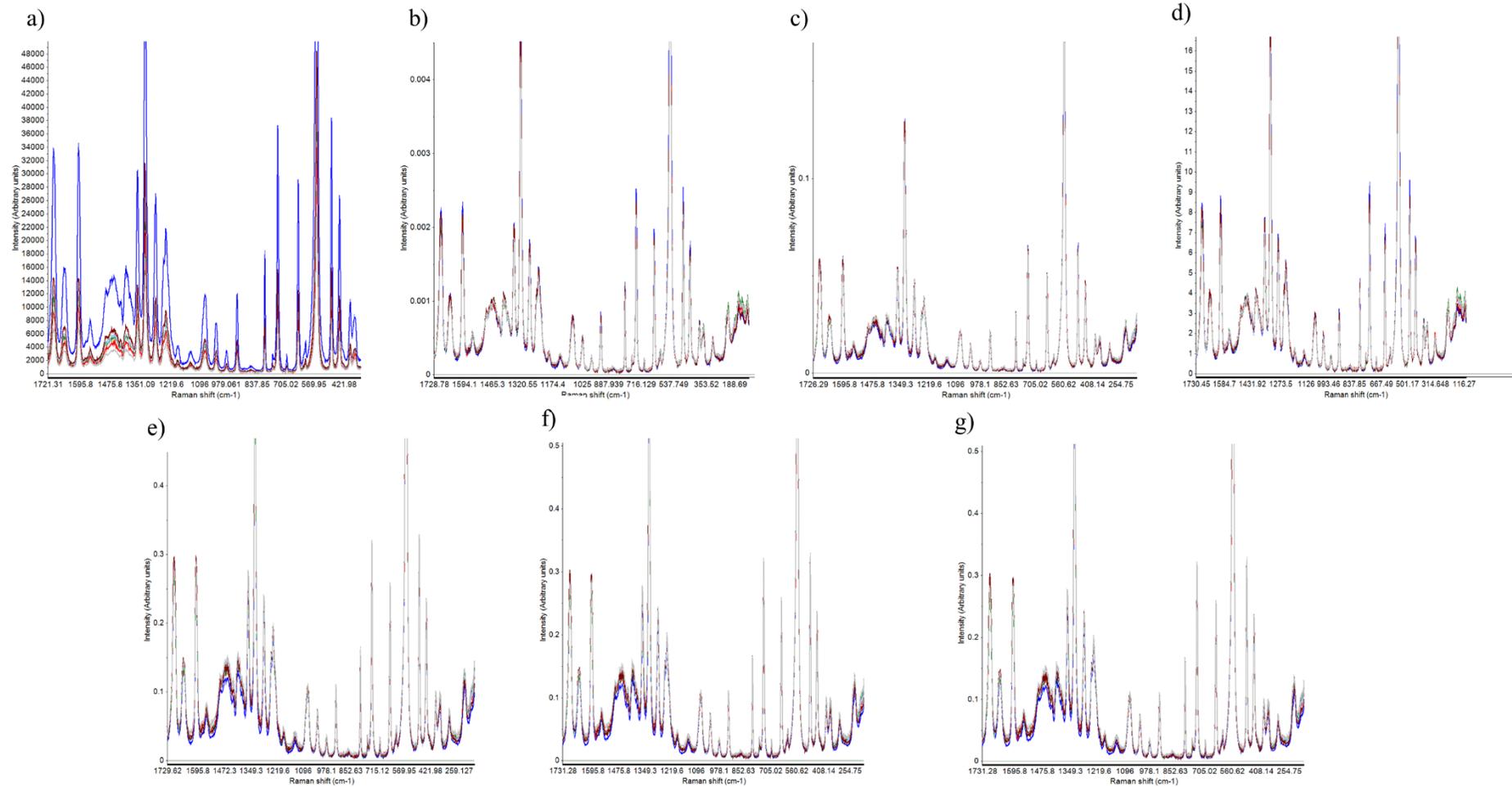


Figure 4. 16: Line plots of normalised and non-normalised CAF spectra (C1 to C6 from Set ‘2’) (Unscrambler® X 10.3). a) Raw spectra, b) Area normalisation, c) Unit vector normalisation, d) Mean normalisation, e) Maximum normalisation, f) Range normalisation and g) Peak normalisation. In all Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.

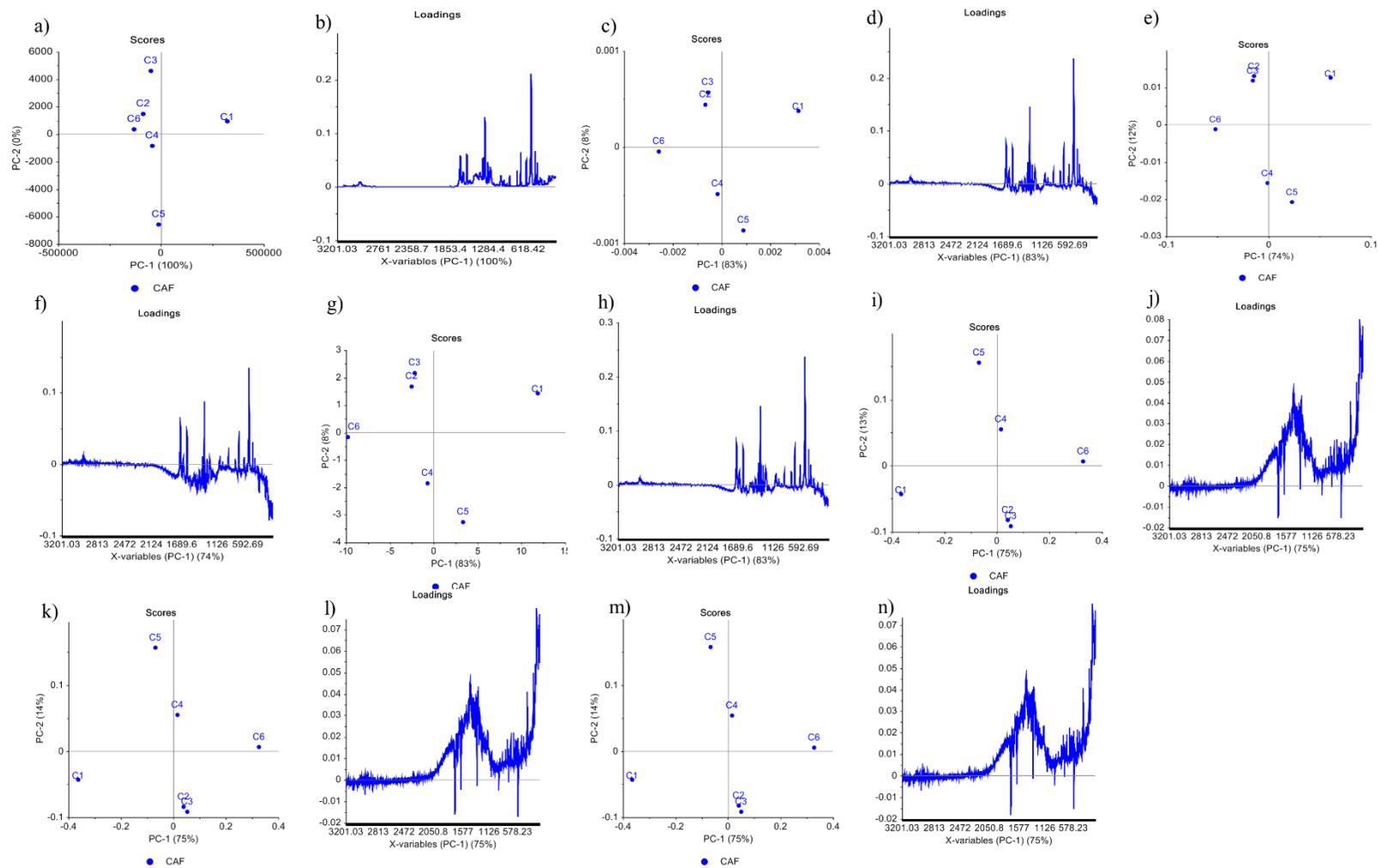


Figure 4. 17: PCA scores and loading plots summarising the normalisation types of six CAF Raman spectra (C1 to C6 from Set ‘2’) in comparison with raw spectra (Unscrambler® X 10.3). (a - b) raw spectra; (c - d) area normalisation; (e - f) unit vector normalisation; (g – h) mean normalisation; (i - j) maximum normalisation; (k - l) range normalisation; (m - n) peak normalisation.

In summary, both area and mean normalisation explained greater calibrated and validated variance. The difference between both values was insignificant in comparison to other normalisation methods. Of all normalisation methods, area normalisation has induced clustering between like substances (replicate CAF spectra), whereas mean normalisation produced the greatest separation between the samples. In contrast, the model produced by unit vector normalisation has been shown to be invalid for new samples due to the large difference between both calibrated and validated variance. Range, maximum and peak normalisation produced similar PCA results since all samples consisted of pure CAF, hence the maximum peak was the same peak employed in range, maximum and peak normalisation. In all the normalisation methods discussed, the influence plot (not shown) with T^2 (PC1) illustrated that all CAF samples fit in the model. The leverage plot (PC1) (not shown) illustrated that both C1 and C6 have a higher leverage than remaining CAF samples, hence have a strong influence on the model than remaining samples.

In conclusion, mean normalisation was found to outperform other normalisation methods for varied datasets (e.g. Set '2') because it normalised individual spectra and produced the largest separation between the samples. In contrast, area normalisation was found to outperform other normalisation methods for replicate samples because it normalises the entire dataset to a constant and reduced the separation between replicate samples. Mean and area normalisation are differentiated with a constant multiplicative factor for the scores plot coordinates [172]. Therefore the spectra for Set '2' were mean normalised and a PCA was generated to evaluate if it would improve the PCA classification of a varied dataset that include replicate spectra (e.g. CAF, BEN and CCL spectra).

Set '2' was mean normalised, then the line plots were visually inspected and compared to that of the raw spectra (Figure 4.18a). They showed that replicate spectra for both CAF and BEN were fully overlaid. However, the CCL spectra were noisy and not fully overlaid, which is expected for mixtures (Figure 4.18b). The PCA scores plot (Figure 4.18c) explained 98 % of the variance in the data, where one PC explained 79 % and 71 % calibrated and validated variance respectively. The model was explained by two PCs before plateauing. All samples lied in the ellipse at 95 % CL. BEN samples were not dominating the model as previously demonstrated (PCA-2, Section 2.2.1). The scores plot showed that the spectra for Set '2' were separated into three groupings: (1) CAF and CCL 1, 2 and 6; (2) B2 and B3; (3) CCL 3-5.

Upon magnifying the scores plot, it could be seen that CAF samples were also separated from CCL mixtures 1, 2 and 6. The scores plot showed that the variance in CAF samples and CCL 3 to 5 mixtures (spectra with higher CAF composition) were explained by PC2. These samples have high positive scores and high values for variables with positive loadings. The variance in both B2 and B3 was explained by PC1. CCL mixtures 3-5 were separated from CCL mixtures 1, 2 and 6 as well as CAF samples. This is explained by the fact that the former mixtures have more prominent peaks for CRE and LAC rather than CAF. The loadings plot (Figure 4.18d) showed the important spectral regions for PC1, which mainly correlated to CAF samples. In the loadings plot (PC1), both B2 and B3 have higher than average values for all the variables which have positive loadings.

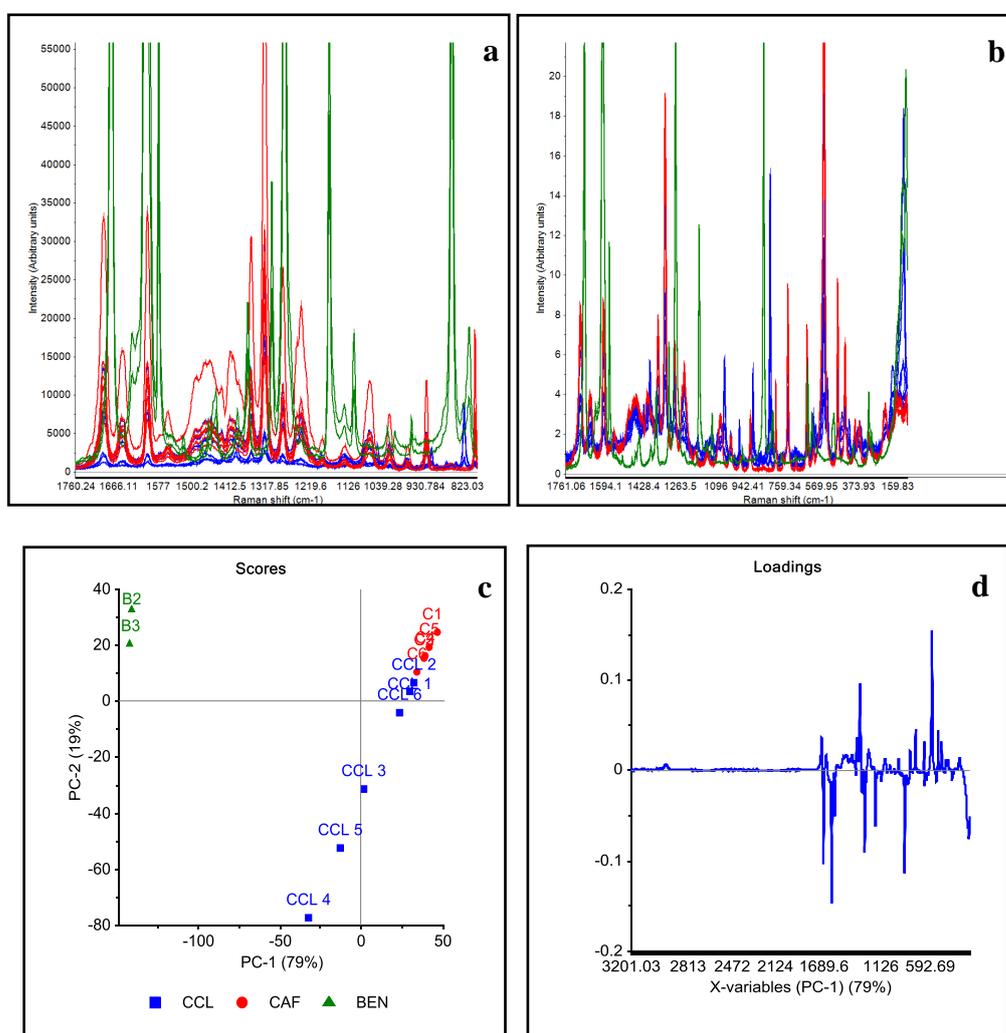


Figure 4. 18: Mean normalisation of Set '2' (Unscrambler® X 10.3). a) line plots of three groupings (CCL (blue), CAF (red) and (BEN (green))) for raw Raman spectra (Set '2'); b) line plots of three groupings (CCL (blue), CAF (red) and (BEN (green))) for mean normalised Raman spectra (Set '2'); c) The PCA 2D-scores and d) the line loading plots.

In this pilot study, mean normalisation was considered in the final combination set of pre-processing methods (Section 4.3.5.) because it explained greater calibrated and validated variance, produced greater separation between groupings and maintained positive correlations between replicate spectra in the PCA model. Close calibration (79 %) and cross-validation (71 %) values showed that the model could be applied to unknown samples.

4.3.4.5.Scatter effects

Descriptive analysis has demonstrated that the spectra exhibit multiplicative scatter effects (MSEs) (Figure 4.9), which could be corrected using standard normal variate (SNV), multiple scatter correction (MSC), normalisation or derivatives [172]. In contrast, spectra with additive scatter effects may benefit from other transformations/ pre-processing such as baseline correction, derivatives, SNV or MSC [172]. In this Section SNV and MSC were investigated for the correction of MSEs.

4.3.4.5.1. Standard Normal Variate (SNV)

SNV corrects for both multiplicative effects which induces fluctuations in absolute intensities, and additive effects, which induces baseline offsets. This is performed by subtracting the average of the spectrum from the data matrix (i.e. making it ‘zero-mean’) and dividing the result by the standard deviation (i.e. making it a ‘unit variance’). In other words, it is a centering and scaling pre-processing tool [263]. Following the application of SNV, pre-processing to both C1 to C6 (Set ‘2’), the visual inspection of CAF spectra before and after pre-processing showed that multiplicative scatter was reduced over the range 1692 - 1377 cm^{-1} and 420 - 100 cm^{-1} (Figure 4.19). However, the baseline between these two regions dropped below zero. The effect shown on the latter ranges may not be a reduction in MSEs but a normalisation effect resulting from the scaling properties of SNV.

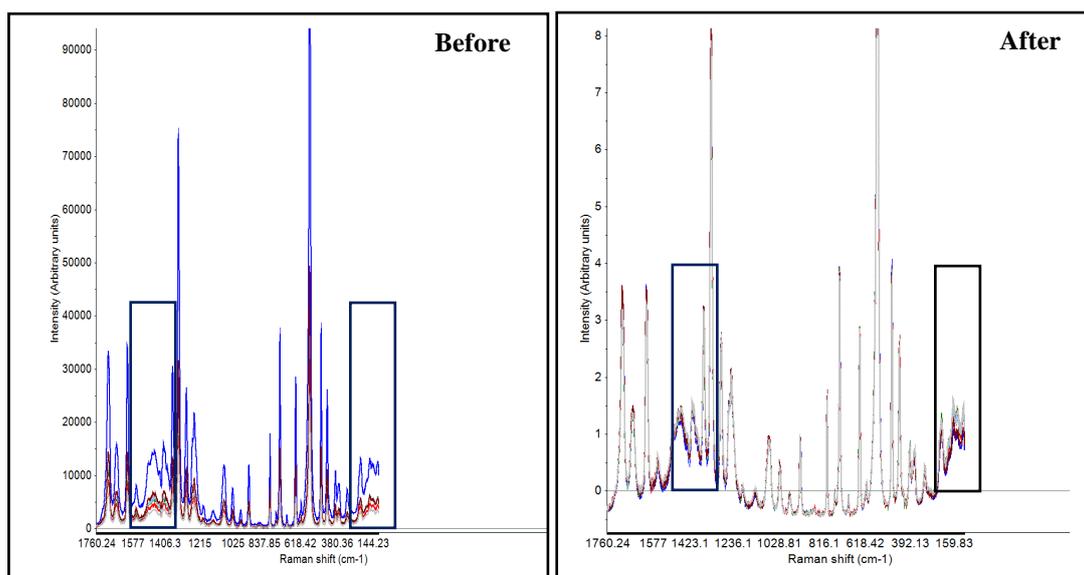


Figure 4. 19: Raman spectra for C1 to C6 (Set ‘2’) before and after pre-processing using SNV (Unscrambler® X 10.3). In both Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.

Set ‘2’ was then pre-processed using SNV. A PCA was generated and compared to that of the raw spectra (PCA-2, Section 4.3.2.1) to evaluate the effect of this pre-processing method. The PCA scores plot (Figure 4.20) showed an improved separation between four groupings: (1) CCL 3 to 5; (2) CCL 1, 2 and 6; (3) C1 to C6; and (4) B2 and B3 as compared to the raw spectra. However, it showed great similarity to the PCA produced by mean normalisation (Figure 4.18). CAF peaks dominated the spectra for grouping 2 and, hence both groupings 2 and 3 were positively correlated. Mixtures CCL 3 to 5 featured peaks for CAF as well as for CRE and LAC, and, hence were grouped together but not in a tight cluster.

The three-PC model explained 99 % of cumulative calibrated variance and 97 % of cumulative cross-validated variance. The variance in grouping 4 was explained by PC1, whereas the variance in grouping 1 was explained by PC2. Both groupings 2 and 3 have not contributed significantly to explaining the variance by PC1 or PC2. Results have demonstrated that SNV improved substance separation using PCA possibly by correcting for MSEs or by its scaling effect. Despite that, SNV transformation did not improve the percentage for both calibrated and validated variance compared to that of the raw spectra (Set ‘2’). The loadings and scores plots showed improved PCA classification and identification of important chemical information. Therefore, SVN would be considered in the final combination set of pre-processing methods.

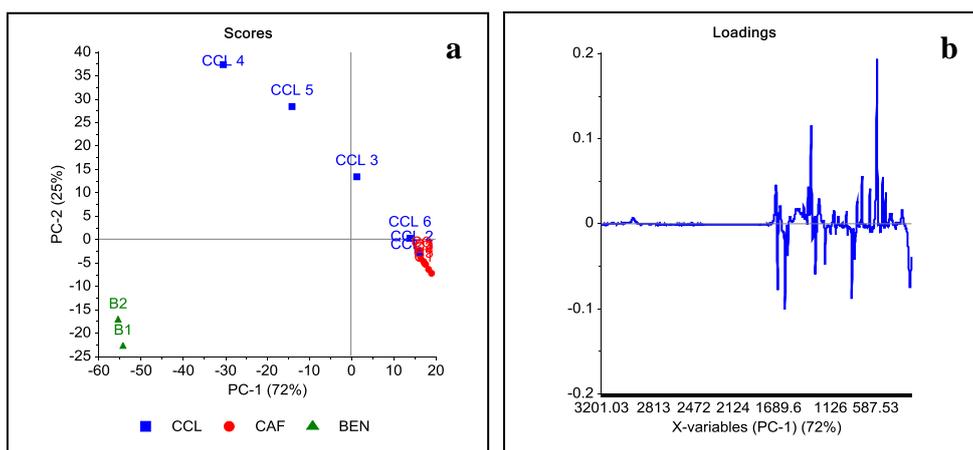


Figure 4. 20: SNV of Set ‘2’. a) 2D-scores and b) line loading plots (Unscrambler® X 10.3).

4.3.4.5.2. Multiple scatter correction (MSC)

MSC corrects for both additive and multiplicative scatter effects. The scatter is usually modelled against a reference spectrum that is calculated from the mean sample [116]. MSC uses a linear regression (with two correlation coefficients: the slope and the intercept) to bring all samples as close to the reference spectrum. MSC may remove important chemical information as regression is based on a reference sample [172] that is automatically selected by the Unscrambler® X 10.3 software from the dataset. Therefore it is preferred to use MSC with replicate spectra rather than varied datasets. The inspection of the line plots of the MSC corrected plots for C1 to C6 from Set ‘2’ showed that slopes were reduced as compared to raw spectra. The visual inspection of CAF spectra before and after pre-processing showed that multiplicative scatter was reduced over the range 1692 - 1377 cm^{-1} and 420 - 100 cm^{-1} (Figure 4.21). The line plots have shown better baseline offset correction in comparison to the pre-processing using SNV to the same spectra.

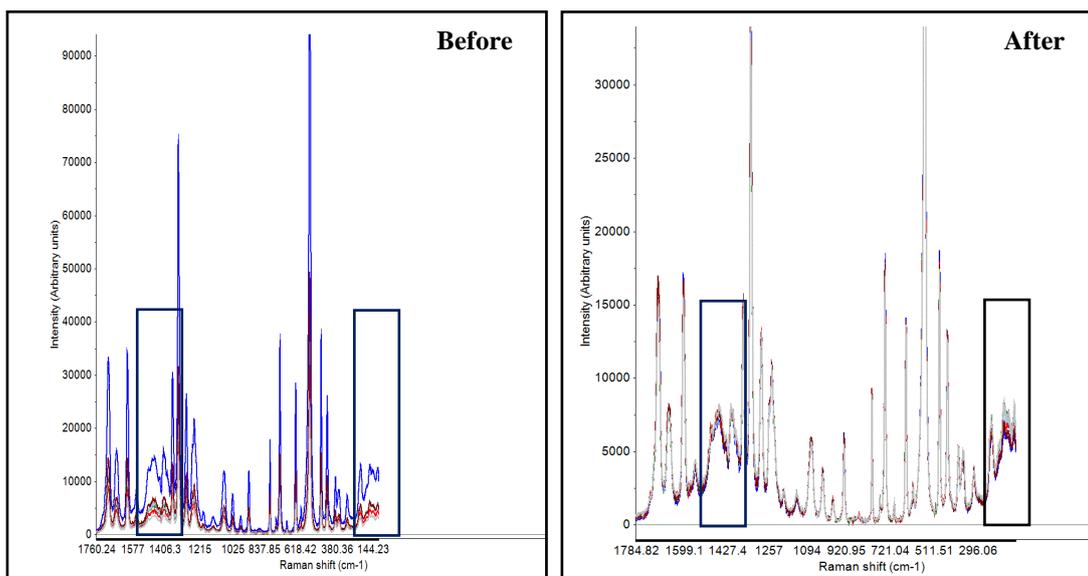


Figure 4. 21: Six CAF spectra before and after pre-processing with MSC (Unscrambler® X 10.3). In both Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.

Following the application of MSC, a PCA for Set ‘2’ was generated and compared to that of the raw data (Figure 4.22). The PCA model results showed that MSC has induced a greater separation between like samples. This was illustrated with the 10,000 fold increase in the coordinates of the scores plot as compared to SNV scores plot (Figure 4.20a). This could be explained by the fact that MSC is not a suitable algorithm for varied datasets since the mean spectrum is not representative for all samples and may be different from the true mean spectrum (i.e. the mean spectrum of CAF would not be representative to BEN samples). Therefore MSC was not considered in the final combined pre-processing methods.

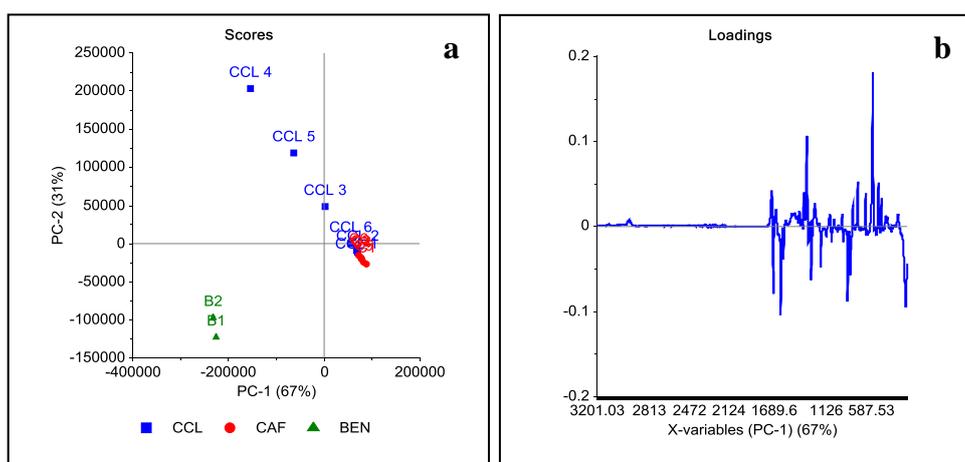


Figure 4. 22: MSC of Set ‘2’. a) 2D-scores and b) line loading plots (Unscrambler® X 10.3).

4.3.5. Combined pre-processing methods

Combined preprocessing methods are commonly undertaken prior to the generation of PCA models for Raman spectra of drug mixtures [154, 184]. In this Section, spectra were examined for regions with baseline offsets, scatter effects and noise after the application of each combined pre-processing method [95]. The effect of combined pre-processing sequences was evaluated if it produces tight clusters of replicate spectra and maximum separation between different samples in the PCA scores plots. The percentage of explained calibrated and validated variance for each model and the closeness between their values was also evaluated. Furthermore, the order in which combined pre-processing methods is used will be assessed.

From the results obtained thus far, it has been established that selected pre-processing was further investigated in combined protocols to correct for noise, baseline offsets, multiplicative scatter effects and variable absolute intensities. These include: smoothing with polynomial order of 1 and smoothing segments of 3 and 5 respectively, baseline offset correction method, first and second derivatives, standard normal variate and mean normalisation.

Optimum performance for both mean and maximum normalisation algorithms is achieved when the values in the data matrix have the same sign, according to the Unscrambler® X 10.3 (2015) software guide (Camo, Norway) [268]. Therefore mean normalisation was preceded by baseline offset as the latter transforms the values in the data matrix into values with positive sign. For this reason, combinations including derivatives instead of baseline offsets correction method would include range normalisation as an alternative to mean normalisation because the data matrix would include values with both positive and negative signs. In addition, range normalisation is an algorithm that is calculated from individual spectra rather than the entire dataset at once. For these reasons, range normalisation is preferred over area normalisation when calculating the first or the second derivatives. Spectra for pure CAF (C1 to C6 from Set '2') were examined for regions with noise, baseline offsets, scatter effects and variable peak intensities after the application of each pre-processing sequence. A PCA was then generated for each method to evaluate substance classification and reduced noise.

Pre-processing sequences investigated are summarized below:

Method 1: Raw data → Smoothing (1-3) → SNV → Baseline offset → Mean normalisation → PCA

Method 2: Raw data → Smoothing (1-5) → SNV → Baseline offset → Mean normalisation → PCA

Method 3: Raw data → SNV → Baseline offset → Mean normalization → PCA

Method 4: Raw data → SNV → Baseline offset → PCA

Method 5: Raw data → Baseline offset → Mean normalization → PCA

Method 6: Raw data → SNV → First derivative → Range normalization → PCA

Method 7: Raw data → SNV → Second derivative → Range normalization → PCA

Compared to the line plots for CAF raw spectra (Figure 4.3a), methods 1-3 and 5 (Figures 23b-d and 23f), gave similar line plots showing overlaid scaled replicate spectra with slightly improved baseline correction and a reduction in MSEs. Method 2 outperformed methods 1, 3 and 5 in reducing the noise. The line plots for methods 1-3 and 5 demonstrated the dominant effect of both baseline offset correction and mean normalisation over smoothing and SNV. In contrast, method 4 showed slight amplified intensities over the regions 1692 - 1377 cm^{-1} and 420 - 100 cm^{-1} , demonstrating the importance of normalisation. Method 6 illustrated the resolution of major peaks but with potential loss of chemical information (absence of peaks after the pre-processing sequence was conducted). Finally, method 7 clearly resolved few major peaks but has been shown to increase the noise in the spectra.

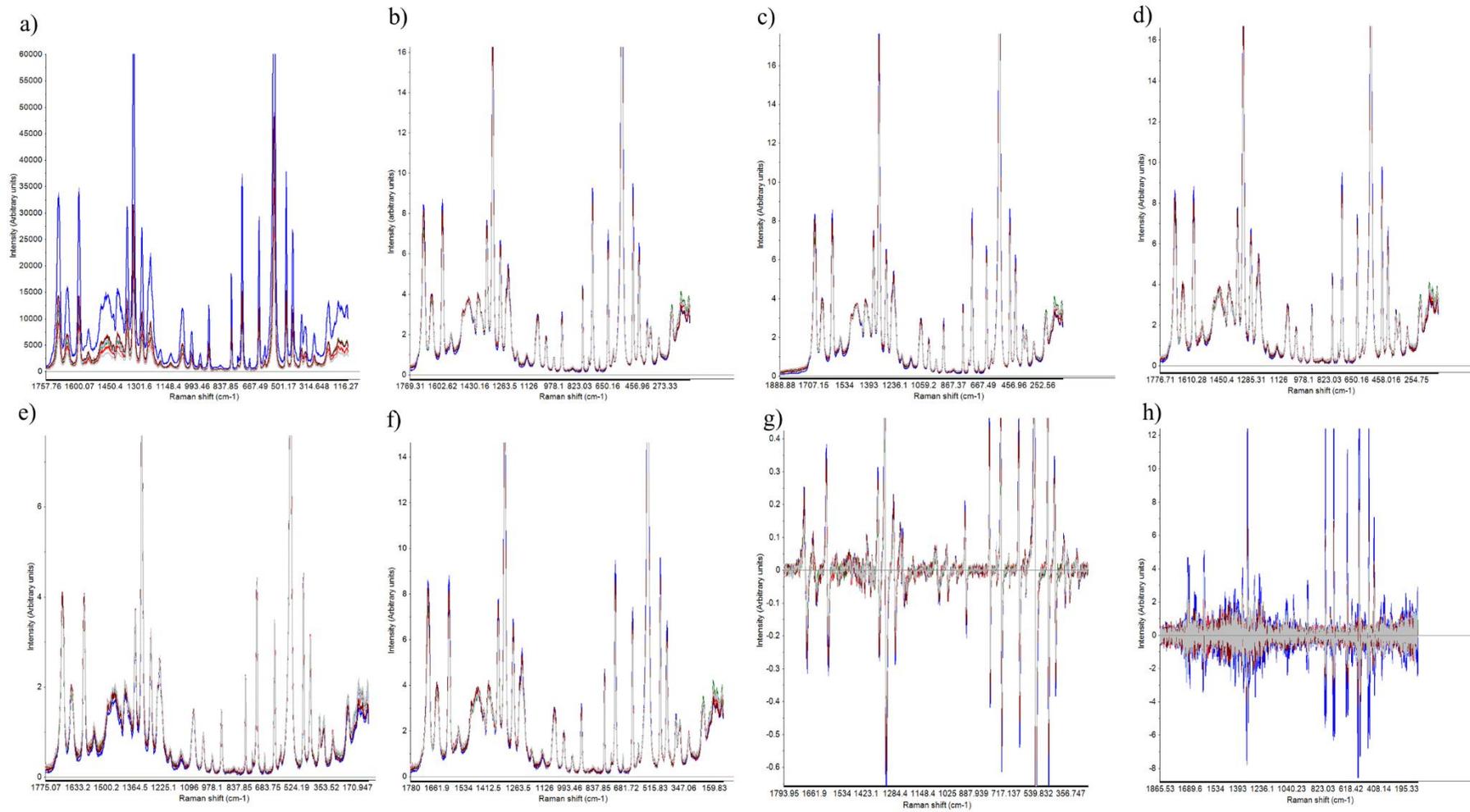


Figure 4. 23: Line plots of combined pre-processing methods using six CAF spectra (C1 to C6 from Set ‘2’) (Unscrambler® X 10.3). a) Raw spectra; b) Method 1; c) Method 2; d) Method 3; e) Method 4; f) Method 5; g) Method 6; and h) Method 7. In all Figures: C1: blue; C2: red; C3: green; C4: light blue; C5: maroon; and C6: grey.

Despite the fact that spectra showed improved flat baselines and less multiplicative scatter as compared to the raw spectra, visual inspection cannot give an accurate judgement on the pre-processing performed. A PCA was generated for each pre-processing sequence for Set '2' to investigate the effect of each sequence on the PCA performance.

Methods 1-3 and 5 gave similar PCA results with 79 % and 19 % of the calibrated variance explained by PC1 and PC2 respectively. The raw data for both scores and explained variance plots was very similar for all four methods. This confirmed that the effect of both baseline offset correction and mean normalisation was more significant than both smoothing and SNV. The scores plots for these four methods illustrated the classification of Set '2' into four groupings: 1) BEN; 2) unclustered grouping for CCL3 to 5; 3) C1-C6; and 4) CCL1, 2 and 6 (Figures 4.24b-d and 4.24f). Groupings 3 and 4 were positively correlated groupings. This is because the Raman spectra in grouping 4 include more prominent CAF peaks than that of LAC and CRE. Smoothing was not a prerequisite in the analysis because the S/N ratio was adequate prior to the application of smoothing transformations. Smoothing has been shown to be more suitable to use in classification studies (i.e. SIMCA etc.) [262]. The scores (Figures 4.24b-d and f) and loadings plots (not shown) illustrated that the variance in BEN samples (i.e. B2 and B3) was explained by PC1, whereas the variance in CCL 3 to 5 was explained by PC2. In these four methods, the samples lied in the ellipse at 95 % CL.

The PCA scores plot for method 4 showed that the samples were similarly grouped into four groupings 1) BEN; 2) unclustered grouping for CCL3 to 5; 3) C1-C6; and 4) CCL1, 2 and 6 (Figure 4.24e). The explained variance plot showed that 68 and 28 % of the calibrated variance was explained by PC1 and PC2 respectively. A large difference between the explained calibrated (68 %) and validated (49 %) variance showed that the model was not valid and may not be applied to new samples. This model emphasised the importance of normalisation for this dataset.

The PCA scores plot for method 6 showed a cluster forming between pure CAF (C1-C6) and the mixtures CCL1, 2 and 6 and a reduction in the distance between the scores (Figure 4.24g). This was demonstrated by the reduced variance and loss of orthogonality between the variables in the 2-D loadings plot (Figure 4.24h). The explained variance plot showed that 68 and 25 % of the calibrated variance was explained by PC1 and PC2 respectively. The variance in BEN samples (i.e. B2 and B3) and CCL 3-5 was explained by PC1 and PC2 respectively. A large

difference between the explained calibrated (68 %) and validated (56 %) variance showed that the model was not valid and may not be applied to new samples.

The PCA scores plot for method 7 showed loss of the groupings (Figure 4.24i). This method was unable to classify the samples according to their properties. This was demonstrated by the reduced coordinates in the scores plot, reduced variance and clustering of the variables around the mean in the 2D-loadings plot (Figure 4.24j). The explained variance plot showed that only 26 and 20 % of the calibrated variance was explained by PC1 and PC2 respectively. The largest difference between the explained calibrated (25.77 %) and validated (0.97 %) variance was observed by applying this method which demonstrates that the model was not valid and may not be applied to new samples. The farther the validation values from the calibration values, the more inaccurate the explanations of the variance by the PCs become [148].

In summary, the optimum pre-processing method for this dataset (Set '2') was method '5' (baseline offset correction, followed by mean normalisation of Raman spectra). The combination of both baseline correction and normalisation in combination with other pre-processing methods corroborates with the research carried out on pharmaceuticals [148, 182, 269] and paint [179]. Pre-processing using baseline offset correction followed by normalization corroborates with the research undertaken by Afseth et al. [261] for biological samples.

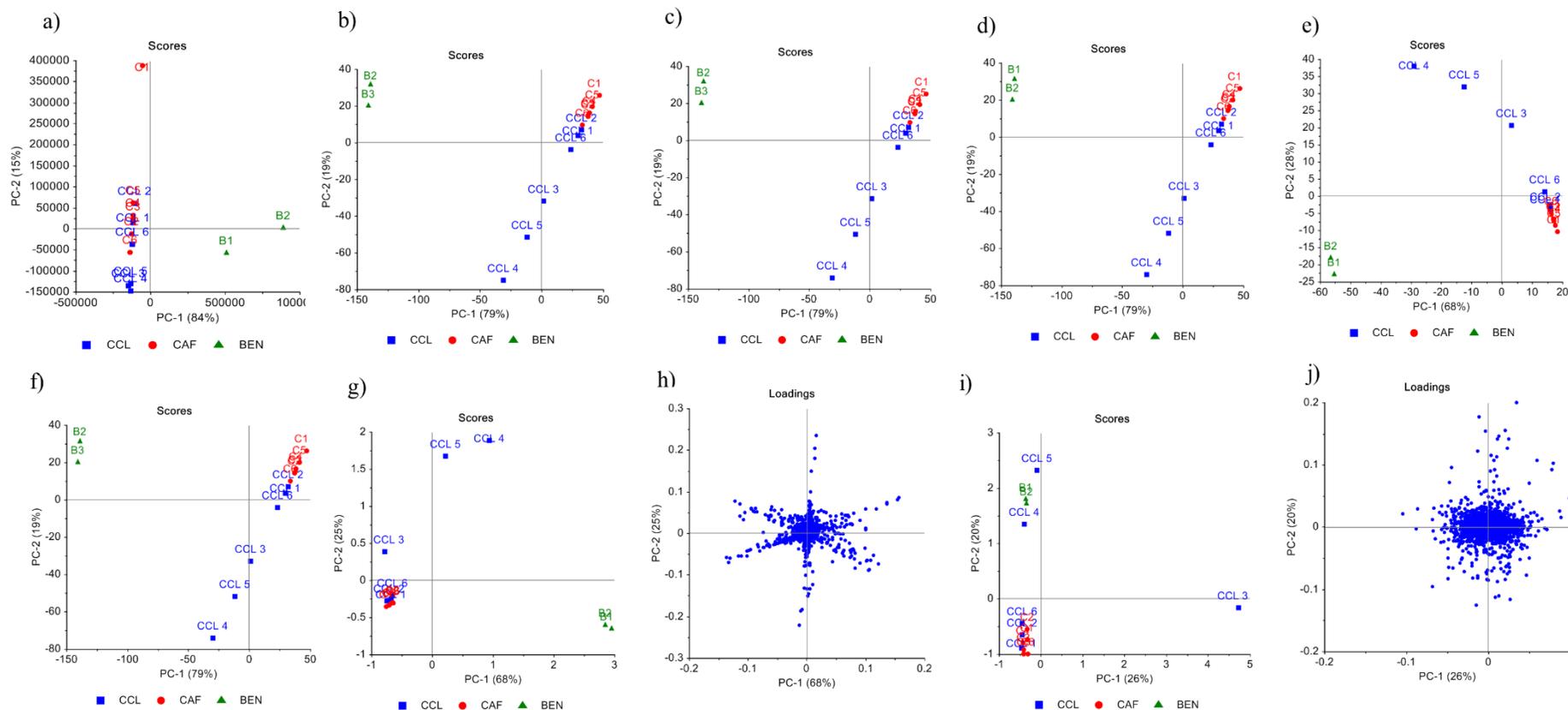


Figure 4. 24: Summary of PCA results for the combined pre-processing methods (Unscrambler® X 10.3). a) the scores plot for the raw spectra (Set '2'); b) the scores plot for method 1; c) the scores plot for method 2; d) the scores plot for method 3; e) the scores plot for method 4; f) the scores plot for method 5; g) the scores plot for method 6; h) the loadings plot for method 6; i) the scores plot for method 7; j) the loadings plot for method 7.

In summary, the best classification for Set '2' was achieved with pre-processing combination methods 1, 2, 3 and 5 with total explained variance of 98 % for PC1/PC2. However, method '5' is considered the optimum method to avoid over-pre-processing since smoothing and SNV had insignificant effects in these sequences for this dataset. Therefore the robustness of method '5' was tested using validation Set '3'.

4.3.5.1. Method validation

Validation of the optimised pre-processing approaches was assessed using an independent subset of samples (i.e. Set '3') in order to ensure that the model was not over-fitted to the data [172, 270]. The line plots for Set '3' (not shown) were inspected using the process described in Figure 4.1. Cosmic rays were zapped from C7, C8 and B8 using WiRE 3.4. BEN 4, 7 and 9 were discarded from the analysis because the detector was saturated during the measurements of these spectra. The line plots showed variations in absolute intensities between replicate spectra and baseline offsets. Both datasets '2' and '3' were interpolated from 3201.03 - 100.55 cm^{-1} to 3200.95 - 100.41 cm^{-1} because of different x-axes. Subsequently, both datasets were imported into the Unscrambler® X 10.3. Descriptive statistics (not shown) showed MSEs in both CAF and BEN test sample groups. It was noted that the spectra in Set '3' showed instrumental, sample, analysis and environmental interferences similar to set '1'. This is because these artefacts are generally common in Raman spectra [188]. In addition, these samples (i.e. Set '3') were measured using the same Raman instrument and reference standard batch (for CAF only). However, the significance of these interferences and the quality of the raw spectra for Set '3' were different from that of set '1'.

Method '5' was employed to pre-process both calibration (Set '2') and validation (Set '3') sets, then a PCA was generated to evaluate whether the developed model (i.e. method '5') will efficiently classify Set '3'. A two-PC model explained 98 and 97 % of the cumulative calibration and validation variance respectively. The 3D-scores plot (Figure 4.25) showed that the dataset was split into three major groupings: 1) BEN validation samples (Set '3') clustered with calibration samples (Set '2'); 2) CAF validation samples (Set '3') clustered between both calibration CAF and CCL 1, 2 and 6 samples (Set '2'); 3) CCL 3-5 (Set '2'). As described above, the spectra of CCL 1, 2 and 6 samples included high composition of CAF peaks, and hence had positive correlations with CAF samples. Little variance was explained by this group i.e. the mixtures CCL 1, 2 and 6 and the CAF samples.

With reference to the scores plot, the loadings plot (not shown) again illustrated that BEN calibration and validation samples had higher than average values for the positive loadings (PC1). In contrast, CCL 3-5 samples had higher than average values for the negative loadings (PC2). All samples lied in the model with 95 % CL except CCL4. This sample was further investigated using the influence and leverage plots (not shown), both of which showed that this sample fits in the model, and, hence is not an outlier. However, it has a high leverage for PC2 and, hence is influencing the model for the explained variance by PC2 (11 %). This is shown by its increased distance from the mean in the scores plot as compared to both CCL 3 and CCL 5. It is not fully understood why CCL4 had a high leverage. CCL had the second lowest S/N in the dataset. Unlike CCL 4 and CCL 5, the Raman spectral analysis of CCL 4 showed a high composition of prominent band for CRE (e.g. torsion vibration $\rho_t(\text{CH}_3)$ 740 cm^{-1} ; (wagging vibration $\rho_w(\text{R-NH}_2)$ 830 cm^{-1}); (asymmetric stretching vibration $\nu_{\text{as}}(\text{CN})$ and/ or stretching vibration $\nu(\text{R-NH}_2)$) 1052 cm^{-1} ; (symmetric stretching vibration $\nu_s(\text{COOH})$, vibration $\nu(\text{CN})$, and/ or deformation $\delta(\text{CN})$) 1397 cm^{-1} ; and deformation $\delta(\text{CH}_2)$ 1424 cm^{-1}) [271]. The PCA scores plot did not show any discrimination between both BEN batches (i.e. from datasets ‘2’ and ‘3’). The overlap between some CCL mixtures and CAF samples showed the limitations of MVA as a sole tool for the classification of complex mixtures.

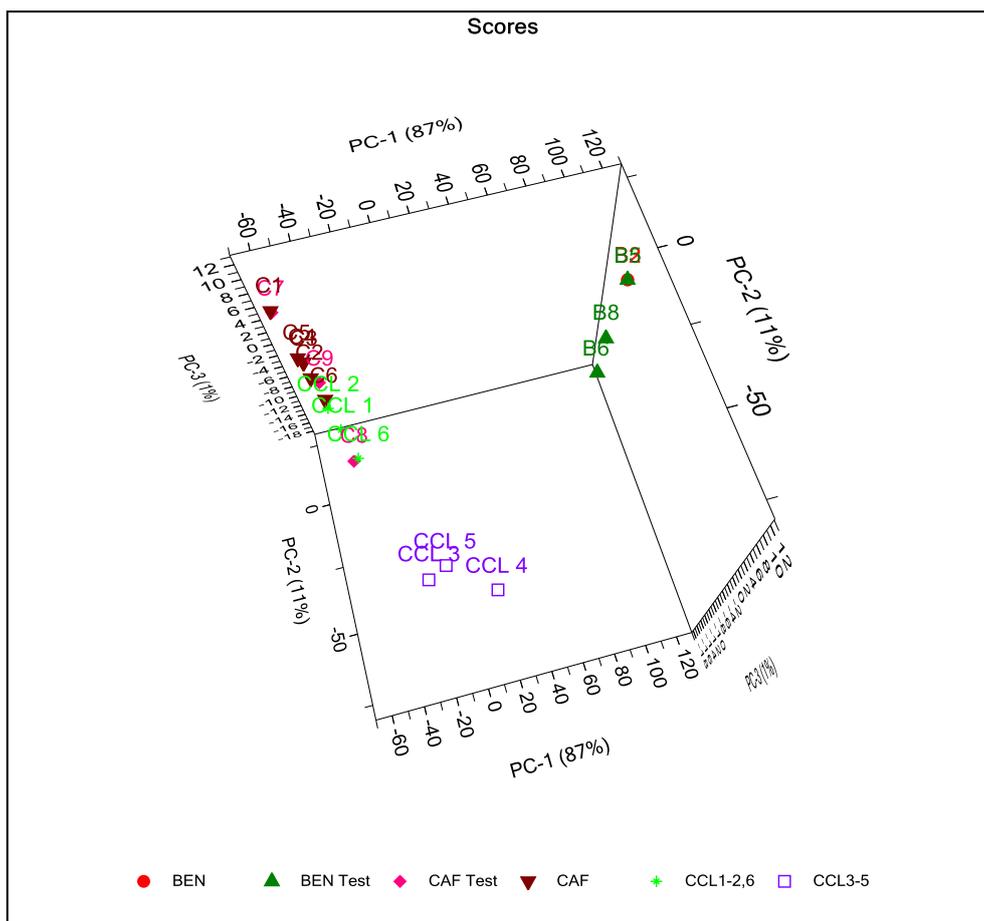


Figure 4. 25: The 3D-scores plot for the validation samples Set ‘3’ following their pre-processing using method ‘5’ (Unscrambler® X 10.3).

In summary, method ‘5’ has correctly classified the validation samples (Set ‘3’) with a two-PC model involving a small percentage of residuals (approximately 2 %). However, more improvement was needed to optimise this method. Therefore a variables’ reduction method was developed to remove/ reduce redundant variables and investigate the robustness of this method in classifying this dataset.

4.3.6. Variables’ reduction

To reduce the size of the data matrix and refine substance classification using the developed pre-processing model, the variables (i.e. data points or wavenumbers) identified by the loading plots throughout this Chapter were further investigated (Table 4.7).

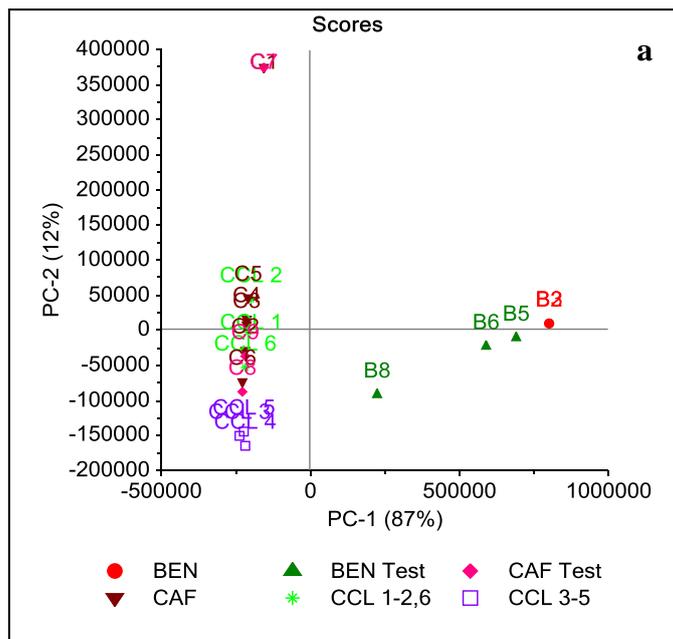
Table 4. 7: Important peaks for BEN, CAF, CRE and LAC ('partial identifiers') identified by PCA loading plots.

Substances	Peaks identified (cm ⁻¹)
BEN	1574, 1679, 1686, 1579
CAF	162, 552, 554, 557, 560, 643, 739, 741, 1170, 1279, 1285, 1328, 1330, 1331, 1399, 1599, 1600, 1603, 1604, 1608, 1694, 1696, 1701, 1703, 2956, 2961,
CRE	740, 830, 1052, 1397, 1424
LAC	860, 863, 1149,

The data matrix (3777 variables per sample) was reduced to seven bands (width 10-15 cm⁻¹) containing the identified variables: [2965-2975]; [2950-2960]; [1690-1705]; [1590-1605]; [1320-1335]; [735-745] and [550-560] cm⁻¹. This new data matrix is named set '4'. Bands were selected rather than peaks to compensate for the slight peak shifts between various pre-processing studies. Method '5' was then employed to pre-process to the new dataset (i.e. set '4') and a PCA was generated to investigate the impact of this method on substance classification.

A two-PC model explained 98.2 and 97.7 % of the cumulative calibrated and validated variance, respectively. The 2D-scores plot (Figure 4.13) showed that the dataset was split into two major groupings: BEN-based and CAF-based samples. The scores plot showed the first separation between both CAF samples and CCL mixtures and an improved clustering of the pure substances (i.e. CAF and BEN). PC1 explained 75 % of the variance for BEN samples (strongest Raman scatterer in the dataset), whereas PC2 explained 23 % of the variance for both CAF samples and CCL mixtures (i.e. CCL 3 and CCL 5). All samples have been shown to lie in the model with 95 % CL. The difference between unprocessed and processed dataset '4' is shown in Figures 4.26a and 4.26b respectively.

Unprocessed



Processed

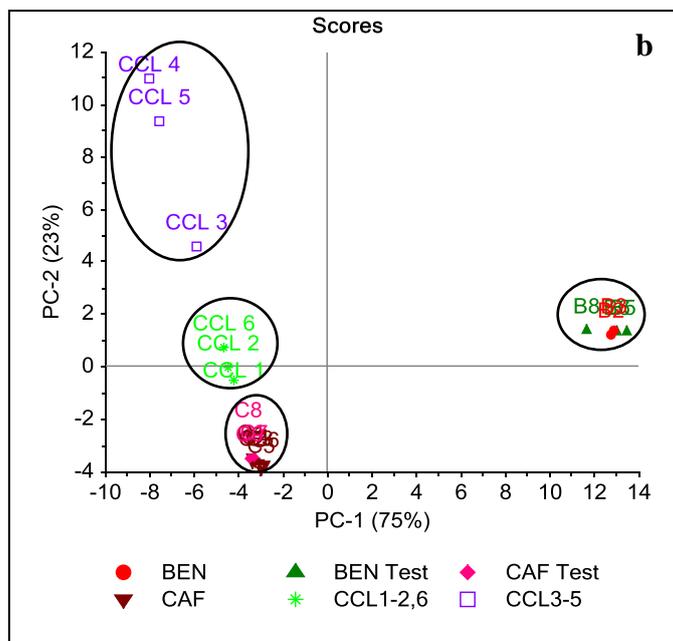


Figure 4. 26: a) The scores plot of the selected-bands spectral data prior to pre-processing, and b) following processing using method ‘5’ (Unscrambler® X 10.3).

The identification of drug constituents in mixtures is commonly achieved by the minimum number of peaks that could be matched to a library or compared to a reference standard [136]. In general, there are no rules to characterise the minimal number of peaks to detect particular substances [267]. In this pilot study, using a simple combined pre-processing sequence (i.e. baseline offset correction method followed by mean normalisation), it was demonstrated that

PCA could be employed to identify and classify drug mixtures using a discrete number of Raman bands/ substance identifiers.

In summary, Raman spectral step-wise pre-processing was carried out to improve the classification of selected pure substances and mixtures of known composition (Figure 4.27). Smoothing methods did not significantly improve the model. This could be because the S/N ratio in all samples was relatively high. The baseline offset method did not improve the model if it was employed as a single pre-processing method. However, it improved the model when it was used in combination pre-processing methods. The linear baseline correction slightly reduced the baseline offset over particular spectral regions at the expense of remaining spectral regions, based on the selected wavenumber range. Using the linear baseline method, the PCA model was not improved as compared to that of the raw data, and, therefore in this study it was not recommended. For this dataset, the first derivative of the Raman spectra showed better PCA classification results than the second derivative. In contrast, the second derivative has been shown to increase the noise in the PCA model and resulted in the loss of chemical information. Pre-processing using the second derivative in combination with other pre-processing methods produced the worst results in terms of sample classification as well as explained and validated variance.

With respect to normalisation, both area and mean normalisation explained greater calibrated and validated variance for Raman spectra. The closeness between both values was greater than with other normalisation methods. Of all normalisation methods, area normalisation has induced clustering between replicate spectra, whereas mean normalisation produced the greatest separation between them. In contrast, the model produced by unit vector normalisation has been shown to be invalid to new samples due to the large variation between both calibrated and validated variance. Range, maximum and peak normalisation have produced similar PCA results for pure CAF. To reduce the scatter effects, SNV as a single pre-processing method improved sample classification by reducing multiplicative scatter effects and scaling variations.

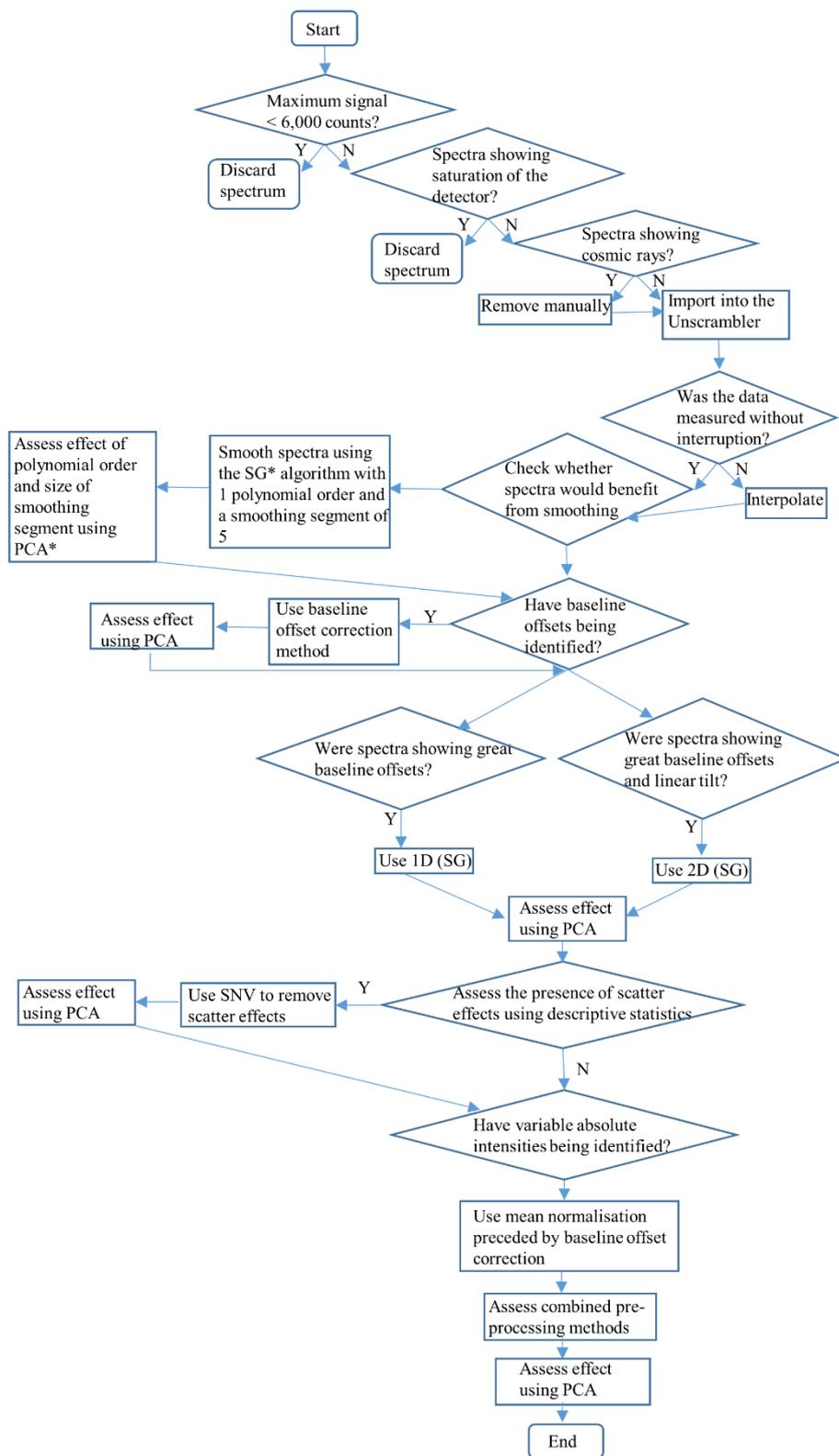


Figure 4. 27: A flow chart showing a protocol for the pre-processing of Raman spectra¹.

¹SG: Savitzky Golay

1D: First derivative

2D: Second derivative

SNV: Standard normal variate

PCA: Principal components analysis

Y and N are 'yes' and 'no' respectively

4.4. Conclusions

In this Chapter, a pre-processing protocol was developed for Raman spectra of NPS-related substances. The optimal combined pre-processing method included a two-step process. This involved the sequential pre-processing using baseline offset correction followed by mean normalisation. The optimal combined pre-processing sequence was successfully applied to a selected Raman dataset for the application of PCA. However, this protocol may not be the optimum option for different datasets or chemometric applications. The developed method included the visual inspection of Raman spectra, rejection of spectra with errors, removal of cosmic spikes, interpolation of the x-axes if required, examination and evaluation of spectra using descriptive statistics and the Unscrambler® software or a relevant software, and, finally, the comparison and assessment of the effect of step-wise pre-processing on Raman spectral data using exploratory PCA. This protocol is recommended for the pre-processing of Raman spectral data of NPS-related substances prior to conducting prediction and classification studies using MVA techniques.

In this study, it was demonstrated that PCA of pre-processed Raman data could be employed as a pre-screening tool to identify substance peak identifiers, which may, in turn assist in optimising the classification of drug mixtures using a discrete number of Raman bands. Spectral pre-processing made it possible to explore and visualise the important differences in the dataset. In this study, unprocessed spectra were mainly separated due to variations in the S/N ratios. Processed spectra have explored the main variations in the dataset due to chemical differences between strong Raman scatterers and heterogeneous mixtures.

Further work should focus on the application of the chemometric protocol on NPS-containing mixtures and larger datasets.

5. Classification of NPS reference standards using Raman spectroscopy and chemometric approaches

5.1. Introduction

This Chapter contains the first study to evaluate the building of a ‘representative’ NPS Raman database that could be employed for the projection and classification of ‘unknown’ previously-seen and previously unseen NPS. This study evaluates a model developed with 53 new psychoactive substance (NPS) reference standards, as ‘representatives’ of 478 NPS, for the identification and classification of 21 test NPS reference standards.

The Raman database was developed based on a study implemented by Zloh et al. [204] in which a hierarchical clustering analysis method combined with chemical similarity indices (i.e. *Tanimoto* similarity index (TSI)) was employed and a ‘representative’ NPS subset was suggested [204]. Hierarchical clustering analysis is a classical data mining technique, which combines similar molecules into groups or clusters and identifies their ‘medoids’ [272]. A ‘medoid’ is a molecule with the highest average similarity to members of the same cluster [204]. Initially, a dataset of 478 NPS, recorded by the EDND (European information system and database on new drugs) [273], was clustered into 21 superclusters, which were, in turn subdivided into 79 clusters *via* a minimum common substructure (MCS) approach [204]. ‘Representative’ NPSs or ‘medoids’ were combined from clusters to form the ‘representative’ database, which represents the chemical space of the original set of 478 NPS. ‘Representative’ NPS were a limited subset, yet retained the maximum chemical and structural diversity of the initial 478 NPS [204].

Since structurally similar compounds are known to have similar physicochemical properties and to exhibit similar spectral features [274], it was postulated that unknown newly emergent NPS, which are structurally similar to ‘representative’ NPSs or ‘medoids’ may exhibit similar structural features and can thus be identified or classified. Therefore based on the above work, ‘representative’ NPS were obtained and evaluated for the identification of ‘unknown’ NPS using Raman spectroscopy coupled with chemometrics. ‘Representative’ NPS used in this Chapter represented superclusters 1 – 13. ‘Representative’ NPS from superclusters 14 – 21 containing single molecules (i.e. singletons) were excluded due to the lack of other NPS with similar structures when this study was conducted. Finally obtained ‘representative’ NPS were subject to commercial availability, cost and popularity in the NPS research field [204].

This work is significant, as there are limited or no ‘representative’ (as defined above) libraries available for newly emerging NPS and there is a great need for the rapid detection of these drugs by law enforcement and front-line staff in emergency settings. The ever-increasing number of NPS, the speed by which they are emerging and the lack of reference standards pose a challenge to their detection. In addition, newly emerging NPS may be undetected using library-based correlation algorithms because of absence from reference libraries and/or limitations of in-built algorithms (see Chapter 3). Interpretation of spectral data for the purpose of structure elucidation requires expertise and may not be feasible if the structure is unknown or due to unknown correlations between structural and spectral information. Therefore developing a chemically diverse ‘representative’ spectral database is significant as it may enable the identification of analogous NPS or near-neighbours to NPS that may emerge in the future.

This Chapter describes building a library of selected chemically diverse NPS reference standards using benchtop Raman spectroscopy. Subsequently, chemometric models were developed following spectral pre-processing, as investigated in Chapter 4, using both exploratory and projection principal component analysis (PCA) models for the identification of unknown NPS against ‘representative’ libraries. To our knowledge, structural similarity studies coupled with Raman spectroscopy, spectroscopic data pre-processing and chemometrics have not been conducted to date for the classification of NPS. The aim is to investigate whether the clustering technique implemented by Zloh et al. [204], the spectral pre-processing method and the use of PCA can be applied for the classification of unknown NPS reference samples, and whether the developed PCA models can categorise the chemical scaffolds of unknown NPS. This study evaluates the complexity associated with NPS with respect to structural and chemical diversity and provides a proof-of-concept for the identification of ‘unknown’ NPS.

5.2. Experimental

5.2.1. Materials

Seventy-four NPS reference standards were commercially obtained from both Chiron AS (Trondheim, Norway) and LGC Group (Teddington, UK). Certificates of analysis of NPS claimed purity of $\geq 98\%$ for all NPS and were used for analysis as supplied. Full characterisation of impurities in supplied NPS reference standards was not performed and is beyond the scope of this work. In this work, the 74 NPS were split into two groups: 53 acting

as a training set (i.e. calibration set) and 21 as a ‘query’ set (i.e. validation set). Initially, 79 ‘representative’ NPS were suggested by Zloh et al. [204]. Only 53 were purchased following the exclusion of singletons and based on availability, cost and popularity in NPS research. The 21 ‘query’ molecules were subdivided into two groups: 17 from the original dataset of 478 NPS provided by the EDND (i.e. previously seen) [204] that were not used to train the model subsequently referred to as ‘test’ molecules, and four, that were not present in the dataset of 478 NPS from the EDND (i.e. previously unseen) referred to as ‘out-of-model (OOM)’ samples (Table 5.1). In order to assist result interpretation, further similarity studies were conducted by Zloh et al. [204] to suggest supercluster/ cluster membership of ‘query’ samples against the clustering model initially developed from 478 NPS (Table 5.1).

5.2.2. Sample preparation

Between 3 - 5 mg of each powdered NPS reference standard were weighed, tapped and flattened on aluminium plates (HSA14521A - Weight dish alum 43 mm, Fisherbrand) for Raman analysis. Two NPS ‘test’ samples (i.e. JWH-018 and MN-18) were obtained as oils (Table 5.1). Oil samples (approximately 10 mg) were recovered by flushing out the containers with approximately 1 mL of acetone to obtain a solution for analysis. The process was repeated and recovered solutions were emptied and aggregated on aluminium plates to form a thin film.

Table 5. 1: NPS reference standards used in this study (training, test and out-of-model (‘OOM’) NPS reference standards).

Training set							
Reference Standard	Sample form	Supercluster	Cluster	Class	Similarity to cluster medoid (%)	Claimed purity	Batch Number
5-MeO-DALT	Powder	1	1.03	Tryptamines	N/A (2 representatives) ²	99.5	15751
5-MeO-MiPT	Powder	1	1.04	Tryptamines	100 (medoid)	99.5	15737
4-HO-DET	Powder	1	1.06	Tryptamines	100 (medoid)	99.5	15727
FDU-PB-22	Powder	1	1.08	Synthetic Cannabinoids	N/A (2 representatives) ²	>99.0	15600
NM-2201	Powder	1	1.09	Synthetic Cannabinoids	N/A (2 representatives) ²	98.5	15706
4-MeO-alpha-PVP	Powder	2	2.01	Cathinones	N/A (2 representatives) ²	99.5	15733
25H-NBOMe	Powder	2	2.05	Phenethylamines	100 (medoid)	99.5	15731
N-Me-2C-B	Powder	2	2.06	Phenethylamines	100 (medoid)	99.5	15661
STP (DOM)	Powder	2	2.08	Phenethylamines	100 (medoid)	99.4	15662
AB-FUBINACA	Powder	3	3.01	Synthetic Cannabinoids	N/A (2 representatives) ²	99.5	14188
AB-PINACA	Powder	3	3.02	Synthetic Cannabinoids	N/A (2 representatives) ²	>99.5	14186
THJ-018 (JWH-018 indazole analogue)	Powder	3	3.04	Synthetic Cannabinoids	N/A (2 representatives) ²	98.0	14686
SDB-006	Powder	3	3.06	Synthetic Cannabinoids	N/A (2 representatives) ²	>99.5	15599
AM-679	Powder	3	3.07	Synthetic Cannabinoids	100 (medoid)	99.5	15728
DPT	Powder	3	3.08	Tryptamines	98	99.5	15942
AM-2201	Powder	4	4.02	Synthetic Cannabinoids	97	98.2±0.5	N/A
JWH-122	Powder	4	4.02	Synthetic Cannabinoids	99	99.5±0.5	11375
UR-144	Powder	4	4.03	Synthetic Cannabinoids	100 (medoid)	99.5	12511
JWH-073	Powder	4	4.04	Synthetic Cannabinoids	100 (medoid)	99.0	9296
5F-APICA	Powder	4	4.05	Synthetic Cannabinoids	86	99.2	13651
Methoxetamine (MXT)	Powder	5	5.01	Arylcyclohexylamines	N/A (2 representatives) ²	99.2	14872
Ketamine	Powder	5	5.02	Arylcyclohexylamines	N/A (2 representatives) ²	99.5	13562
4F- α -PVP	Powder	5	5.03	Cathinones	100 (medoid)	99.5	15273
4-Me-N-ethylnorpentedrone	Powder	5	5.04	Cathinones	100 (medoid)	99.5	15729
Phenazepam	Powder	5	5.05	Benzodiazepines	N/A (2 representatives) ²	99.3	4097
Flubromazepam	Powder	5	5.06	Benzodiazepines	N/A (2 representatives) ²	99.5	14560

Afloqualone	Powder	5	5.07	Quinazolines	N/A (2 representatives) ²	99.5	15810
Mebroqualone	Powder	5	5.08	Quinazolines	N/A (2 representatives) ²	98.6	15652
3-MeO-PCE	Powder	5	5.09	Arylcyclohexylamines	N/A (2 representatives) ²	N/A	14477
DL-4662	Powder	5	5.10	Cathinones	N/A (2 representatives) ²	99.5	15730
5F-APINACA	Powder	5	5.11	Synthetic Cannabinoids	83	99.5	13652
JWH-015	Powder	5	5.12	Synthetic Cannabinoids	100 (medoid)	99.5	13618
4-MeO-PCP	Powder	5	5.13	Arylcyclohexylamines	100 (medoid)	99.4	14717
α -PVP	Powder	5	5.14	Cathinones	76	99.5	15373
2-AI	Powder	5	5.15	Aminoindanes	N/A (2 representatives) ²	98.0	13108AH
N-Me-2-AI	Powder	5	5.16	Aminoindanes	N/A (2 representatives) ²	>99.6	15127
Dimethocaine	Powder	5	5.17	Anaesthetics	N/A (2 representatives) ²	>99.5	15597
Mephedrone (4-MMC)	Powder	5	5.18	Cathinones	92	98.0	051M4701V
N-PB-22 (PB-22 indazole analogue)	Powder	6	6.01	Synthetic Cannabinoids	N/A (2 representatives) ²	98.9	15622
PB-22	Powder	6	6.02	Synthetic Cannabinoids	N/A (2 representatives) ²	99.5	13648
4-acetylpsilocin fumarate (4-AcO-DMT)	Powder	6	6.03	Tryptamines	N/A (2 representatives) ²	98.6	11288
2-MAPB	Powder	7	7.01	Arylalkylamines	N/A (2 representatives) ²	>99.5	15598
5-EAPB	Powder	7	7.02	Arylalkylamines	N/A (2 representatives) ²	98.8	15563
6-MAPB	Powder	7	7.03	Arylalkylamines	N/A (2 representatives) ²	99.5	15601
N-ethyl-amphetamine	Powder	7	7.04	Phenethylamines	100 (medoid)	99.5	15569
Etizolam	Powder	8	8.01	Benzodiazepines	N/A (2 representatives) ²	N/A	16225
Flubromazolam	Powder	8	8.02	Benzodiazepines	N/A (2 representatives) ²	99.5	14821
α -PBT	Powder	9	9.01	Cathinones	95	>99.5	15596
MPA	Powder	9	9.02	Arylalkylamines	N/A (2 representatives) ²	N/A	6985
GHB	Powder	10	10.01	γ -hydroxybutyrate	N/A (2 representatives) ²	98.1	13762
Methylone	Powder	11	11.00	Cathinones	100 (medoid)	99.5	15590
trans-CP 47,497-C8	Powder	12	12.00	Synthetic Cannabinoids	100 (medoid)	99.5	15949
Zopiclone	Powder	13	13.00	Others	19	99.5	11179
Test							
β k-2C-B	Powder	2	2.06	Cathinones	78		
5-APB	Powder	2	2.08	Alkylarylamines	35		

6-APB	Powder	2	2.08	Alkylarylamines	34
MN-18	Oil	3	3.03	Synthetic Cannabinoids	N/A (2 representatives) ²
JWH-018	Oil	4	4.02	Synthetic Cannabinoids	100 (medoid)
Adrafinil	Powder	5	5.14	Others	22
Ethylphenidate (EPD)	Powder	5	5.14	Piperidine	36
Phenibut	Powder	5	5.14	Others	44
5-IAI	Powder	5	5.15	Aminoindanes	N/A (2 representatives) ²
Flephedrone (4-FMC) ¹	Powder	5	5.18	Cathinones	68
Mephedrone (4-MMC) ¹	Powder	5	5.18	Cathinones	92
Mexedrone	Powder	5	5.18	Cathinones	Not present
5F-PB-22	Powder	6	6.02	Synthetic Cannabinoids	N/A (2 representatives) ²
Pyrazolam	Powder	8	8.02	Benzodiazepines	N/A (2 representatives) ²
5,6-MDAI	Powder	11	11.00	Aminoindanes	75
BB-22	Powder	12	12.00	Synthetic Cannabinoids	21
Dextromethorphan HBr (DXM)	Powder	12	12.00	Others	37
Out-of-model ('OOM')					
Methylphenidate (MPD)	Powder	2	2.30	Piperidine	46
		5	5.40		
		5	5.13		
		5	5.14		
S-Cathinone	Powder	5	5.18	Cathinones	83
Methamphetamine	Powder	7	7.40	Phenethylamines	91
MDMA	Powder	11	11.00	Phenethylamines	81

¹Both flephedrone and mephedrone 'test' samples were synthesised in-house by both Dr J. Ward and Dr K.E. Kellett.

²Clusters containing 2 representative NPS is a cluster that did not contain a medoid and one of both representatives was selected based on availability, cost and level of interest in NPS research

5.2.3. Raman spectroscopy

Spectra were acquired using a benchtop Renishaw inVia™ Raman microscope (Serial No. 29U091) employing a laser with excitation wavelength (λ_{ex}) of 785 nm and a 5 % laser output power of 10.6 mW (5.8 mW at sample using a x20 objective lens), a high sensitivity ultra-low noise RenCam CCD detector, held in a vacuum of -70 °C, an ultra-high precision diffraction grating of 1200 lines/ mm. Parameters employed for both calibration of the Raman instrument and measurements of all samples were a 10 sec. exposure time, 1 accumulation, 5 % laser power, x20 objective lens (spot radius = 1.2 μm). The instrument was calibrated regularly prior to analysis using a static silicon wafer reference. Calibration was carried out in order to ensure wavenumber accuracy and adjusting the wavenumber to $520 \pm 0.6 \text{ cm}^{-1}$. Ten spectra were acquired for each NPS reference standard, at ten random spots denoted in this Chapter as R1 to R10, over a spectral range of 3200 - 100 cm^{-1} . In order to optimise the signal and acquire spectra with adequate signal to noise (S/N), the signal for each replicate measurement was evaluated by focussing the laser on different sample regions i.e. bright, dull or dark regions, yet this procedure was sample-dependent. For example, bright spots may worsen the signal in samples exhibiting an intense fluorescent background but may give a better S/N in poor Raman scatterers. Spectral data were exported as WiRE (Windows-based Raman Environment) version 3.4 (.wxd) files which were converted to text file format (.txt), then saved as Microsoft Excel 2013 files (.xlsx), imported into Origin Pro 2016 software (OriginLab, USA) for data pre-processing, then finally imported as Excel (.xlsx) files into the Unscrambler® X10.4 software for chemometric analysis. Raman data was collected in collaboration with Drs E. Samaras and J. Calvo-Castro.

5.2.4. Raman spectral pre-processing

The pre-processing of raw spectral data by means of chemometrics has been shown to impact the classification of chemically diverse compounds [275]. Raman spectral pre-processing was carried out using WiRE 3.4 (Renishaw, UK), Microsoft Excel 2013 and Origin Pro 2016 (OriginLab, USA) softwares. Cosmic rays were zapped using WiRE 3.4. Raman spectra, saved as Excel (.xlsx) files, were then inspected for anomalies using both overlaid and stacked line plots of the raw data as well as spectral correlation coefficients. Anomalies were evaluated using PCA (Unscrambler® x 10.4) to test their impact on sample classification and whether they represent true outliers, being removed from the analysis if this was found to be the case. The Raman spectral region below 250 cm^{-1} was truncated. Raman spectra were then imported

into Origin Pro 2016 software for initial pre-processing (i.e. smoothing and baseline subtraction). Smoothing was performed using the Savitzky Golay (SG) algorithm, 4th polynomial order and a smoothing window of 21 data points. Spectral baselines were subtracted using the 2nd derivative, adjacent-averaging smoothing with a window size of 3 points, a threshold of 0.05 and 200 points to find the baseline anchor points and point interpolation. Pre-processed data were then exported back to Microsoft Excel (2013) to remove negative data points (negative data points were zeroed using the 'IF' function). Individual spectra were then maximum normalised. Spectra were finally re-inspected following pre-processing using line plots and the calculation of correlation coefficients between spectra. A summary of combined pre-processing steps used in the analysis is shown in Figure 5.1.

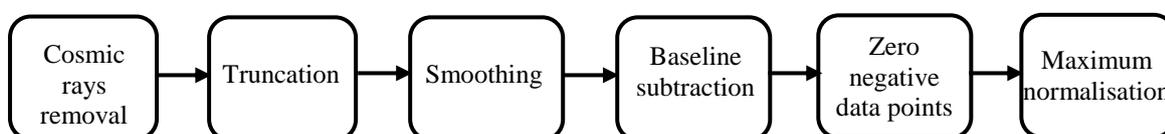


Figure 5. 1: A schematic outlining combined pre-processing steps of spectral Raman data for NPS reference standards.

5.2.5. Principal component analysis (PCA)

PCA was performed using the Unscrambler® X 10.4 software (CAMO PROCESS AS, Oslo, Norway) using the NIPALS (Non-linear iterative projections by alternating least squares) algorithm. PCA parameters were a maximum of 20 principal components (PCs) to calculate the PCA, outlier identification and mean centring of the data, conducting uncertainty tests, using the full validation method with 1000 iterations to allow the calculation of a greater number of PCs (≥ 3).

5.2.6. Prediction via PCA projection

Following the generation of a PCA model from the training samples (53 NPS x 10 measurements), prediction of the chemical scaffolds of 'query' samples was performed by projecting developed PCA models to latent space over the spectral range 1750 – 1300 cm^{-1} . The dataset for the training samples consisted of 530 rows and 522 variables, whereas the dataset for the 'query' samples consisted of 210 rows and 522 variables. PCA Projection was generated using the Unscrambler® X 10.4 software for three PCs. Similarity between 'query' (i.e. 17 'test' and 4 'OOM' pure NPS) and library substances (i.e. 53 training pure NPS) was evaluated using a) PCA projection plots; b) correlation coefficients (i.e. r^2 values) between overlapping spectra over the spectral range 1750 - 1300 cm^{-1} ; c) structural similarity data

obtained from Zloh et al. [204] (i.e. TSI values); d) common substructures between ‘representative’ and ‘query’ NPS; e) Raman spectra; f) line loading plots. Three-dimensional (3D) plots represented in this Chapter were drawn using Origin Pro 2016 (USA).

Summary of the design of experiments is illustrated in Figure 5.2.

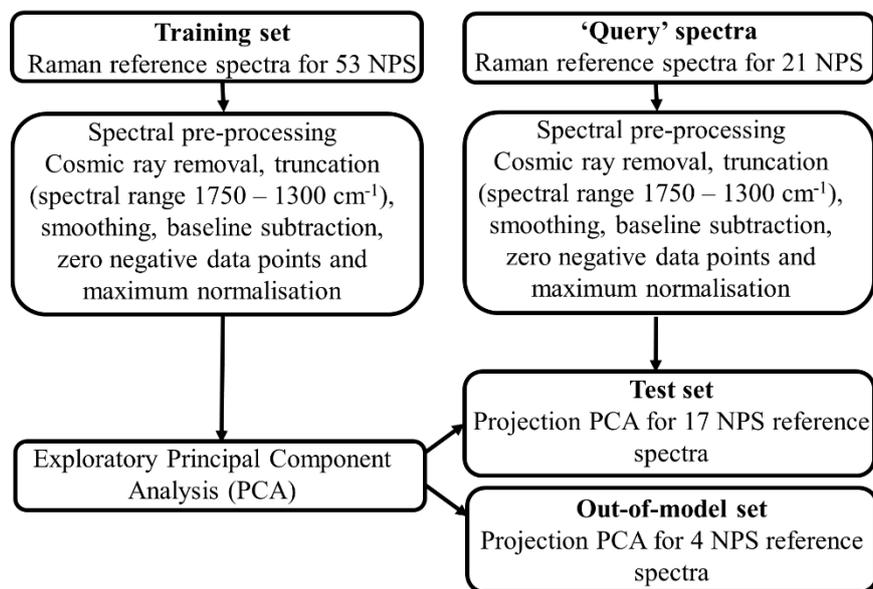


Figure 5. 2: A schematic describing summary of experiments. It includes a summary of both training and ‘query’ sets (i.e. test and ‘OOM’ sets), combined pre-processing steps and details of projection PCA studies.

5.3. Results and Discussion

5.3.1. Method development of Raman analysis and chemometrics

5.3.1.1. Raman analysis method development

A ‘representative’ NPS database was generated using benchtop Raman spectroscopy. Raman instrumental parameters (i.e. laser power, objective lens, exposure time and number of accumulations) employed were changed from Chapter 4. It was important to ensure that the spectra of the ‘representative’ NPS Raman library are of high quality and to preserve sample integrity by preventing its degradation or burning. Taking this into account, a lower laser power 5 % (5.8 mW at sample) was initially used to prevent inadvertent sample burning. Laser power was reduced to 0.5 or 1 % if saturation of the detector occurred and also in samples with high fluorescent backgrounds. In contrast, laser power was increased to 10 % (10.6 mW at sample) in samples with very poor S/N. In an attempt to capture less variability due to physicochemical properties between replicate measurements, the objective lens x20 (spot radius = 1.2 μm) was employed instead of the x50 lens, known for its high spatial resolution (spot radius = 638 nm).

Exposure time was 10 sec. as defaulted by the instrument (lowest). Therefore it was not feasible to reduce exposure time with samples that saturated the detector or exhibited a high fluorescent background. Number of accumulations was 1 (default) to reduce analysis time since more than 10 spectra were collected for most samples. Number of accumulations was increased to 5 for samples with poor S/N.

5.3.2. Pre-processing method development

Spectra of the 53 training NPS were acquired using a benchtop Raman instrument (10 measurements/ NPS). Raw spectral data are often pre-processed before they are deemed useful for conducting multivariate data analysis. As stated in Chapter 4, initial inspection of Raman spectra was conducted to evaluate the extent to which instrument artefacts influence the quality of replicate spectra and to establish the necessity and types of possible pre-processing methods that would need to be undertaken. Using exploratory PCA and visual inspection of Raman spectra, several issues were identified. These included: 1) cosmic rays; 2) large number of variables; 3) shot/ residual noise; 4) fluorescent background; 5) data points with positive and negative signs; 6) variable absolute Raman intensities. Therefore pre-processing methods including cosmic ray removal, truncation, smoothing, baseline subtraction, zeroing of negative data points and normalisation were suggested for investigation. Pre-processing methods were developed using single spectra of 53 NPS reference standards, assessed using PCA, then applied to the full data matrix of 530 reference spectra. The Section below describes the systematic method development for Raman analysis, pre-processing of spectral data and variable reduction techniques, the purpose of which was to improve the S/N, extract data from analysis artefacts and increase the PCA percentage explained variance, while maintaining the model robustness.

5.3.2.1. Cosmic ray removal

Cosmic rays were removed using the ‘zapping’ function using WiRE 3.4 as explained in Chapter 4.

5.3.2.2. Truncation

The first step undertaken in data pre-processing was truncating the spectral region below 250 cm^{-1} , a cut-off that is commonly adopted by analysts [193]. The truncated region is related to lattice vibrations in crystals and longitudinal acoustic modes, which can be used to determine the length of straight chain systems (outside the scope of this study) [276]. In this region, Raman modes with low frequencies may also be related to metal-ligand bonds which are not

of interest in this work. In general, this cut-off is important to avoid instrument-related features arising from the holographic notch filters [193]. The in-house Renishaw inVia™ Raman microscope is equipped with an edge filter, which did not efficiently block the high intensity Rayleigh scattering near the laser line in the low frequency range (ca. $< 400 \text{ cm}^{-1}$). Intense Rayleigh scattering may cause interferences with Raman bands of interest and may affect the reliability of data pre-processing e.g. for normalisation. Subsequently, pre-processing of all datasets analysed in this Chapter was performed using the spectral range $3200 - 250 \text{ cm}^{-1}$.

5.3.2.3. Smoothing

Smoothing using the SG algorithm is a common pre-processing method employed to reduce/filter high frequency noise in spectral data by fitting a polynomial to a segment of data points [172]. Larger segments and lower polynomial orders lead to more smoothing being generated. In this dataset, following smoothing, the cumulative calibrated explained variance for the first three PCs decreased by 2%. In other words, explained calibrated variance was reduced from 12, 9 and 7 % to 12, 8 and 6 % respectively for PC1/PC2/PC3 for non-smoothed and smoothed averaged spectra. This was expected due to reduced explained variance related to shot/ residual noise. The impact of smoothing average spectra on supercluster classification (Figure 5.3) and on model robustness was minimal (closer calibration and validation values in smoothed spectra (58 and 31.5 respectively) as compared to non-smoothed spectra (59.8 and 31 respectively) for PC1/PC2/PC3). However, it was postulated that since smoothing acts as a low pass filter it may remove high frequency noise through combined pre-processing, which may have a greater impact on the classification of the more complex dataset of 530 spectra. In this respect, smoothing was performed using the Savitzky Golay (SG) algorithm as in Chapter 4 but with a higher polynomial order and a larger smoothing window (i.e. 4th polynomial order and a smoothing window of 21 data points) (Origin Pro 2016).

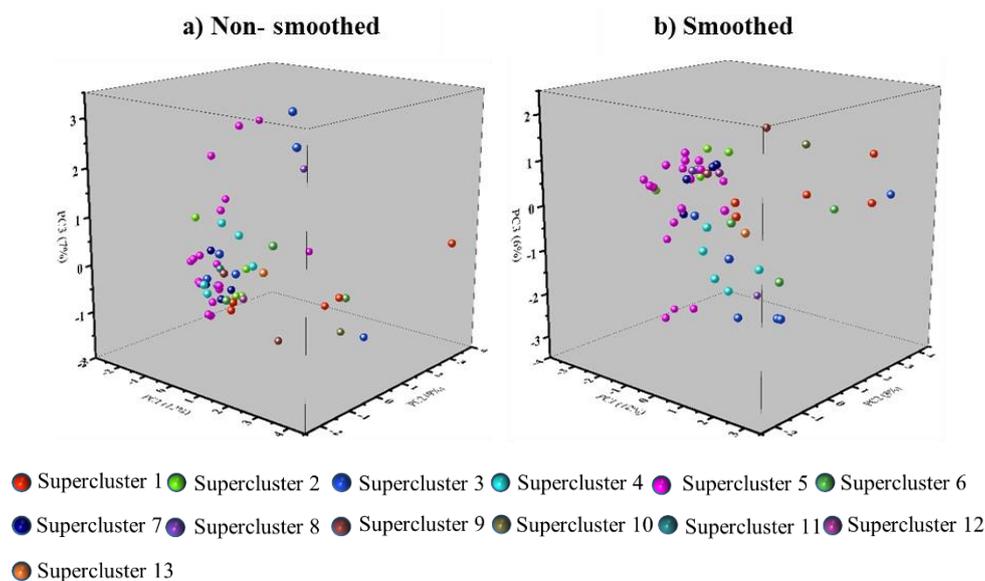


Figure 5. 3: 3D-PCA scores plots of average a) non-smoothed and b) smoothed average spectra of 53 NPS reference standards over the spectral range 3200 – 250 cm^{-1} . Average spectra were sample-grouped based on their supercluster membership [204] (Table 5.1).

5.3.2.4. Baseline subtraction

The majority of NPS reference standards analysed in this work were claimed to have purity $\geq 98\%$. Nevertheless, Raman analysis resulted in a wide range of S/N and spectra with steep baselines and raised fluorescent background (Figure 5.4), which may have been the result of minor impurities using the 785 nm laser wavelength. Conversely, minor baseline offsets may have resulted from instrument artefacts. In contrast to the method developed in Chapter 4, using the Unscrambler® X 10.3 software, the baseline offset method has produced insignificant difference to NPS classification (PCA results not shown) of the full dataset (530 spectra). In this dataset, baseline subtraction was important to prevent misclassification of unknown NPS due to major baseline offsets. Using the Origin Pro 2016 software, different parameters were investigated in order to develop a consistent method for various NPS samples. A user-defined method was developed using 200 points to find the baseline anchor points. Remaining parameters were used as defaulted by the software. The number of points was increased from the default of 8 points to provide a larger window that allows the software to customise a greater number of anchor points to pull the baseline down to zero. Optimum points have not exceeded 50 points in the calculation of baseline subtraction of the 530 NPS reference spectra. However, this method has produced negative data points (Figure 5.4f), which may impact the reliability of spectral normalisation and consequent ‘query’ NPS classification. For this reason, smoothed baseline subtracted spectra were exported back to Microsoft Excel (2013) to zero negative data points.

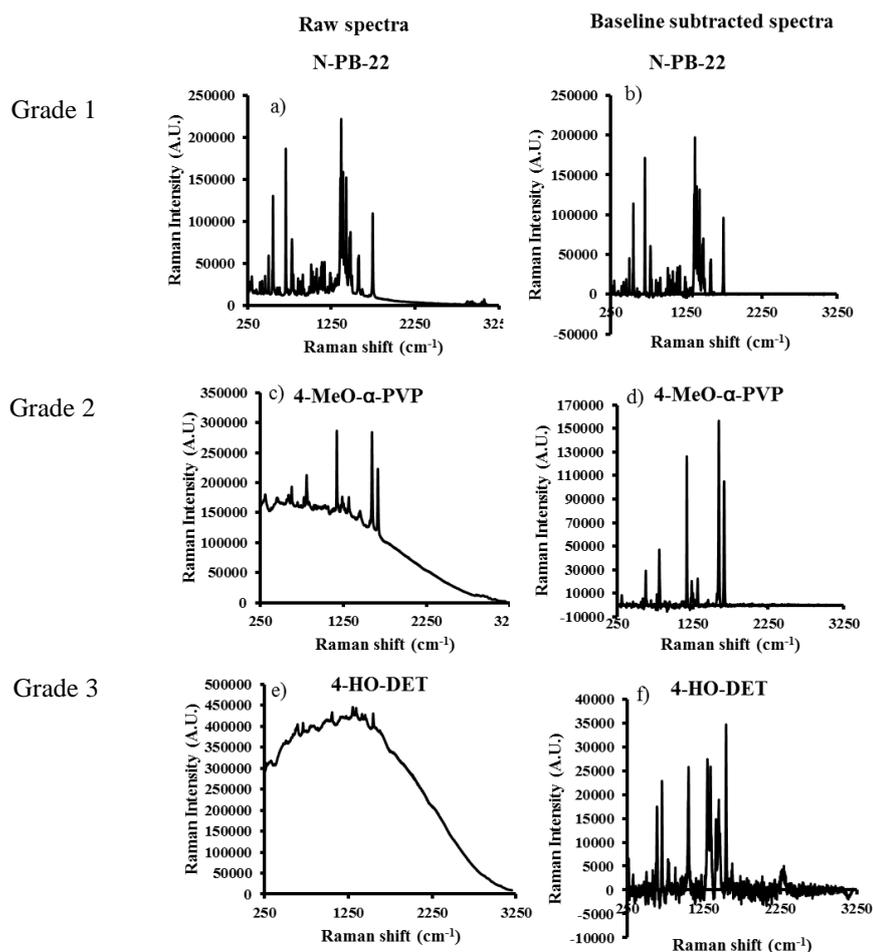


Figure 5. 4: Raman spectra of selected NPS reference standards (spectral range 3200 – 250 cm^{-1}). The line plots show different grades of baseline offsets in Figures 5.4a, c and e. Spectra were pre-processed using smoothing and baseline subtraction in Figures 5.4b, d and f respectively (Origin Pro 2016).

5.3.2.5. Normalisation

Absolute intensity can provide useful information about the Raman strength of different NPS. However, variation in absolute intensities is known to be common in Raman spectra [37], which may impact analysis. In this study, variations in absolute intensities and a wide S/N range were initially identified and have been shown to impact grouping of replicate spectra and, hence sample classification using PCA (see Chapter 4). Therefore normalisation of spectra was recommended because it enhances the variance between samples by correcting for interfering scaling variations. Mean normalisation was the normalisation type of choice for the dataset analysed in Chapter 4. Due to the high chemical diversity between substances investigated in this Chapter normalisation types (i.e. mean, maximum and range normalisation), which normalise individual spectra rather than the whole dataset were investigated using the Unscrambler® X 10.3 software (plots not shown).

Although different algorithms were used to calculate the three types of normalisation, relative intensities were retained in the three types of normalised spectra. Although a greater percentage variance was explained by mean normalisation (96.2 %) of average spectra of 53 NPS in comparison with both range and maximum normalisation (90.3 %) for the first three PCs, 70 % of the validated variance was explained by range and maximum normalisation as opposed to 54.6 % by mean normalisation. This demonstrated that the use of range or maximum normalisation is more robust in this dataset as it is likely to perform better with unknown samples. Since maximum normalisation is commonly used [176, 184], it was carried forward and was calculated for the entire datasets using Microsoft Excel (2013).

5.3.2.6. Variable reduction and determination of the size of the data matrix

A high number of variables has been shown to increase the risk of degrading the accuracy of the PCA model [183]. The variable reduction approach employed in Chapter 4 was not suitable for this study because of the large size of the dataset. In this work, variable reduction techniques were investigated through two methods: truncation of the spectral range and the computation of ‘bins’ (see below). The first method aimed at reducing the variable number from 3777 (3200 – 100 cm^{-1}). Initially, spectra were truncated below 250 cm^{-1} (Section 5.3.2.2.), which reduced the number of variables to 3642. In order to reduce the number of variables further, the optimum spectral range containing essential spectral features was selected. This was assessed using exploratory PCA plots: 1) Explained variance plots to ensure optimisation of model robustness; 2) Scores plots to assess classification of NPS superclusters, the percentage of explained variance by the first three PCs, and the presence of outliers at 95 % confidence limit (CL); 3) Loading plots to investigate the main variables, which have high loadings. Raman spectral ranges used with pure and street NPS samples varied in the literature and were dependent on the NPS investigated [68, 113, 114, 120, 132, 152, 277]. In this Chapter spectral ranges investigated were 3200 – 250, 1750 – 250, 1750 – 1000 and 1750 – 1300 cm^{-1} . The Section below describes the evaluation of results obtained using the latter spectral ranges.

1) Explained variance plots: As explained in Chapter 4, closeness between the values of calibrated and the validated explained variance is indicative of method robustness, reliability and applicability to ‘unknown’ samples (i.e. not included in the original model). The spectral range 1750 – 1300 cm^{-1} , of pre-processed (cosmic rays removed, truncated, smoothed, baseline subtracted, negative data points zeroed and maximum normalised) single spectra of 53 NPS reference standards, has generated closer values between

cumulative calibrated and validated explained variance for 20 PCs (91 and 64 % respectively) and, hence is indicative of method robustness as compared to remaining suggested spectral ranges (Figure 5.5).

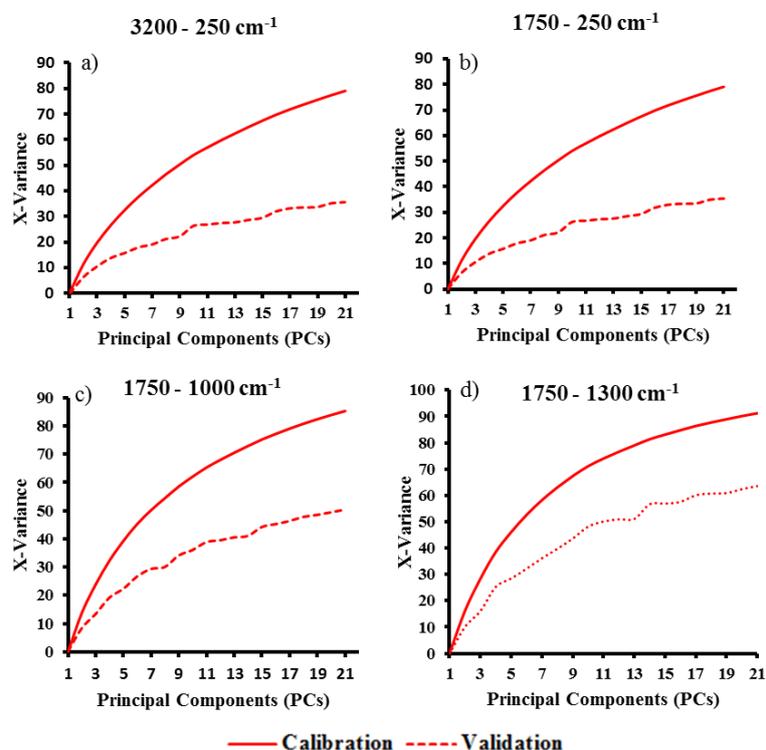


Figure 5. 5: PCA explained variance plots over the investigated spectral ranges.

The region 1750 – 1300 cm^{-1} is important for organic compounds as it includes vibrations such as carbonyl stretches $\nu(\text{C}=\text{O})$ at 1750 - 1680 cm^{-1} (medium to weak) as in cathinones; olefinic groups $\nu(\text{C}=\text{C})$ at 1750 - 1500 cm^{-1} (strong); amine NH deformation vibrations for amines, amine salts and amide substances at 1660 – 1500 cm^{-1} (weak); aromatic and heteroaromatic rings at 1620 – 1420 cm^{-1} (medium to weak); methyl and methylene deformation vibrations at 1500 – 1300 (weak to medium) such as $\delta(\text{CH}_3)$ at 1380 cm^{-1} (medium); $\delta(\text{CH}_2)$ and $\delta(\text{CH}_3)$ asym at 1470 - 1400 cm^{-1} (medium); and $\nu(\text{CC})$ related to aromatic ring chain vibrations at 1580 and 1600 cm^{-1} (strong), 1450 and 1500 cm^{-1} (medium) [125]. These functional groups are very common in NPS as identified by their chemical structures and, hence this region is of utmost importance in identifying and classifying NPS. Explained variance plots of non-pre-processed spectra plateaued after 1 PC and, therefore did not reflect the difference between explained variance plots of selected spectral ranges.

2) Loading plots: The PCA loading plots of non-pre-processed single spectra of 53 NPS reference standards, showed that variables with high loadings (Figures 5.6a, b and c) were those associated with noise (spectral range 3200 – 1750 cm^{-1}). Figure 5.6b is the correlation

loading plot, where the outer and inner ellipses indicate 100 and 50 % explained variance respectively. Variables with high loading are closer to the outer ellipse (Figure 5.6c) (explained variance by PC1 = 94 %).

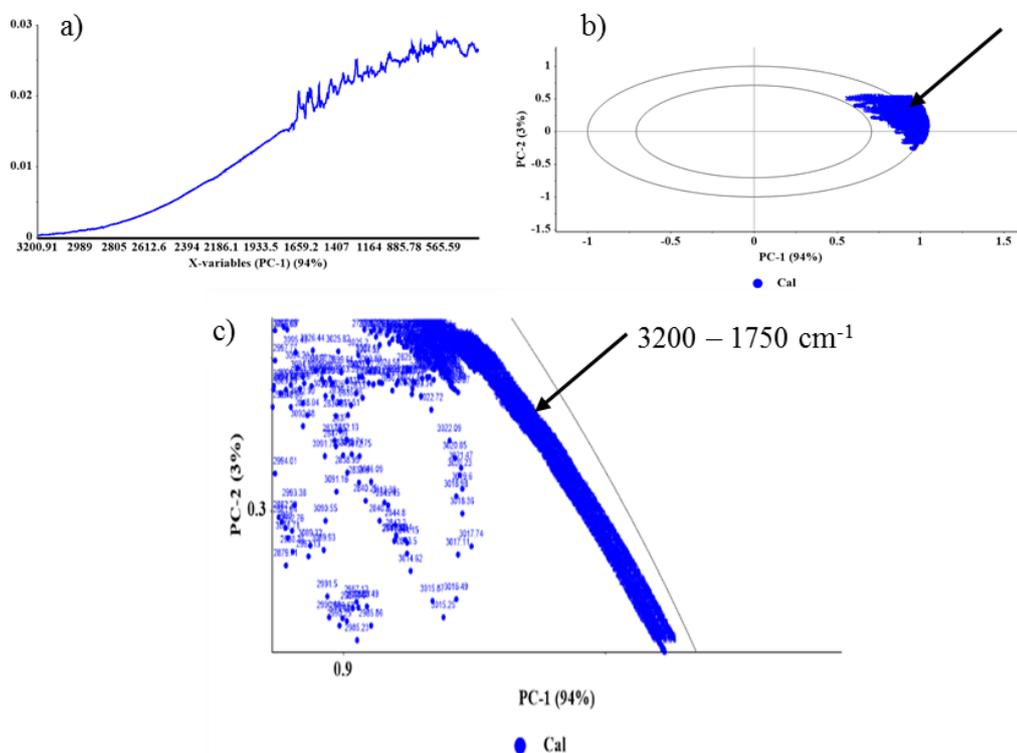


Figure 5. 6: PCA a) line loading plot, b) correlation loading plot and c) zoomed-in correlation loading plot over the spectral range 3200 – 250 cm^{-1} .

Weak Raman bands were observed in the range 3000 – 2800 cm^{-1} for a few NPS (Figure 5.4a). This region is commonly associated with strong Raman $\nu(\text{C-H})$ stretching vibrations in aliphatic chains [125, 278]. PCA analysis showed that these bands do not have high loadings nor improve the percentage explained variance by PCA (plots not shown). Furthermore, most Raman spectral features in overlaid single spectra of 53 NPS reference standards were spread across the range 1750 – 250 cm^{-1} (plot not shown). Therefore the spectral range 3200 – 1750 cm^{-1} was also truncated (Figure 5.6a).

3) Scores plots: The PCA scores plots of the selected spectral ranges investigated for pre-processed single spectra of 53 NPS reference standards, have shown greater explained variance by the first three PCs over 1750 – 1300 cm^{-1} .

Therefore the scores plots of 10 replicate spectra for all 53 spectra were inspected. Results showed that the explained variance for both spectral ranges 3200 – 250 and 1750 – 250 cm^{-1} remained unchanged with calibrated explained variance of 11, 9 and 6 % for PC1/PC2/PC3

respectively. A PCA was then performed on a smaller spectral range (1750 – 1000 cm^{-1}), which contained prominent spectral features and also, which included a reduced number of variables. The calibrated explained variance has slightly increased to 13, 10 and 8 % for PC1/PC2/PC3 respectively. Reducing the number of variables further to 522 over the spectral range 1750 – 1300 cm^{-1} resulted in a greater calibrated explained variance of 16, 12 and 9 % for PC1/PC2/PC3 respectively (Figure 5.7). The scores plot showed a better grouping of replicate spectra and an improved classification between different samples over the spectral range 1750 – 1300 cm^{-1} .

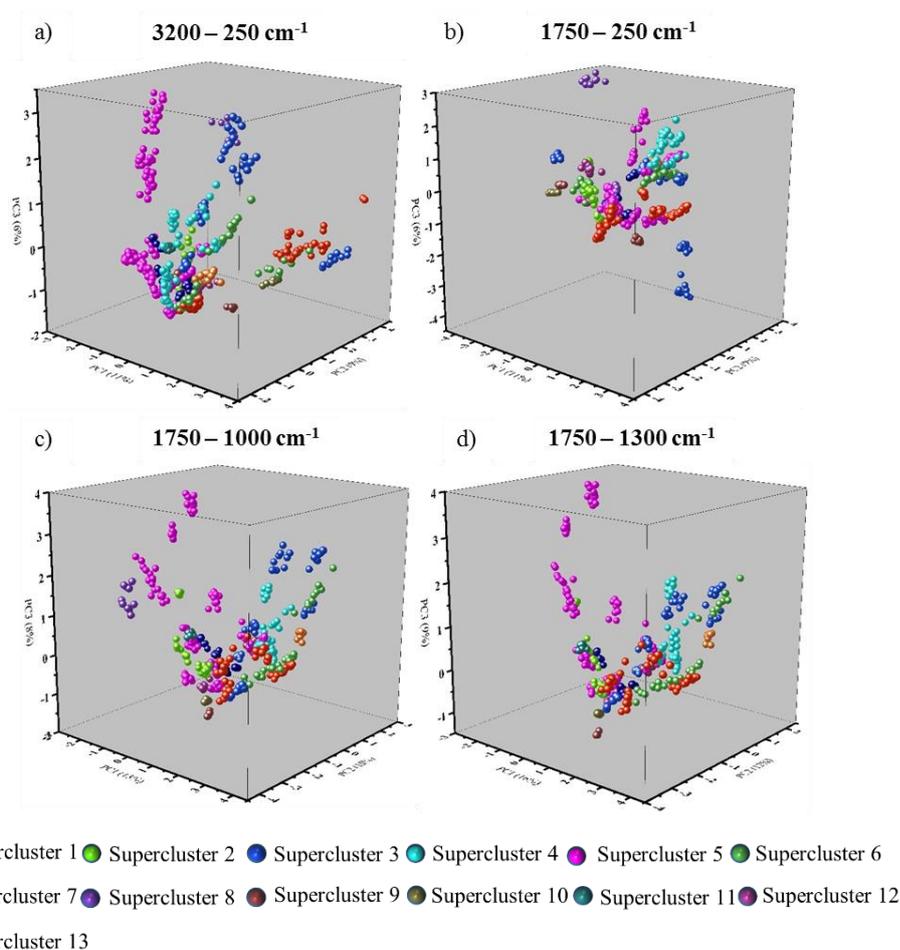


Figure 5. 7: PCA scores plots of 10 replicate spectra of 53 NPS reference standards over the spectral ranges a) 3200 – 250; b) 1750 – 250; c) 1750 – 1000; d) 1750 – 1300 cm^{-1} . Replicate spectra were sample-grouped based on their supercluster membership [204].

The second variable reduction method aimed at reducing the number of variables by means of ‘bins’. ‘Binning’ of spectral data is a known data reduction technique used in different types of spectroscopies such as NMR and has been applied to various data matrices including biological samples [279]. A bin is a value given to small spectral ranges of just a few wavenumbers (i.e. variables). In this respect, two methods were investigated: 1) area under the

curve (AUC) and 2) sum of intensities of given bins. Bin sizes were also investigated. For example, a bin size of 300 cm^{-1} would result in 10 bins (spectral range (i.e. 3000 cm^{-1})/ bin size (i.e. 300 cm^{-1})). Reducing the initial number of variables from 3642 to 10 risks reducing the explained variance and generating an imbalanced data matrix. Therefore the maximum number of bins that could practically be generated (i.e. 300 and 600 bins) was investigated, which represent spectral ranges $3200 - 250$ and $1750 - 250\text{ cm}^{-1}$ respectively.

For this purpose, calculating bins of average spectra was performed as a test prior to a wider application on the entire dataset (i.e. 530 spectra). For the first method (i.e. AUC), the spectra were exported from WiRE 3.4 to Microsoft Excel 2013 for maximum normalisation of individual spectra, then re-exported back to WiRE 3.4 to calculate AUC based on peak centres. AUC values were then exported back to Microsoft Excel 2013 where bin values (sum of AUC in 5 cm^{-1} ranges) were calculated using the 'IF' function, again based on peak centres.

Due to the small bin size, slight peak shifting between spectra of the same reference NPS resulted in bin values of zero, which was not true as these bins have values $>$ zero if method 2 was applied (i.e. sum of intensities). Furthermore, AUC estimates areas, where there are no counts (i.e. exceeding the instruments' resolution), which may not be accurate. This limitation was due to the default WiRE 3.4 algorithm used to calculate AUC based on peak centres. Subsequently, the second method, sum of intensities of peaks within individual bins was calculated from maximum normalised spectra using the 'IF' function (Microsoft Excel 2013). Results were maximum normalised again, then carried forward for PCA analysis. Figure 5.8 shows the scores plots of 600 and 300 bins using the 'sum of intensities' method. Explained variance for both plots was 12, 9 and 6 % for PC1/PC2/PC3 respectively.

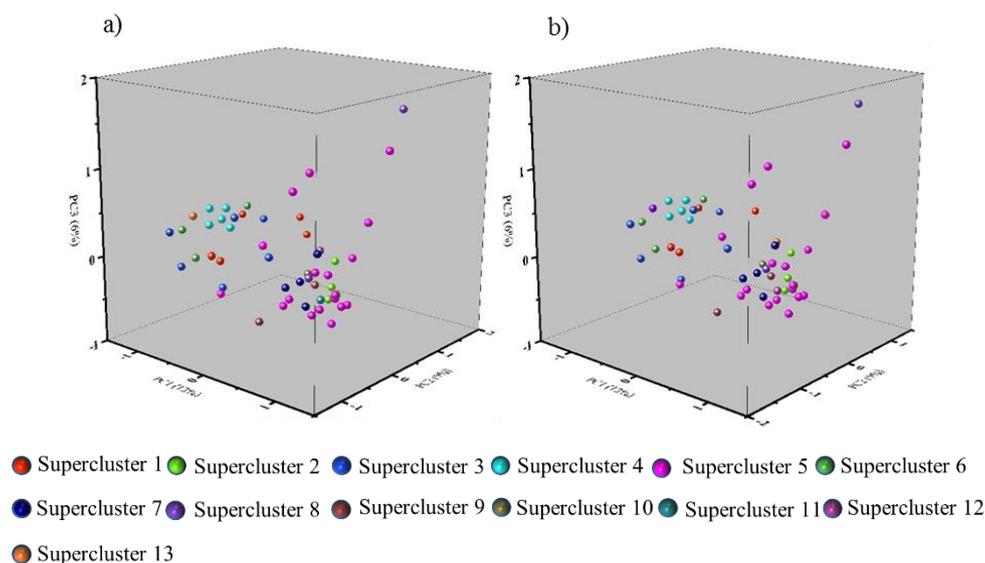


Figure 5. 8: PCA scores plots of average spectra of 53 NPS reference standards for a) 600 bins, b) 300 bins over the spectral ranges 3200 – 250 and 1750 - 250 cm^{-1} respectively. Average spectra were sample-grouped based on their supercluster membership [204].

Binning using 600 bins was applied to the entire dataset (i.e. 530 spectra) and resulted in reduced explained variance and worsening of model robustness as compared to average spectra of 53 NPS reference standards (plots not shown).

In summary, the first variable reduction method aimed at reducing the variable number from 3777 (3200 – 100 cm^{-1}) to 522 (1750 – 1300 cm^{-1}) data points/ spectrum. This method generated more robust PCA models and improved classification of NPS spectra as demonstrated by scores and explained variance plots. In contrast, the second method involving the use of bins reduced the explained variance and model robustness and, hence were not used in this thesis.

The size of the data matrix was also investigated through the use of average (average of 10 replicate spectra), single, three, five or ten spectra. Results were assessed using PCA over the spectral range 1750 - 1300 cm^{-1} . Figure 5.9 showed that the calibrated explained variance for PC1/PC2/PC3 of single, 3, 5 and 10 spectra improved with the increase in the number of replicate spectra. The disadvantage of using average spectra (Figure 5.3b) in the analysis of a highly diverse dataset was that average spectra may mask variations between the 10 replicate spectra. It was important to capture this variation to enable developed models to be reliable in the field. Furthermore, the use of a single spectrum (average or single) did not given a clear insight on overlap between samples or sample superclusters/ clusters. The use of 10 replicate spectra generated the highest explained variance in the model, demonstrating model robustness

(explained variance plot not shown) and provided a better insight into sample projected location on the calculated PCs in the chemical space of selected NPS.

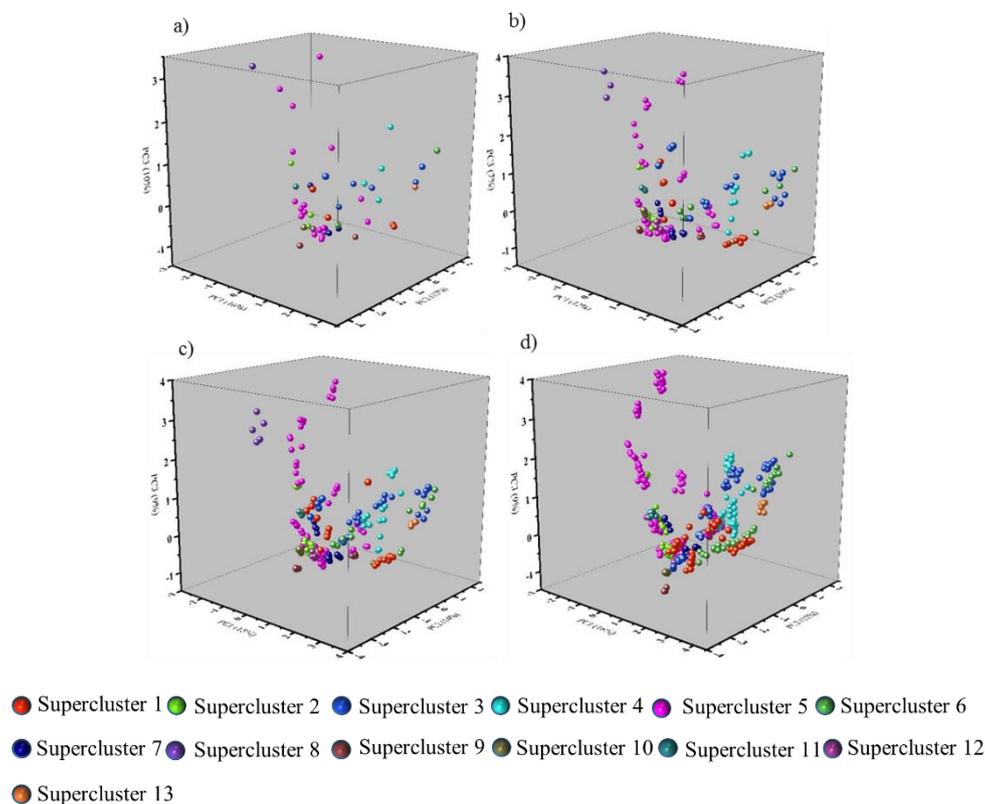


Figure 5. 9: PCA scores plots of a) 1 single spectrum, b) 3 replicate spectra, c) 5 replicate spectra and d) 10 replicate spectra of 53 NPS reference standards over the spectral range 1750 - 1300 cm^{-1} . Spectra were sample-grouped based on their supercluster membership [204].

Hitherto, spectral data were investigated for anomalies using line plots and preliminary correlation coefficient tables between spectra (Microsoft Excel 2013). Cosmic rays were manually ‘zapped’ using WiRE 3.4. Spectral data were then truncated below 250 cm^{-1} and the final spectral range used was 1750 to 1300 cm^{-1} . Truncation was followed by combined pre-processing including smoothing using the SG algorithm, 4th polynomial order and a smoothing window of 21 data points, followed by baseline subtraction using the 2nd derivative, adjacent-averaging smoothing with a window size of 3 points, a threshold of 0.05 and 200 points (Origin Pro 2016). The latter step was followed by transforming the dataset with values with positive signs (by zeroing the negative data points) and maximum normalisation (Microsoft Excel 2013).

5.3.3. Raman and chemometric analysis

5.3.3.1. Overview of PCA classifications

Prior to conducting projection studies, an overview of the dataset was performed using an exploratory PCA of pre-processed NPS Raman spectra (53 NPS x 10 measurements) over the spectral range 1750 – 1300 cm^{-1} . The cumulative calibrated and validated explained variance for 20 PCs was 89.9 and 88.1 % respectively. The first three PCs explained approximately 37 % of the total explained variance, (16/12/9 % for PC1/PC2/PC3). Despite the fact that the dataset includes chemically and structurally diverse NPS, the total percentage variance explained by the first three PCs is very low. This could possibly be because the samples contain Raman active functional groups, which exhibit similar Raman bands over the relatively short spectral region 1750 – 1300 cm^{-1} . In such a case, the inclusion of a large number of PCs risks incorporating explained variance due to residual noise. Therefore the first three PCs only were considered for analysis (Figure 5.10).

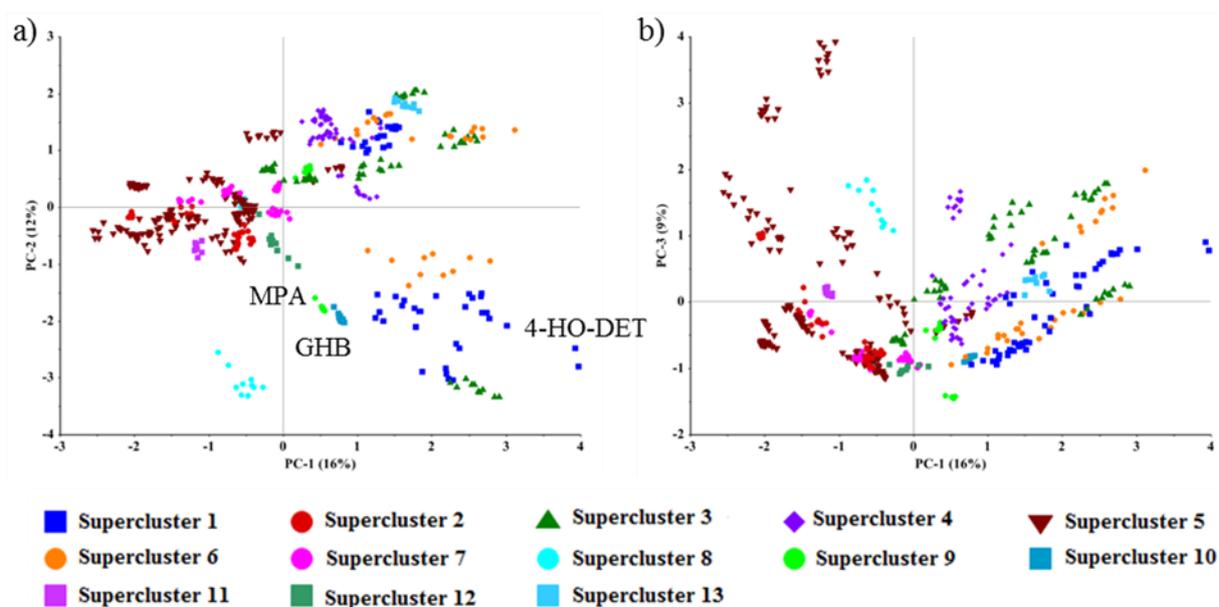


Figure 5. 10: PCA two-dimensional (2D)-score plots for 530 NPS reference spectra over the spectral range 1750 - 1300 cm^{-1} , specifically 2D-scores plot for a) PC1/PC2 and b) PC1/PC3. The scores for 4-HO-DET, GHB and MPA are marked on Figure 5.10a.

The 2D-score plots outline the classification of ‘representative’ NPS reference standards based on superclusters (Figure 5.10 and Table 5.1). The plots showed that replicate spectra of individual NPS grouped together, yet were discriminated from different NPS, demonstrating the efficiency of the pre-processing method as compared to raw unprocessed data. For example, replicate spectra of both MPA (light green – supercluster 9) and GHB (turquoise – supercluster

10) grouped together, yet both MPA and GHB were clearly discriminated from each other (Figure 5.10a).

Exceptions to this observation included pre-processed spectra of NPS that displayed intense fluorescent background, such as 4-HO-DET (dark blue – supercluster 1) (Figure 5.10a). Despite the high chemical diversity between NPS in this dataset, there is a degree of delineation between superclusters. The 2D-score plot for PC1/PC2 showed that members of superclusters 2, 5 (except 5F-APICA), 8 and 11 were dispersed over the upper and lower left quadrants, whereas superclusters 1, 3 (except AM-679), 4, 6, 9, 10 and 13 were dispersed over the upper and lower right quadrants. The plots showed the scatter of superclusters over more than one quadrant, yet NPS belonging to individual EMCDDA/EDND classes occupied distinct regions of the plots, usually in a single quadrant. It is important to note that some classes such as quinazolines and anaesthetics are classified as ‘other’ according to the EMCDDA/EDND classification. However, in this work, they were labelled by their pharmacological or chemical class to enable analysis and interpretation of results.

It was evident that the scores plot displayed grouping of NPS according to their classes. The upper and lower left quadrants contained cannabinoids, tryptamines, γ -hydroxybutyrate and ‘other’ categories of NPS. In contrast, the upper and lower right quadrants contained the cathinones, phenethylamines, arylalkylamines, benzodiazepines, arylcyclohexamines, quinazolines, aminoindanes and anaesthetics. This illustrates the potential efficiency and discriminatory power of the developed model in classifying NPS based on structural diversity.

The only exception to this division (see above) was NPS members of supercluster 9, which were found in the right quadrants but contained NPS from two different EMCDDA/EDND classes; an arylalkylamine (i.e. MPA) and a cathinone (i.e. α -PBT). These two NPSs are different from other compounds belonging to their EMCDDA/EDND classifications but structurally similar to one another in that both incorporate thiofuran moieties. This may be the reason they group together and not according to their EMCDDA/EDND classifications. Figure 5.11 illustrates this finding and shows the structural similarity between α -PBT and MPA. It also showed the difference in structure between α -PBT and another cathinone α -PVP.

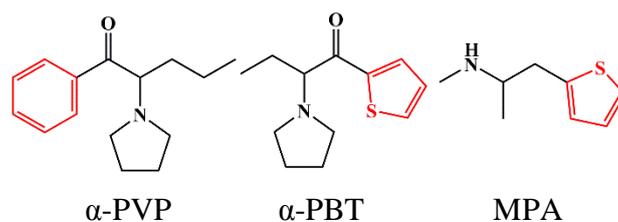


Figure 5. 11: The chemical structures of α -PVP, α -PBT and MPA. The molecular structure for α -PVP is represented as an example of a cathinone analogue with a typical β k-aromatic group.

AM-679 (cluster 3.07) did not cluster with members of supercluster 3 as expected, possibly because the structural similarity with other cannabinoid members of supercluster 3 was relatively poor (TSI values ranging from 29 to 57 %) [204]. The 3D-scores plot (Figure 5.12) shows the classification of NPS based on superclusters, with the largest percentage variance in the dataset explained by PC1 (16 %). Interestingly, the ‘Origin projections’ of supercluster 5 (pink) illustrated a greater explained variance along the planes PC1/PC3 and PC2/PC3. This is possibly because supercluster 5 is the largest supercluster in the dataset containing 18 chemically diverse NPS cluster representatives. ‘Origin projections’ (Origin Pro 2016) represent a different function from ‘PCA projections’ (the Unscrambler® X 10.4). This function plots selected scores on the three planes (PC1/PC2/PC3) to allow the visualisation and confirm closeness of scores to each other on 2D-planes.

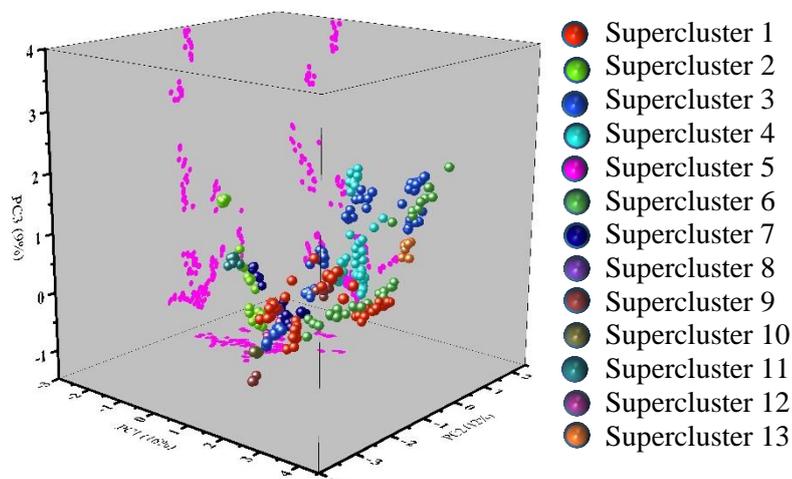


Figure 5. 12: 3D-scores plot of 530 reference spectra. Projections of supercluster 5 are illustrated in pink simultaneously along the three planes PC1/PC2/PC3 (Origin Pro 2016).

5.3.3.2. Prediction of ‘query’ NPS (‘test’ and ‘OOM’) reference standards *via* PCA projection (NPS reference standards model)

In this Section, ‘query’ spectra (i.e. ‘test’ and ‘OOM’ spectra) were projected on the PCA model developed with 53 training NPS (530 NPS reference spectra). This model is termed the ‘NPS reference standards/benchmark’ model. Zloh et al. [204] have calculated similarity

coefficients (TSI values) between ‘query’ samples and anticipated cluster assignments for these samples (Table 5.1). In this Chapter similarity between ‘query’ (i.e. 17 ‘test’ and 4 ‘OOM’ pure NPS) and library substances (i.e. 53 training pure NPS) was evaluated using PCA projection plots (Figure 5.13), correlation coefficients between spectra over the spectral range 1750 - 1300 cm^{-1} (i.e. r^2 values), structural similarity data (i.e. TSI values) if previously calculated by Zloh et al. *via* hierarchical clustering analysis [204], common substructures, Raman spectra and line loading plots.

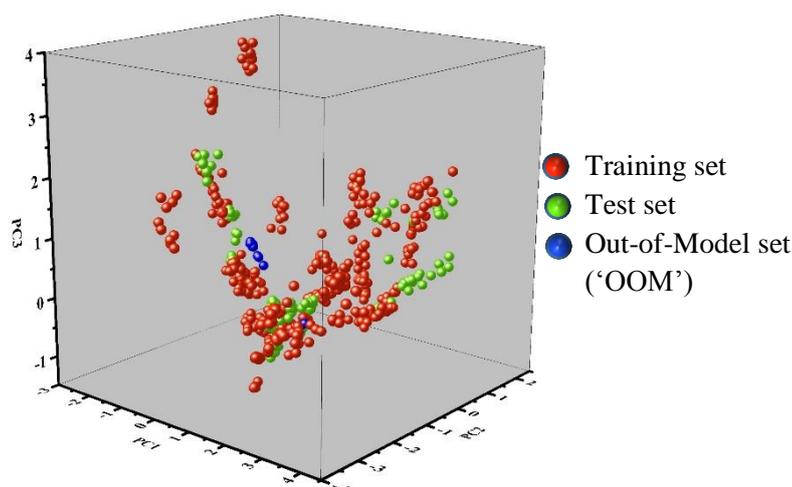


Figure 5. 13: PCA projection plot illustrating training, ‘test’ and ‘OOM’ sets. ‘Query’ spectra (i.e. ‘test’ and ‘OOM’ spectra) are projected on the PCA model developed with 53 training NPS (530 NPS reference spectra).

Table 5.2 showed close calibration and validation values for the training set. This demonstrated the success of the ‘full’ validation method employed, where each sample from the training set was tested against remaining samples as an ‘unknown’. The Table also showed the projection values of the ‘query’ samples, which are close to the training values for three PCs, demonstrating the applicability of the model to unknown samples (i.e. not part of the original model generated from the training samples). This illustrated the optimum performance of the developed method, its applicability to unknown NPS and the potential assignment/ classification/ correlations of newly emerging NPS with their near-neighbours.

Table 5. 2: PCA projection results for 21 unknown pure NPS against the PCA model developed using 53 pure NPS

	PC-0	PC-1	PC-2	PC-3
Calibration (training)	0.022	0.018	0.016	0.014
Validation (training)	0.022	0.019	0.016	0.014
Projection (test and ‘OOM’)	0.020	0.017	0.015	0.013

All training and projected samples were included in the Hotelling T^2 ellipse at 95 % CL, except a few replicate spectra for 4HO-DET, DL-4662 and afloqualone (ellipse not shown in Figure 5.13). However, the examination of these spectra using the influence plot (plot not shown) demonstrated that they are not true outliers (i.e. not associated with high F-residuals and Hotelling T^2 values (see Chapter 4)), hence are described by the calibration model.

Because of the large size of the dataset used in this study, it was challenging to visualise and confirm overlap between ‘query’ and training NPS on the PCA projection scores plots. Therefore in order to evaluate the performance of the PCA projection model, the spectral correlation between ‘query’ samples and superclusters/ clusters/ EMCDDA-EDND classes and position in 3D-PCA projection scores plots was evaluated for each supercluster and for clusters within each supercluster. This was performed by retaining the position of each supercluster from the PCA projection plots. *Tanimoto* similarity indices (TSI) values were calculated for ‘query’ samples against the original clustering model to anticipate supercluster membership of query NPS [204]. In this respect, the projection of ‘query’ samples, which were anticipated to correlate with supercluster 2 for example, were evaluated against the training members of supercluster 2, while retaining their position in the PCA projection plots. In this study, there were no ‘query’ samples to challenge superclusters 1, 9, 10 and 13 and, therefore these superclusters will not be discussed further in the following Section.

Supercluster 2

Supercluster 2 is composed of four training samples including a cathinone (4-MeO- α -PVP) and three phenethylamines (25H-NBOMe, N-Me-2C-B and STP). Both N-Me-2C-B and STP were medoids in clusters 2.06 and 2.08 respectively within supercluster 2. This supercluster was challenged with three test samples i.e. two arylalkylamines (i.e. 5-APB and 6-APB) and one phenethylamine (i.e. β k-2C-B) and one ‘OOM’ sample i.e. the piperidine MPD. STP (training) and both 5-APB and 6-APB (test) are members of cluster 2.08. In contrast, both N-Me-2C-B (training) and β k-2C-B (test) are members of cluster 2.06 (Figure 5.14a). Common substructures between ‘representative’ NPS in supercluster 2 are presented in Figure 5.14d.

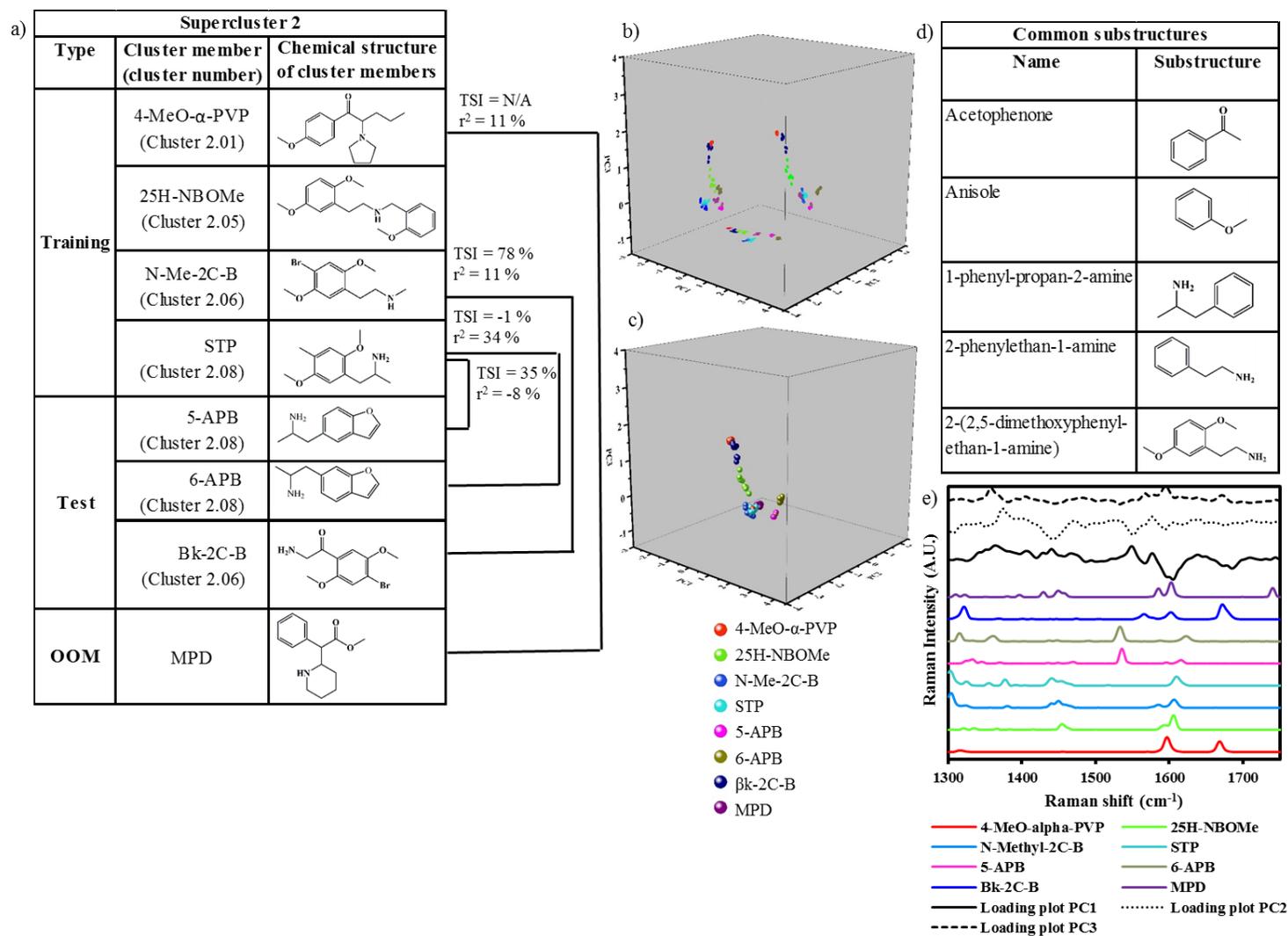


Figure 5. 14: a) NPS training members of supercluster 2 with the relevant test and ‘OOM’ samples, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructures; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.

The PCA projection scores plots (Figures 5.14b and 5.14c) showed that the test samples were in close proximity to the training samples. Both β k-2C-B and N-Me-2C-B were in close proximity along PC1/PC2 as expected, being both 2C compounds as named by Alexander and Anna Shulgin [18]. This is also because both are members of cluster 2.06 with TSI value of 78 %. However, β k-2C-B clustered closer to 4-MeO- α -PVP along PC1/PC2/PC3. This was not surprising as both share a cathinone substructure (i.e. acetophenone) [280]. This was confirmed *via* line plots, which showed that both β k-2C-B and 4-MeO- α -PVP share a Raman band at ca. 1600 cm^{-1} , potentially attributed to $\nu(\text{CC})$ aromatic ring quadrant stretching vibrations. This band correlated with a high negative loading along PC1. Both NPS also exhibit a Raman band at ca. 1670 cm^{-1} , potentially attributed to olefinic ($\text{C}=\text{C}$) stretching vibrations (Figure 5.14e).

β k-2C-B was also close to 25H-NBOMe (cluster 2.05) along PC1/PC2, possibly because both share an anisole group. Again, this was confirmed *via* line plots, which showed that both NPS have a Raman band at ca. 1605 cm^{-1} , potentially attributed to $\nu(\text{CC})$ aromatic ring quadrant stretching vibrations, which was associated with a high negative loading along PC1. For STP and both 5-APB and 6-APB, members of cluster 2.08 with a common 1-phenyl-propan-2-amine substructure (Figure 5.14d), a poor structural similarity was observed (TSI = 35 and 34 % respectively). In addition to poor spectral similarity over $1750 - 1300\text{ cm}^{-1}$ and lack of correlations to variables with high loadings, they were in close proximity on the projection scores plots along PC1/PC2/PC3. This could be due to a shared band at ca. 1620 cm^{-1} , potentially attributed to NH_2 bending vibrations (Figure 5.14e) [281]. It was noted that training and test phenethylamines (see above) grouped with 4-MeO- α -PVP (cathinone) - all share a common 2-phenylethan-1-amine substructure (Figure 5.14d). This could be because 4-MeO- α -PVP, 25H-NBOMe, β k-2C-B, N-Me-2C-B and STP share Raman bands at ca. 1601 , 1609 , 1605 , 1611 and 1615 cm^{-1} , which correlated with high negative loadings along PC1 (Figure 5.14e).

Supercluster 2 was also challenged with MPD, an 'OOM' NPS sample, which was not included in the initial clustering analysis of 478 NPS molecules. MPD was tested against the clustering model developed by Zloh et al. [204] and was assigned to supercluster 2 with a 46 % similarity to the medoid of cluster 2.03 [204]. Therefore the similarity between MPD and supercluster 2 in general and cluster 2.03 in particular was evaluated. MPD clustered tightly with STP (medoid of cluster 2.08) along PC1/PC2/PC3 as well as with N-Me-2C-B (cluster 2.06) (Figure 5.14b). This is potentially because all three NPS exhibited a deformation Raman band at ca.

1443 cm^{-1} , which correlated with weak positive loading along PC1 and negative loading along PC2 (Figure 14e). Although no training samples were included from cluster 2.03, MPD ('OOM') grouped well with members of the parent supercluster, which were in close proximity on the scores plot, due to shared structural features that exhibited Raman active bands in the range 1750 – 1300 cm^{-1} and/ or correlated to important variables in the loadings plots, thus creating a unique pattern that discriminates them from other NPS.

Based on the interpretation provided above, it was concluded that 'test' and 'OOM' samples in supercluster 2 were consistent with training samples of supercluster 2, demonstrating the great potential and degree of specificity of the developed model. Both 5-APB and 6-APB were in close proximity to the training sample STP - member of the same cluster (2.08). β k-2C-B grouped well with members of the parent supercluster, whereas the phenethylamines (training and test) grouped well together as well as with the cathinone training sample 4-MeO- α -PVP due to their structural similarities.

Supercluster 3

Supercluster 3 is composed of five synthetic cannabinoids (AB-FUBINACA, AB-PINACA, THJ-018, SDB-006 and AM-679) and one tryptamine (DPT). AM-679 was the medoid in clusters 3.07. This supercluster was challenged with one test sample MN-18, a synthetic cannabinoid with proposed structural similarity to cluster 3.03, a cluster with two representative NPS [204]. None of the training samples is a member of cluster 3.03, which represents a new challenge to the PCA model (Figure 5.15a).

The Origin projection plot showed that training and test NPS are relatively close in proximity along PC1/PC2/PC3, except for SDB-006, which was delineated along PC1/PC2 and PC2/PC3 (Figure 5.15b). This could possibly be because this supercluster has a larger number of training NPS samples than previously discussed supercluster 2 and or because it contained a larger number of synthetic cannabinoids, which are known to be very chemically and structurally diverse [282].

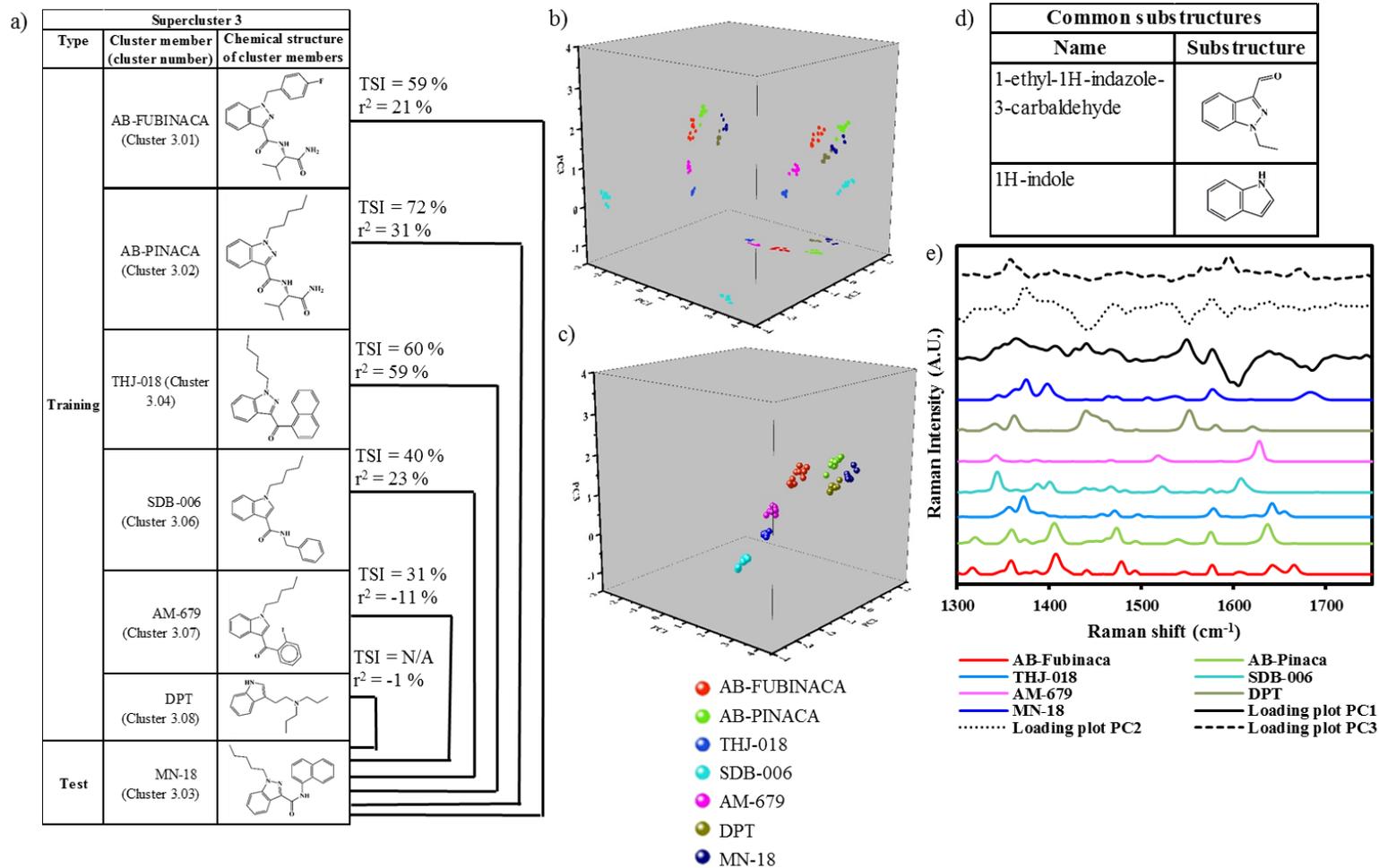


Figure 5. 15: a) NPS members of supercluster 3 with relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructures; e) Raman spectra and PC line loading plots.

The projection plot showed that MN-18 spectra were in close proximity to both DPT and AB-PINACA. MN-18 has a larger common substructure with AB-PINACA (TSI = 72 %) than with DPT (i.e. 1-ethyl-1H-indazole-3-carbaldehyde compared to 1H-indole) (Figure 5.15b). Both MN-18 and AB-PINACA exhibit a Raman band at ca. 1578 cm^{-1} , potentially attributed to a quadrant stretch and corresponding to a moderate positive loading along PC1/PC2. The next NPS in proximity to MN-18 was AB-FUBINACA (cluster 3.01; TSI = 59 %; $r^2 = 21$ %) along PC1/PC2/PC3, despite poor spectral correlation over the range 1750 – 1300 cm^{-1} . The farthest from MN-18 was SDB-006 (cluster 3.06; TSI = 40 %; $r^2 = -11$ %). This was possibly due to the Raman band in SDB-006 at ca. 1610 cm^{-1} , which correlated to a high negative loading leading to its separation from remaining training samples.

It was noted that the structural similarity between MN-18 and cannabinoid training samples was lowest with AM-679 (TSI = 31 %). Structural similarity between AM-679 and remaining training cannabinoids samples was also relatively poor i.e. 29, 29, 43 and 57 % with AB-FUBINACA, AB-PINACA, THJ-018 and SDB-006 respectively. The line plots showed that training cannabinoids, except AM-679, share a peak at ca. 1580 cm^{-1} corresponding to moderate positive loadings along PC1/PC2. This may explain why AM-679 was the only NPS from supercluster 3, which was located in the left quadrant of the PCA scores plot for all the training set 53 NPS as opposed to remaining cannabinoids (see Figure 5.10 and Section 5.3.3.1).

Based on the interpretation provided above, the MN-18 test sample correlated with training samples (i.e. synthetic cannabinoids) of supercluster 3, but slightly delineated from SDB-006 because of a unique Raman band with high loading. This was expected as MN-18 is also a synthetic cannabinoid. Despite the fact that none of the training samples were members of cluster 3.03, the MN-18 test sample grouped closely with the NPS (AB-PINACA) to which it had the highest structural similarity (TSI = 72 %) due to shared Raman bands with moderate loading over the range 1750 – 1300 cm^{-1} .

Supercluster 4

Supercluster 4 is composed of five synthetic cannabinoids (AM-2201, JWH-122, UR-144, JWH-073 and 5F-APICA), which grouped together across PC1/PC2/PC3, as expected. This supercluster was challenged with one test sample JWH-018 (a synthetic cannabinoid) which is a member of cluster 4.02 (Figure 5.16a).

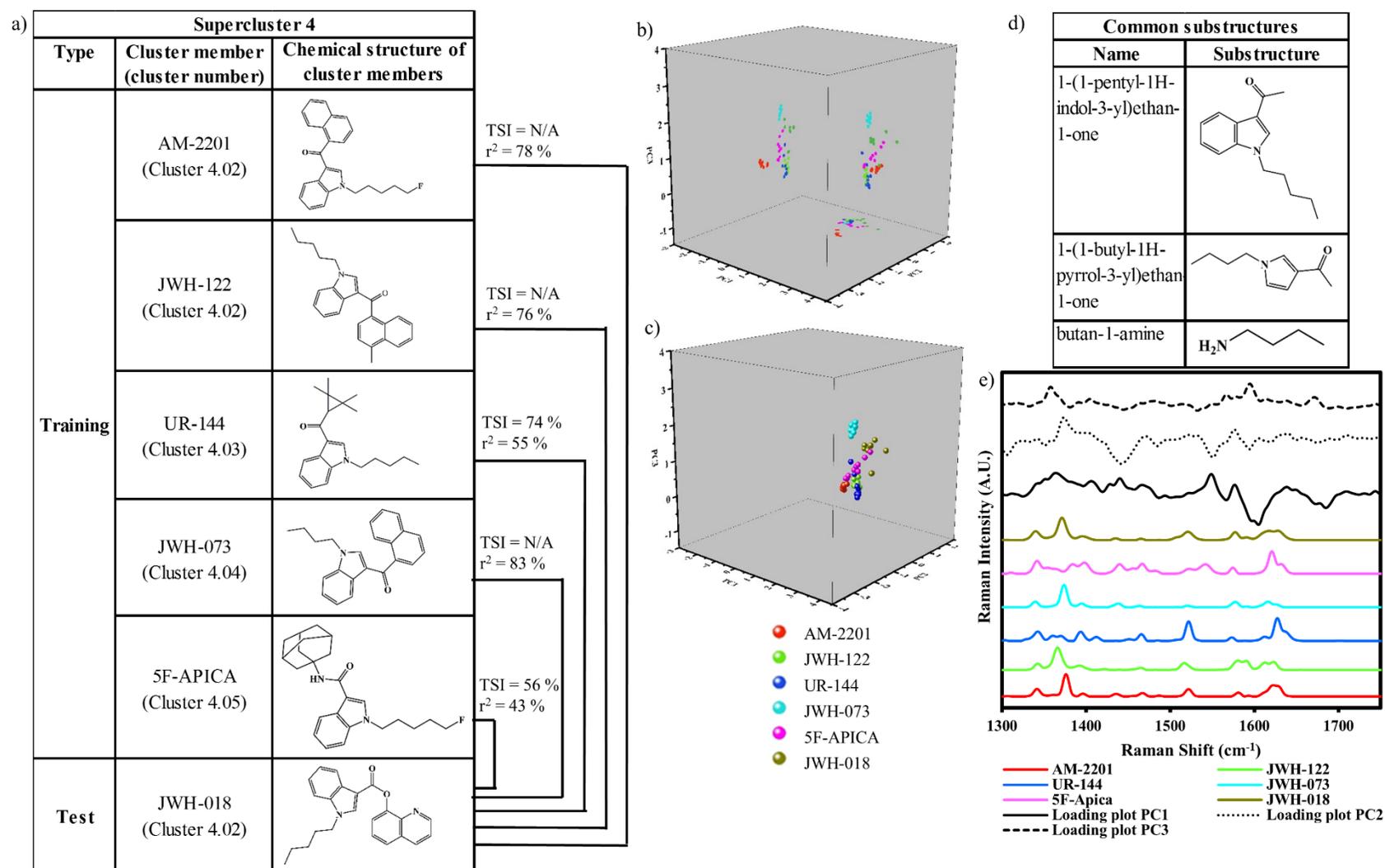


Figure 5. 16: a) NPS training members of supercluster 4 with the relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructures; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.

Both training samples AM-2201 and JWH-122 were members of cluster 4.02, with a common substructure 1-(1-pentyl-1H-indol-3-yl)ethan-1-one (Figure 5.16d). The Raman analysis of JWH-018 (an oil sample) was very challenging and this may explain why the 10 replicates are not tightly grouped on the scores plot. The projection scores plots showed that JWH-018 grouped with all training samples, but it was closer to 5F-APICA (TSI = 56 %; $r^2 = 43$ %) and most delineated from AM-2201 ($r^2 = 78$ %).

The line plots showed a moderate Raman band at ca. 1373, 1376, 1369 and 1379 cm^{-1} for JWH-018, JWH-073, JWH-122 and AM-2201 respectively (Figure 5.16e). This peak, potentially attributed to CH_2 or CH_3 on hydrocarbons, was shifted to a higher frequency in AM-2201 due to adjacent electronegative fluorine atom. This shift could be the reason why both JWH-018 and AM-2201 were delineated on the scores plot.

Based on the above interpretation, the cannabinoid JWH-018 test sample was correlated with all training samples (synthetic cannabinoids) of supercluster 4, as expected. The nature of the sample (i.e. oil) may have resulted in a poorer correlation of the test sample with cluster 4.02.

Supercluster 5

Supercluster 5 is composed of 18 training NPS reference standards, where each sample is a representative of a different cluster (5.01 – 5.18). Supercluster 5 is the largest in this study because members of this supercluster share common substructures, which can be simple moieties such as benzene rings. Therefore it was important to include a large number of ‘representative’ NPS of the wider chemical diversity between clusters. Supercluster 5 is composed of four arylcyclohexamines (3-MeO-PCE, 4-MeO-PCP, ketamine and methoxetamine), five cathinones (4F- α -PVP, 4-Me-N-ethylnorpentadrone, α -PVP, DL-4662 and mephedrone), two benzodiazepines (flubromazepam and phenazepam), two quinazolines (afloqualone and mebroqualone), two synthetic cannabinoids (5F-APINACA and JWH-015), two aminoindanes (2-AI and N-Me-2-AI) and one anaesthetic (dimethocaine) (Figure 5.17).

Four of these NPS i.e. 4F- α -PVP, 4-Me-N-ethylnorpentadrone, 4-MeO-PCP and JWH-015 were medoids of clusters 5.03, 5.04, 5.13 and 5.12 respectively (Figure 5.17 and Table 5.1). The PCA projection scores plot showed specific trends (Figures 5.17, 5.18a and 5.18b). For example, the aminoindanes training samples grouped together, as did the arylcyclohexamines, benzodiazepines, cannabinoids and cathinones (except DL-4662). The anaesthetic dimethocaine grouped with the cathinones, the group ‘other’ (i.e. zopiclone) grouped with

cannabinoids and the quinazolines were delineated along PC3, but closer to each other along PC1/PC2. These observations were further investigated.

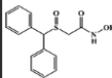
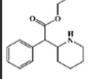
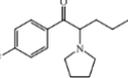
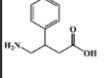
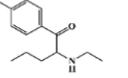
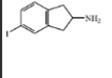
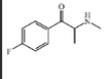
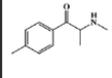
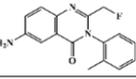
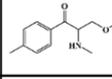
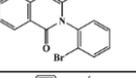
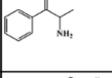
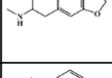
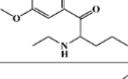
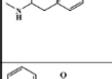
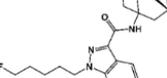
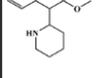
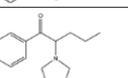
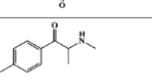
Supercluster 5						
Training	Methoxetamine (Cluster 5.01) (Arylcyclohexamine)		Test	Adrafinil (Cluster 5.14)		
	Ketamine (Cluster 5.02) (Arylcyclohexamine)			Ethyphenidate (EPD) (Cluster 5.14)		
	4F- α -PVP (Cluster 5.03) (Cathinone)			Phenibut (Cluster 5.14)		
	4-Me-N-ethynorpentadone (Cluster 5.04) (Cathinone)			5-IAI (Cluster 5.15)		
	Phenazepam (Cluster 5.05) (Benzodiazepine)			Flephedrone (4-FMC) (Cluster 5.18)		
	Flubromazepam (Cluster 5.06) (Benzodiazepine)			Mephedrone (4-MMC) (Cluster 5.18)		
	Afloqualone (Cluster 5.07) (Quinazoline)			Mexedrone (Cluster 5.18)		
	Mebroqualone (Cluster 5.08) (Quinazoline)			OOM	S-cathinone (Cluster 5.04 and 5.18)	
	3-MeO-PCE (Cluster 5.09) (Arylcyclohexamine)				MDMA (Cluster 5.10)	
	DL-4662 (Cluster 5.10) (Cathinone)				Methamphetamine (Cluster 5.15)	
	5F-APINACA (Cluster 5.11) (Cannabinoid)		Methylphenidate (MPD) (Cluster 5.04; 5.13 and 5.14)			
	JWH-015 (Cluster 5.12) (Cannabinoid)					
	4-MeO-PCP (Clusters 5.13) (Arylcyclohexamine)					
	α -PVP (Cluster 5.14) (Cathinone)					
	2-AI (Cluster 5.15) (Aminoindane)					
	N-Me-2-AI (Cluster 5.16) (Aminoindane)					
	Dimethocaine (Cluster 5.17) (Anaesthetic)					
	Mephedrone (4-MMC) (Cluster 5.18) (Cathinone)					

Figure 5. 17: NPS training members of supercluster 5 with the relevant test and ‘OOM’ samples, cluster membership, TSI and r^2 values.

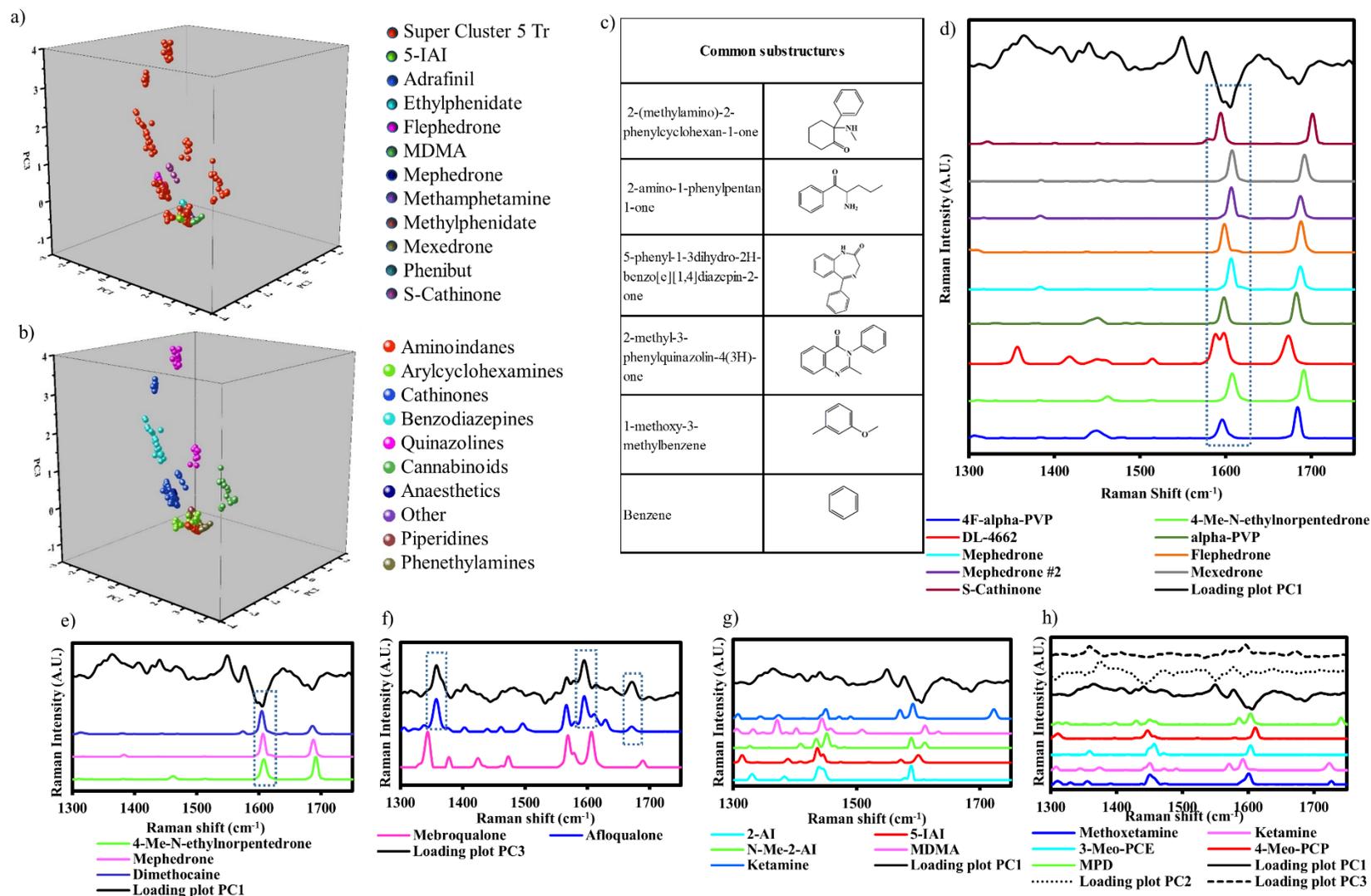


Figure 5. 18: Plots representing the analysis relevant to supercluster 5. a) PCA projection plots (Training samples (red) versus individual test and ‘OOM’ samples; b) training, test and ‘OOM’ samples classified as per EMCDDA/EDND classification; c) common substructures; d-h) selected Raman spectra and PC line loading plots.

In depth analysis showed that DL-4662 did not group well with the cathinones, possibly because DL-4662 spectra exhibit two Raman bands with high negative loadings at ca. 1600 and 1605 cm^{-1} . Each cathinone in supercluster 5 had one peak or the other. DL-4662 was the only cathinone, which had both peaks and, hence is different from the others in this region. It was noted that DL-4662 was most structurally similar to 4F- α -PVP, 4-Me-N-ethylnorpentadone, α -PVP and mephedrone (TSI = 49, 56, 54 and 45 % respectively) (Figure 5.18d).

The anaesthetic dimethocaine grouped with the cathinones along PC1/PC2/PC3, possibly because they shared moderate Raman bands at 1605, 1609 and 1596 cm^{-1} for dimethocaine, 4-Me-N-ethylnorpentadone and mephedrone respectively, which were associated with high negative loadings along PC1 (Figure 5.18e). Dimethocaine also shared a Raman band with mephedrone at 1688 cm^{-1} , potentially attributed to C=O stretches. Although carbonyl stretches usually occurs above 1770 cm^{-1} , conjugation may lower its frequencies. In which case it occurs at 1688 cm^{-1} due to its conjugation to aromatic rings [283, 284]. It was noted that dimethocaine was structurally dissimilar to 4F- α -PVP, 4-Me-N-ethylnorpentadone, DL-4662, α -PVP and mephedrone with TSI values of 22, 23, 22, 23 and 25 % respectively. However, spectral similarity (r^2) between dimethocaine and both 4-Me-N-ethylnorpentadone and mephedrone over the range 1750 – 1300 cm^{-1} was 76 and 90 % respectively, possibly accounting for its position on the scores plot. In contrast to superclusters 1 – 3, where r^2 values between spectra over the designated range were relatively poor (i.e. -11 – 70 %), NPS in superclusters 4 and 5 showed closeness between scores, where r^2 was approximately > 75 %.

Interestingly, mebroqualone (cluster 5.07) and afloqualone (cluster 5.08) shared a common substructure '2-methyl-3-phenylquinazolin-4(3H)-one' (Figure 5.18c), yet a poor r^2 value (i.e. 15 %) in the region 1750 – 1300 cm^{-1} . On the PCA projection plot, both were separated along PC1/PC3 and PC2/PC3, but were closer to each other along PC1/PC2 (largest explained variance), as expected because of their structural similarity. In fact, this is explained *via* line plots, which showed that afloqualone exhibited Raman bands at ca. 1676, 1569 and 1360 cm^{-1} with corresponding high positive loadings along PC3, supporting their separation across PC3 (Figure 5.18f).

Surprisingly, the PCA scores plot showed that zopiclone was in close proximity to THJ-018 (a JWH-018 indazole analogue (cluster 3.04)) and PB-22 (a JWH-018 quinoline carboxylate analogue (cluster 6.02)) along PC1/PC2/PC3, with structural similarity of only 24 and 25 %

respectively. This is because the three NPS showed shared Raman bands over the spectral range 1750 – 1300 cm^{-1} (Figure 5.19). Medium vibrational frequency shifts $\delta(\text{CH}_3)$ were observed in all three spectra at 1374, 1365 and 1373 cm^{-1} for zopiclone, PB-22 and THJ-018 respectively [125]. When alkane groups have identical substituents, then the CH_2 ‘scissors’ deformation and CH_3 out-of-phase deformation are found in the same frequency region 1480 – 1430 cm^{-1} [125]. When the CH_2 and CH_3 are on hydrocarbons, this band is near the 1460 cm^{-1} as is the case with both THJ-018 and PB-22, where weak doublet Raman bands with similar relative intensities, were observed at 1459 and 1469 cm^{-1} for PB-22 and at 1460 and 1473 cm^{-1} for THJ-018. When the CH_2 is near an unsaturated group (piperazine), this band is near the 1450 cm^{-1} as is the case with zopiclone, where weak doublet Raman bands, with similar relative intensities are observed at 1444 and 1466 cm^{-1} . Moderate and weak stretching bands for pyrazine and pyridine were observed at 1581 and 1576 cm^{-1} for zopiclone and PB-22 respectively. A moderate aromatic ring quadrant stretch vibrations (rings in both naphthoyl and indazole groups) is observed in the same region at 1579 cm^{-1} . It is interesting to note that zopiclone is a bioisostere of both THJ-018 and PB-22.

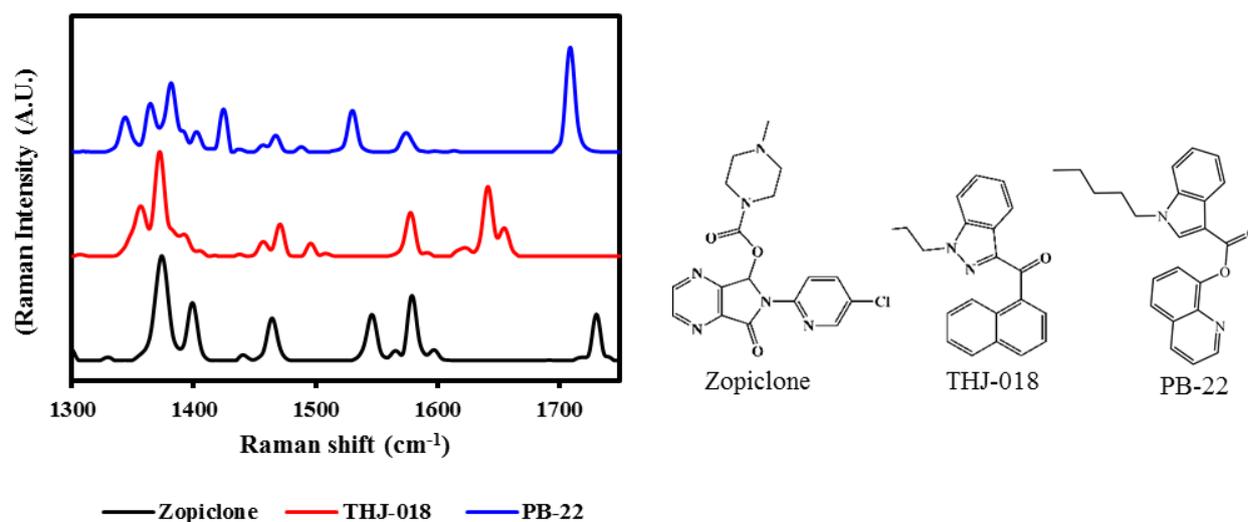


Figure 5. 19: The Raman spectra of zopiclone, THJ-018 and PB-22 and their chemical structures.

Supercluster 5 was challenged with seven test samples (adrafinil, EPD, phenibut, 5-IAI, 4-FMC, 4-MMC and mexedrone) and four ‘OOM’ samples (S-cathinone, MDMA, methamphetamine and MPD). Similarity studies performed by Zloh et al. [204] for the test samples against the clustering model suggested the assignment of adrafinil, EPD and phenibut to cluster 5.14; 5-IAI to cluster 5.15; and flephedrone, mephedrone#2 and mexedrone to cluster 5.18 (Figure 5.17) [204]. Similarity studies performed by Zloh et al. [204] for the ‘OOM’

samples against the clustering model suggested the assignment of S-cathinone to clusters 5.04 and 5.18; MDMA to cluster 5.10; methamphetamine to cluster 5.15 and MPD to clusters 5.04, 5.13 and 5.14 (Figure 5.17) [204]. In the Section below, the grouping/ classification of training, test and ‘OOM’ samples of the same cluster was examined first.

For cluster 5.04, the projection scores plot showed that training and ‘OOM’ samples (i.e. 4-Me-N-ethylnorpentadron, S-cathinone and MPD) grouped together along PC1/PC2. Despite relatively poor spectral correlation coefficients over the spectral region 1750 – 1300 cm^{-1} (i.e. $r^2 = 6 - 27\%$), all compounds had a common ‘benzaldehyde’ substructure and shared Raman bands at 1610, 1596 and 1605 cm^{-1} , associated with high negative loadings along PC1 (line plots not shown). This may account for their proximity and their delineation from other NPS on the scores plot along PC1 (Figure 5.18a).

In contrast, cluster 5.10 training and ‘OOM’ samples (i.e. DL-4662 and MDMA respectively) did not group together, despite a relatively large TSI value (61 %) and common membership to cluster 5.10. A high TSI value does not necessarily reflect common Raman features along the designated spectral range as explained above. Both DL-4662 and MDMA were in close proximity along PC1/PC2 and delineated along PC1/PC3. This is because DL-4662 exhibited Raman bands at ca. 1676, 1598 and 1360 cm^{-1} associated with positive loadings along PC3, leading to their delineation across PC3 (line plots not shown). In general, MDMA was in close proximity to aminoindanes training and test samples (2-AI (cluster 5.15), N-Me-2-AI (cluster 5.16) and 5-IAI (cluster 5.15)) and ketamine (cluster 5.02). This was confirmed *via* line plots, which showed that all spectra have a weak to moderate doublet at ca. 1440 – 1460 cm^{-1} , with varied relative intensities possibly attributed to semi-circle stretches of aromatic rings (Figure 5.18g).

The projection scores plot also showed that MPD grouped tightly with arylcyclohexamines (3-MeO-PCE, 4-MeO-PCP, ketamine and methoxetamine) along PC1/PC2/PC3, except for 4-MeO-PCP. The structural similarity between 4-MeO-PCP and 3-MeO-PCE, ketamine and methoxetamine was 67, 45 and 59 % respectively. This is possibly because all compounds except for 4-MeO-PCP exhibited Raman bands at ca. 1600 or 1605 cm^{-1} , both of which are associated with high negative loadings across PC1 (Figure 5.18h).

The same analysis methodology was employed to test the selectivity of the PCA projection results in identifying/ classifying the test and ‘OOM’ samples for clusters 5.14, 5.15 and 5.18.

Results showed that cluster 5.14 training and test samples (i.e. α -PVP, Adrafinil, EPD, and phenibut) and cluster 5.15 training, test and ‘OOM’ samples (i.e. 2-AI, 5-IAI and methamphetamine) each grouped together, whereas cluster 5.18 training, test and ‘OOM’ samples were in close proximity across PC1/PC2/PC3. Results also showed that MPD grouped well with cathinones. All findings were confirmed by examining and evaluating the PCA projection plots, correlation coefficients between spectra over the spectral range 1750-1300 cm^{-1} (i.e. r^2 values), structural similarity data (i.e. TSI values) if available in Zloh et al. [204], common substructures, Raman spectra and line loading plots.

Based on the interpretation above, PCA projection plots showed that ‘test’ and ‘OOM’ samples used to challenge supercluster 5 correlated with training members of supercluster 5. In fact, ‘test’ and ‘OOM’ samples, members of clusters 5.04, 5.13, 5.14, 5.15 and 5.18 were consistent with members of these clusters, with the exception of cluster 5.10, where the ‘OOM’ sample MDMA correlated with the parent supercluster instead. As stated above, similarity of ‘query’ samples to members of the parent supercluster in this study was mainly linked to spectral features over the spectral range 1750 – 1300 cm^{-1} associated with important variables with high loadings, demonstrating the significance of this spectral range in enhancing the variance between NPS, yet improving their classification to ‘representative’ NPS and, hence superclusters/ clusters.

Supercluster 6

Supercluster 6 is composed of two synthetic cannabinoids (N-PB-22 and PB-22) and one tryptamine (4-acetylpsilocin fumarate (4-AcO-DMT)). Supercluster 6 was challenged with one test sample 5F-PB-22 (a synthetic cannaboinoid) (Figure 5.20a). Both training and test samples (i.e. PB-22 and 5F-PB-22 respectively) are members of cluster 6.02, sharing a common ‘phenyl acetate (quinolin-8-ylacetate)’ substructure (Figure 5.20d). As expected, the projection scores plot showed that both PB-22 and 5F-PB-22 grouped tightly together along PC1/PC2/PC3 ($r^2 = 89\%$). This was confirmed *via* examination of line plots, which showed that 5F-PB-22 and PB-22 shared a number of peaks including peaks at ca. 1715 and 1746 cm^{-1} . The latter peak, possibly attributed to C=O, was shifted to the higher end of the carbonyl region due to its conjugation to 1H-indazole rather than 1H-indole as is the case for both PB-22 and 5F-PB-22 [283, 284].

Both 5F-PB-22 and PB-22 shared a triplet at ca. 1347, 1367 and 1383 cm^{-1} , possibly attributed to the aliphatic chain vibrations. The aliphatic chain in both NPS is attached to 1H-indole as

opposed to 1H-indazole, as in the case of N-PB-22. This triplet was associated with a positive medium loading along PC1. Furthermore, the cannabinoids 5F-PB-22, PB-22 and N-PB-22 also were in close proximity to each other on the PCA projection plot. This is potentially because they shared common peaks at ca. 1580 and 1430 cm^{-1} , associated with a medium positive loadings along PC1/PC2 (Figure 5.20e).

Results showed that 4-AcO-DMT was delineated from PB-22 and 5F-PB-22 along PC1/PC2 and PC2/PC3 but was closer (but still delineated) along PC1/PC3 (Figure 5.20c). This could be due to a peak (strong) at ca. 1551 cm^{-1} exhibited by 4-AcO-DMT, but absent in both PB-22 and 5F-PB-22 and possibly attributed to an NH deformation band. However, these are usually weak Raman bands, which occur between 1660 and 1500 cm^{-1} [125]. 4-AcO-DMT incorporates a 1H-indole group, which is not attached to an aliphatic chain as in the case of both PB-22 and 5F-PB-22. 4-AcO-DMT is the only NPS in supercluster 5, among training and test samples, which does not have a quinolone group (results in a similar ring stretching between 1660 and 1500 cm^{-1}). Therefore this should not be confused with this peak as this group is not present in this molecule. The peak at 1551 cm^{-1} correlated with a medium positive loading along PC1 and negative loading along PC2. This may be why it is delineated from both PB-22 and 5F-PB-22.

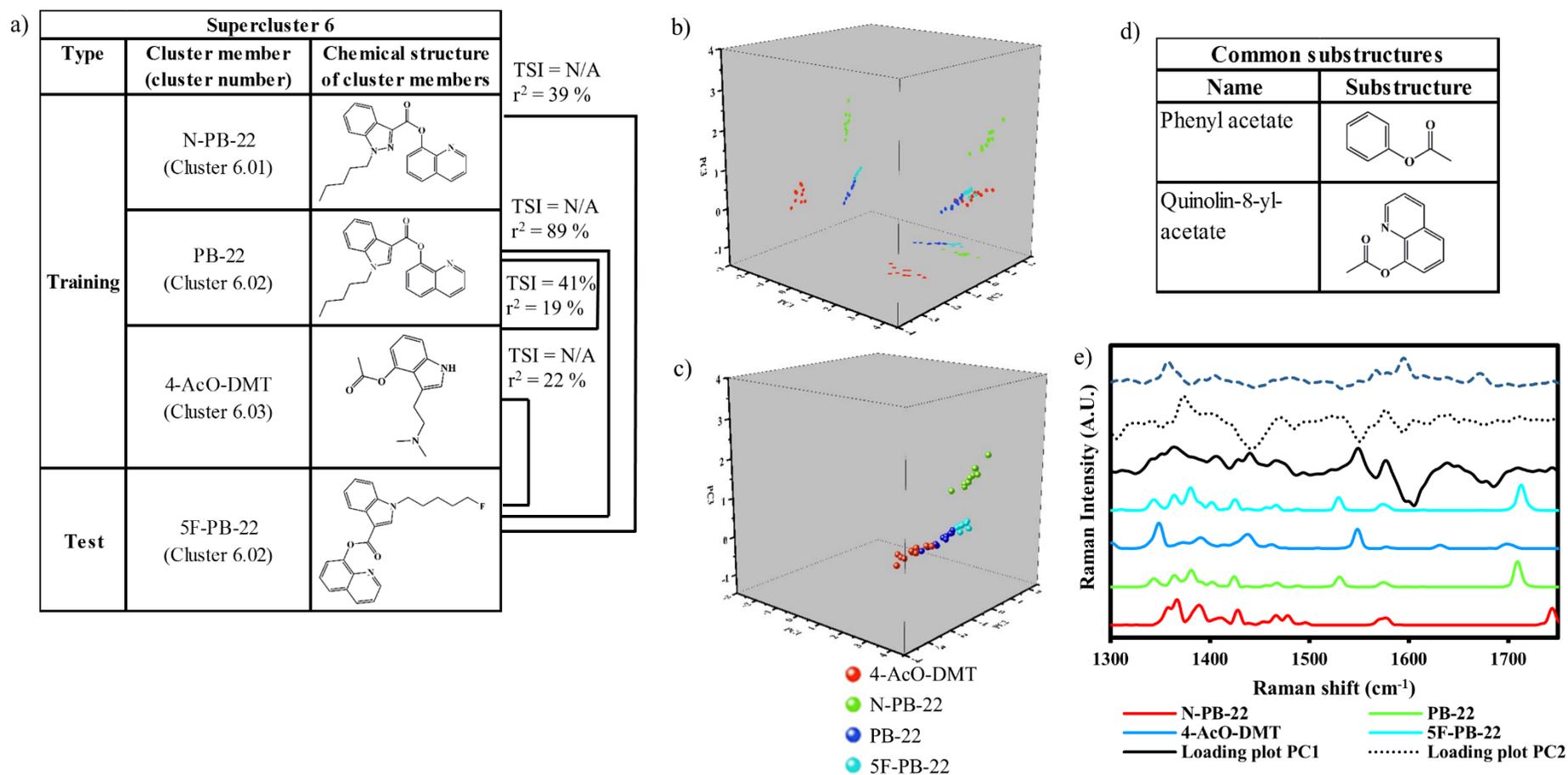


Figure 5. 20: a) NPS training members of supercluster 6 with the relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X10.4); d) common substructures; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.

Furthermore, another peak unique to 4-AcO-DMT is at ca. 1444 cm^{-1} (medium), associated with positive loadings along PC1 and high negative loadings along PC2, again highlighting why 4-AcO-DMT is delineated from both PB-22 and 5F-PB-22 along PC1 and PC2. As described above, both 5F-PB-22 and PB-22 share a band at ca. 1367 cm^{-1} , absent in 4-AcO-DMT and associated with a positive loading along PC3. These findings explained the discrimination between both 5F-PB-22 and PB-22 and 4-AcO-DMT along the three PCs.

Supercluster 7

Supercluster 7 is composed of three arylalkylamines (2-MAPB, 5-EAPB and 6-MAPB) and one phenethylamine (N-ethylamphetamine). Supercluster 7 was challenged with one ‘OOM’ sample methamphetamine (a phenethylamine), which is a member of cluster 7.04 (Figure 5.21a). The training sample N-ethylamphetamine is also a member of cluster 7.04 and they both shared a common substructure ‘butan-2-amine’ (Figure 5.21d).

The projection scores plot showed that the ‘OOM’ sample methamphetamine grouped tightly with N-ethylamphetamine, 5-EAPB and was in close proximity with 6-MAPB. 2-MAPB was delineated from remaining NPS, which grouped together. These findings were confirmed along PC1/PC2/PC3 (Figure 5.13).

The ‘OOM’ sample methamphetamine grouped tightly with N-ethylamphetamine as expected. This is possibly because both NPSs are phenethylamines members of the same cluster with a common butan-2-amine substructure (Figure 5.21c). Both substances also shared numerous common peaks and this may explain why they grouped tightly on the projection scores plot. The first one is a strong peak at ca. 1605 cm^{-1} , correlated with high positive loadings along PC1 and possibly attributed to aromatic quadrant stretch vibrations. Another moderate peak shifted at ca. 1585 cm^{-1} , potentially attributed to aromatic quadrant stretch vibrations. The Raman effect for aromatics has been shown to exhibit bands at ca. 1620 – 1565 cm^{-1} regardless of the symmetry or the position of substitution on the ring, as opposed to IR [283, 284]. Finally, a peak at ca. 1430 cm^{-1} (weak), associated with a weak positive loading along PC1.

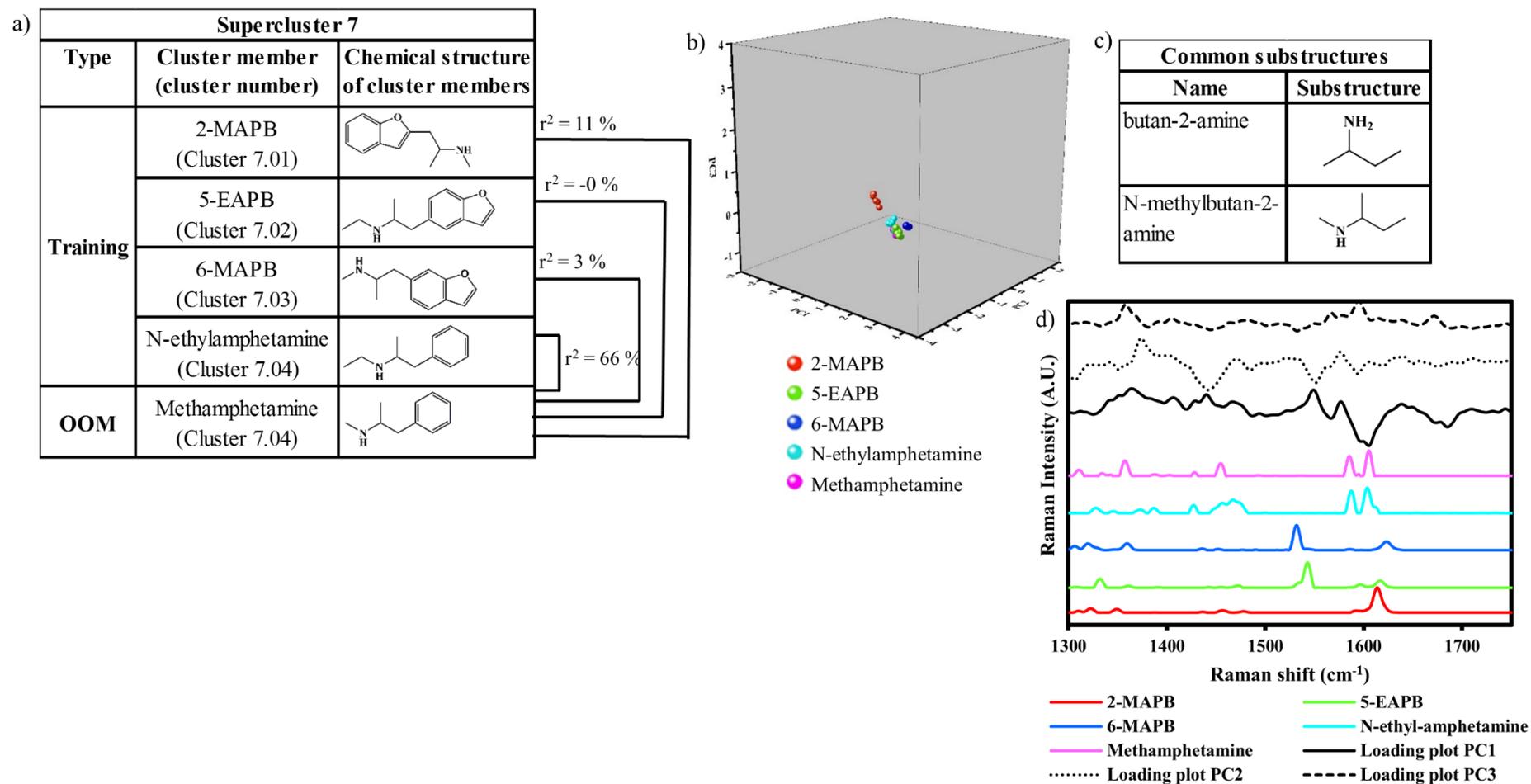


Figure 5. 21: a) NPS training members of supercluster 7 with the relevant ‘OOM’ sample, cluster membership numbers and r^2 values and chemical structures; b) PCA projection plot (Unscrambler® X10.4); c) common substructures; d) Raman spectra and PC line loading plots, PCA projection plot for PC1/PC2/PC3.

By examining the chemical structures (Figure 5.21a), 2-MAPB is the only compound in the group with an aliphatic chain attached to a furan group: the other compounds have the aliphatic chain attached to a benzene ring in the benzofuran group. However, the lack of correlation to important variables with high loadings across the range $1750 - 1300 \text{ cm}^{-1}$ failed to explain why 2-MAPB is delineated from remaining NPS on the PCA projection plot. In contrast, the structural similarity between N-ethylamphetmine and the arylalkylamines in the training set was 30, 41 and 47 % with 2-MAPB, 5-EAPB and 6-MAPB respectively. 6-MAPB was in close proximity to N-ethylamphetamine on the projection scores plot, possibly because they share a peak at ca. 1360 cm^{-1} (weak), possibly attributed to CH_3 in-phase ('umbrella') deformation that is dependent on the electronegativity of the adjacent atom. CH_3 was 1 carbon away from N with N-ethylamphetamine, whereas it was conjugated directly to N with 6-MAPB. Usually, the more electronegative the atom, the higher the frequency with a range ($1470 - 1250 \text{ cm}^{-1}$), which was not the case here (Figure 5.21d).

Supercluster 8

Supercluster 8 is composed of two benzodiazepines (flubromazolam and etizolam). Supercluster 8 was challenged with one test sample pyrazolam (a benzodiazepine), which is a member of cluster 8.02. The training sample flubromazolam is also a member of cluster 8.02 (Figure 5.22a). All training and test samples shared a common substructure '3,7-dimethyl-9H-[1,2,4]triazolo[4,3-a][1,4]diazepine' (Figure 5.22d). Most benzodiazepines include a 7-membered ring, an additional benzene ring and an electron attracting group at position 7 of the fused heterocyclic rings to ensure biological activity. Benzodiazepines are sub-categorised according to the functional group attached to the 7-membered ring, which may include keto, hydroxyl, imidazo or triazolo groups. The three NPS (training and test) in this supercluster have a triazolo group as the functional group.

The PCA projection scores plot showed that etizolam was delineated from both pyrazolam (test sample) and flubromazolam not only because the latter two NPS belong to the same cluster (8.02) but also possibly because etizolam does not have the additional benzene ring or an electron attracting group at position 7 on the fused heterocycle. Etizolam was closer to both pyrazolam and flubromazolam along PC1/PC2. In contrast, both pyrazolam and flubromazolam were in close proximity along PC1/PC2 and PC1/PC3 but they were delineated along PC2/PC3 (Figure 5.22b). These findings were confirmed by examination of the line plots (Figure 5.22e).

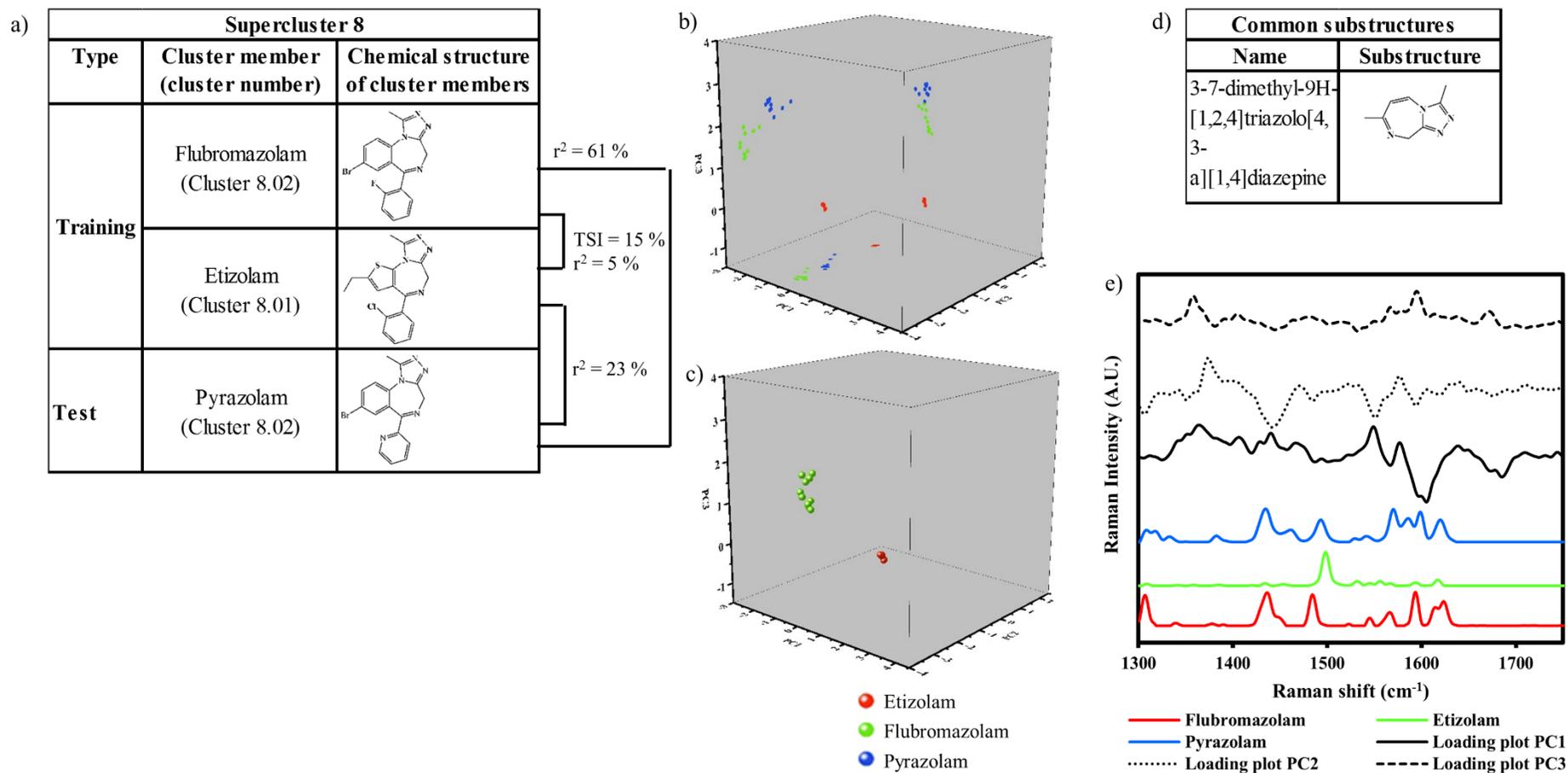


Figure 5. 22: a) NPS training members of supercluster 8 with the relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4) (Pyrazolam scores are behind flubromazolam scores); d) common substructure; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.

The Raman spectra showed that both flubromazolam and pyrazolam shared a strong peak at ca. 1600 cm^{-1} (absent in etizolam) with high negative loadings along PC1 and medium positive loadings along PC3. They also shared a peak (weak) at ca. 1572 cm^{-1} (absent in etizolam) associated with a weak positive loading along PC2. In contrast, flubromazolam (strong), pyrazolam (strong) and etizolam (very weak) shared a peak at ca. 1444 cm^{-1} , which was associated with weak positive loadings along PC1 and strong negative loadings along PC2.

Supercluster 11

Supercluster 11 is composed of one training cathinone sample (methydone (medoid)) because it comprises only one cluster and was challenged with one test sample 5,6-MDAI (an aminoindane) and one 'OOM' sample MDMA (a phenethylamine), suggested to be members of the same supercluster (Figure 5.23a and Table 5.1) [204]. Supercluster 11 comprises one cluster, where the common substructure for all members is 'benzo[d][1,3]dioxol' (Figure 5.23d). Testing the PCA model for supercluster 11 with only one training NPS sample represented a new challenge to the projection model. The PCA projection plot showed that both 5,6-MDAI and MDMA are grouped together but methydone is slightly delineated from them along PC1/PC2 and PC1/PC3. They were closer together along PC2/PC3 (Figures 5.23b and 5.23c).

The Raman spectra showed that all three NPS shared a peak at ca. 1446 cm^{-1} , correlated with a weak peak with positive loading along PC1 and a strong peak with negative loadings along PC2. In addition, this peak had different intensities (normalised), where it was strong for MDMA, moderate and slightly shifted to 1449 cm^{-1} for methydone and weak broad for 5,6-MDAI. This explained why the three NPS were in close proximity along PC2/PC3, yet delineated along PC1/PC2 and PC1/PC3 (Figure 5.23e). Methydone also exhibited a strong peak at ca. 1679 cm^{-1} , which is correlated with a moderate negative loading along PC1. This explained why it was delineated from both 5,6-MDAI and MDMA along PC1.

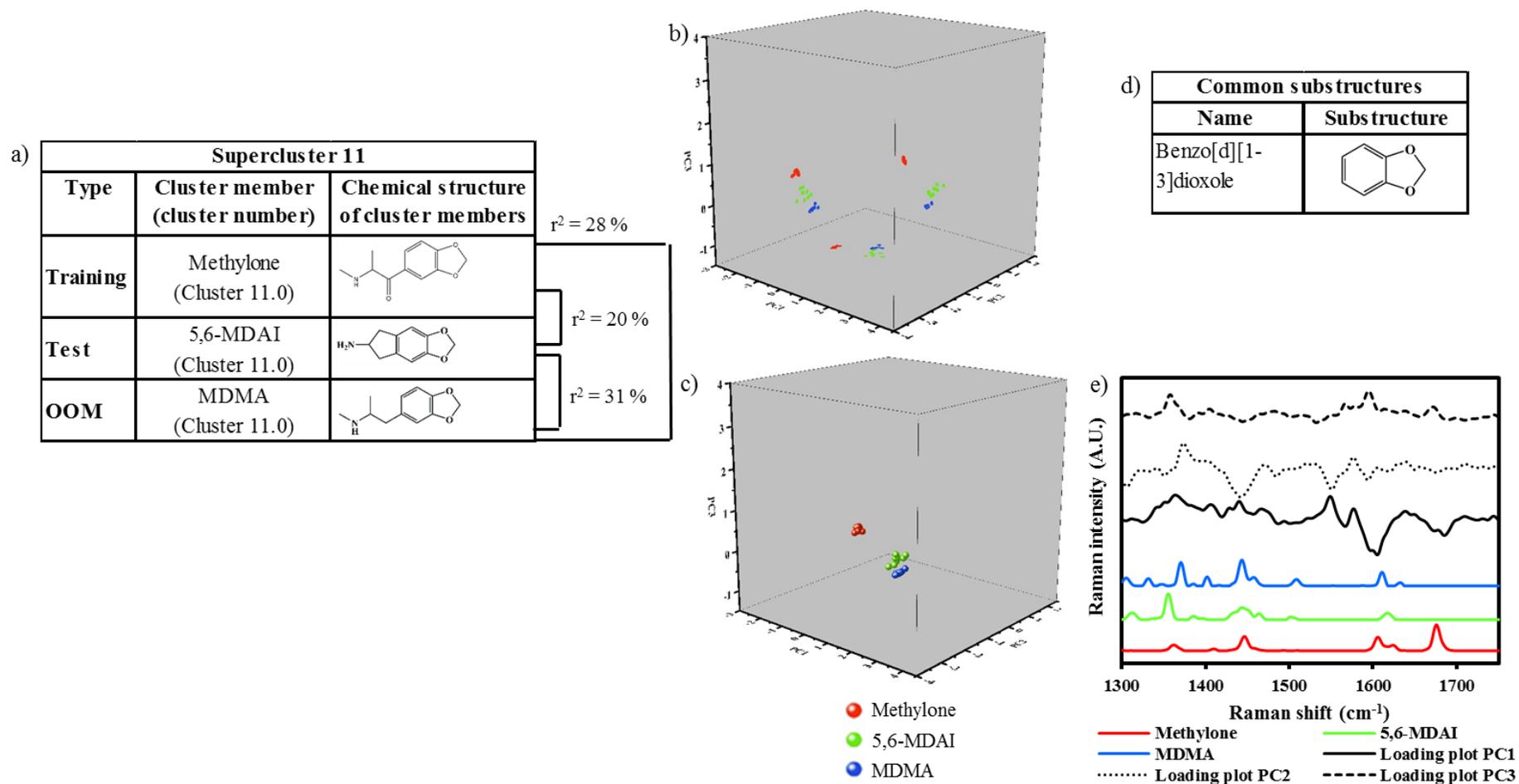


Figure 5. 23: a) NPS training members of supercluster 11 with the relevant test sample, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructure; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.

In conclusion, both test and ‘OOM’ samples were grouped together but the training sample was marginally delineated from them along PC1/PC2 and PC1/PC3, yet in close proximity along PC2/PC3. A relatively poor correlation of the test sample was observed, in this Section due to numerous reasons. Firstly, the raw spectra of the test sample (i.e. 5,6-MDAI) exhibited high fluorescent background, resulting in poor correlation coefficient values across the full spectral range 3200 – 250 cm^{-1} (i.e. an average r^2 between replicate spectra of 74 %), leading to a relatively poor classification. Harkai and Putz [161] highlighted that the pre-processing of spectra displaying intense fluorescence signals risks precluding the weaker Raman signals, which, in turn hinders the identification of the ‘unknown’ sample, and this could be the reason 5,6-MDAI did not group with the training sample methylene. Similarity values (TSI) between MDMA and both 5,6-MDAI and methylene were not initially calculated by Zloh et al. [204], since MDMA was not included in the initial clustering model. If MDMA had higher TSI values for 5,6-MDAI as opposed to methylene, this could be a reason why both MDMA and 5,6-MDAI were in close proximity to each other than with methylene on the projection PCA scores plot. This demonstrates the challenge of classifying and identifying a previously unknown NPS such as MDMA, despite its relative high similarity with the medoid of supercluster 11 (i.e. 81 %). Training only one sample from supercluster 11 is a limitation of this dataset.

Supercluster 12

Supercluster 12 includes one training synthetic cannabinoid sample (C8-CP 47,497) and was challenged with two test samples BB-22 (a cannabinoid) and DXM (an NPS classified as ‘other’), which were suggested to be members of supercluster 12 (Figure 5.24a). Supercluster 12 is made of one cluster, where the common substructure for all members is ‘cyclohexane’ (Figure 5.24d).

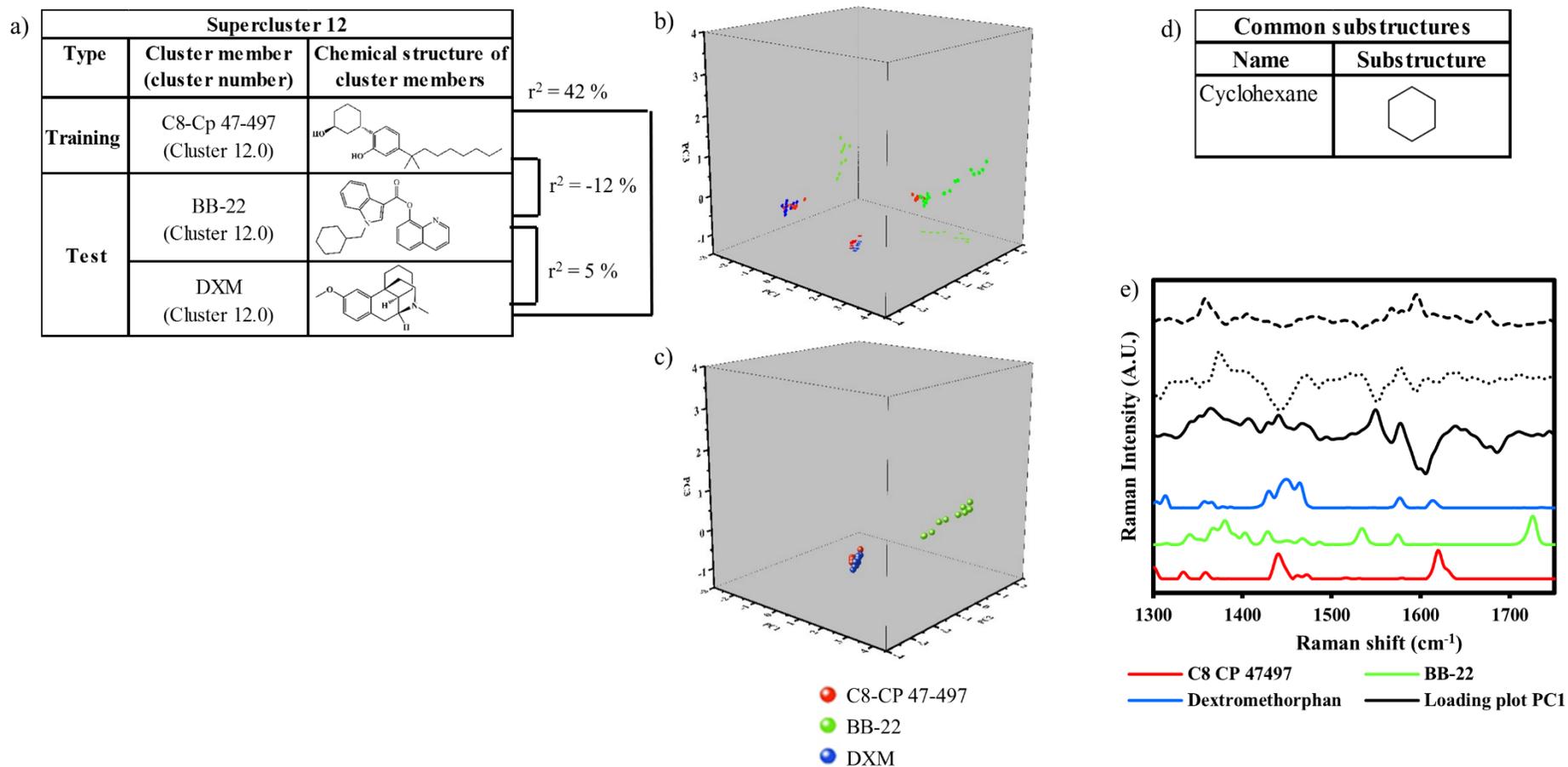


Figure 5. 24: a) NPS training members of supercluster 12 with the relevant test samples, cluster membership, TSI and r^2 values and chemical structures; b) PCA projection plots (Origin Pro 2016); c) PCA projection plot (Unscrambler® X 10.4); d) common substructure; e) Raman spectra and PC line loading plots for PC1/PC2/PC3.

The projection PCA scores plot showed that both C8-CP 47,497 and DXM grouped together along PC1/PC2/PC3 ($r^2 = 42\%$) (Figures 5.24b and 5.24c), possibly because they shared common Raman bands at ca. 1) 1623 cm^{-1} (weak); a doublet at ca. 2) 1474 and 3) 1462 cm^{-1} (weak); and 4) 1443 cm^{-1} (strong) for C8-CP 47,497 and at ca. 1) 1616 cm^{-1} (strong); a doublet at ca. 2) 1466 and 3) 1453 cm^{-1} (moderate); and 4) 1434 cm^{-1} (weak) for DXM. These peaks were shifted towards the higher frequency shift range in C8-CP 47,497 as compared to DXM, possibly due to the presence of two hydroxyl groups (increased electronegativity) in C8-CP 47,497 (Figure 5.24e).

In contrast, BB-22 was delineated from both C8-CP 47,497 and DXM along PC1/PC2/PC3. This is possibly because over the spectral range $1750 - 1300\text{ cm}^{-1}$, BB-22 exhibited Raman bands at ca. 1731 (strong) and 1538 (moderate) cm^{-1} (both absent in C8-CP 47,497 and DXM spectra) but these were not associated with important variables along the loading plots. Hence, the delineation was not fully understood. However, in Figure 5.10a, C8-CP 47,497 was the only cannabinoid that did not group with remaining cannabinoids in the right quadrants of the PCA scores plots. The validation results of the ‘hierarchical clustering’ model by Zloh et al. suggested that both DXM and BB-22 are 37 and 21 % structurally similar to the medoid (i.e., C8-CP 47,497). Therefore, this may suggest why both DXM and BB-22 were closer to each other than with the training sample C8-CP 47,497 [204]. In contrast to most cannabinoids, which contain aromatic rings and heteroatoms, C8-CP 47,497 contains a cyclohexane ring and hydroxyl groups, whereas BB-22 contains conjugated rings, heteroatoms and ester and carboxyl groups, and this could be another reason why the test samples did not group with the training sample (Figure 5.24a).

5.4. Conclusions

Preliminary PCA projection results demonstrated that the combined use of similarity studies, Raman spectroscopy and chemometrics has the potential of successfully suggesting the chemical scaffolds of ‘unknown’ NPS. The optimum performance of the developed ‘NPS reference standards/benchmark’ model was achieved following robust pre-processing methods. In this study, the feasibility of suggesting the chemical scaffolds of ‘query’ NPS to their near-neighbours from ‘representative’ NPS Raman libraries was achieved with a 76 and 75 % success rate for test and ‘OOM’ NPS reference standards respectively. However, suggesting

the chemical scaffolds of 'OOM' samples was more challenging, possibly because they exhibit structural similarity to multiple superclusters/ clusters.

Exploratory PCA showed that thirteen superclusters were classified over right (stimulants and depressants) and left (hallucinogens and cannabinoids) quadrants of the PCA scores plot and comprised NPS reference samples from distinct EMCDDA/EDND classes. Thirteen superclusters were also classified such that replicate NPS spectra grouped together, yet they were delineated from other NPS.

Results showed that 76 % (n = 13/17) of the 'test' samples correlated with their anticipated cluster. Two test samples (i.e. β k-2C-B and MN-18) correlated with parent superclusters instead. However, for MN-18, there were no training samples from the same cluster (i.e. 3.03). Both 5,6-MDAI and BB-22 were not successfully aligned with the expected cluster or supercluster. The raw spectra of 5,6-MDAI exhibited intense fluorescent background, which may have hindered the accurate correlations of this substance with pre-processed spectra of training samples members of supercluster 11. In contrast, the test sample BB-22 was very structurally different from the single training sample C8-CP 47,497 in supercluster 12.

Similarity studies conducted by Zloh et al. [204] against initial clustering model have suggested the possible assignment of the 'OOM' samples to numerous superclusters/ clusters. Results showed that 75 % (n = 3/4) of the 'OOM' samples correlated with suggested clusters and class. S-cathinone successfully grouped with cluster 5.18 (cathinones), whereas methamphetamine successfully consistent with clusters 5.15 (2-AI, an aminoindane and a modified phenethylamine) and 7.04 (the phenethylamine N-ethylamphetamine). It was also consistent with members of the parent supercluster i.e. N-Me-2-AI, an aminoindane analogue, member of cluster 5.16 and ketamine, member of cluster 5.02. In addition, the piperidine MPD successfully correlated with the piperidine analogue EPD member of cluster 5.14. MPD also correlated with the arylcyclohexamines members of the parent supercluster (clusters 5.01, 5.02 and 5.09) expect with 4-MeO-PCP (cluster 5.13). In contrast, inaccurate correlations were observed with MDMA, which was not consistent with clusters 5.10 or 11.0 or specific members of the parent superclusters. This preliminary work has shown that NPS were delineated or grouped together based on common Raman active bands over the designated spectral range, which may have been associated with significant loadings.

Further work should focus on larger number of training samples representing individual clusters. Future work should include larger number of ‘test’ and ‘OOM’ samples in general and particularly to challenge major superclusters such as supercluster 5. More work is needed to evaluate the model’s robustness by challenging the superclusters with NPS, which have relatively poor similarity to cluster medoids (as opposed to the work performed in this Chapter), non-NPS samples and NPS complex mixtures.

6. Classification of NPS internet samples using NPS, adulterants and cutting agents reference standards *via* Raman spectroscopy and chemometric approaches – a proof-of-concept

6.1. Introduction

This Chapter evaluates the use of a Raman database of ‘53 representative’ new psychoactive substances (NPS) reference standards (see Chapter 5 and Zloh et al. [204]) and 22 cutting agents/ adulterants for the identification of 25 NPS products purchased from the internet. In this Chapter spectra for cutting agents/ adulterants were added to the training set to build complexity to the PCA model and evaluate the feasibility of categorising the chemical scaffolds of NPS in ‘street-like’ samples using the new ‘NPS and non-NPS model’.

Using handheld Raman instruments, in-field identification of NPS is often dependent on prior knowledge of the ‘query’ NPS, efficiency of in-built algorithms and availability of reference spectra in in-built libraries [12, 128]. The latter is usually dependent on the availability of reference standards. However, this is not always possible because updated libraries and synthesis of reference standards are always lagging behind the continuous emergence of NPS. In addition, the heterogeneity of NPS products in terms of composition, concentrations, number of constituents and the presence of unknown constituents or impurities, which may emit fluorescence signals may constitute other tactics used by drug designers to hinder identification [12, 78, 140].

The use of Raman spectroscopy with a long laser excitation wavelength (λ_{ex}) (i.e. 1064 instead of 785 nm) for improving the identification of NPS internet products has been investigated [12, 161]. In Chapter 3, results showed improved identification of NPS by 30 % due to reduced fluorescence. However, limitations of in-built libraries, algorithms and products’ complexity hindered the identification of 31/60 NPS products. Chemometric methods were developed and have been shown to improve classification of simple mixtures (Chapter 4). Chemometric methods were then developed using a chemically diverse ‘representative’ spectral NPS database (Chapter 5) and has successfully suggested the chemical scaffolds of 76 and 75 % test and ‘out-of-model’ NPS reference standards.

In this Chapter a ‘new’ PCA model is generated from the ‘representative’ NPS Raman database, previously generated in Chapter 5, as well as cutting agents/ adulterants commonly

incorporated in seized NPS samples. The feasibility of discriminating between NPS and non-NPS substances i.e. cutting agents/ adulterants using the ‘new’ PCA model is evaluated. Method validation is carried out to evaluate the classification of NPS internet products (‘query’ samples) by projecting the Raman spectra of the internet products onto the ‘new’ PCA model (training samples). Internet samples were selected such that claimed NPS on the label or class analogues are included in the training set, selected NPS are of sufficient diversity representing various NPS superclusters and chemical classes (Chapter 5) and characterised with a wide range of HPLC purity profiles

This study evaluates the effect of Raman spectral data pre-processing (Chapters 4 and 5) on the classification and/ or assignment to previously described superclusters/ clusters (Chapter 5) of known pure substances and suggestion of chemical scaffolds of NPS contained in mixtures of unknown composition using PCA. This work is novel and significant because it investigates the feasibility of improving the identification of ‘unknown’ NPS products from Raman spectral libraries containing the ‘representative’ NPS Raman database, common cutting agents/ adulterants and chemometric methods.

6.2. Materials and Methods

6.2.1. Reagents and chemicals

In this Chapter, an initial PCA model is developed using 53 NPS (see Chapter 5), eight adulterants (i.e. benzocaine (BEN), caffeine (anhydrous) (CAF), diltiazem hydrochloride (DIL), lidocaine hydrochloride (LID), paracetamol (PAR), phenacetin (PHE), procaine hydrochloride (PRO) and theophylline (THEO)) and 14 cutting agents (i.e. α -lactose monohydrate (LAC), α -D-glucose anhydrous (GLU), calcium carbonate (CaCO₃), creatine monohydrate (CRE), dextrose monohydrate (DEX), d-mannitol (MAN), L(+)-glutamic acid monosodium salt monohydrate (GLUT), L-tyrosine (L-TYR), maize starch (STA), microcrystalline cellulose (MCC), niacinamide (NIA), sucrose (SUC), talc (TAL) and taurine (TAU)) (Appendix A Table 2.1 and Appendix D Tables A6.1). In contrast to Chapters 2 – 4, a greater variety of cutting agents (e.g. MAN, GLUT and NIA) were included in this study based on the most recent seizures’ reports in the UK and NPS literature [38, 213]. In this study 25 NPS products were purchased from the internet for method validation (Table 6.1).

Table 6. 1: NPS products purchased from the internet.

Product No.	Sample name	Supplier (website)	Supercluster no. ¹	Cluster no. ¹
P1	2-AI	Smokeys Chemsite	5	5.15
P2	2-AI	Smokeys Chemsite	5	5.15
P3	AB-FUBINACA	EU Chemicals	3	3.01
P4	AB-PINACA	EU Chemicals	3	3.02
P5	AB-PINACA	EU Chemicals	3	3.02
P6	Etizolam	rc-lab	8	8.01
P7	Etizolam	benzo_fury.me	8	8.01
P8	Etizolam	chemicalwire	8	8.01
P9	Etizolam	rc-lab	8	8.01
P10	Flubromazepam	Buckledbonzi	5	5.06
P11	Flubromazepam	Buckledbonzi	5	5.06
P12	Flubromazepam	High Store	5	5.06
P13	Flubromazepam	EU Chemicals	5	5.06
P14	Flubromazepam	Chemical Powder Shop	5	5.06
P15	Flubromazepam	RCNet Chemicals	5	5.06
P16	MDAI	EU Chemicals	11*	N/A ²
P17	MDAI	EU Chemicals	11*	N/A ²
P18	MPA	buyanychem.com/rearch-chemicals	9	9.02
P19	MPA	benzo_fury.me	9	9.02
P20	N-Me-2-AI	Chemical Powder Shop	5	5.16
P21	Pink Champagnes	Herbal High	N/A ³	N/A ³
P22	Pink Champagnes	Herbal High	N/A ³	N/A ³
P23	Pink Champagnes	Herbal High	N/A ³	N/A ³
P24	Magic beans	Herbal High	N/A ⁴	N/A ⁴
P25	THJ-018	Buckledbonzi	3	3.04

¹is the supercluster and cluster membership of claimed NPS on product labels (see Chapter 5); ^{*}Test sample, not used in generating the initial clustering model developed by Zloh et al. [204] (see Chapter 5); ²Supercluster 11 is composed of one cluster only; ³'Aminoindane' is quoted on the label. No particular aminoindane analogue is specified in P21 – P23; ⁴Claimed ingredients on the label do not include an NPS.

6.2.2. Methods

In this Section, details of the Raman analysis of the new training samples (i.e. cutting agents and adulterants) and 'query' samples (i.e. NPS internet products) is presented. The composition of internet samples was investigated using three analytical techniques. Qualitative analysis was carried out using: 1) library-based correlations with handheld Raman spectroscopy and 2) confirmatory gas chromatography – mass spectrometry (GC-MS). Quantitative analysis was undertaken using high performance liquid chromatography (HPLC). A full characterisation of NPS products was not performed and is beyond the scope of this work.

6.2.2.1. Raman analysis

6.2.2.1.1. Sample preparation

Approximately 3 - 5 mg of each cutting agent and adulterant reference standard and NPS internet product were weighed, in powder form, tapped and flattened on aluminium plates for benchtop Raman analysis. To ensure homogeneity of NPS internet product, a Vortex Genie 2 (Scientific Industries, Inc.) was employed for two minutes prior to sampling.

6.2.2.1.2. Analysis of cutting agents, adulterants and NPS internet products using benchtop Raman Spectroscopy

Spectra for training and ‘query’ samples i.e. cutting agents, adulterants and NPS internet products were acquired using a benchtop Renishaw inViaTM Raman microscope (see Chapter 5 for full method details). Benchtop Raman spectra for both internet samples P12 and P24 could not be obtained with a laser excitation wavelength (λ_{ex}) of 785 nm due to intense fluorescent background. In this respect, the laser power was reduced to 1.0 and 0.5 % (2.2 and 1.1 mW respectively) in an attempt to reduce fluorescence and acquire spectra but was deemed unsuccessful.

6.2.2.2. Qualitative analysis of NPS internet products using handheld Raman Spectroscopy

Two handheld Raman instruments, ProgenyTM and FirstGuardTM (SciMed Ltd, Rigaku, USA) with laser λ_{ex} of 1064 nm, were employed for the identification of NPS content in internet products using in-built algorithms. Specifications of both instruments are shown in Table 2.3. The NPS internet products were first analysed using the ProgenyTM instrument. Four methods were used to collect Raman spectra based on the substances’ nature: method A (2000 ms exposure time; 490 mW laser power; 10 averages), method B (2000 ms exposure time; 200 mW laser power; 10 averages), method C (2000 ms exposure time; 100 mW laser power; 10 averages); method D (2000 - 5 ms exposure time; 50 mW - 1 mW laser power; 10 averages).

All samples were initially analysed using method A, but methods B and C were employed for samples that were burned or were prone to burning from high laser power (i.e. coloured samples). Method D was developed in an attempt to collect Raman signals from challenging samples, which exhibited intense fluorescent background and/or burned with method C by adopting an iterative approach to reducing both the laser power and exposure time. All methods used an in-built baseline correction function for each measurement. The instrument was calibrated each day immediately before analysis using a benzonitrile reference standard. All samples were analysed directly through glass vials after optimisation of the vial holder attachment with respect to the focal point. The spectra from the cutting agents and adulterants were uploaded into on-board libraries. NPS products were automatically compared to the on-board reference library and reported a percentage correlation using the in-built ‘Rigaku mixtures algorithm (RMA)’. The mean \pm the standard deviation of the highest hit was calculated from the triplicate measurements and reported (Table 6.2).

Table 6. 2: Summary of qualitative and quantitative analysis of NPS internet products using handheld Raman spectroscopy, GC-MS and HPLC

Product No.	Sample name	Formulation	Qualitative Analysis						Quantitative Analysis	
			Handheld Raman Analysis ($\lambda_{ex} = 1064 \text{ nm}$)			GC-MS Analysis			HPLC Analysis	
			Raman ID ¹	Algorithm	Valid hit (%)	MS ID ¹	Base peak (m/z)	RT ²	Average purity (%)	% RSD ³
P1	2-AI	Powder	2-AI	RMA ⁴	97.2 ± 0.4	2-AI	133	6.3	98 ± 5	5
P2	2-AI	Powder	2-AI	RMA	97.0 ± 0.0	2-AI	133	6.3	91 ± 4	4
P3	AB-FUBINACA	Powder	Inconsistent correlations			AB-FUBINACA	109	14.7	77 ± 3	4
P4	AB-PINACA	Powder	AB-PINACA	RMA	86 ± 3	AB-PINACA	215	12.4	49 ± 4	8
P5	AB-PINACA	Powder	AB-PINACA	RMA	91 ± 3	AB-PINACA	215	12.4	54 ± 2	4
P6	Etizolam	Tablets	Data acquisition failed			Etizolam	342	23.1	0.85 ± 0.04	4.71
P7	Etizolam	Tablets	MCC	RMA	86 ± 4	Etizolam	342	23.2	0.66 ± 0.03	4.55
P8	Etizolam	Tablets	MCC	RMA	92 ± 3	Etizolam	342	23.2	0.66 ± 0.03	4.55
P9	Etizolam	Tablets	MCC	RMA	88 ± 2	Etizolam	342	23.1	0.68 ± 0.03	4.41
P10	Flubromazepam	Powder	Flubromazepam	RMA	97.2 ± 0.4	Flubromazepam	305	13.2	95 ± 2	2
P11	Flubromazepam	Pellets	Inconsistent correlations			Flubromazepam	305	13.1	6.1 ± 0.2	3.3
P12	Flubromazepam	Pellets	Data acquisition failed			Flubromazepam	305	13.1	3.9 ± 0.1	2.6
P13	Flubromazepam	Pellets	MCC	RMA	77 ± 4	Flubromazepam	305	13.1	6.6 ± 0.3	4.5
P14	Flubromazepam	Pellets	Inconsistent correlations			Flubromazepam	305	13.1	7.8 ± 0.2	2.6
P15	Flubromazepam	Pellets	MCC	RMA	86 ± 4	Flubromazepam	305	13.1	2.4 ± 0.1	4.2
P16	MDAI	Powder	5-IAI	RMA	85 ± 0	5,6-MDAI	160	8.5	95 ± 2	2
P17	MDAI	Powder	5,6-MDAI	RMA	82 ± 1	5,6-MDAI	160	8.5	98.0 ± 0.1	0.1
P18	MPA	Powder	MPA	RMA	98.6 ± 0.8	MPA	58	8.7	107 ± 1	1
P19	MPA	Powder	MPA	RMA	99 ± 0	MPA 6-APB CAF 5-MeO-DALT	58 44 194 110	8.8 12.0 14.4 17.4	93.8 ± 0.8	0.9
P20	N-Me-2-AI	Powder	N-Me-2-AI	RMA	83 ± 5	N-Me-2-AI	147	6.0	100.1 ± 0.5	0.5
P21	Pink Champagnes	Powder	No correlations			2-AI/CAF	133, 194	9.2, 14.4	15.0 ± 0.1	0.7
P22	Pink Champagnes	Powder	CAF	HQI ⁵	54.2 ± 0.2	2-AI/CAF	133, 194	9.4, 14.4	14.6 ± 0.1	0.7
P23	Pink Champagnes	Powder	Phosphorous	HQI	70 ± 3	2-AI/CAF	133, 194	9.3, 14.5	14.0 ± 0.1	0.7
P24	Magic beans	Powder	Inconsistent correlations			2-AI/CAF	133, 194	9.4, 14.4	15.8 ± 0.1	0.6
P25	THJ-018	Powder	THJ-018	RM	97.4 ± 0.8	THJ-018	127	14.2	100.7 ± 0.6	0.6

¹ID: identification, ²RT: retention time, ³RSD: relative standard deviation, ⁴RMA: Rigaku mixtures algorithm (ProgenyTM), ⁵HQI: hit quality index (FirstGuardTM). Miss V. Guarino, a visiting Erasmus student, has contributed to the GC-MS analysis and has conducted the HPLC analysis under my supervision.

If the correlations between triplicate analyses were inconsistent, it was reported as ‘inconsistent correlations’ instead of a mean value. Using the ProgenyTM instrument, no data were obtained for P21 – P24. Therefore the FirstGuardTM % ‘hit quality index (HQI)’ correlations, documented in Chapter 3 for products P53 – P55 and P51, were reported instead (Tables 3.7 and 6.2) for comparison.

6.2.2.3. Qualitative analysis of NPS internet products using gas chromatography - mass spectrometry

Gas chromatography – Mass spectrometry (GC-MS) analysis was used to confirm the identity of the NPS claimed on the label of the purchased internet products. The method was adapted from [236]. The mass spectra obtained were compared to purchased reference standards and EI spectral libraries: NIST (v. 1.0.2.2), SWGDRUG MS (v. 2.1 (2015)) and Cayman (v. 04292014). GC-MS experiments were conducted under my supervision by a visiting Erasmus student, Miss V. Guarino (Appendix D Table A6.2).

6.2.2.4. Quantitative analysis of NPS internet products using high performance liquid chromatography

High performance liquid chromatography (HPLC) analysis was employed to identify and quantify the NPS claimed on the label of the purchased internet products. Analysis was performed with a reverse phase HPLC equipped with an integrated HPLC system Liquid Chromatograph (USA), fitted with an in-line degasser, auto-injector and a SPD-M20A photo diode array detector (USA). The stationary phase (UltraCore 5 SuperC18, 150 mm × 4.6 mm i.d. 5 µm particle size) was obtained from Advanced Chromatography Technologies Limited (Scotland). The column was fitted with a guard cartridge (ACE) and maintained at 25 °C. Data analysis was carried out using Lab Solution for LC (v. 5.54 SP2) software (USA). The flow rate was 1.0 - 1.5 mL min⁻¹ with an injection volume of 5 - 15 µL (Appendix D Table A6.3). Three replicate injections of each calibration standard were performed.

Prior to HPLC analysis, screening of maximum UV-visible (UV-VIS) absorption of NPS internet samples was investigated by absorption spectroscopy using an Agilent Technologies spectrophotometer (Cary 100 UV-Vis), equipped with cuvette holders, over the range 800 – 200 nm. NPS sample solutions in 50:50 methanol:deionised water (DW) or acetonitrile (ACN) (HPLC grade, Sigma Aldrich, UK) were illuminated in a quartz cuvette by a 350 W tungsten-halogen visible source with quartz window, deuterium arc UV source. The blank reference was 50:50 methanol:DW or ACN. The UV profiles of samples were compared to the spectra

collected from the blank. All spectra were recorded with a bandwidth of 0.2 – 4.0 nm, 0.1 nm steps, motor-driven with a response time of 0.1 s and a scan speed of 600.000 nm min⁻¹ at 20°C. Data were processed using the software Cary WinUV (v. 4.20 (468)). Both HPLC and UV-VIS experiments were conducted under my supervision by a visiting Erasmus student, Miss V. Guarino.

6.2.2.4.1. Preparation of the aqueous buffer, mobile phases and calibration standards for HPLC analysis

Ultrapure DW was used and the pH of the solution was adjusted to pH 2.0 ± 0.1 by dropwise addition of orthophosphoric acid (HPLC electrochemical grade, Sigma Aldrich, UK) (Hanna Instruments Ltd, PH 209 PH Meter, Bedfordshire, UK). Mobile phases were prepared by separately mixing the aqueous buffer and ACN in various proportions, specific to each NPS sample (Appendix D Table A6.3). Prior to use, mobile phases were vacuum filtered through Whatman 0.22 µm pore nylon membranes (Fisher Scientific, UK) and degassed in an ultrasonic bath for 10 min at 25 °C. For the calibration standards, ca. 2 mg of each NPS reference standard were weighed accurately into a 25 mL glass volumetric flask and diluted to volume with mobile phase to give a solution containing 80 µg mL⁻¹ of NPS. This solution was further diluted with mobile phase to give calibration standards containing 57, 39, 26, 17 and 11 µg mL⁻¹ of NPS.

6.2.2.4.2. NPS internet sample preparation for HPLC analysis

Approximately 8 mg of each NPS internet sample were weighed accurately into a 100 mL glass volumetric flask and diluted to volume with 50:50 methanol:DW or ACN to give a solution containing about 80 µg mL⁻¹ of NPS. Solutions were then vortex-mixed (Vortex Genie 2 - Scientific Industries, Inc.) for two min. and ultrasonicated (Ultrasound - Fisherbrand FB15055, Belgium) for five min. Solutions were then re-diluted to give solutions containing a final concentration of ca. 40 – 80 µg mL⁻¹ of NPS (Appendix D Table A6.3).

6.2.2.5. Raman spectral pre-processing

Methods developed in Chapter 5 were applied and investigated in this Chapter as a proof-of-concept (see Chapter 5 Sections 5.2.4. and 5.3.2. for full method details).

6.2.2.5.1. Principal Components Analysis (Unscrambler® X 10.4)

The PCA parameters employed in developing the initial PCA model in Chapter 5 are applied to datasets analysed in this Chapter (see Chapter 5 Section 5.2.5. for full method details).

A schematic outlining the Raman/ chemometrics experimental protocol is described in Figure 6.1.

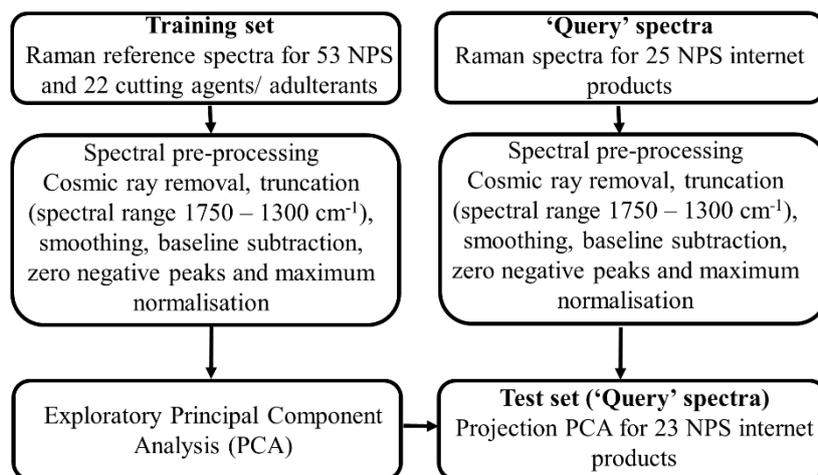


Figure 6. 1: Raman/ chemometrics experimental protocol. Schematic outlining the details of the training and ‘query’ sets, steps of combined spectral pre-processing and chemometrics.

6.3. Results and Discussion

6.3.1. Selection of cutting agents/ adulterants and NPS internet products

Fourteen cutting agents and eight adulterants were selected such that the majority of these substances were commonly reported in seized NPS street samples or included as a result of being detected during direct in-house internet sample analysis evaluated in this Chapter. The ‘new’ training sample subset (i.e. cutting agents and adulterants) includes pharmaceutically active substances and excipients to ensure a wide chemical diversity in the training set and to evaluate whether they hinder the identification of NPS through PCA analysis. A review of the literature, forensic and government reports (2009 - 2015) was conducted on Google Scholar and Scopus to identify cutting agents/ adulterants incorporated in NPS mixtures [38]. Words and phrases used in the search included ‘analysis of new/ novel psychoactive substances’, ‘new psychoactive substance mixtures’, ‘new psychoactive substance cutting agents’ and ‘new psychoactive substance adulterants’.

Results of the search revealed that seized NPS products, purchased over the internet for research purposes or documented in government reports were found to contain BEN, CAF, CRE, GLU, LID, PHE, PRO, SUC *inter alia* cutting agents and adulterants [73, 79, 121, 132, 217, 247-253]. Forensic reports described common cutting agents identified in UK seizures to include bulking agents such as CRE, GLU, GLUT, LAC, MAN, SUC and TAU. Other cutting agents and adulterants such as LAC, L-TYR, MCC and NIA were selected in this study because

they were identified through GC-MS and handheld Raman analysis of NPS internet products (Chapter 3 and Table 6.2). In 2014, adulterants over which law enforcement have gained power to seize if they suspected that they were used for cutting NPS include BEN, LID and PHE [218]. Both CRE and GLUT have been shown to be the most popular cutting agents found in NPS seizures in the UK from 2013 – 2016 [38, 208, 209, 213, 285].

In general, in a similar manner to illicit drugs [286], NPS products may be cut with bulking agents and fillers which may be purposely added. Adulterants such as BEN, CAF, LID and PRO are co-added to NPS because they may mimic, enhance and potentiate the effects of controlled drugs of abuse [247]. LID is a local anaesthetic and may be incorporated in NPS mixtures for its numbing effect, whereas BEN and PRO have been used as adulterants for their cocaine-like effect [247]. Both GLUT and TAU are known food supplements, whereas both GLU and SUC give a sweet taste. Excipients which may be co-added to NPS to counteract their side-effects include CaCO₃, DIL, PAR, PHE and THEO [247, 287]. For example, CaCO₃ may counteract stomach acidity [287], DIL may counteract tachycardia and THEO is used for its bronchodilating activity, hence facilitating the smoking of NPS. Benzodiazepines were found to be co-ingested with the stimulant MDPV to counteract its excitatory effect [59]. Furthermore, excipients may have common chemical substructures (e.g. NIA and cathinone (Appendix D Table A6.1 and Figure 5.17)) to target drugs in order to add to the complexity of the mixture and hinder identification (e.g. BEN, a very strong Raman scatterer, may mask NPS Raman signals) [12, 95, 135, 192].

6.3.2. Overview of classifications of new datasets using PCA

In this Section, an overview of three datasets is performed using exploratory PCA: Dataset 1: pre-processed Raman spectra of 22 cutting agents/ adulterants, Dataset 2: 25 NPS internet products, Dataset 3: pre-processed Raman spectra of ‘53 representative’ NPS and 22 cutting agents/ adulterants. Dataset 3 is the ‘new’ PCA model ‘NPS and non-NPS/benchtop’ model against which pre-processed Raman spectra of 23 NPS ‘query’ samples are projected using PCA projection (Unscrambler® X 10.4). For the first two datasets, grouping, overlap, correlations, delineation or outliers between scores are evaluated using PCA scores plots, common chemical substructures, shared Raman bands across the range 1750 – 1300 cm⁻¹ and/ or association with variables with high loadings and PCA scoresplot ellipse at 95 % CL. In addition to this, the third dataset is evaluated using PCA projection scores plots.

6.3.2.1. Overview of the classification of cutting agents/ adulterants (Dataset 1)

Prior to undertaking a PCA analysis, initial inspection of Raman spectra was conducted (Chapter 5 and Figure 6.2). Pearson's correlation coefficient (r^2) between replicate raw spectra of each cutting agent/ adulterant was $\geq 90\%$, except for GLU, MCC and LID, where r^2 was 80, 83 and 75 % respectively. For MCC, raw spectra exhibited a high fluorescent background with poor Raman signal relative to noise. Unlike raw data, pre-processed MCC spectra resulted in noisy spectra with flat baselines. Harkai and Putz emphasised that intense fluorescence may hamper the identification of NPS even after the pre-processing of spectra [161]. For both GLU and LID, initial raw spectra exhibited good S/N and relatively flat baselines. However, the instrument's artefacts resulting in variable relative intensities between replicate spectra may have resulted in reduced r^2 values. The line plots were also inspected to ensure the suitability of the spectral range $1750 - 1300\text{ cm}^{-1}$ used in the initial model (Chapter 5).

By examining the Raman spectra, Raman bands for most cutting agents were spread along the spectral ranges $2990 - 2890$ and $1500 - 250\text{ cm}^{-1}$ (Figure 6.2a), whereas Raman bands for most adulterants were spread along the spectral range $1700 - 250\text{ cm}^{-1}$ (Figure 6.2b). By including the Raman bands over the spectral range $2990 - 2890\text{ cm}^{-1}$, insignificant changes to the classification of cutting agents and adulterants occurred using exploratory PCA, specifically to the PCA scores and explained variance plots. Hence the spectral range $1750 - 1300\text{ cm}^{-1}$ was used for consistency.

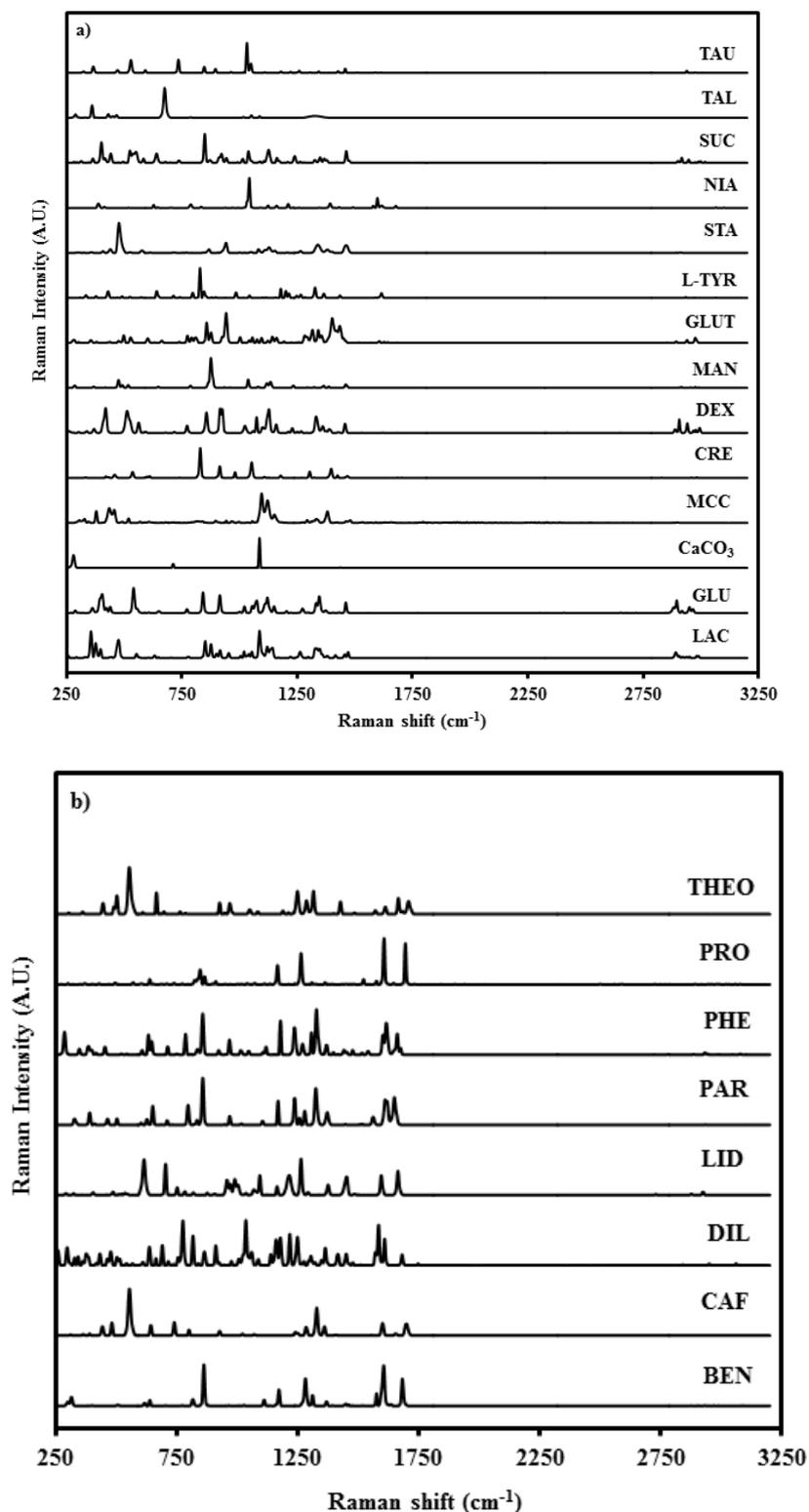


Figure 6. 2: Pre-processed average Raman spectra of a) 14 cutting agents and b) eight adulterants over the spectral range 3200 – 250 cm⁻¹.

To obtain an overview of dataset 1, an exploratory PCA was performed using 10 pre-processed spectra for each substance over the spectral range 1750 – 1300 cm⁻¹. The dataset consisted of 220 rows and 522 variables. Samples were categorised based on their chemical classes as amino

acids (CRE, GLUT, L-TYR and TAU), anaesthetics (BEN, LID and PRO), carbohydrates (MCC and STA), food supplements (NIA), inorganic substances (CaCO₃), minerals (TAL), pharmaceuticals (DIL, PAR, PHE and THEO), stimulants (CAF) and sugars (LAC, GLU, DEX, MAN and SUC). The proposed categorisation in this work is used as an example of numerous possible categorisations which could be applied.

The first three PCs explained approximately 59 % of the total explained variance, (24/19/16 % for PC1/PC2/PC3) (Figure 6.3a). By examining the 2D-scores plots, PC1/PC2 showed that the explained variance is mainly due to PAR/PHE, then to a lesser extent due to BEN/PRO, all of which are very strong Raman scatterers compared to remaining cutting agents/ adulterants (Figure 6.3b). This was confirmed through the scores and loadings bi-plot (plot not shown), where the variables explaining greater than 50 % of the total variance was associated with functional groups present in these substances. The line loadings plot also confirmed these findings as the peaks with high loadings were associated with these substances. Examples of these variables include 1685, 1608, 1580 and 1314 cm⁻¹ for BEN; 1697 and 1609 cm⁻¹ for PRO; 1652, 1623, 1614, 1374 and 1328 cm⁻¹ for PAR; and 1663, 1620, 1605, 1372, 1330 and 1310 cm⁻¹ for PHE. Bands in these regions are typically due to carbonyl-containing compounds such as amide groups (below 1700 cm⁻¹), amine groups (1660 – 1500 cm⁻¹), quadrant stretches (1620 – 1420 cm⁻¹) and methyl deformation vibrations (1500 – 1250 cm⁻¹).

The 2D-scores plots, PC1/PC2 showed that carbohydrates and sugars grouped together in the lower left quadrant, which was expected as carbohydrates contain one or more units of sugars. Replicates of the analgesics PAR and PHE were scattered in the lower right quadrant, despite the fact that the r^2 between replicate spectra for both PAR and PHE was ≥ 90 %. However, they were in close proximity to each other, possibly because they share a common substructure (N-(4-(λ^1 -oxidanyl)phenyl)acetamide) (Figure 6.3d) and Raman active functional groups (see above) over the region 1750 – 1300 cm⁻¹. This was confirmed *via* line plots (Figure 6.3e) showing that PAR and PHE share common doublets: one with different relative intensities at ca. 1623 and 1614 cm⁻¹ for PAR and at ca. 1620 and 1605 cm⁻¹ for PHE; and another at ca. 1374 and 1328 cm⁻¹ for PAR and at ca. 1372 and 1330 cm⁻¹ for PHE. Furthermore, both BEN and PRO clustered close to DIL. DIL shares a peak with BEN at ca. 1685 cm⁻¹ and with both BEN and PRO at ca. 1610 cm⁻¹, possibly due to carbonyl stretching and amine deformation vibrations respectively.

In contrast to pharmaceuticals and anaesthetics, GLUT is an alkyl chain conjugated with carboxylic acids on both ends (Appendix D Table A6.1). The variance along PC3 (Figure 6.3c), was mainly explained by GLUT, where two Raman bands at ca. 1436 and 1404 cm^{-1} (Figure 6.2a) were associated with two peaks with medium loading along PC3, which may be attributed to CH_3 in-phase deformation. The band at higher frequency is likely to be adjacent to the most electronegative atom e.g. oxygen in this molecule.

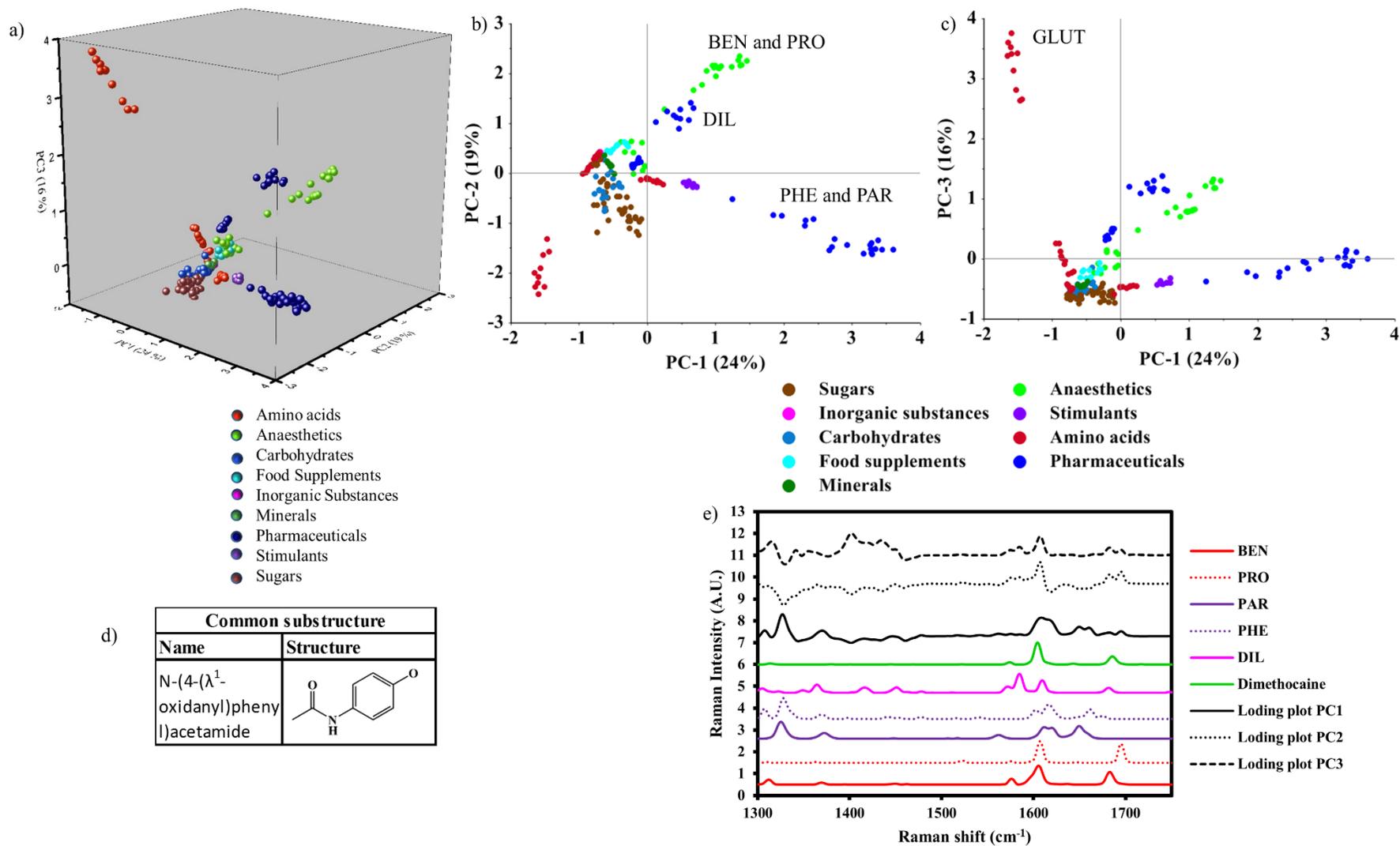
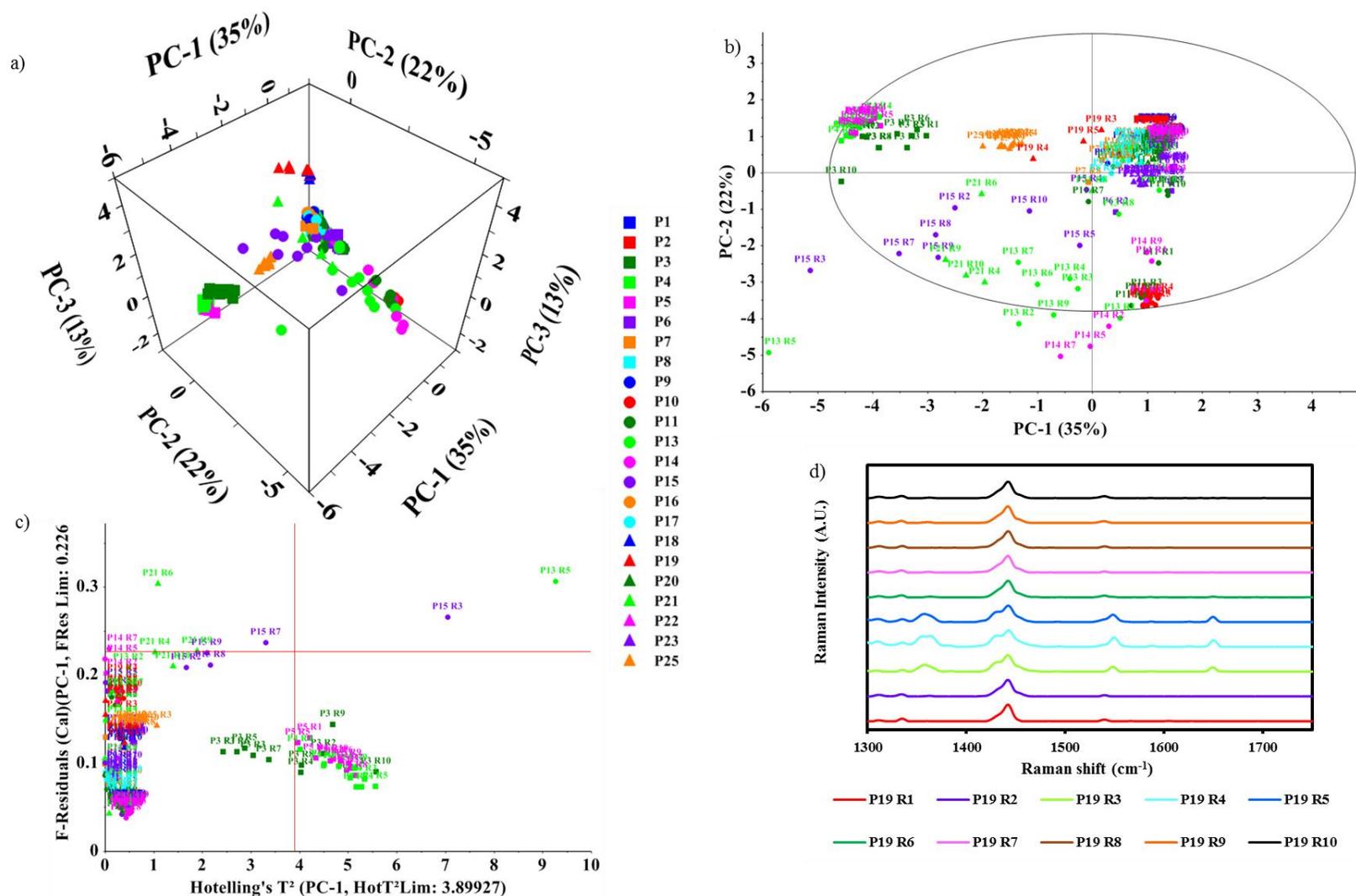


Figure 6. 3: a) Three-dimensional (3D)-scores plot of 22 cutting agent/ adulterant reference spectra; b) Two-dimensional (2D)-scores plots for PC1/PC2; c) 2D-scores plots for PC1/PC3; d) Common substructure between PAR and PHE; d) Pre-processed Raman spectra of BEN, PRO, PAR, PHE, DIL, dimethocaine and the line loading plots for PC1/PC2/PC3.

6.3.2.2. Overview of the classification of test 23 NPS internet products (Dataset 2)

Twenty-five NPS products were purchased from 11 websites. They consisted of aminoindanes (i.e. P1, 2, 16, 17 and 20), arylalkylamines (i.e. P18 and P19), benzodiazepines (i.e. P6 – P18), branded products (i.e. P21 – P24) and synthetic cannabinoids (i.e. P3 – P5 and P25). Claimed NPS in these products were members of previously described superclusters (Chapter 5 and Table 6.2). The r^2 values between raw replicate spectra of each product was ≥ 90 %. This showed that r^2 values can be relatively high between replicate spectra equally exhibiting high fluorescent backgrounds, as was the case with numerous internet samples. To obtain an overview of the classification of 25 NPS internet products, an exploratory PCA was performed using pre-processed 10 replicate spectra for each product over the spectral range 1750 – 1300 cm^{-1} . The data matrix consisted of 230 rows and 522 variables. Both P12 and P24 were not included in the PCA analysis as no Raman spectra were obtained due to high fluorescent backgrounds with a 785 nm λ_{ex} .

The calibrated cumulative explained variance of the model was 70 % (35/22/13 % for PC1/PC2/PC3) (Figure 6.4a). The 2D-scores plot (Figure 6.4b) showed that a few spectra (i.e. P13 R2 and R5; P14 R2, R5 and R7; and P15 R3) fall outside the ellipse at 95 % CL. These spectra were investigated further and have been shown to be for flubromazepam pellets with very poor S/N. Quantitative analysis using HPLC has shown that the purity of these samples was 2.4 ± 0.1 to 7.8 ± 0.2 %. The influence plots for PC1/PC2 (Figure 6.4c) showed that only P13 R5 and P15 R3 are outside of the model with high F-residuals Hotelling T^2 , confirming the impact of poor S/N in these spectra. The PCA of 25 NPS internet products was re-calculated after removing these two spectra. The generated PCA showed insignificant changes to sample classification on PCA scores and total explained variance plots by the first three PCs (plots not shown). Therefore these spectra were not rejected from the analysis but were re-visited in the PCA projection analysis.



¹One legend is used for Figure 6.4a, 6.4b and 6.4c

Figure 6. 4: a) 3D-scores plot of 23 NPS internet samples; b) 2D-scores plots for PC1/PC2 with an ellipse at 95 % CL; c) 2D-influence plot for PC1/PC2; d) pre-processed replicate Raman spectra for P19¹.

Claimed aminoindanes (i.e. P1, 2, 16, 17, 20-23) were grouped together in the upper right quadrant of the 2D-scores plot for PC1/PC2, suggesting that the products were relatively pure. P1 and P2 were two 2-AI powders purchased from the same website and both have been shown to contain 2-AI using GC-MS [226], HPLC and handheld Raman. P1 and P2 were grouped together with P20 (N-Me-2-AI, a class analogue of 2-AI). The latter was purchased from a different website and has been shown to contain N-Me-2-AI using GC-MS [288], HPLC and handheld Raman. P16 and P17, labelled as MDAI and purchased from the same website, grouped together. Both have also been shown to contain 5,6-MDAI using both GC-MS and HPLC, but, as for P16, handheld Raman analysis results were inconsistent but correlated mostly to another aminoindane i.e. 5-IAI (% RMA = 85 ± 0) (Table 6.2).

Claimed arlyalkylamines (P18 and P19) grouped together in the upper right quadrant of the 2D-scores plot for PC1/PC2, except replicate spectra P19 R3 - R5. P18 and P19 were purchased from different websites but have both been shown to contain MPA (Table 6.2). GC-MS analysis for P19 showed that this sample was also adulterated with 6-APB [240], CAF [224, 225] and 5-MeO-DALT [219]. In fact, the inspection of replicate spectra for P19 shows that spectra R3 - R5 exhibited Raman bands at ca. 1652, 1552 and 1360 cm^{-1} that are absent from remaining replicate spectra (Table 6.2 and Figure 6.4d). This demonstrated the importance of taking Raman measurements at multiple powder areas in complex NPS street mixtures because it may assist in identifying multiple constituents.

Claimed benzodiazepines (etizolam P6 – P9 and flubromazepam P10, 11, 13 – 15) were obtained as coloured tablets and pellets respectively except for P10, which was a white powder. Etizolam tablets were purchased from three websites, whereas flubromazepam products were purchased from five different websites. For P12, fluorescent background was more intense than the Raman peaks. Hence no signal was obtained for P12 and, therefore was excluded from the PCA analysis. Replicate spectra of P6 – P10 grouped together, whereas replicate spectra of P11 – P15 were scattered on the PCA 2D-scores plot (PC1/PC2), possibly due to poor S/N in pre-processed spectra but also showing that they may potentially have different compositions of various components (Figure 6.4a and 6.4b). The etizolam products (P6 – P9) grouped in the upper right quadrant of the 2D-scores plot (PC1/PC2). The GC-MS analysis of P6 – P9 showed that they all contained etizolam [289] but the HPLC analysis showed very low etizolam content in these products (0.66 ± 0.03 to 0.85 ± 0.04 %). Library correlations using handheld Raman ($\lambda_{\text{ex}} = 1064 \text{ nm}$) identified MCC as the top hit in P7 – P9 (% RMA = 86 ± 4 to 92 ± 3 %) (Table

6.2). Handheld Raman ($\lambda_{\text{ex}} = 1064 \text{ nm}$) data acquisition for P6 failed due to high fluorescent background potentially resulting from dark coloured tablet coating (dark turquoise). Replicate spectra for P10 grouped in the lower right quadrant of the PCA 2D-scores plot (PC1/PC2) and were confirmed to contain flubromazepam using GC-MS [290], HPLC and handheld Raman. Flubromazepam-containing pellets (P11, P13 – P15) claimed to contain 8 mg/ pellet. GC-MS analysis confirmed that they all contained flubromazepam, with HPLC purity ranging from 2.4 ± 0.1 to 7.8 ± 0.2 %. Raman analysis for P11 and P14 gave inconsistent results. However, handheld Raman analysis ($\lambda_{\text{ex}} = 1064 \text{ nm}$) for P13 and P15 correlated to MCC (% RMA = 77 ± 4 to 86 ± 4) (Table 6.2).

Branded products (P21 – P24) were obtained as capsules-containing maroon powders which were emptied in glass vials for analysis. The benchtop Raman analysis ($\lambda_{\text{ex}} = 785 \text{ nm}$) of these powders resulted in spectra with a high fluorescent background. For P24, fluorescent background was more intense than the Raman peaks and hence no signal was obtained. Consequently, P24 replicates were excluded from the PCA analysis. Replicate spectra for P21 - P23 were scattered on the PCA 2D-scores plot, possibly due to poor S/N in pre-processed spectra but also due to their relative levels of impurity. GC-MS analysis confirmed the presence of 2-AI and CAF in these samples. Using HPLC, the purity of 2-AI in these products ranged from 14.0 ± 0.1 to 15.0 ± 0.1 %. Handheld Raman analysis ($\lambda_{\text{ex}} = 1064 \text{ nm}$) resulted in no correlation for P21, correlated to CAF for P22 and to phosphorous for P23 (see Chapter 3 and Table 6.2).

Claimed synthetic cannabinoids (P3 - P5 and P25) grouped in the upper left quadrant of the 2D-scores plot for PC1/PC2. P3 (AB-FUBINACA) replicate spectra were confirmed to contain AB-FUBINACA using both GC-MS [39] and HPLC. Raman analysis gave inconsistent correlations to other analogues of synthetic cannabinoids 5F-APINACA and AB-PINACA. Both P4 and P5 (AB-PINACA) replicate spectra have been confirmed to contain AB-PINACA using GC-MS [291], HPLC and handheld Raman. P25 (THJ-018) replicate spectra have been confirmed to contain THJ-0.18 using GC-MS [37], HPLC and handheld Raman.

6.3.2.3. Overview of the classification of 53 NPS reference standards and 22 cutting agents/ adulterants (Dataset 3)

To obtain an overview of the classification of NPS (i.e. 53 NPS reference standards) versus non-NPS substances (i.e. 22 cutting agents/ adulterants), an exploratory PCA was performed using pre-processed 10 replicate Raman spectra for each substance. The data matrix consisted

of 750 rows and 522 variables. The calibrated cumulative explained variance of 36 % for the first three PCs was 16/11/9 % for PC1/PC2/PC3 (Figure 6.5a). Classification of the 53 NPS spectra on the 2D-scores plot according to superclusters and EMCDDA/EDND classifications followed similar trends as observed in Chapter 5 Section 5.3.3.1. (Figures 5.10, 6.5b and 6.5c). Possible overlap/ correlations between 13 NPS previously described superclusters representing the NPS chemical space of 478 NPS (Chapter 5) and cutting agents/ adulterants' proposed categories were evaluated along computed PC1/PC2/PC3. To achieve this, average Raman reference spectra (i.e. average of 10 spectra/ substance) over the range 1750 – 1300 cm^{-1} (Figures 6.6a-6.6e), common substructures and r^2 values (Tables not shown) between overlapping spectra were examined.

On the PCA scores plot (Figure 6.5a), overlap between the scores representing the Raman spectra for NPS and active adulterants such as anaesthetics, pharmaceuticals, stimulants and food supplements is expected because they share common substructures and functional groups such as aromatic rings and carbonyl groups. In addition, some adulterants may belong to the same class as library NPS such as benzocaine (adulterant) and dimethocaine (training NPS sample), both of which are anaesthetics. Overlap between NPS and cutting agents is expected to a lesser extent because of greater structural dissimilarity with NPS. However, unlike internet samples, where Raman signals from cutting agents may mask that of NPS in mixtures [12] (Chapter 3), in this PCA model, the positions of pure cutting agents, on computed PCs, are compared to pure NPS. Therefore this masking effect is not relevant, unless it is intrinsic to specific substances. An in-depth analysis was conducted to compare each NPS supercluster (Chapter 5) to each cutting agent/ adulterant class to investigate possible overlap (Table 6.3). Only the significant trends identified are discussed in this Chapter.

From supercluster 2, both N-Me-2C-B (cluster 2.06) and STP (cluster 2.08) overlapped with a few LID spectra along PC1/PC2. This is possibly because of common Raman bands for all three substances at ca. 1455 cm^{-1} (weak to strong), associated with weak positive loadings along PC2 (Figure 6.6a). Another Raman band at ca. 1378 cm^{-1} (weak to medium) was observed for STP and LID and associated with weak positive loadings (PC1).

Moreover, 25H-NBOMe (cluster 2.05) overlapped with the strong Raman scatterer BEN along PC1/PC2. Common Raman bands were observed at ca. 1607 and 1605 cm^{-1} (strong) for 25H-NBOMe and BEN respectively (Figure 6.6b). This band was associated with a high negative loading along PC1. Pearson's correlation coefficient between spectra for 25H-NBOMe and

BEN over 1750 – 1300 cm^{-1} was 69 %, which indicates possible overlap with an NPS containing-BEN would be unsurprising. This is because this band is likely to be present in other NPSs as well.

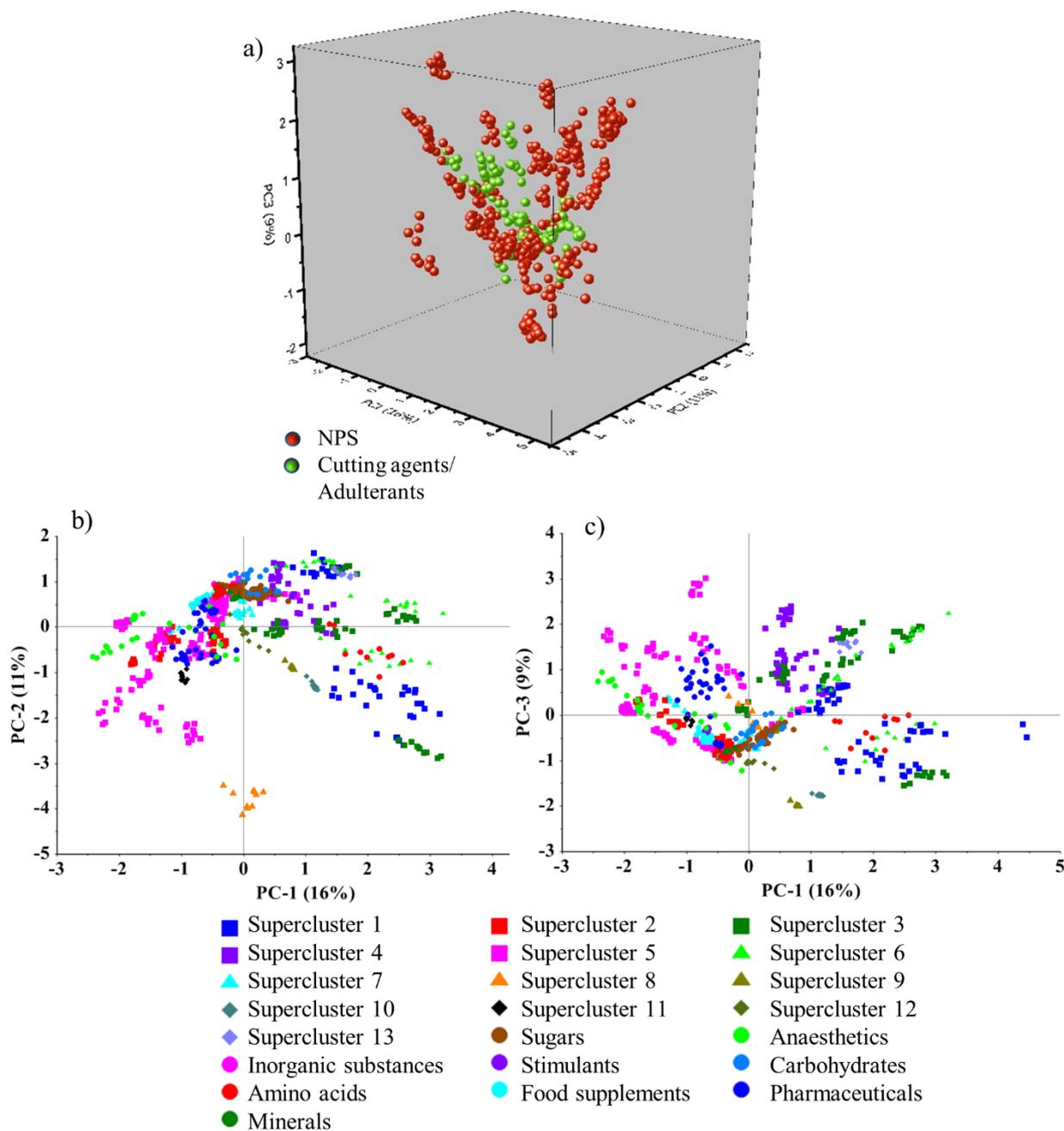


Figure 6. 5: The NPS and non-NPS model. a) 3D-scores plot of 53 NPS and 22 cutting agent/adulterant reference spectra; b) 2D-scores plots for 13 NPS superclusters and cutting agent/adulterant classes (PC1/PC2); c) 2D-scores plots for 13 NPS superclusters and cutting agent/adulterant classes (PC1/PC3).

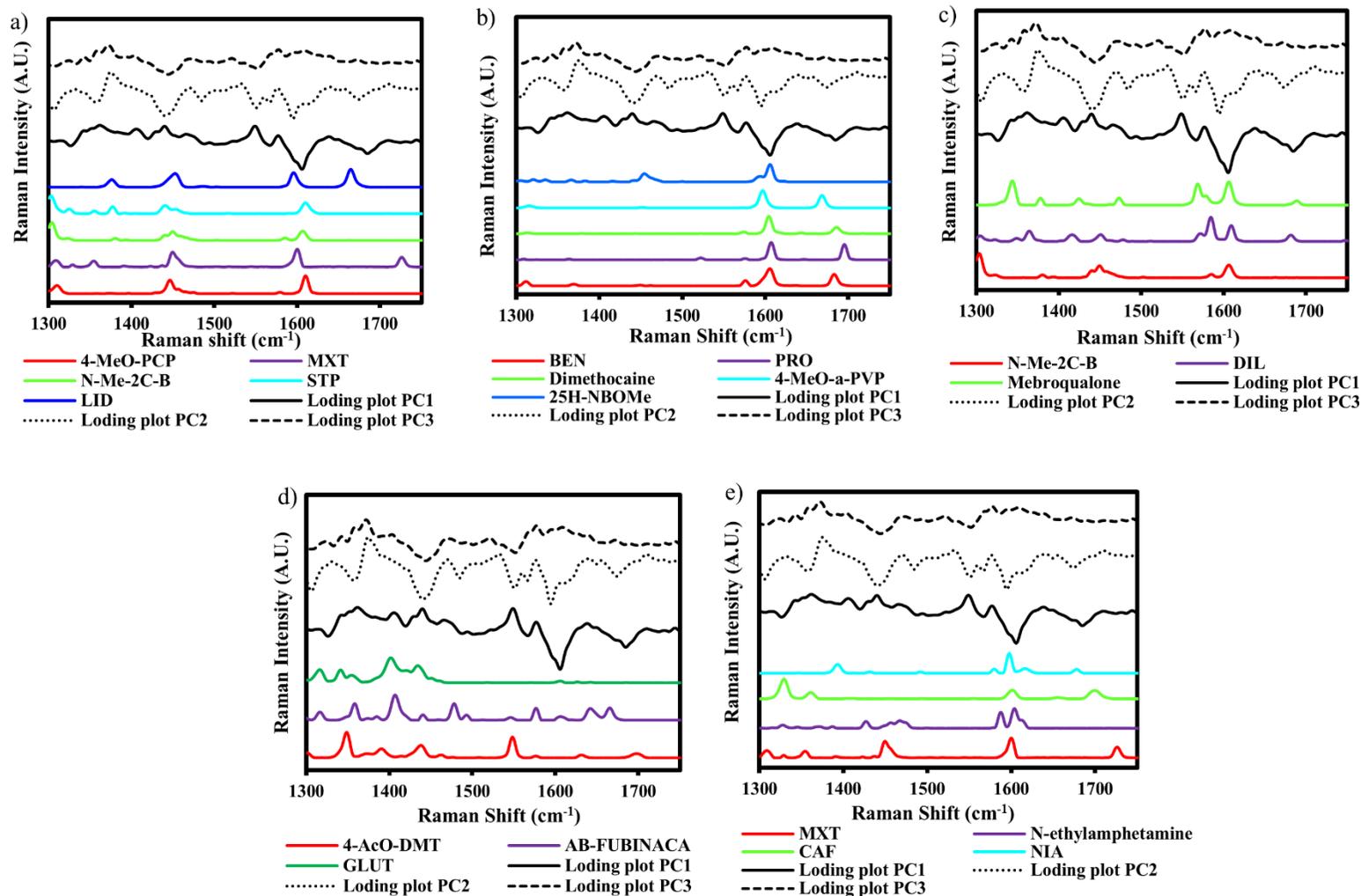


Figure 6. 6: Pre-processed Raman spectra for: a) 4-MeO-PCP, MXT, N-Me-2C-B and STP; b) BEN, PRO, dimethocaine, 4-MeO- α -PVP and 25H-NBOMe; c) N-Me-2C-B, DIL and mebroqualone); d) 4-AcO-DMT, AB-FUBINACA and GLUT; e) MXT, N-ethylamphetamine, CAF and NIA and line loading plots (PC1/PC2/PC3).

Table 6. 3: Analysis of potential overlap between reference spectra for 53 NPS and 22 cutting agent/ adulterants along the PCA scores plot

Supercluster No.	Overlap along PCA scores plot		Principal components (PCs)	Approximate wavenumber of common Raman bands (cm ⁻¹)	Loadings (Retrieved from PCA loading plots)
	NPS	Cutting agent/adulterant(s)			
2	N-Me-2C-B and STP	LID	PC1/PC2	1455 (weak to strong)	Weak positive loadings along PC2
	STP	LID	PC1/PC2	1378 (weak to medium)	Weak positive loadings along PC1
	25H-NBOMe	BEN	PC1/PC2	1607 (strong) for 25H-NBOMe and 1605 (strong) for BEN	High negative loadings along PC1
	4-MeO- α -PVP	BEN and PRO	PC1/PC3	1599 (strong) for 4-MeO- α -PVP, 1608 (strong) for BEN and 1609 (strong) for PRO	N/A
	N-Me-2C-B	DIL	PC1/PC2	1612 and 1586 (medium to strong) and 1456 (weak) for N-Me-2C-B; 1610 and 1590 (strong to weak) and 1453 (weak) for DIL	N/A
3	AB-FUBINACA	GLUT	PC1/PC2	1445 (weak) and 1407 (strong) for AB-FUBINACA; 1439 (weak) and 1406 (strong) for GLUT	N/A
5	Mebroqualone	DIL	PC1/PC2	1690 (weak to medium)	Weak positive loadings along PC1 Medium positive loadings along PC2/PC3
				1608 (strong to medium)	High positive loadings along PC1/PC2/PC3
				1582 and 1570 for mebroqualone and 1585 and 1575 for DIL	N/A
	MXT	LID	PC1/PC2	1598 (strong)	N/A
	4-MeO-PCP and MXT	LID	PC1/PC2	1455 for LID, 1448 for 4-MeO-PCP and 1452 for MXT (strong)	N/A
	MXT	CAF	PC1/PC2	1605 (strong)	High negative loadings along PC1
	Dimethocaine	BEN	PC1/PC3	1689 (strong)	Weak positive loadings along PC1
	Dimethocaine	BEN and PRO	PC1/PC2/PC3	1607 (strong)	High positive loadings along PC1/PC2/PC3
6	4-AcO-DMT	GLUT	PC1/PC2	1443 and 1439 (weak)	N/A
7	N-ethylamphetamine	CAF and NIA	PC1/PC2	1605 (strong)	High negative loadings along PC1
				1589 (strong) for N-ethylamphetamine; 1583 (weak) for NIA	N/A

The NPS 4-MeO- α -PVP (cluster 2.01) overlapped with both BEN and PRO along PC1/PC3, possibly due to common Raman bands at ca. 1599, 1608 and 1609 cm^{-1} (strong) for 4-MeO- α -PVP, BEN and PRO respectively (Figure 6.6b). As indicated earlier, the full structure of BEN is a substructure of PRO. However, the r^2 value of the spectra between 4-MeO- α -PVP and both BEN and PRO surprisingly showed correlation was greater for BEN (38 %) than for PRO (10 %) over 1750 – 1300 cm^{-1} , thus showing the limitation of the selected spectral range in this model. Relatively high r^2 values have previously been shown to indicate overlap on the scores plot between substances with strong Raman activity exhibiting Raman bands associated with high loadings across the range 1750 – 1300 cm^{-1} . Furthermore, N-Me-2C-B overlapped with DIL ($r^2 = 33$ %) along PC1/PC2, possibly due to shared Raman bands at ca. 1612 cm^{-1} (medium), 1586 cm^{-1} (strong) and 1456 cm^{-1} (weak) for N-Me-2C-B and at ca. 1610 cm^{-1} (strong), 1590 cm^{-1} (weak) and 1453 cm^{-1} (weak) for DIL (Figure 6.6c). Different relative intensities was not associated with high loadings, hence have not contributed to the explained variance (Table 6.3).

From supercluster 3, AB-FUBINACA (cluster 3.01) overlapped with GLUT along PC1/PC2, possibly due to common Raman bands at ca. 1445 and 1439 cm^{-1} (weak) for AB-FUBINACA and GLUT respectively (Figure 6.6d). These bands are potentially attributed to CH_2 vibrations near unsaturated groups. Two equal intensities Raman bands (strong) at ca. 1407 and 1406 cm^{-1} were also observed for AB-FUBINACA and GLUT respectively. These bands are potentially attributed to CH_2 vibrations but also to C-F stretches for AB-FUBINACA (Table 6.3).

From supercluster 5, an overlap was observed between mebroqualone (cluster 5.08) and DIL along PC1/PC2 ($r^2 = 35$ %). The line plots showed shared Raman bands at ca. 1690 cm^{-1} (weak to medium) potentially attributed to carbonyl stretching (Figure 6.6c). This band was associated with weak positive loadings along PC1 and medium positive loadings along PC2/PC3 and, hence contributed to explained variance and indicated delineation from other training samples along PC1/PC2/PC3. Another common Raman band between the two substances was observed at ca. 1608 cm^{-1} (strong to medium), potentially attributed to quadrant stretches. This band was associated with high positive loadings along PC1/PC2/PC3. A doublet observed at ca. 1582 and 1570 cm^{-1} for mebroqualone and at ca. 1585 and 1575 cm^{-1} for DIL, was potentially attributed to quadrant stretches. The latter doublets exhibited different relative intensities, without any significant contribution to the explained variance (Table 6.3).

Additionally, an overlap was observed between both 4-MeO-PCP (cluster 5.13) and MXT (cluster 5.01) and LID along PC1/PC2 ($r^2 = 23$ and 50% respectively), possibly due to Raman bands at ca. 1598 cm^{-1} (strong) for both LID and MXT and at ca. 1455 , 1448 and 1452 cm^{-1} (strong) for LID, 4-MeO-PCP and MXT respectively (Figure 6.6a). The latter bands were not associated with important variables as identified by the loading plots, and, hence have not significantly contributed to the explained variance along PC1/PC2/PC3. An overlap between MXT and CAF was observed along PC1/PC2 ($r^2 = 16\%$), possibly due to a common Raman band at ca. 1605 cm^{-1} (strong) (Figure 6.6e). This band was associated with high negative loadings along PC1 (Table 6.3).

An overlap was observed between dimethocaine (supercluster 5) and both BEN and PRO along PC1/PC3 ($r^2 = 91$ and 65% respectively) (Figure 6.6b). High r^2 values are possibly due to common substructures over $1750 - 1300\text{ cm}^{-1}$ and indicates likely overlap with NPS containing-BEN and PRO (two strong Raman scatterers). The line plots for the three anaesthetics showed Raman bands at ca. 1689 cm^{-1} (strong) for both dimethocaine and BEN potentially attributed to carbonyl stretching. This band was associated with a weak positive loading along PC1. The line plots also showed a Raman band at ca. 1607 cm^{-1} (strong) for the three substances potentially attributed to quadrant stretching. This band was associated with high positive loading along PC1/PC2/PC3 (Table 6.3).

From supercluster 6, an overlap was observed between 4-AcO-DMT (cluster 6.03) and GLUT ($r^2 = 32\%$) along PC1/PC2, possibly due to common Raman bands at ca. 1443 and 1439 cm^{-1} (weak) for 4-AcO-DMT and GLUT respectively (Figure 6.6d). These bands are potentially attributed to CH_2 vibrations near unsaturated groups. Finally, from supercluster 7, an overlap was observed between N-ethylamphetamine (cluster 7.04) and both CAF and NIA along PC1/PC2 ($r^2 = 20$ and 26% respectively). The line plots (Figure 6.6e) showed Raman bands at ca. 1605 cm^{-1} (strong) for all three substances. These bands are potentially attributed to quadrant stretches and were associated with a high negative loading along PC1. Other bands that may also be attributed to quadrant stretches were observed at ca. 1589 cm^{-1} (strong) and 1583 cm^{-1} (weak) in N-ethylamphetamine and NIA spectra respectively (Table 6.3).

In conclusion, based on the interpretation above, an overlap between NPS and non-NPS reference spectra was observed on the PCA scores plot. The substances, which overlapped with NPS included the anaesthetics (i.e. BEN, LID and PRO), pharmaceuticals (i.e. DIL), stimulants (i.e. CAF), food supplements (i.e. NIA) and amino acids (i.e. GLUT). The correlation

coefficient r^2 values between overlapping spectra was $\geq 20\%$ over the spectral region $1750 - 1300\text{ cm}^{-1}$. The NPS, which overlapped with cutting agents/ adulterants were mainly members of superclusters 2, 3, 5, 6 and 7. Overlapping substances shared common strong Raman active bands potentially attributed to quadrant and semi-circle stretches as well as CH_2 and CH_3 deformation vibrations over the designated spectral range. A greater overlap was observed between training NPS and adulterants, which are members of the same chemical class. This model may be employed as a tool to suggest further investigation of seized samples containing the cutting agents/ adulterants that have been shown to overlap with NPS.

6.3.3. Prediction of test NPS internet samples using PCA projection (Unscrambler X 10.4)

Following the generation of the ‘NPS and non-NPS/benchtop’ PCA model from the training samples (i.e. ‘53 representative’ NPS, 14 cutting agent and eight adulterant reference standards x 10 measurements (dataset 3)), prediction of the chemical scaffolds of claimed NPS in ‘query’ samples was performed by projecting developed PCA models to latent space over the spectral range $1750 - 1300\text{ cm}^{-1}$.

The combined analysis of 25 NPS internet products (Table 6.2) showed that they contained a number of claimed and unclaimed NPSs (2-AI, 5-MeO-DALT, 5,6-MDAI, 6-APB, AB-FUBINACA, AB-PINACA, etizolam, flubromazepam, MPA, N-Me-2AI and THJ-018). Analysis also showed that a few samples were adulterated with CAF (i.e. P21 – P24) and a few may contain cutting agents such as MCC (i.e. P7 - 9, 13 and 15). Prior knowledge of NPS content in the products through multi-chemical analysis assisted interpretation of results. Confirmation of similarity/ overlap/ closeness/ correlation between ‘query’ (i.e. 23 NPS internet products) and training substances is evaluated using numerous methods including PCA projection plots (Figure 6.7a), correlation coefficients between overlapping spectra over the spectral range $1750 - 1300\text{ cm}^{-1}$ (i.e. r^2 values), common substructures between training and test samples (Chapter 5 and Appendix D Table A6.1), Raman spectra and line loading plots.

In this dataset, the aim of the analysis was to evaluate the ability to identify the claimed NPS contained in NPS internet products particularly those that are highly adulterated. The PCA projection 2D-scores plot for PC1/PC2 was examined (Figure 6.6b). Analysis is presented such that NPS internet products are grouped into three categories with increasing complexity: 1) a group for pure products, where purity and identification were confirmed using HPLC and GC-

MS respectively 2) a group, where the NPS content in the sample was not included in the training set and 3) a group for complex brands and formulations (i.e. tablets and pellets).

PCA projection of 23 NPS internet samples (20/9/16 % for PC1/PC2/PC3) onto the PCA ‘NPS and non-NPS/benchtop’ model (16/11/9 % for PC1/PC2/PC3) developed from training samples showed close projection values for PC1/PC2, slightly lower values for PC3 (Figure 6.7a and Table 6.4). This showed that the model has performed well in identifying the ‘query’ compounds along PC1/PC2. The residual variance plot showed how well the ‘query’ samples are described by the calculated PCs of the original model (Figure 6.7c). The projection line (green) plummeted to zero indicating a good description of the projected samples.

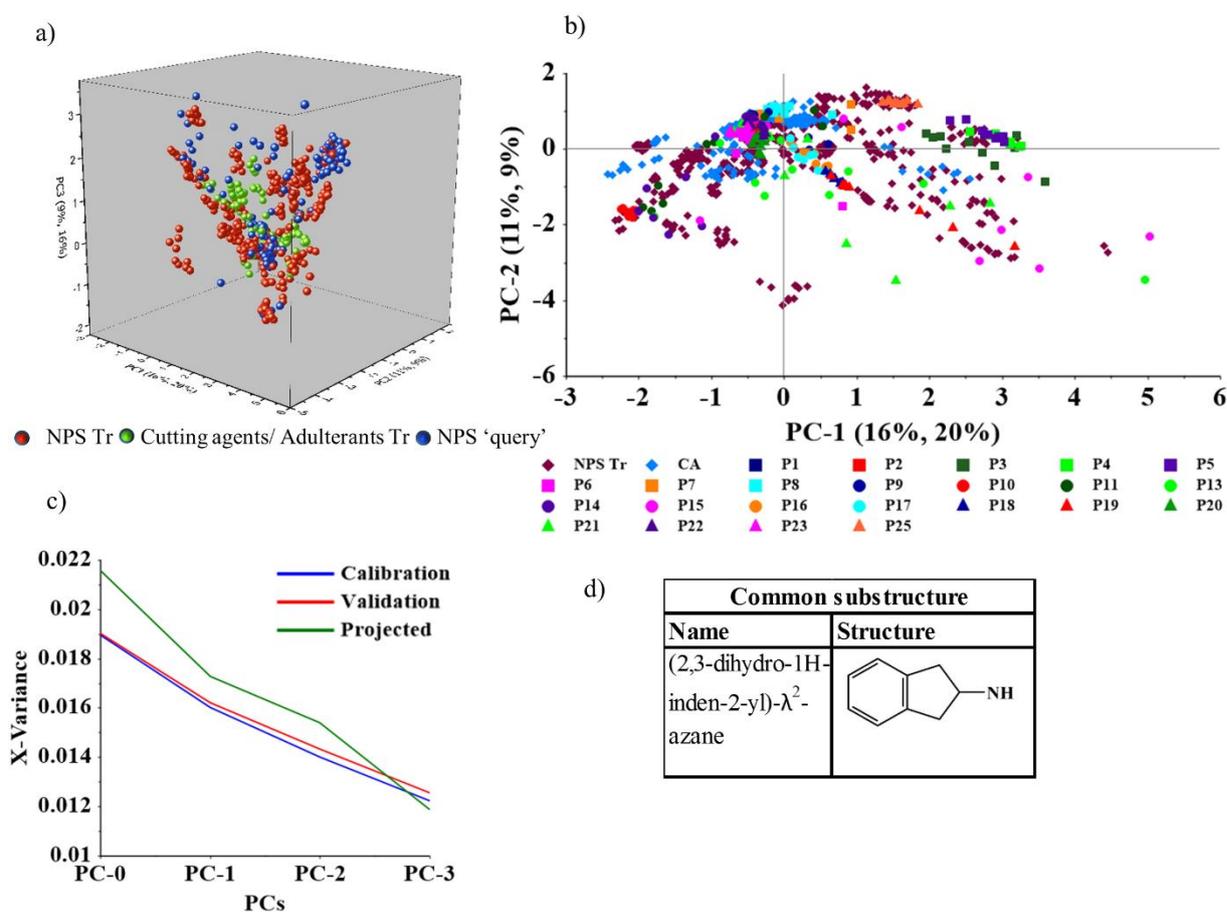


Figure 6. 7: a) 3D-PCA projection scores plot; b) 2D-PCA projection scores plot. a and b are pre-processed Raman spectra of 23 NPS internet products projected onto a PCA model generated from training samples (‘53 representative’ NPS and 22 cutting agents/ adulterants reference standards). The first and second percentage values represent the percentage explained variance by PC1/PC2/PC3 for initial and projected samples respectively; c) PCA projection residual variance plot. This plot shows the calibration and validation plots of the training samples (blue and red respectively) and the projection plot of the projected ‘query’ samples; d) Common substructure between 2-AI and N-Me-2-AI ((2,3-dihydro-1H-inden-2-yl)-12-azane).

PCA projection results showed close calibration and validation values for the first three PCs as described in Table 6.4.

Table 6. 4: PCA projection analysis results

	PC-0	PC-1	PC-2	PC-3
Calibration (training)	0.019	0.016	0.014	0.012
Validation (training)	0.019	0.016	0.014	0.013
Projection (test)	0.022	0.017	0.015	0.012

All training and projected samples were included in the Hotelling T^2 ellipse at 95 % CL, except a few replicate spectra for 5-MeO-MiPT and 4HO-DET reference standards and the NPS internet spectra P13 R5, P15 R3 and P21 R6 (ellipse not shown in Figure 6.7b). However, the examination of these spectra using the influence plot (plot not shown) demonstrated that they are not true outliers except for P13 R5, P15 R3 and P21 R6, which were associated with high F-residuals and Hotelling T^2 values (see Chapter 4), hence were rejected by the projection model because they were not well described by the calibration model.

The first group of the NPS internet products in this analysis includes the relatively pure NPS products P1-5, 10, 18-20 and 25 (see Section 6.3.2.2.). By examining the PCA 2D-scores plot, replicate Raman spectra for both P1 and P2, labelled as 2-AI and confirmed to contain 2-AI, have been shown to group together in the upper left quadrant with 2-AI and N-Me-2-AI reference standards as expected. Both 2-AI and N-Me-2-AI are aminoindanes with (2,3-dihydro-1H-inden-2-yl)-12-azane as a common substructure (Figure 6.7d). Replicate Raman spectra for P1 and P2 were also in close proximity to P21 and P22, confirmed to contain 2-AI (Table 6.2). Similarly, replicate Raman spectra for both P4 and P5 labelled as AB-PINACA and confirmed to contain AB-PINACA (Table 6.2), grouped together in the upper right quadrant with AB-PINACA reference standards. Both products were easily classified, despite being of relatively medium purity as shown through HPLC analysis (ca. 50 %). Pearson's correlation coefficient suggested high spectral similarity between both 2-AI and AB-PINACA reference spectra and replicate spectra for P1 - P2 and P4 - P5 respectively over the spectral range 1750 – 1300 cm^{-1} , with r^2 values ≥ 90 % except for P2 R7 ($r^2 = 84$ %). However, in internet mixtures, in contrast to NPS reference standards, r^2 values correlate to specific measurements (powder areas) and may not be representative if the original batch was not homogeneous.

Similarly, replicate Raman spectra for P10 labelled as flubromazepam and confirmed to contain flubromazepam (Table 6.2) grouped together in the lower left quadrant with flubromazepam reference standards and with P11, P13 and P14, known to contain flubromazepam (Table 6.2). Again, r^2 values between flubromazepam reference spectrum and replicate spectra for P10 was $\geq 90\%$, suggesting the presence of flubromazepam in this product.

Examination of the average Raman reference spectra plots (not shown) for 2-AI, AB-PINACA and flubromazepam plots against all ten replicate spectra for P1, P2, P4, P5 and P10 respectively showed that replicate spectra were very similar to each other and to reference spectra. However, P2 R5 (green spectrum) exhibited Raman bands at ca. 1644 and 1374 cm^{-1} , which were absent from the average reference spectrum for 2-AI (Figure 6.8). This is not unexpected as these samples may contain low amounts of unclaimed ingredients. A full characterisation analysis was not performed on these products.

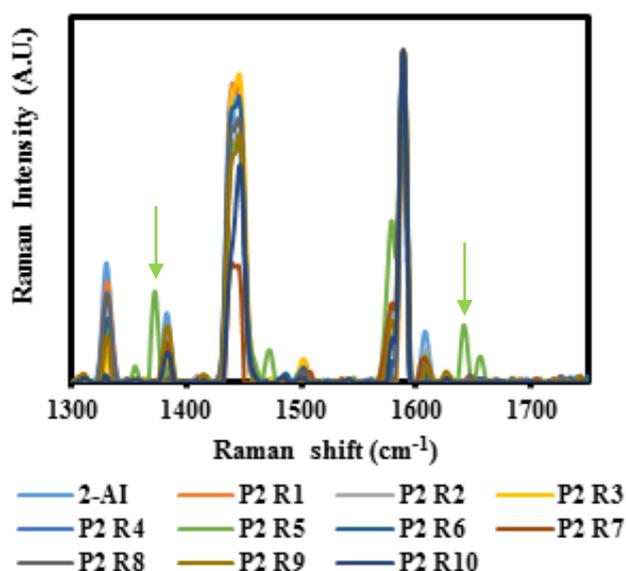


Figure 6. 8: Overlaid line plots for 2-AI reference standard and replicate spectra for P2. Raman peaks at ca. 1644 and 1374 cm^{-1} are present in P2 R5 (green) but absent in remaining replicate spectra for P2.

Replicate Raman spectra for both P18 and P19, labelled as MPA and confirmed to contain MPA, grouped together in the lower left quadrant with MPA reference standards. However, P18 R6 and R10 were close to P17 R1 and R10 respectively. P17 is known to contain 5,6-MDAI ($r^2 \geq 80\%$) (Table 6.2). The line plots (Figure 6.9a) for P18 R6 and R10, and P17 R1 and R10 showed that there were no common Raman bands between P18 and both the product and reference standard 5,6-MDAI, except a very weak band at 1316 cm^{-1} , which did not confirm the presence of 5,6-MDAI in P18 ($r^2 = 30 - 40\%$). However, this could simply be because the

correlation r^2 values between average spectra of both reference arylalkylamines 5,6-MDAI and MPA is 40 %. Projection plot (Origin Pro 2016) (Figure 6.9b) confirmed the grouping between the MPA reference spectrum and both P18 R6 and R10 and confirmed the delineation of both P17 spectra (i.e. R1 and R10) along PC1/PC2/ PC3.

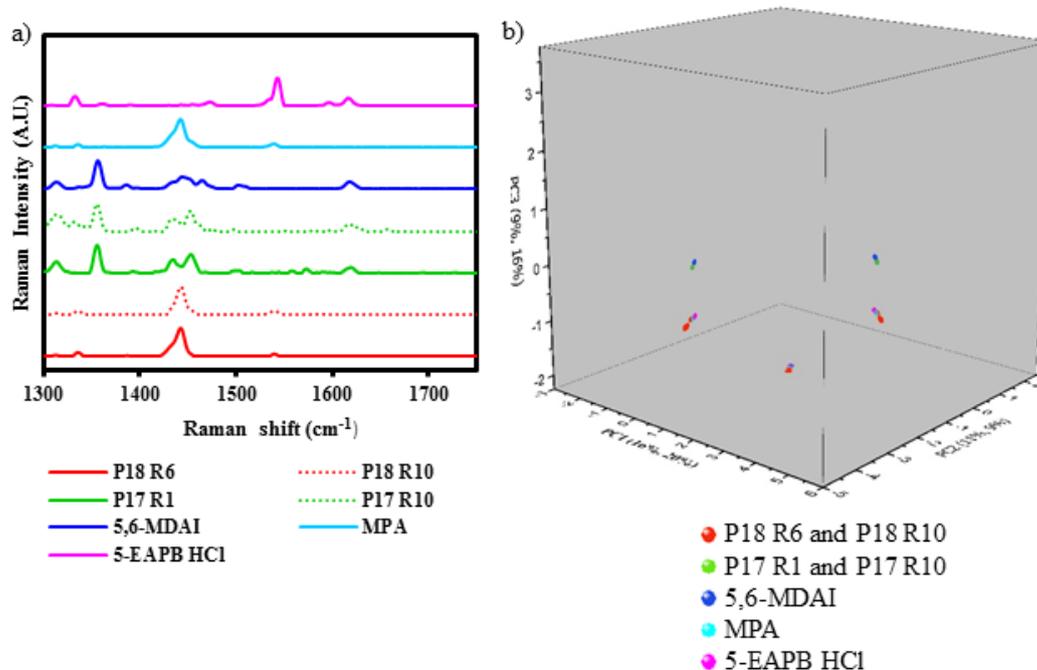


Figure 6. 9: a) Stacked line plots for average Raman reference spectra for the arylalkylamines 5,6-MDAI, 5-EAPB HCl and MPA, and NPS internet spectra P17 R1 and R10, and P18 R6 and R10; b) 3D-PCA projection scores plot showing the latter reference spectra and samples projected onto the three planes PC1/PC2/PC3 (Origin Pro 2016).

For P19, r^2 values between the product's spectra and MPA reference spectrum were greater than 99 %, except for P19 R3 - R5, where, r^2 values were 88, 71 and 83 % respectively. All three replicate spectra showed high spectral correlation coefficient (i.e. ≥ 67 %) with the tryptamines 5-MeO-DALT and 5-MeO-MiPT. GC-MS analysis of P19 confirmed that this product was adulterated with 6-APB, CAF and 5-MeO-DALT. 6-APB was not included in the training set and, therefore no spectra for 6-APB were included in the initial PCA model. Unlike all replicate spectra for P19, which overlapped with MPA reference spectra, R3 - R5 were in close proximity to the tryptamines 4-HO-DET, 5-MeO-DALT, DPT and 5-MeO-MiPT. They were closer to the latter two NPS (Figure 6.10a). The line plots (Figure 6.10b) confirmed that P19 R3 - R5 share common peaks with 5-MeO-DALT at 1652, 1551 cm⁻¹ and a doublet at ca. 1357 and 1367 cm⁻¹, which is clearly observed in P19 R4 but is exhibited as a broad band in both P19 R3 and R5. Figure 6.10a shows that P19 R4 is closer to 5-MeO-MiPT and DPT rather than 5-MeO-DALT, which was confirmed using the projection plot (Figure 6.10c), and so their relative proximities are understandable. Figure 6.10 showed that the Raman band at ca. 1550

cm^{-1} is exhibited by all the tryptamines, with slight shifts between them, which, in turn explained why P19 R3 - R5 were non selectively in close proximity with the tryptamines. The line plots showed that all replicate spectra shared a common strong Raman band with MPA at ca. 1445 cm^{-1} , hence suggesting the presence of MPA as expected. It was noted that the spectral range $1750 - 1300 \text{ cm}^{-1}$ used in this analysis do not confirm the presence of 6-APB or CAF.

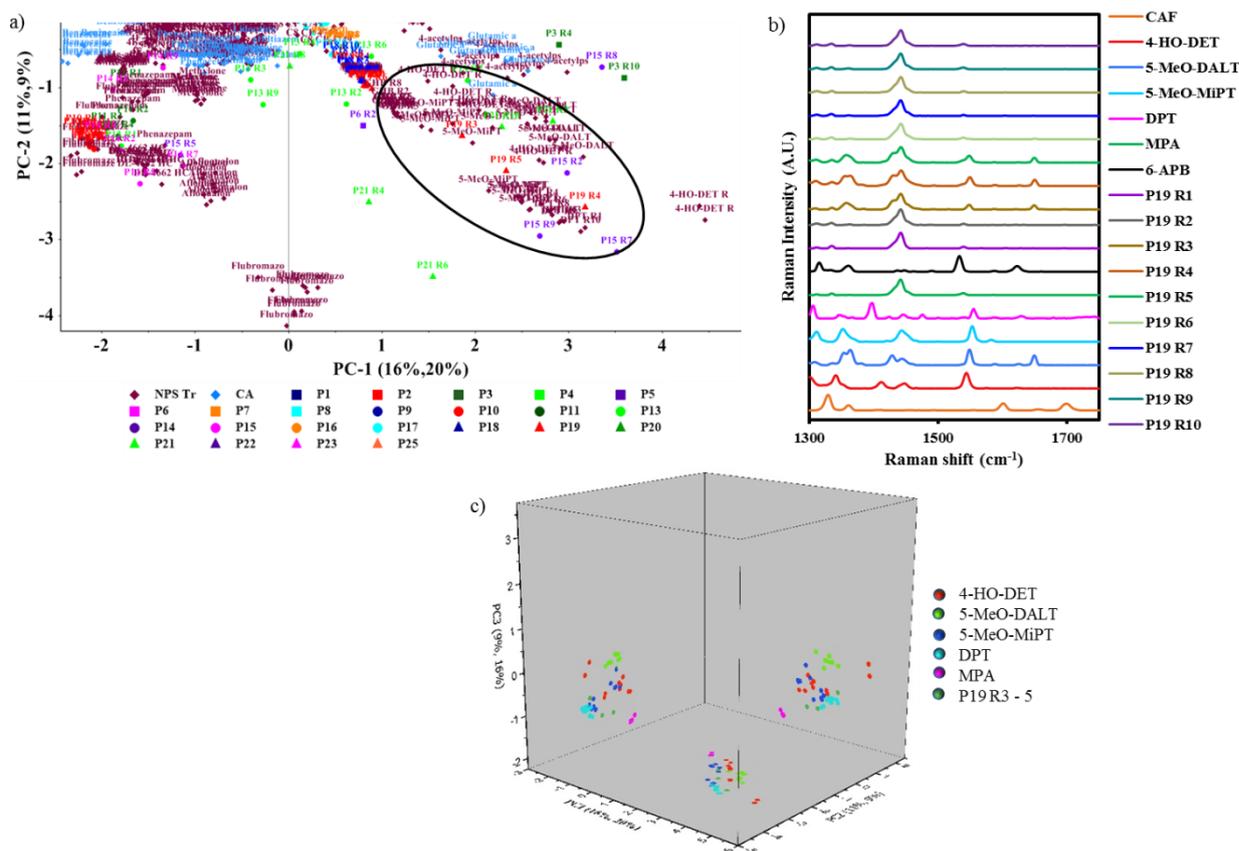


Figure 6. 10: a) The 2D-PCA projection of 23 internet products onto the PCA model generated from the training samples (53 NPS reference standards and 22 cutting agents/ adulterants) (Unscrambler[®] X10.4). P19 R3, R4 and R5 and the reference tryptamines 4-HO-DET, 5-MeO-DALT, 5-MeO-MIPT and DPT are highlighted with an ellipse; b) Line plots for replicate spectra of P19; c) 3D-PCA projection plot of average reference spectra of tryptamines (i.e. 4-HO-DET, 5-MeO-DALT, 5-MeO-MIPT and DPT) and MPA, and P19 R3 – R5 (Origin Pro 2016).

Replicate Raman spectra for P20, labelled and confirmed as containing N-Me-2-AI, were scattered in the left quadrants of PCA projection scores plot, close to 2-AI reference standard and both P1 and 2 (confirmed to contain 2-AI) (Figure 6.7b). However, P20 R3 and R4 were the only spectra, which were in close proximity to N-Me-2-AI reference standard. Remaining replicate spectra were closer to 2-AI reference standard, 3-MeO-PCE, ketamine, LID and STP. Replicate spectra for P20 and N-Me-2-AI average reference spectrum (Figure 6.11a) have been shown to share two doublets: 1) at ca. 1613 and 1591 cm^{-1} and 2) 1456 and 1438 cm^{-1} ($r^2 \geq 87$

%). Common bands at ca. 1613, 1456 and 1438 cm^{-1} were also observed in P20 and 2-AI average spectrum ($r^2 \geq 63\%$). Despite the fact that no common bands were observed with 3-MeO-PCE, ketamine, LID and STP, the projection plot displayed closeness between them, with the exception of few LID spectra (Figure 6.11b).

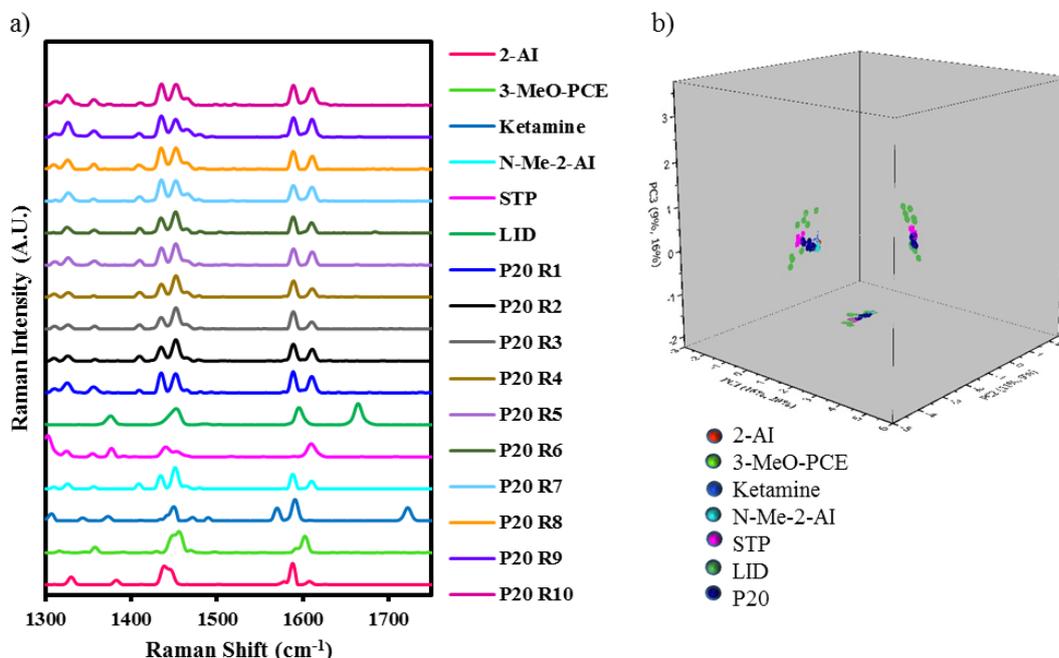


Figure 6. 11: a) Line plots for average reference spectra of 2-AI, 3-MeO-PCE, ketamine, N-Me-2-AI, STP and LID, and replicate spectra of the product P20; b) 3D-PCA projection plot of the latter average reference spectra and samples (Origin Pro 2016).

Finally, replicate Raman spectra for P25, labelled and confirmed to contain THJ-018 using GC-MS, HPLC and Raman spectroscopy, grouped together in the upper right quadrant with THJ-018, as expected, and zopiclone reference standards. Clustering of zopiclone with synthetic cannabinoids was previously addressed in Chapter 5 Section 5.3.3.1.

In contrast to the previous samples, which readily grouped with the NPS reference standard that correspond to the main NPS identified in the mixture (Table 6.2), the product P3, labelled and confirmed to contain AB-FUBINACA using GC-MS and HPLC, yielded inconsistent correlations *inter alia* to other synthetic cannabinoid analogues such as 5F-APINACA and AB-PINACA using handheld Raman spectroscopy. Replicate Raman spectra for AB-FUBINACA were scattered between the upper and lower right quadrants of the 2D-PCA projection scores plot, located between both AB-PINACA and N-PB-22 on one side and AB-FUBINACA on the other side. They were closer to the former than the latter. They were very close to both P4 and P5, which were both known to contain AB-PINACA. The line plots showed that all ten

replicates shared common bands with AB-FUBINACA reference standard i.e. at ca. 1670, 1646, 1580, 1498, 1481, 1410, 1361 and 1320 cm^{-1} ($r^2 \geq 64\%$) (Figures 6.12a and 6.12b). AB-PINACA also exhibited these bands except the band at ca. 1670 cm^{-1} ($r^2 \geq 60\%$), whereas N-PB-22 exhibited bands at ca. 1580 and 1481 cm^{-1} only ($r^2 \geq 33\%$). The overlaid spectra of the AB-PINACA and AB-FUBINACA reference standards and the replicate spectra for P4 and P5 (Figure 6.12) showed that this could be because of a band at ca. 1666 cm^{-1} that is present in the spectrum for AB-FUBINACA but absent in the products' spectra. In addition, this could also be due to two shifted bands at ca. 1643 and 1476 cm^{-1} . Surprisingly, the Raman spectra clearly showed a greater similarity between the products' spectra and AP-PINACA, as opposed to AB-FUBINACA. This result, which appears to contradict the confirmatory analysis, could simply be because of sample heterogeneity. This example illustrated the complexity of NPS products.

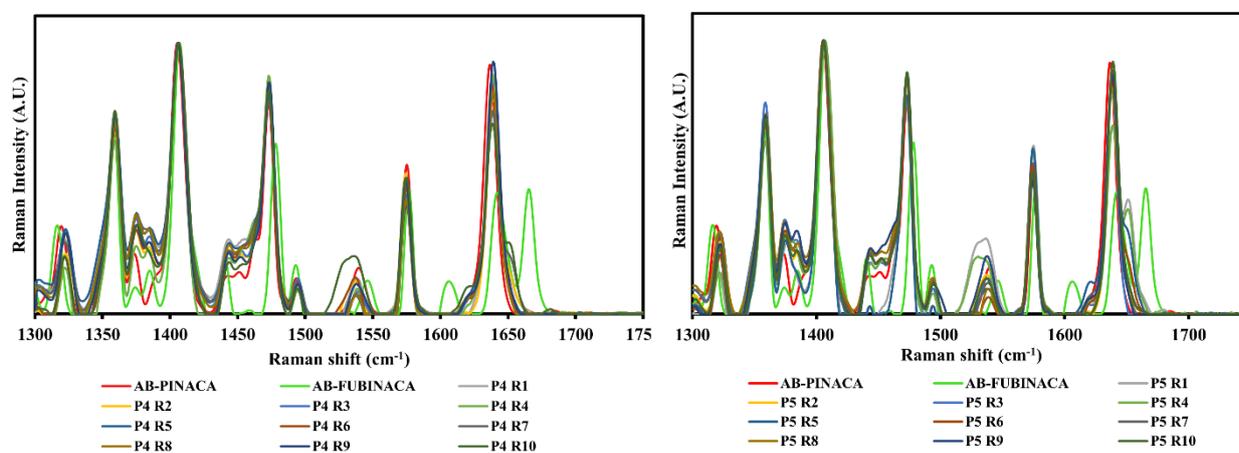


Figure 6. 12: Line plots for average reference spectra of AB-PINACA (red), AB-FUBINACA (light green) and replicate spectra for a) P4; b) P5.

The analysis of the second group of NPS internet products addressed the ability of the PCA model to identify ‘unknown’ samples, for which there is no reference substance in the training set. In this respect, replicate Raman spectra for both P16 and 17 were analysed. The latter products were labelled and confirmed to contain 5,6-MDAI using both GC-MS and HPLC. However, using handheld Raman spectroscopy, 5-IAI (an aminoindane class analogue) was identified as the main component in both products. Both 5,6-MDAI and 5-IAI were not included in the PCA model training set and, therefore no spectra for either standards were included in the initial PCA model. The NPS with greatest structural similarity to 5,6-MDAI is the medoid methylone (Chapter 5, supercluster 11). However, P16 and 17 did not group with methylone, as may be expected, possibly because of unknown impurities in these products and possibly due to poor S/N in the 5,6-MDAI reference spectra.

Both P16 and 17 grouped together near the centre of the scores plot, in close proximity to SDB-006, C8 CP 47-497 and 5-EAPB. In the initial PCA projection, where, 21 test pure NPS were projected onto a PCA model of 53 pure NPS (Chapter 5 Section 5.3.3.2.), 5,6-MDAI was also projected in close proximity to both 5-EAPB and C8-CP, 47-497. The line plots (Figure 6.13a) showed that replicate spectra for P16 and P17 were very similar, all exhibiting Raman bands at ca. 1624 and 1358 cm^{-1} ($r^2 \geq 78\%$), consistent with 5,6-MDAI, a doublet at ca. 1333 and 1316 cm^{-1} consistent with C8 CP 47-497 ($r^2 \geq 8\%$) and a band at 1620 cm^{-1} consistent with 5-EAPB ($r^2 \geq -0.8\%$). These findings were confirmed using the projection plot, where these samples grouped together along PC1/PC2/PC3 except SDB-006 (dark blue) (Figure 6.13b). As described earlier, these results show that r^2 values do not necessarily reflect grouping or delineation on the scores plot because not all of the variance is explained or accounted for in the projections, whereas it is in the calculation of r^2 values.

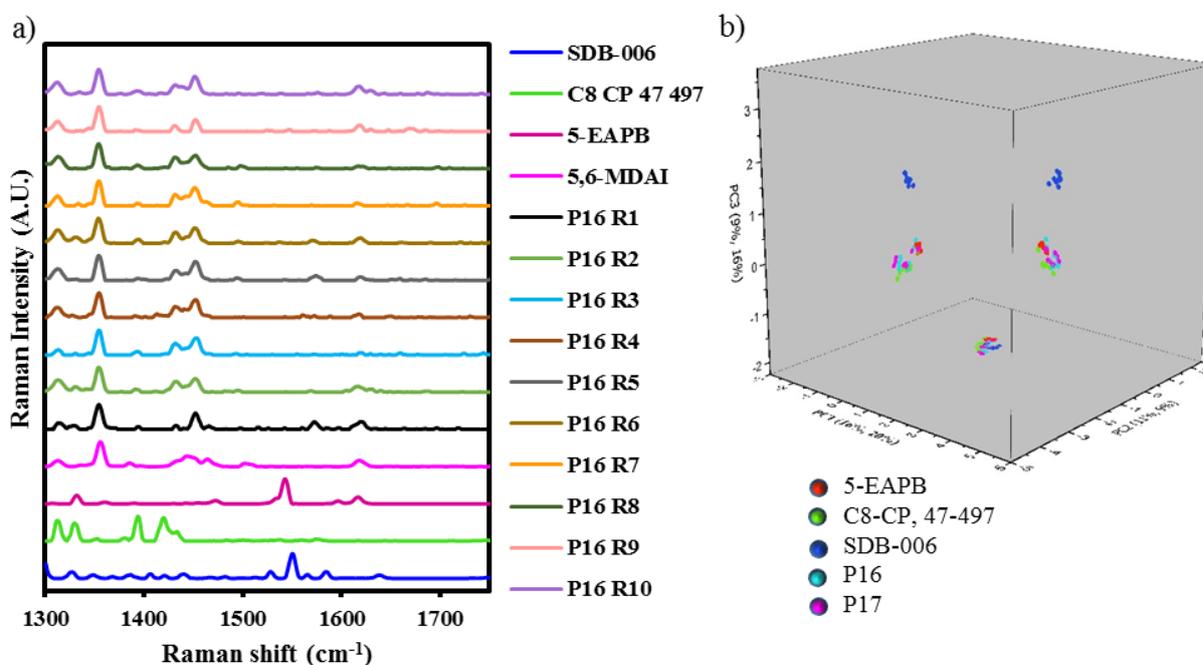


Figure 6. 13: a) Line plots for average reference spectra of 5-EAPB, 5,6-MDAI, C8 CP 47-497 and SDB-006, and replicate spectra of P16; b) 3D-PCA projection plot of average reference spectra of 5-EAPB, 5,6-MDAI, C8 CP 47-497 and SDB-006, and replicate spectra of both P16 and P17 (Origin Pro 2016).

The analysis of the third group of the NPS internet products evaluates the ability of the model to identify complex mixtures e.g. branded products, formulations and products, which exhibited very low purity with HPLC analysis. In this respect, P6-9, 11-15, 21-23 are discussed. Products 6 – 9 are etizolam tablets the replicates of which have mostly scattered in the upper left quadrant. Raw spectra exhibited very poor S/N and intense fluorescent background in all

replicate spectra, which, in turn resulted in pre-processed spectra with poorly pronounced Raman bands as described by Harkai and Putz [161] and relatively poor average r^2 values between replicate spectra. For P6, r^2 was ≥ 95 %, except for R2, where r^2 was ≥ 25 %. However, for P7 - P9, average r^2 was ≥ 71 , 80 and 33 % respectively. Poor quality of spectra, even after pre-processing may hamper identification of these substances. This was observed with P7, which was in close proximity to CRE, MAN, TAL, TAU, L-TYR, NM-2201 and UR-144; P8 was in close proximity to SUC, JWH-015 and JWH-073 and P9 was in close proximity to TAU, JWH-015, JWH-122, SDB-006 and P21R7. These substances may not be false positives as a full characterisation of internet NPS products was not performed.

Despite the fact that P6 - P9 were confirmed to contain etizolam using both GC-MS and HPLC, using PCA projection, a high correlation coefficient was only observed between the etizolam average reference spectrum and replicates P6 R1, R3 - R10 ($r^2 \geq 87$ %) and P9 R1, R2 and R4 ($r^2 \geq 23$ %) (Figure 6.14a). P7 shared common Raman bands with NM-2201 ca. 1383 and 1340 cm^{-1} with most replicates, explaining why grouping between P7 and NM-2201 occurred. Despite the fact that MCC was identified as the top hit in P7 – P9 using handheld Raman spectroscopy, a high correlation coefficient between spectra of P8 and MCC ($r^2 \geq 68 - 96$ %) was only observed (Figure 6.14b). P8 also grouped in close proximity to JWH-015 and JWH-073 possibly because they exhibited common bands at ca. 1383 and 1340 cm^{-1} across all ten replicates. It is not unexpected that etizolam was not identified in P7 – P9: firstly, because its purity with HPLC was < 1 %. Secondly, because of the presence of unknown fillers and impurities in the tablets, which contributed to the poor signal obtained with a 785 nm λ_{ex} . Thirdly because of the limitation of the pre-processing method used such as the limited spectral range. Fourthly because of the quality of the library spectra, where the purity of NPS reference standards was ≥ 98 % with a 2 % of unidentified impurities; and finally due to the intrinsic nature of cutting agents in hindering the identification of low amounts NPS.

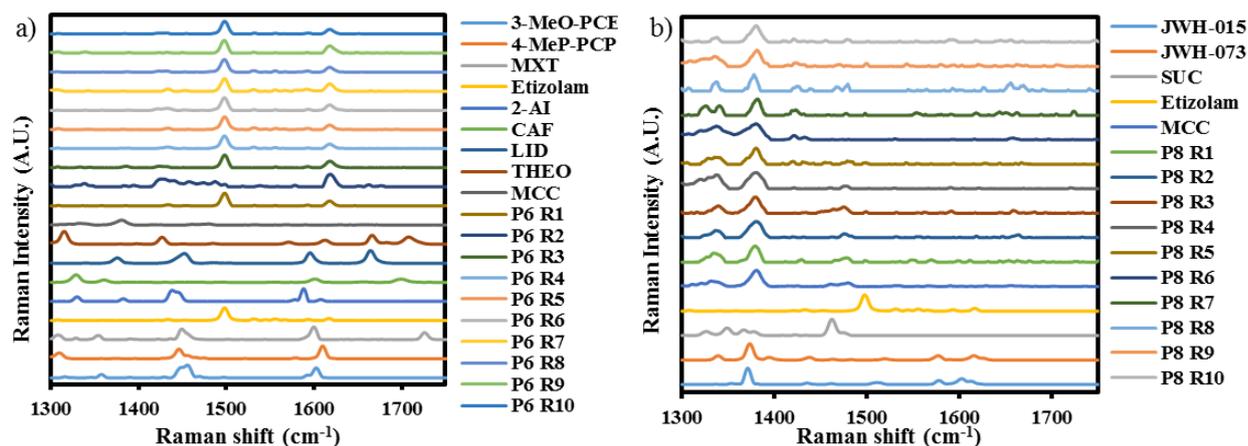


Figure 6. 14: Pre-processed average Raman spectra for: a) 3-MeO-PCE, 4-MeO-PCP, MXT, etizolam, 2-AI, CAF, LID, THEO, MCC and P6; b) JWH-015, JWH-073, SUC, etizolam, MCC and replicate spectra of P8.

Products 11-15 were all coloured pellets containing flubromazepam, where the colour was incorporated within the pellets' content rather than merely being used as a coating. These products were revealed to be impure *via* HPLC analysis with average percentage purity ranging from 2.4 ± 0.1 to 7.8 ± 0.2 . The line plots of P11 replicate spectra (Figure 6.15a) showed that R1 – R4 and R9 – R10 exhibited Raman bands similar to the flubromazepam reference spectrum ($r^2 \geq 95\%$). Replicate spectrum P11 R8 showed partial similarity to the flubromazepam reference spectrum ($r^2 \geq 69\%$) but replicate spectra P11 R5 – R7 were very different with r^2 values $\leq 43\%$. By examining the 2D and 3D-PCA projection scores plots, P11 R1 – R4 grouped in the lower left quadrant in close proximity to flubromazepam reference standard, whereas P11 R8, R9 and R10 grouped next to P13 and P14, other flubromazepam products of the same colour (intense green) purchased from different websites, possibly indicating a similar supply chain (Figures 6.7b and 6.15b).

Replicate spectra for P13 exhibited intense fluorescent backgrounds, very poor S/N and a wide range of correlation coefficients between replicate spectra (27 - 99 %). P13 R1 – R4, R6 and R8 – R10 shared common Raman bands with the flubromazepam reference spectrum ($r^2 = 47 - 93\%$). However, only P13 R1 grouped alongside the flubromazepam reference spectrum in the PCA model, P13 R2 – R4, R6, R9 and R10, were scattered and delineated from other NPS. This was expected because of poor S/N in these spectra. P13 R8 grouped near P11 and P14 was confirmed to contain flubromazepam. Poor correlation coefficient were observed between P13 R2 – R4, R6, R9 and R10 Raman spectra (not shown) and the flubromazepam standard. This was shown *via* r^2 values and the nearest substances on the projection plot were 5-EAPB, DIL and MPA. P13 R5 was in close proximity to 4-HO-DET ($r^2 = 48\%$). This is potentially a

misclassification as both spectra exhibited high fluorescent backgrounds in raw spectra and very poor S/N in pre-processed spectra. Subtraction of fluorescent backgrounds is usually a common approach to overcome this problem despite the fact that it does not remove the shot noise from the spectra [101]. Subtraction of fluorescent backgrounds has also been shown to increase the risk of subtracting weak Raman peaks, which may be useful in identifying an ‘unknown’ substance [243]. In addition, the diversity of excipients and impurities in NPS may impact the performance of the chemometric model for the identification of the ‘analyte’ of interest in the mixture. This was previously described by Mainali & Seelenbinder for the identification of illicit drugs using vibrational spectroscopy and a universal chemometric model in the presence of multiple adulterants [118].

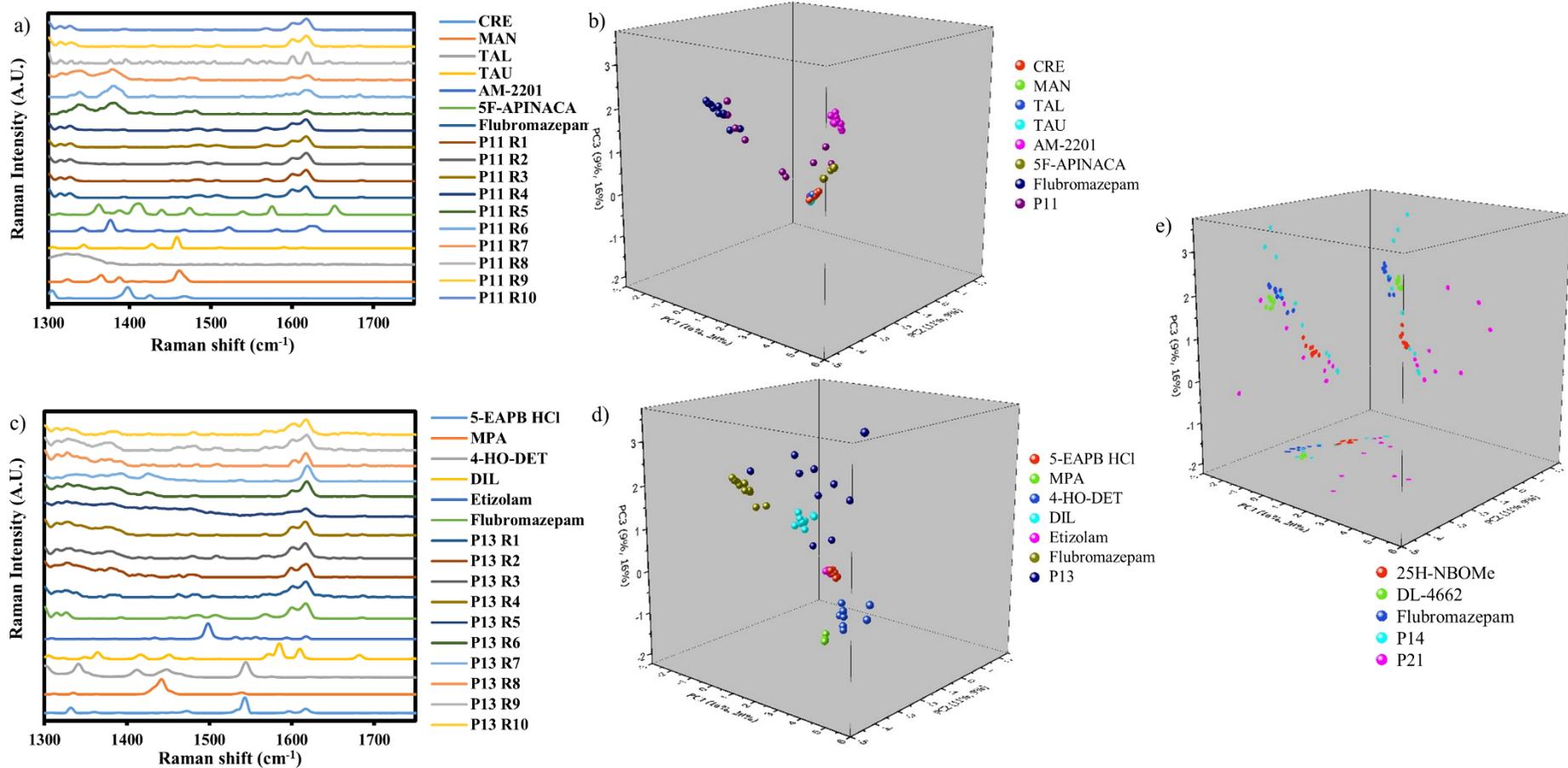


Figure 6. 15: a) Pre-processed average Raman spectra for CRE, MAN, TAL, TAU, AM-2201, 5F-APINACA, flubromazepam and replicate spectra of P11; b) 3D-PCA projection scores plot for CRE, MAN, TAL, TAU, AM-2201, 5F-APINACA, flubromazepam and replicate spectra of P11; c) Pre-processed Raman spectra for 5-EAPB, MPA, 4HO-DET, DIL, etizolam, flubromazepam and replicate spectra of P13; 3D-PCA projection scores plot for: d) 5-EAPB, MPA, 4HO-DET, DIL, etizolam, flubromazepam; and e) 25H-NBOMe, DL-4662, flubromazepam, replicate spectra of both P14 and P21.

Examination of the Raman spectra for P13 and potential overlapping NPS standards (Figure 6.15c) and 3D-projection scores plot (Figure 6.15d) showed that poor Raman signal and fluorescent backgrounds hindered the identification of NPS content in coloured pellets. In similar cases, spectra should be examined first prior to suggesting chemical scaffolds of content using PCA projection plots.

Replicate spectra of P14 exhibited raised baselines with visible sharp peaks. Pearson's correlation coefficients between replicate spectra varied from 47 to 97 %. The line plots of P14 replicate spectra R1 - R10 (plots not shown) showed that R1, R2, R4, R5, R7 and R9 exhibited Raman bands similar to the flubromazepam reference spectrum ($r^2 = 85 - 99 \%$). By examining the 2D and 3D-PCA projection scores plot, P14 R1, R2, R4, R5, R7 and R9 grouped in the lower left quadrant in proximity to the flubromazepam reference standard (Figures 6.7b and 6.15e). However, most replicate spectra grouped with other NPS and were evaluated individually.

Both P14 R5 and R9 were closer to DL-4662 and 25H-NBOMe respectively than the flubromazepam standard. By examining the line plots, no common Raman bands were observed, despite that r^2 between 25H-NBOMe and P14 R1 - R10 being between 43 - 54 %, demonstrating the negative impact of raised baselines. Furthermore, both P14 R6 and R10 grouped with P11, another flubromazepam product. The latter spectra possibly have similar adulteration to P11 (unknown Raman bands not clearly identifiable from line plots - not shown). Pearson's correlation coefficients between both P14 R6 and R10 and flubromazepam reference spectrum is 60 and 70 % respectively. The same applies to P14 R3 grouping close to P11 and P13. Both were confirmed to contain flubromazepam ($r^2 = 92 \%$). P14 R8 clustered near P21 R3 (confirmed to contain CAF) and the CAF reference standard. Pearson's correlation coefficients between both P14 R8 and flubromazepam and CAF reference spectra was 87 and 38 % respectively. No shared Raman bands were observed to confirm if this sample contained CAF. The projection plot (Figure 6.15e) showing the different positions of P14 replicate spectra on the PCA projection plot with respect to flubromazepam as well as the proximities of a few P14 replicate spectra with other NPS such as 25H-NBOMe, DL-4662 and the product P21 demonstrated the complexity of these samples and the limitation of PCA projection to fully identify and classify NPS street-like samples.

Replicate spectra of P15 (pink pellets) exhibited high fluorescent backgrounds with very poor S/N in all replicate spectra except for R2, R7 and R9, where defined Raman features were observed. There was no Raman bands observed in P15 R1, R3 and R6. This is an example of a product, where a few replicate spectra displayed Raman signals whilst others did not due to intense fluorescence possibly resulting from impurities. Therefore the latter spectra were not discarded because sharp Raman peaks were observed on top of the fluorescent background. Pearson's correlation coefficients between pre-processed replicate spectra reflected this difference with a wide correlation range from -18 to 98 %. The Raman spectra and 3D-PCA projection scores plot of P15 replicate spectra R1 - R10 (Figure 6.16a) showed that only R5 displayed defined Raman bands similar to the flubromazepam reference spectrum ($r^2 = 59\%$). By examining the 2D-PCA projection scores plot, R5 grouped in close proximity to P14 (confirmed to contain flubromezepam) and flubromazepam reference standard (Figure 6.6b). P15 R2, R7 and R9 grouped in the lower right quadrant in proximity to the tryptamines and in particular to the DPT reference standard. By comparing the line plots of P15 R2, R7 and R9 and the DTP reference spectrum (Figure 6.16a) ($r^2 = 48 - 53\%$), it was observed that they shared common peaks at ca. 1623, 1452, 1365 and 1341 cm^{-1} , explaining their relative proximities on the scores plot. Remaining spectra were close to other NPS reference standards such as UR-144. The 3D-projection scores plot showed that only one spectrum was in close proximity to UR-144 (Figure 16b). This is most likely a false positive due to the very poor S/N in these spectra.

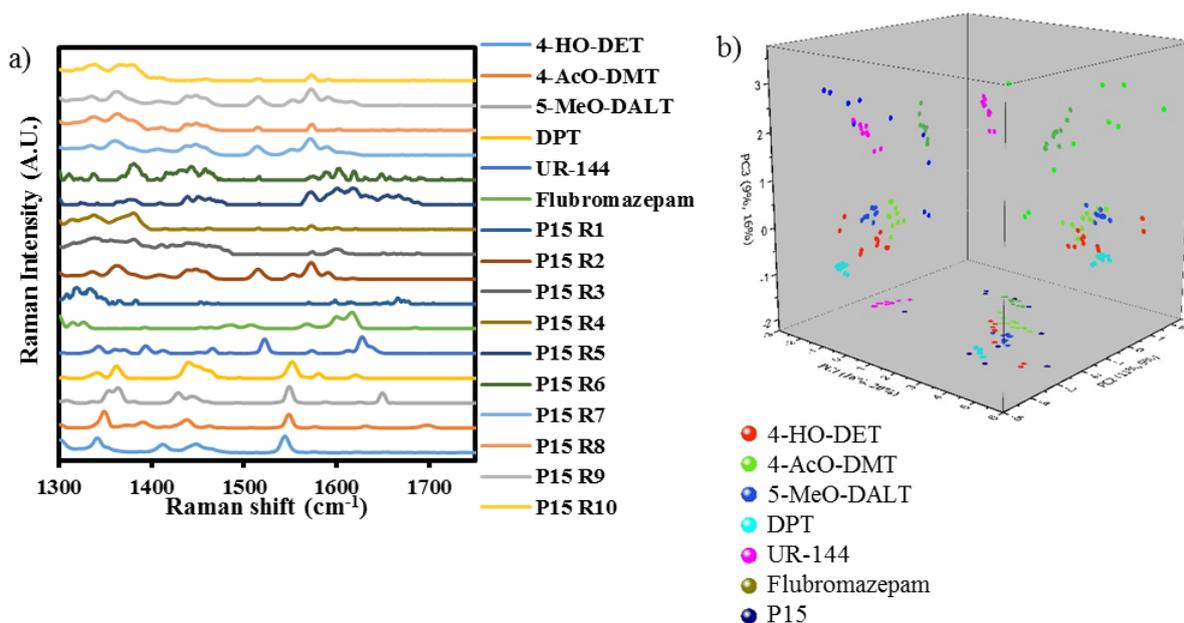


Figure 6. 16: a) Line plots for average reference spectra of 4-HO-DET, 4-AcO-DMT, 5-MeO-DALT, DPT, UR-144, flubromazepam and replicate spectra of P15; b) 3D-PCA projection plot

of HO-DET, 4-AcO-DMT, 5-MeO-DALT, DPT, UR-144, flubromazepam and replicate spectra of P15.

Replicate spectra of P21 (brown powders) exhibited high fluorescent background with very poor S/N between replicate spectra ($r^2 = 6$ to 96 %). P21 R1 - R3 and R5 were the only replicate spectra which showed distinct peaks. Overlap between replicate spectra of P21 with NPS reference standards and products (i.e. 2-AI, CAF, 5-MeO-DALT, MXT, LID and P6) was carefully analysed due to the poor quality of pre-processed spectra (Figure 6.17a).

The r^2 values between the 2-AI average reference spectrum and all replicates was greatest for both R1 and R2 ($r^2 \geq 95$ %), confirming their proximities on the PCA projection plot. In contrast, R2 – R6, R9 and R10 overlapped with CAF reference spectra with a wide range of r^2 values (i.e. 20 to 91 %), which suggested that because CAF is a relatively strong Raman scatterer, smaller values of r^2 may be indicative of its presence in the sample. It may also indicate that CAF may hinder low amounts of NPS in internet samples. CAF was confirmed in this product using GC-MS. Despite the fact that no common Raman bands were observed between P21 and 5-MeO-DALT, MXT, LID or etizolam average reference spectra, they clearly overlapped on the PCA project plot.

In contrast, spectra with relatively better quality were collected for P22. Distinct Raman bands were defined in all replicate spectra and were consistent with 2-AI ($r^2 = 72 - 98$ %), N-Me-2-AI ($r^2 = 33 - 62$ %) and CAF ($r^2 = -1 - 43$ %) average reference spectra (Figure 6.17b). By inspecting the 2D-PCA projection scores plot (Figure 6.7b), replicate spectra were in close proximity to 2-AI, N-Me-2-AI, CAF and N-ethylamphetamine reference spectra. However, the line plots showed that there was no common bands between N- ethylamphetamine and these substances over the range 1750 - 1300 cm^{-1} (Figure 6.17b). Similarly for P23, replicate spectra displayed defined Raman peaks, which were consistent with CAF ($r^2 = 95 - 99$ % for P23 R2 to R10) and 2-AI ($r^2 = 64$ % for P23 R1 only) (Figure 6.17c). Again, no common Raman bands were observed between P23 and ketamine, etizolam, PAR and THEO average reference spectra and, hence overlap with these spectra was not explained.

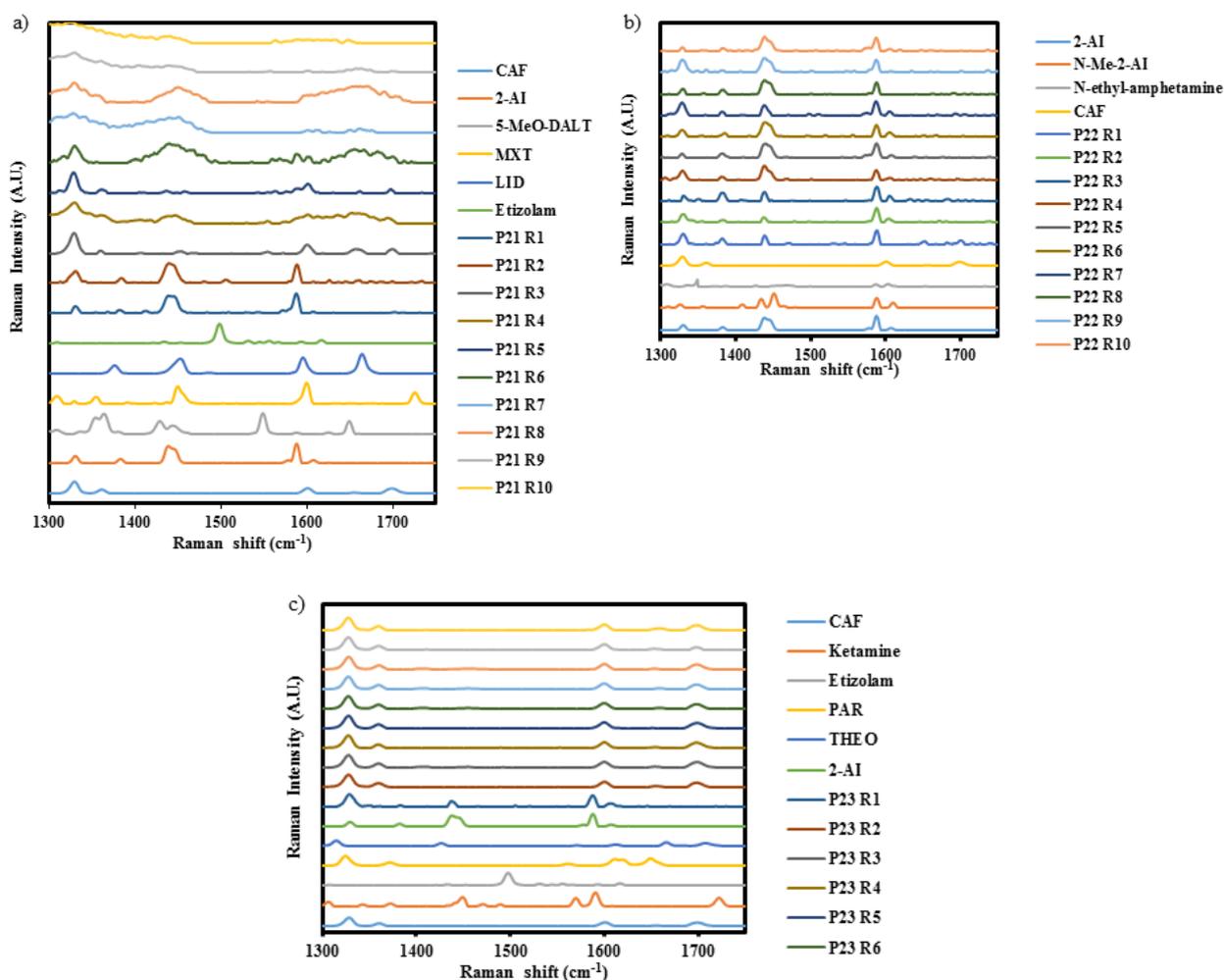


Figure 6. 17: Line plots for average reference spectra of: a) 2-AI, 5-MeO-DALT, CAF, etizolam, LID and MXT and replicate spectra of P21; b) 2-AI, N-Me-2-AI, N-ethylamphetamine, CAF and replicate spectra of P22; c) 2-AI, CAF, etizolam, ketamine, PAR, THEO and replicate spectra of P23.

6.4. Conclusions

In this Chapter, the feasibility of suggesting the chemical scaffolds of NPS in NPS internet products was evaluated using the ‘NPS and non-NPS model/benchmark’ model, which consisted of the ‘NPS representative’ database and common excipients. Using this model, ‘query’ samples were well described by the calculated PCs of the calibration model and indicated a good description of the projected ‘query’ sample. Results demonstrated the great potential of the chemometric model and the ‘NPS representative’ database in classifying complex mixtures of NPS.

Results showed that identification of NPS in these products was influenced by the purity of the products and the quality of the spectra. Relatively pure products, grouped well with relevant NPS reference spectra, as expected. NPS products, where the claimed NPS content was not

included in the training set overlapped with class analogues. In complex branded and formulated products, PCA projection was limited in confirming the content of the mixtures, possibly due to high F-residuals and Hotelling T^2 values, poor S/N in pre-processed spectra, fluorescing impurities and number of constituents in the mixtures.

In addition, the overview of the new datasets (Sections 6.3.2.1. to 6.3.2.3.) gave an insight into the potential impact of the signals resulting from impurities and common excipients on the performance of the model. The first PCA model generated from the excipients' dataset showed that the variance was mainly explained by adulterants with strong Raman scattering activity. It also showed that adulterants and cutting agents were grouped according to their assigned chemical classes.

The second PCA model generated from NPS internet products explained a greater calibrated cumulative variance (i.e. 70 % for PC1/PC2/PC3) as compared to the pure substances, in Chapter 5, reflecting the complexity of the samples and the variance between replicate spectra of individual samples. This PCA also reflected the purity of the internet samples, such that replicate spectra of samples with greater purity grouped together, whereas replicate spectra of samples with lower purity (branded products and pellet formulations) were scattered on the PCA scores plot.

The PCA analysis of the third dataset for 53 NPS reference standards and 22 cutting agents/ adulterants) illustrated the potential overlap between NPS and non-NPS substances, which included the anaesthetics (i.e. BEN, LID and PRO), pharmaceuticals (i.e. DIL), stimulants (i.e. CAF), food supplements (i.e. NIA) and amino acids (i.e. GLUT) and NPS members of superclusters 2, 3, 5, 6 and 7. This overlap may be employed as a tool to suggest further investigation of samples containing these cutting agents/ adulterants. A greater overlap occurred between training NPS and adulterants, which were members of the same chemical class, such as anaesthetics.

Future work should focus on testing a wider variety of 'query' NPS to evaluate the suggested overlap between NPS representing all previously described superclusters and non-NPS substances. Future work should also focus on refining the pre-processing methods and evaluate the proof-of-concept method using handheld Raman spectroscopy.

7. Classification of NPS internet samples using handheld Raman spectroscopy, a 1064 nm laser source and chemometric approaches – a proof-of-concept

7.1. Introduction

This Chapter contains the first report to evaluate the use of a handheld Raman instrument, equipped with a laser excitation wavelength (λ_{ex}) of 1064 nm and a spectral database of ‘53 representative’ new psychoactive substances (NPS) reference standards and 22 cutting agents/adulterants for the identification of 21 NPS reference standards and 25 NPS products purchased from the internet (see Chapters 5 and 6 and Zloh et al. [204]). Chapters 5 and 6 have demonstrated the feasibility of using principal component analysis (PCA) in conjunction with Raman spectral data for the identification of NPS. However, these proof-of concept studies were initially undertaken with a benchtop Raman instrument, providing a great performance owing to its high resolution and signal to noise, excellent optics and confocal microscope. In order to investigate the in-field applications of the developed models, the use of handheld Raman spectroscopy was evaluated. Chapters 2 and 3 demonstrated that using handheld Raman spectroscopy equipped with a longer λ_{ex} (i.e. 1064 instead of 785 nm) and a mixtures algorithm improved the identification of NPS in NPS ‘street-like’ mixtures and reduced false positive and false negative rates. Therefore, in this Chapter, the chemometric models developed in Chapters 5 and 6 were applied using a handheld Raman ‘ProgenyTM’ instrument to evaluate the feasibility of classifying NPS in NPS internet mixtures by suggesting their chemical scaffolds and/ or correlating them with their near neighbours.

With handheld Raman spectroscopy, in-field identification of NPS is often dependent on the accuracy and effectiveness of the in-built algorithms and/ or availability of reference spectra in in-built libraries, where the latter is dependent on the availability of appropriate reference standards. However, this is usually problematic because updated libraries and synthesis of reference standards often lag behind the continuous emergence of NPS. In addition, the heterogeneity of NPS products [12, 78, 140], the diversity of the incorporated excipients and impurities and the presence of fluorescing impurities hinder the identification of NPS when using a standard 785 nm laser λ_{ex} (see Chapters 2 - 6) [12, 161]. Therefore, this work is novel and significant because it investigates the feasibility of improving the identification of ‘unknown’ NPS, in their pure form or in ‘street-like’ mixtures, by combining the use of a longer

λ_{ex} (i.e. 1064 nm) and chemometrics with Raman spectral libraries containing the ‘representative’ NPS Raman database, common cutting agents/ adulterants (Chapter 6).

In this study, the application of the proof-of-concept studies (Chapters 5 and 6) lies in the accurate classification, or otherwise, of an independent test set (i.e. NPS reference standards and internet samples) using handheld Raman spectroscopy equipped with a 1064 nm laser λ_{ex} against known NPS reference spectra including the spectra of the ‘53 representative’ NPS previously discussed in Chapter 5. To achieve this, a PCA model is generated from pre-processed Raman spectra (Unscrambler® X 10.4). The training set used in generating the PCA model contained the ‘representative’, structurally and chemically diverse NPS Raman database, as well as cutting agents/ adulterants commonly incorporated in seized NPS samples (Chapter 6).

The two-fold aim of this study is to evaluate the classification of ‘query’ NPS firstly by projecting the Raman spectra of both NPS reference standards and internet products onto the PCA model generated from Raman reference spectra of NPS alone and secondly by projecting onto a PCA model derived from the Raman spectra of NPS and cutting agents/ adulterants. Hence this study evaluates the effect of using a longer λ_{ex} (i.e. 1064 nm) on a handheld platform in conjunction with Raman spectral data pre-processing using the Unscrambler® software on the classification and/ or assignment to the previously described EMCDDA/EDND classes or superclusters/ clusters (Chapter 5) of known pure substances and NPS contained in mixtures of unknown composition using PCA projection.

7.2. Materials and Methods

7.2.1. Reagents and chemicals

Seventy-four NPS reference standards were commercially obtained from both Chiron AS (Trondheim, Norway) and LGC Group (Teddington, UK) (see Chapter 5). In this work, the 74 NPS were split into two groups: 53 acting as a training set and 21 as a test set. The 21 test set molecules were subdivided into two groups: 17 from the original dataset of 478 NPS provided by the EMCDDA/EDND [204, 273] that were not used to train the model subsequently referred to as ‘test’ or ‘query’ molecules, and four, that were not present in the dataset of 478 NPS from the EMCDDA/EDND referred to as ‘out-of-model (OOM)’ samples (see Chapter 5 Table 5.1). The Raman spectra of the eight adulterants and fourteen cutting agents previously purchased (see Chapter 6 Section 6.2.1 and Chapter 2 Table 2.1 for full details) were used in this Chapter

as additional training samples in the generation of the PCA model. Twenty-five NPS products were purchased from the internet (Table 6.1) to evaluate the ability of the PCA model, generated using a handheld platform with a 1064 nm laser λ_{ex} , to project the Raman spectra of these products and correctly classify the NPS contained therein.

7.2.2. Raman analysis

7.2.2.1. Sample preparation

For building the library, due to limitation in the amount of sample procured, ca. 3 - 5 mg of each NPS reference standard were weighed, in a powder form, tapped and flattened on aluminium plates (HSA14521A - Weight dish alum 43 mm, Fisherbrand), covered with glass microscope cover slips (22 x 22 x 0.13/ 0.17 mm) and sellotaped for handheld Raman analysis (Figure 7.1).

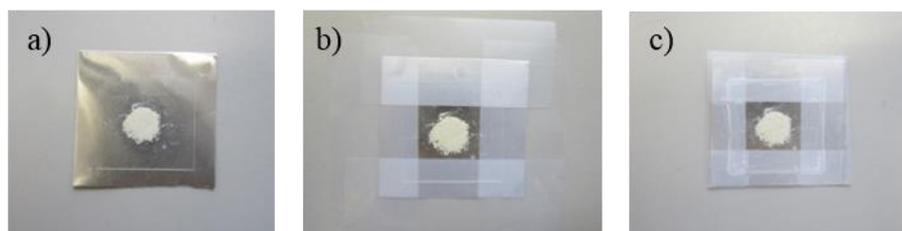


Figure 7. 1: Sample preparation prior to handheld Raman analysis (ProgenyTM, Rigaku, USA). a) Sample placed between aluminium plate and cover slip; b) Attaching aluminium plate to cover slip with removable sellotape; c) Final format for sample presentation to the handheld instrument for Raman analysis.

As described in Chapter 5, two NPS ‘test’ samples (i.e. JWH-018 and MN-18) were obtained as oils. Oil samples (ca. 10 mg) were recovered onto aluminium plates by flushing out the containers with approximately 1 mL of acetone to obtain a solution for analysis. These samples were not suitable for handheld analysis due to insufficient amounts/ concentrations of sample presented to the handheld Raman instrument and, therefore were not included in the subsequent PCA.

Powders of cutting agent and adulterant reference standards were placed into clear glass vials (Kimble Chase vial screw thread with PTFE cap, China) for Raman analysis. These spectral data were also added to the training sample set. NPS internet products were obtained in the form of either powders, capsules or tablets. Powders and capsules were emptied into glass vials in the same manner as for cutting agents and adulterants prior to Raman analysis. Tablets were crushed using an agate mortar and pestle before the resultant powders were transferred into

glass vials as above. Vials were vortex-mixed (30 sec.) and shaken before collection of each replicate spectrum using a VORTEX-GENIE2 (Scientific industries, Inc., USA).

7.2.2.2. Analysis of NPS reference standards and internet products using handheld Raman Spectroscopy

A handheld ProgenyTM Raman instrument (Rigaku, USA) with laser λ_{ex} of 1064 nm, was used to obtain Raman spectra of the internet products. These spectra were used alongside PCA in an attempt to identify the NPS present in the products *via* projection onto the ‘pure NPS’ and ‘pure NPS plus cutting agents/ adulterants’ PCA models. Instrument specifications are shown in Chapter 2 Table 2.3. Four methods were used to collect Raman spectra based on the nature of the substance being investigated: Method A (2000 ms exposure time; 490 mW laser power; 10 averages), Method B (2000 ms exposure time; 200 mW laser power; 10 averages), Method C (2000 ms exposure time; 100 mW laser power; 10 averages); Method D (2000 - 5 ms exposure time; 50 mW - 1 mW laser power; 10 averages). All NPS reference standards were analysed using method A.

Initially, Method A was used as a default to obtain spectra for all of the internet samples. However, for samples that were burned using Method A or thought to be prone to burning from high laser power (i.e. coloured samples), Methods B and C were used. Method D was developed by iteratively reducing laser power and exposure time in order to collect Raman signals from challenging samples that caused saturation of the detector and/or burned with Method C. Irrespective of the method used, the instrument was calibrated each day immediately before analysis using a benzonitrile reference standard (Rigaku, USA). All samples were analysed directly through glass vials after optimisation of the vial holder attachment with respect to the focal point.

Ten replicate spectra were measured for the training (NPS and cutting agents/ adulterants reference standards), test and ‘OOM’ NPS reference samples. Five replicate spectra were measured for the NPS test internet samples. No spectral data was collected for NPS internet samples P6, 12 and 21 – 24 due to saturation of the detector and sample burning/ degradation. Using alternative Methods B – D did not generate any useful Raman data due to poor S/N and, therefore samples P6, 12 and 21 – 24 were not included into the subsequent PCA.

7.2.2.3. Raman spectral pre-processing

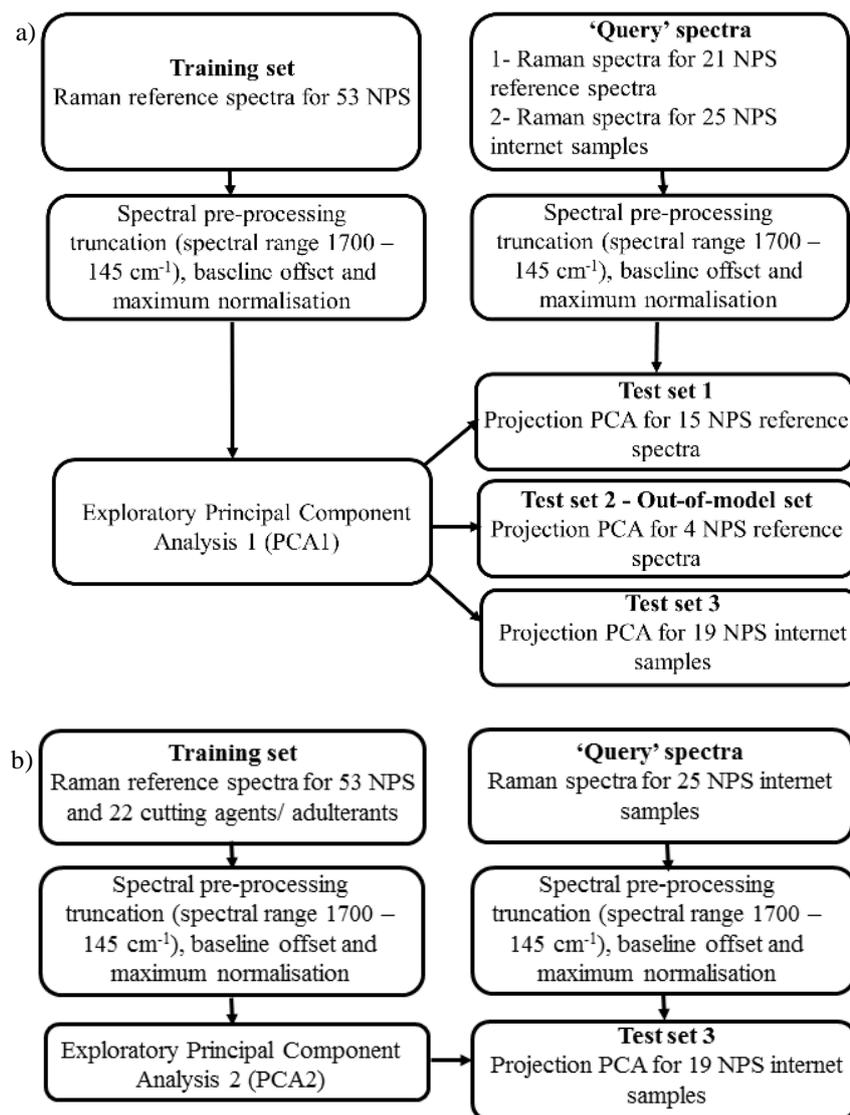
In Chapters 4 and 5, pre-processing methods were developed based on high resolution spectral data generated using a benchtop Raman instrument with a 785 nm laser λ_{ex} . In this Chapter Raman spectra obtained with a 1064 nm laser λ_{ex} were of comparatively poor resolution. This is due, in part, to the use of a less sensitive InGaAs detector and space restriction, where a holographic volume phase grating (VPG) replaced the moving grating of the benchtop instrument, leading to a shorter spectral range of 2355 cm^{-1} (3100 cm^{-1} for the benchtop) and reduced number of data points of 512 (3777 for the benchtop). In addition, the Renishaw laser spot size was 1.2 μm , whereas for the ProgenyTM, it was 20 μm . This difference shows that the Renishaw provided a very high spatial resolution even with a x20 objective lens and possibly no more than a single particle was detected at any time, whereas, using the ProgenyTM, Raman scattering was collected from a greater powder surface area. In light of these differences, suggested pre-processing methods were modified (see method development in Section 7.3.1 of this Chapter). In this Chapter, pre-processing methods include spectral truncation of data points over 1700 - 2500 cm^{-1} , followed by baseline offset and maximum normalisation (previously evaluated in Chapter 4) (Unscrambler® X 10.4).

7.2.2.4. Principal Components Analysis (Unscrambler® X 10.4)

The PCA parameters employed in developing the initial PCA model in Chapter 5 are applied to datasets analysed in this Chapter (See Chapter 5 Section 5.2.5 for full method details).

7.2.3. Prediction via PCA projection

Prediction of the chemical scaffolds of ‘query’ samples was performed by projecting developed PCA models to latent space over the spectral range 1700 – 145 cm^{-1} . PCA Projection was generated using the Unscrambler® X 10.4 software (CAMO PROCESS AS, Oslo, Norway) for three PCs. A schematic outline of the Raman/ chemometrics experimental protocol is described in Figure 7.2.



¹‘Query’ samples constitute the overall number of samples investigated. ‘Test’ samples are part of the ‘query’ samples but they constitute the actual number of samples taken into analysis and projected against the PCA models. The difference constitutes the samples, which did not yield analysable Raman spectra.

Figure 7. 2: Raman and chemometrics experimental protocol: a) three test sets including Raman spectra for 15 NPS ‘test’ reference standards, 4 ‘OOM’ reference standards and 19 NPS internet samples projected onto PCA1 derived from ‘53 representative’ NPS reference standards; b) one test set including Raman spectra for 19 NPS internet samples projected onto PCA2 derived from ‘53 representative’ NPS plus 22 cutting agent/ adulterant reference standards¹.

7.3. Results and Discussion

7.3.1. Handheld Raman method development

As discussed in Chapters 4 and 5, initial inspection of Raman spectra was conducted to evaluate the extent to which artefacts of the instrument influence the quality of replicate spectra and to

establish the necessity for and nature of the pre-processing methods that would need to be used in the analysis of the data.

Using exploratory PCA and visual inspection of Raman spectra, several issues were identified. These included: 1) lower number of variables than data collected from the benchtop instrument, 2) poor S/N, 3) shot/ residual noise, 4) values in the dataset with positive and negative signs and 5) variable absolute Raman intensities.

In general, for the application of MVA in a well-controlled dataset, the number of samples should be much larger than the number of variables. In such cases, the number of degrees of freedom is small. However, this is not the case with Raman spectral data (i.e. > 500 variables in typical handheld Raman instruments). Therefore, feature selection or variable reduction techniques are often required prior to conducting MVA [191]. In Chapters 4 and 5, a number of variable reduction methods were evaluated. In Chapters 5 and 6, analysis was performed over a limited spectral range (i.e. 1750 – 1300 cm^{-1}) and the number of variables or data points was ca. 522. By contrast, using the same spectral range with the handheld instrument, the number of variables was reduced to ca. 90, which was insufficient to reliably calculate the PCA projection using the Unscrambler® X 10.4. Unlike the methods adopted with the benchtop instrument, larger spectral ranges were evaluated for the handheld instrument in order to efficiently increase the number of variables.

Unlike Chapters 5 and 6, the reduction or filtering of high frequency noise was not addressed *via* smoothing methods for two reasons. The first is to preserve small Raman peaks and the second is to prevent a further reduction in the number of variables as discussed above. However, the region 1700 - 2500 cm^{-1} (shot and residual noise) was truncated as it does not contain any spectral data and may impact classification of ‘query’ samples (Chapter 5 Sections 5.3.2.2. and 5.3.2.6.).

In addition, values in the dataset with positive and negative signs may impact the reliability of calculating maximum normalisation using the Unscrambler® X 10.4, where a single sign should be used (Chapter 4). Finally, variable absolute Raman intensities have been shown to impact sample classification and consequently, normalisation of spectra was essential to scale the spectral data and ensure an even distribution of the variance between samples (see Chapter 4). Therefore, pre-processing methods including truncating the spectral range 2500 – 1700 cm^{-1} (region without any spectral data), baseline offset method to remove negative data points

and maximum normalisation were suggested for investigation (Unscrambler® X 10.4). It is expected that raised baselines observed in Chapters 5 and 6 are reduced with a longer laser λ_{ex} (i.e. 1064 nm) and, therefore baseline subtraction was not accounted for in this method.

The impact of using a larger spectral range with a reduced number of variables on the accurate identification of ‘query’ samples was investigated. The selected spectral region 1700 – 145 cm^{-1} includes vibrations attributed to diverse functional groups, which may improve the identification of larger molecular moieties, as compared to the benchtop instrument, where a shorter spectral range (i.e. 1750 – 1300 cm^{-1}) was used in the analysis (see Chapters 5 and 6).

The region 1700 – 145 cm^{-1} includes vibrations attributed to carbonyl stretches C=O at 1700 - 1680 cm^{-1} (medium to weak) as in cathinones; olefinic groups C=C at 1750 - 1500 cm^{-1} (strong); amine NH deformation vibrations for amines, amine salts and amide substances at 1660 – 1500 cm^{-1} (weak); aromatic and heteroaromatic rings at 1620 – 1420 cm^{-1} (medium to weak); methyl and methylene deformation vibrations at 1500 – 1300 (weak to medium) such as CH₃ at 1380 cm^{-1} (medium); CH₂ and CH₃ asym at 1470 - 1400 cm^{-1} (medium); and CC related to aromatic ring chain vibrations at 1580 and 1600 cm^{-1} (strong), 1450 and 1500 cm^{-1} (medium); C-F stretches at 1350 – 1000 cm^{-1} (weak to moderate) as in 4F- α -PVP, 5F-APICA, 5F-APINACA, 5F-PB22, afloqualone, AM-2201, flephedrone, flubromazepam, flubromazolam and NM-2201; C-O stretching vibrations at 1300 – 750 cm^{-1} , where, Raman bands with variable intensities for carbonyl compounds occur at 1300 - 1100 cm^{-1} for esters and at 1310 – 980 cm^{-1} for anhydrides; CH wag vibrations of olefinic and acetylenic compounds (weak) at 1000 – 600 cm^{-1} ; aromatic ring vibrations involving in-plane 2,4,6 stretches (very strong) at 1290 – 990 cm^{-1} ; aromatic ring CH wag vibrations (weak) at 900 – 700 cm^{-1} ; halo-alkane stretching involving F, Cl, Br and I (strong) at 850 – 480 cm^{-1} [125] as in 5-IAI, flubromazepam, flubromazolam, ketamine, mebroqualone, N-Me-2C-B, phenazepam and pyrazolam.

It is hypothesised that larger molecular moieties identified may improve classification of pure NPS but may hinder NPS identification in mixtures *via* PCA projection owing to poorer resolution.

7.3.2. Overview of PCA1 generated from ‘53 representative’ NPS reference standards

Prior to predicting the ‘query’ samples, an overview of the training samples (i.e. ‘53 representative’ pure NPS), was performed using an exploratory PCA (i.e. PCA1) over the

spectral range 1700 – 145 cm^{-1} (Figure 7.3). The cumulative calibrated and validated explained variance for 20 PCs was 82 and 78 % respectively. The first three PCs explained approximately 34 % of the total variance (15/12/7 % for PC1/PC2/PC3). The total % variance explained by the first three PCs was low. This is partly because the dataset is chemically and structurally diverse as previously investigated in Chapter 5 Section 5.3.3.1. In addition, the relatively poor Raman scattering efficiency of the 1064 nm laser λ_{ex} (Raman scattering is proportional to the fourth power of the frequency of the laser source [120]) and sample presentation for handheld analysis, which has eliminated variations between replicate reference spectra when compared to the benchtop instrument.

In Figure 7.3, the training samples were classified into previously described superclusters (Chapter 5). The data matrix consisted of 530 rows and 512 variables. The PCA scores plot for PC1/PC2 showed that replicate spectra of individual NPS grouped together, yet were discriminated from different NPS, demonstrating the efficiency of the pre-processing method in removing artefacts not pertinent to the investigated substances. For example, the spectra of five reference NPS (i.e. 4-HO-DET, 5-MeO-DALT, 5-MeO-MiPT, FDU-PB-22 and NM-2201), all members of supercluster 1 (red boxes), were clustered into five distinct groups in the upper quadrants of the scores plot for PC1/PC2. Each group represented replicate spectra of individual NPS reference standard. Both FDU-PB-22 and NM-2201 (upper left quadrant) are in very close proximity to each other, which is not unexpected since both are structurally similar synthetic cannabinoids, yet do not show any overlap between their replicate spectra upon zooming in, supporting the fact that the model is sufficiently discriminatory between structurally similar NPS (Figure 7.3).

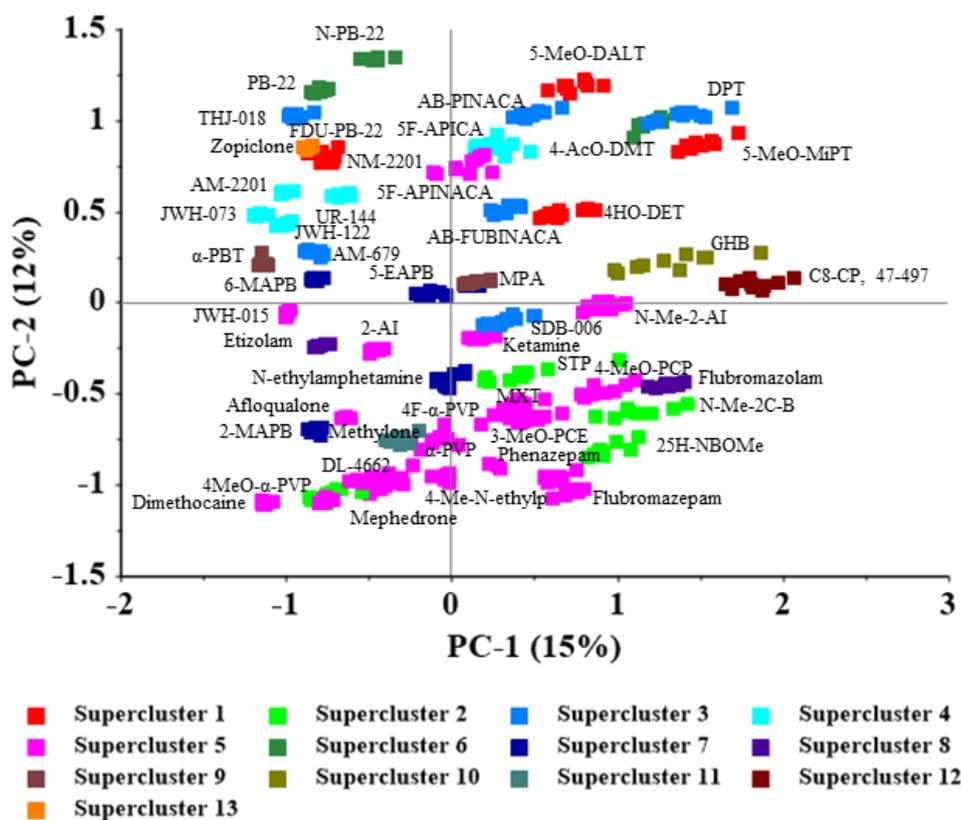


Figure 7. 3: 2D-PCA scores plot (PC1/PC2) illustrating the classification of ‘53 representative’ NPS reference standards classified in previously described superclusters (Chapter 5). Raman spectra were collected using Progeny™ (Rigaku, USA).

Similar pure NPS classification trends were observed as previously described in Chapter 5 Section 5.3.3.1. using the benchtop instrument. In PCA1 (Figure 7.3), results showed that the scores plot displayed grouping of NPS in discrete regions of the PCA 2D-scores plot, mostly according to their EMCDDA/EDND classes. For example, for PC1/PC2, the tryptamines (4-AcO-DMT, 4HO-DET, 5-MeO-DALT, 5-MeO-MiPT and DPT) grouped together in the upper right quadrant, whereas the phenethylamines (25H-NBOMe, N-Me-2C-B and STP) grouped in the lower right quadrant. Moreover, the cathinones (4F- α -PVP, 4-MeO- α -PVP, α -PVP, 4-Me-N-ethylnorpentadone, DL-4662, mephedrone and methylone) grouped in the lower half of the scores plot, whereas the cannabinoids (5F-APICA, 5F-APINACA, AB-FUBINACA, AM-679, AM-2201, JWH-073, JWH-122, NM-2201, N-PB-22, PB-22 and THJ-018) grouped in the upper half of the scores plot. PCA1 results (PC1/PC2) showed that etizolam was delineated from the benzodiazepines flubromazolam and pyrazolam (supercluster 8), whilst, C8-CP, 47-497 was delineated from remaining cannabinoids, as observed and previously discussed in Chapters 5 and 6. This illustrates the potential efficiency and discriminatory power of the developed model in classifying NPS based on structural diversity using handheld Raman

spectroscopy ($\lambda_{\text{ex}} = 1064 \text{ nm}$) and chemometrics (Figure 7.3). Consequently, these methods were applied to subsequent PCAs.

7.3.3. Prediction of ‘test’ and ‘out-of-model (OOM)’ pure reference samples *via* PCA projection onto the ‘NPS reference standards/handheld’ model (test sets 1 and 2)

In this Section, prediction of ‘test’ and ‘OOM’ NPS reference standards was performed by projecting the Raman spectra of 15 ‘test’ and 4 ‘OOM’ pure NPS onto PCA1 generated from ‘53 representative’ NPS reference standards. Similarity between test sets 1 and 2 (i.e. 15 ‘test’ and 4 ‘OOM’ pure NPS respectively) and library/ training substances (i.e. ‘53 representative’ pure NPS) was evaluated using PCA projection plots (Unscrambler® X 10.4) (Figure 7.4).

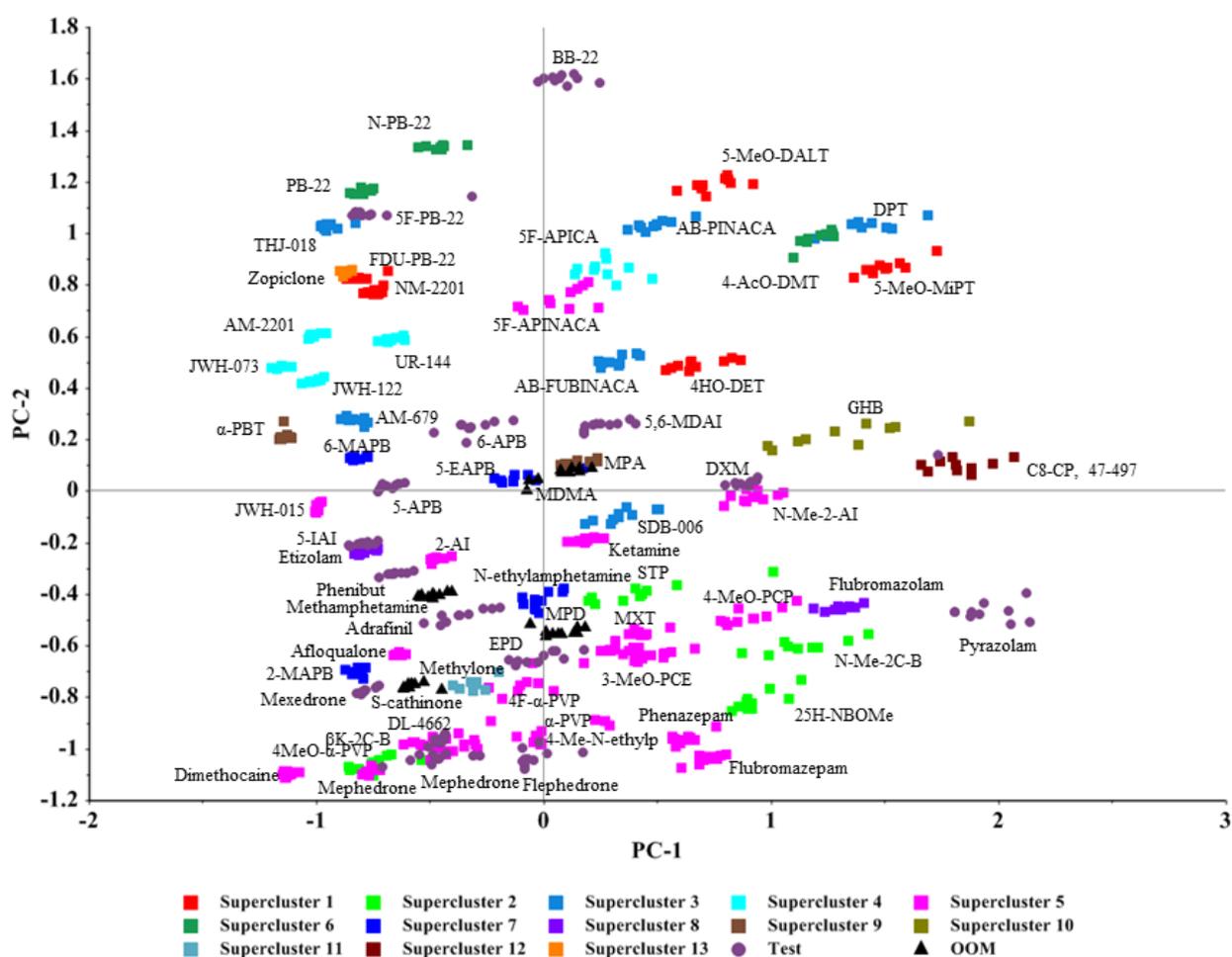


Figure 7. 4: 2D-PCA projection plot (PC1/PC2) illustrating training, ‘test’ and ‘OOM’ sets. Test sets 1 and 2 (i.e. ‘test’ and ‘OOM’ reference spectra) are projected onto the PCA model developed with 53 training NPS, classified in previously described superclusters (Chapter 5). Raman spectra were collected using Progeny™ (Rigaku, USA).

When ‘test’ and ‘OOM’ samples were individually projected onto PCA1, the variance explained for ‘test’ and ‘OOM’ samples was 12/12/4 and 4/7/2 % for PC1/PC2/PC3

respectively (Figures 7.5a and 7.5b). The total explained variance by the PCA projection model for PC1/PC2/PC3 described how well the variance in the ‘query’ samples is explained by the calibration model.

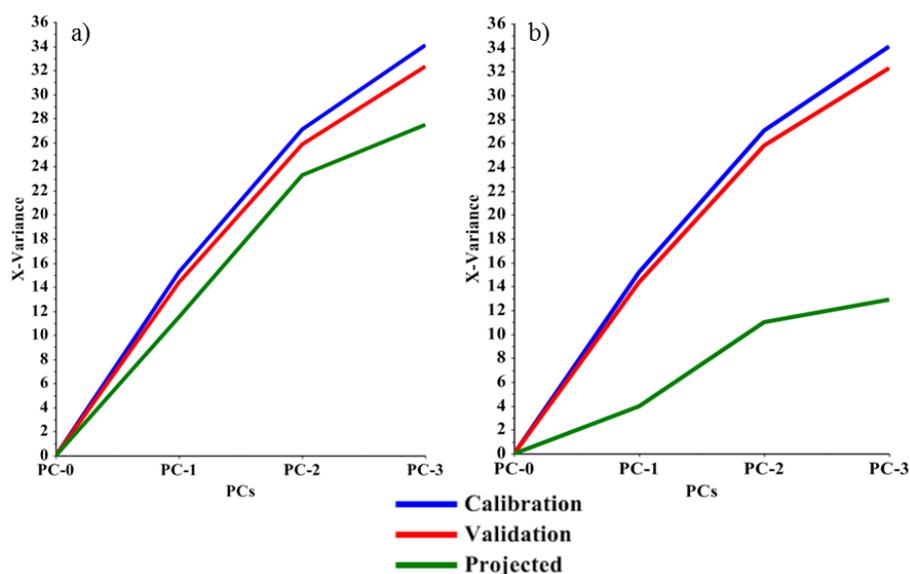


Figure 7. 5: PCA projection explained variance plots for: a) ‘test’ (test set 1) and b) ‘OOM’ (test set 2) sets. This is the variance explained by the projection of ‘test’ and ‘OOM’ samples onto PCA1.

The variance explained by PCA projection was reduced to 10/11/4 % for PC1/PC2/PC3 when all ‘test’ and ‘OOM’ reference samples were projected together onto PCA1. Poorer projection explained variance for ‘OOM’ samples was expected since they were not included in the initial model of 478 NPS, used to produce the ‘representative’ database through ‘hierarchical cluster analysis’ [204]. Hence they have reduced the model’s efficiency in prediction *via* PCA projection (Chapter 5).

PCA projection results for both test sets 1 and 2 were relatively poor ($\leq 25\%$) for the first three PCs owing to the complexity and chemical diversity of the dataset and the reduced number of variables owing to relatively poor resolution of the instrument. This also demonstrates the limitations of the calibration model PCA1 in predicting the ‘query’ samples. However, the PCA projection explained variance plot (Figure 7.5a) shows relatively close calibration and validation values between training and test reference samples for three PCs demonstrating the potential of successfully suggesting the chemical scaffold or EMCDDA/EDND class of a pure ‘unknown’ sample, previously part of the 478 NPS employed to develop the initial model (i.e. the ‘53 representative’ NPS) (Chapter 5). Unlike random validation, the PCA method was designed to implement a full validation to enhance model robustness. In contrast to ‘test’

samples, PCA projection values for ‘OOM’ samples were relatively poor for three PCs suggesting sub-optimal potential for the successful classification of an ‘unknown’ newly emerging to the market (Figure 7.5b), yet a definitive conclusion cannot be made due to the limited number of ‘OOM’ samples tested (i.e. four) and also because most of the ‘OOM’ samples are known drugs of abuse (i.e. MDMA, MPD and methamphetamine) and, therefore may not fully represent newly emerging NPS.

All training and projected samples were included in the Hotelling T^2 ellipse at 95 % CL, except a few pyrazolam replicate spectra (ellipse not shown in Figure 7.4). However, the examination of these spectra using the influence plot (plot not shown) demonstrated that they are not true outliers (i.e. not associated with high F-residuals and Hotelling T^2 values (see Chapter 4)), hence are described by the calibration model.

In this analysis a successful grouping/ overlap or delineation between scores was confirmed *via* direct comparison to the results obtained in Chapter 5 (Figure 7.6). Grouping/ overlap or delineation of both ‘test’ and ‘OOM’ samples was evaluated for each of the previously described superclusters (Chapter 5).

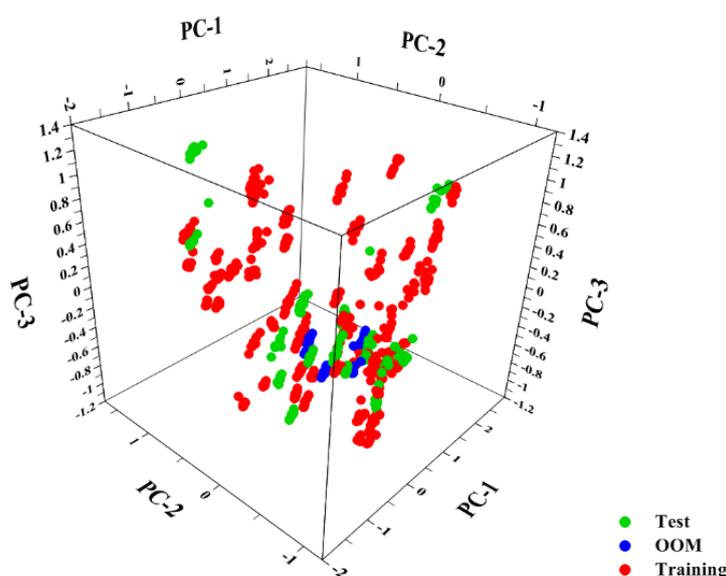


Figure 7. 6: 3D-PCA projection plot illustrating training, ‘test’ and ‘OOM’ sets. Test sets 1 and 2 (i.e. ‘test’ (green) and ‘OOM’ (blue) spectra) are projected on the PCA model developed with ‘53 representative’ training NPS (n = 10/ NPS (red)).

Superclusters 1, 2, 4, 9, 10 and 13 were not challenged with any ‘test’ or ‘OOM’ samples due to high costs and lack of availability of reference standards when this study was conducted. This is one of the limitations of the dataset.

For supercluster 2, ‘test’ samples 5-APB and 6-APB grouped in close proximity to other arylalkylamines (i.e. 5-EAPB and MPA) on PC1/PC2 (Figure 7.4). The ‘test’ sample β k-2C-B (cluster 2.06) was not aligned with N-Me-2C-B (cluster 2.06) as expected, possibly because of the very poor S/N in N-Me-2C-B replicate spectra obtained with the handheld instrument. However, β k-2C-B grouped in close proximity to the cathinones (modified structures of phenethylamines) 4-MeO- α -PVP, DL-4662 and mephedrone. On PC1/PC3, β k-2C-B grouped in close proximity to the phenethylamine methamphetamine (plot not shown). Supercluster 2 was challenged with the ‘OOM’ sample MPD which, unsurprisingly grouped with EPD, another piperidine (Figure 7.4). The replicate spectra for MPD clustered together in the same plot region as all the cathinones. Both cathinones and piperidines are modified structures of phenethylamines. It is important to note that EPD was not one of the training samples, but one of the ‘test’ samples included in the analysis to challenge supercluster 5 (Figure 7.4). For supercluster 2, results were similar to that obtained with a high resolution benchtop instrument. ‘Test’ and ‘OOM’ spectra correlated with structurally similar members of their EMCDDA/EDND class or near-neighbour. In the case of ‘OOM’ samples, where a member of the same EMCDDA/EDND class was not part of the training set, ‘OOM’ samples were aligned with structurally similar NPS.

In contrast to supercluster 2, where ‘test’ and ‘OOM’ samples grouped in proximity to their EMCDDA/EDND class analogues, ‘test’ and ‘OOM’ samples employed to challenge supercluster 5, grouped in close proximity to training, other test or ‘OOM’ samples, which are members of the same supercluster/ cluster or class. For example, ‘test’ samples adrafinil (classified as ‘other’) and phenibut (classified as ‘other’) and ‘OOM’ sample MPD, all members of cluster 5.14 (Chapter 5), were grouped in the same region of the 2D-scores PCA projection plot (Figure 7.4). The ‘test’ sample phenibut was in close proximity to adrafinil (cluster 5.14), as explained above, but was closer to 2-AI and methamphetamine, both members of cluster 5.15.

In addition, ‘test’ cathinone samples (i.e. flephedrone, mephedrone and mexedrone) and ‘OOM’ sample S-cathinone grouped in the same plot region as all other cathinones irrespective of their cluster membership. The latter ‘test’ and ‘OOM’ cathinones are all members of cluster 5.18.

The ‘OOM’ sample MDMA grouped in close proximity to the arylalkylamines 5-EAPB and MPA and the aminoindane 5,6-MDAI. Unlike results obtained in Chapter 5 Section 5.3.3.2.,

the examination of the average Raman spectra in conjunction with the line loading plots for PC1/PC2 demonstrated that grouping of MDMA in close proximity to the arylalkylamines may be considered as a misclassification. This demonstrated the challenge evidenced by the ‘OOM’ samples (Figure 7.7). In contrast, 5,6-MDAI and MDMA were in the same plot region possibly because the former is a ‘test’ sample and the latter is an ‘OOM’ samples, both anticipated members of supercluster 11 with a common methylene dioxide group. However, the grouping of 5,6-MDAI was expected to occur with other aminoindane class analogues from the training set (e.g. 2-AI) or alternatively with samples with structural similarity, which are training members of the same supercluster (e.g. methylone (supercluster 11)). The reason for this grouping is not fully understood. However, it could possibly be attributed to the peak at ca. 1679 cm^{-1} , which is present in methylone but absent in both MDMA and 5,6-MDAI Raman spectra (Figure 5.23). In such cases, newly emerging NPS may group together due to structural similarity amongst themselves without an indication to their identity from the training set.

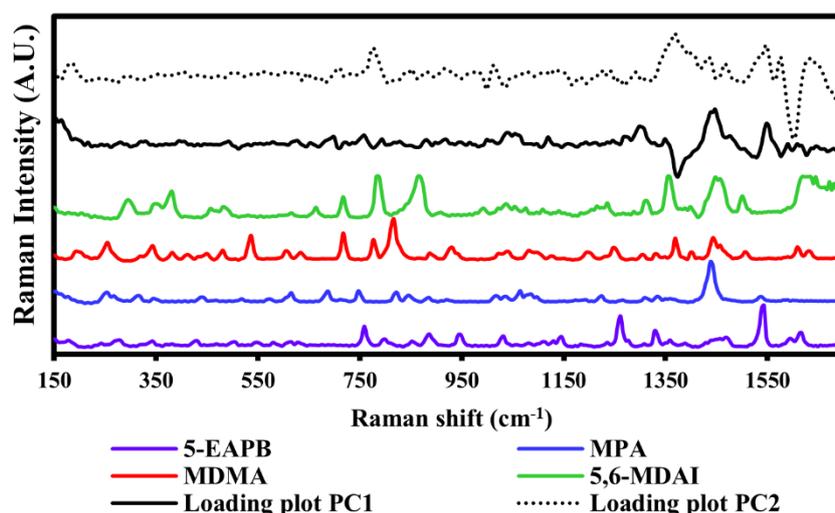


Figure 7. 7: Average reference Raman spectra for 5-EAPB, MPA, MDMA and line loading plots for PC1 and PC2.

Finally, the ‘OOM’ sample methamphetamine (cluster 7.04) grouped with its class analogue N-ethylamphetamine (cluster 7.04) (Figure 7.4). Based on the interpretation provided for supercluster 5, ‘test’ samples correlated with members of the expected supercluster, cluster or class analogue. In contrast, ‘OOM’ samples correlated with class analogues or members of the anticipated cluster when a class analogue was a member of the training set. However, where a class analogue was not included in the training set, ‘OOM’ samples were aligned with structurally similar training NPS.

For supercluster 6, the ‘test’ sample 5F-PB-22 (cluster 6.02) grouped with PB-22 (cluster 6.02) and THJ-018 (cluster 3.04), another cannabinoid. It was also in proximity to N-PB-22 (cluster 6.01) (Figure 7.4). Again, the ‘test’ sample correlated with class analogues and members of the supercluster and cluster.

For supercluster 8, the ‘test’ sample pyrazolam was in close proximity to the training sample flubromazolam. Both are fluorinated benzodiazepines, members of cluster 8.02. Both were delineated from etizolam (observation previously described in Chapter 5). In this case, the ‘test’ sample correlated with a class analogue, a member of the same supercluster and cluster.

For supercluster 12, the ‘test’ sample DXM was in the same plot region as the training sample C8-CP, 47-497. In contrast, the test sample BB-22 was delineated from the training sample C8-CP, 47-497 (observation previously described in Chapter 5). This is also a challenging case because supercluster 12 is composed of one cluster only, with a simple moiety ‘cyclohexane’ as a common substructure, which may, in turn, misalign/ misclassify newly emerging NPS. In addition, the ‘test’ sample DXM is classified as ‘other’ according to the EMCDDA/EDND classification and, hence it is challenging to investigate whether it correlate with other class analogues.

In conclusion, for both ‘test’ and ‘OOM’ samples, on PC1/PC2, nine samples grouped with their EMCDDA/EDND class analogues, five grouped with members of their expected supercluster/ cluster and three clustered with both EMCDDA/EDND class analogues and members of their supercluster/ cluster. Two ‘test’ samples 5-IAI and β k-2C-B deviated from both the EMCDDA/EDND or supercluster/ cluster classifications and grouped with etizolam and the cathinones respectively. However, β k-2C-B correlated with its class analogues on PC1/PC3. In summary, 89 % of both test and ‘OOM’ correlated with their EMCDDA/EDND classes and/ or superclusters/ clusters, demonstrating the potential of the model to classify/ align new pure samples.

7.3.4. Prediction of test NPS internet samples (test set 3) via PCA projection onto PCA1

This model was not previously described with the benchtop instrument. However, it was applied in this Chapter as a preliminary step due to the complexity of data interpretation. This is because suggestion of the chemical scaffolds of NPS in internet samples is more challenging using handheld Raman instruments. In addition, this PCA projection study provides valuable

insight into the content of the training set (i.e. NPS only or NPS and non-NPS). Unlike the benchtop Raman instrument, which is equipped with confocal microscope lenses allowing the irradiation of micron size samples (spot radius = 1.2 μm) enabling the identification of mixture components, the handheld instrument has poor selectivity through irradiation of a larger sample size (spot radius = 20 μm), making the identification of mixture components more challenging. Furthermore, since the Raman scattering is proportional to the fourth power of the frequency of the laser sources, the S/N in Raman spectra analysed with the benchtop instrument ($\lambda_{\text{ex}} = 785 \text{ nm}$) is greater than that of the handheld instrument ($\lambda_{\text{ex}} = 1064 \text{ nm}$). The impact of these challenges/ limitations were evaluated through PCA projection of NPS internet samples onto PCA1 developed with '53 representative' reference NPS training samples (Figure 7.2a) and PCA2 developed with 75 training samples (i.e. '53 representative' NPS and 22 cutting agents/ adulterants reference samples) (Figure 7.2b).

The combined analysis of 25 NPS internet products (Chapter 6, Table 6.2) showed the complexity of the internet mixtures. Overlap/ closeness of projections of test samples (i.e. 19 NPS internet products) to training samples (i.e. 53 NPS reference standards) were evaluated using PCA projection plots (Figure 7.8) and direct comparison to results obtained in Chapters 5 and 6. In this Section the aim of the analysis was to evaluate the transferability of methods onto a handheld platform for the identification of claimed NPS in NPS products. This was achieved by projecting Raman spectra of 19 NPS internet products onto PCA1 derived from the Raman spectra of '53 representative' NPS reference standards (Figure 7.2a).

The cumulative calibrated and validated explained variance by PCA1 for 20 PCs was 82 and 78 % respectively. The first three PCs explained approximately 34 % of the total explained variance, (15/12/7 % for PC1/PC2/PC3). The variance explained by PCA projection of internet samples was 22/2/7 % for PC1/PC2/PC3 (Figure 7.8a and 7.8b).

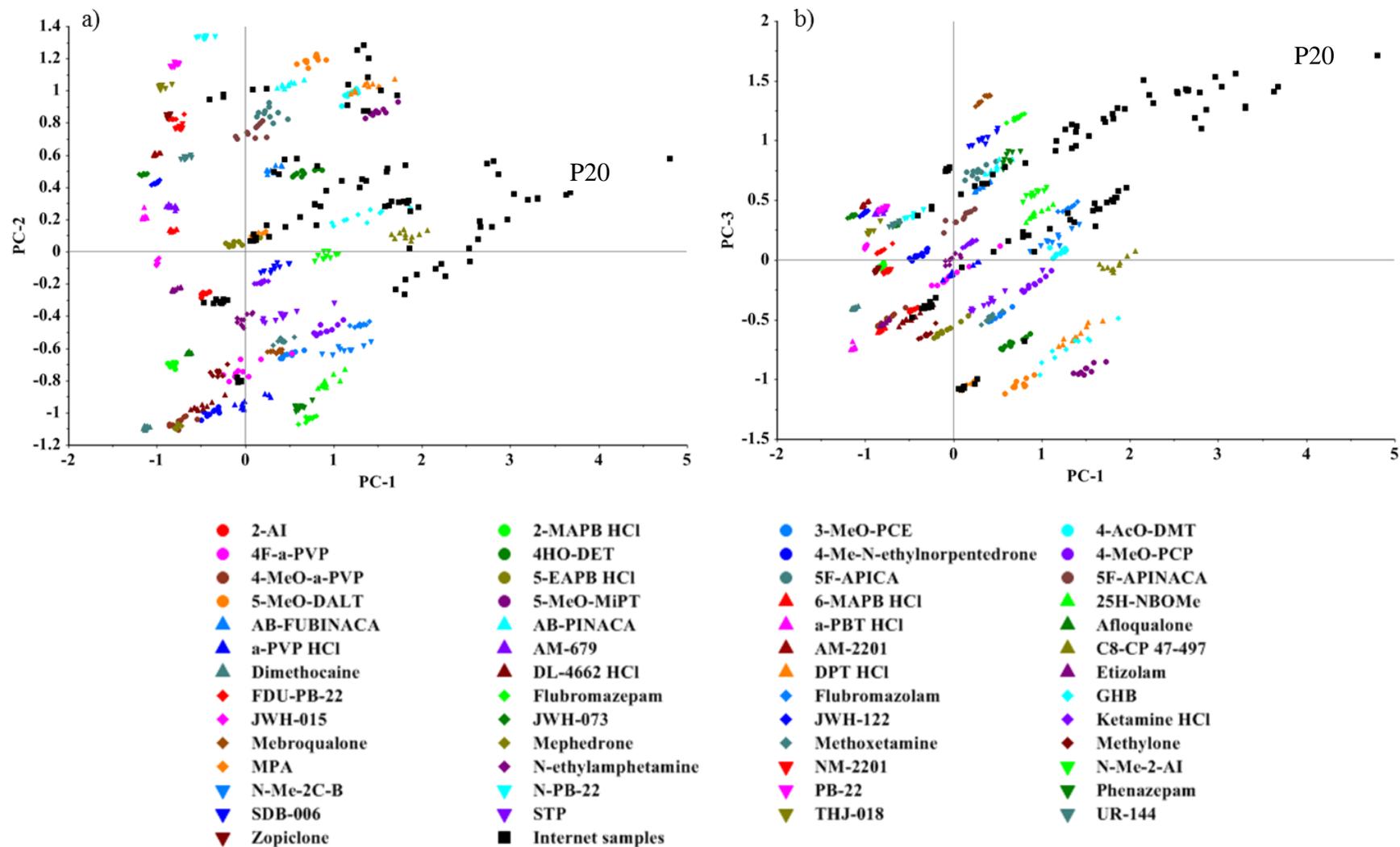


Figure 7. 8: a) 2D-projection PCA scores plot for PC1/PC2; b) 2D-projection PCA scores plot for PC1/PC3. a and b are pre-processed Raman spectra of 19 NPS internet products (test set 3) projected onto PCA1 generated from training samples ('53 representative' NPS reference standards).

PCA projection of 19 NPS internet samples (19 NPS internet samples x 5 measurements) showed relatively poor projection values ca. 31 % for PC1/PC2/PC3 (< 100 %) (Figure 7.9). This demonstrated that the ‘query’ samples were not well described by the computed PCs of the original calibration model. This could be due to the high complexity of the ‘query’ samples, the limitations of the technique and the imbalance of the data matrix (number of samples versus number of variables) [191].

The PCA projection was computed for three PCs only and the line plots (Figure 7.9) showed that the projection line (green) is getting closer to both calibration and validation lines and is rising upwards. This demonstrated the potential improvement of the projection model possibly if more PCs were computed. However, a greater number of PCs was not computed because the explained variance with the first three PCs was very low and, therefore computing more PCs carries the risk of incorporating explained variance due to residual noise. This was in fact demonstrated by lower projection values than calibration and validation values on PCs 2 and 3. Furthermore, the use of a longer wavelength ($\lambda_{\text{ex}} = 1064 \text{ nm}$) has significantly reduced fluorescent backgrounds exhibited by NPS internet samples. However, raised baselines were still observed for some samples such as P11, 13-15 and 20, which accounted for most of the variance described by PC1 (Figure 7.8a and 7.8b). The baseline offset function using the Unscrambler® software does not bring the baseline down to zero i.e. does not subtract raised baselines or correct for fluorescent backgrounds and, therefore pre-processing methods employed in this Chapter may need to be optimised by evaluating a baseline subtraction step.

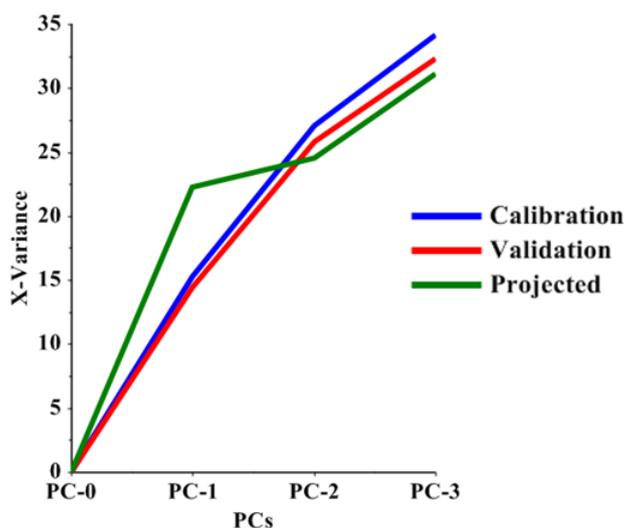


Figure 7. 9: PCA projection explained variance plot. This plot shows the calibration (blue) and validation (red) plots of the training samples (i.e. ‘53 representative’ NPS) and the projection plot (green) of projected ‘query’ samples (i.e. 19 NPS internet products).

As discussed in Chapter 6, PCA projection analysis of NPS internet products was presented such that the products were grouped into three levels of complexity: 1) a group for pure products, where purity and identification were confirmed using HPLC and GC-MS respectively, 2) a group, where the NPS content in the sample was not included in the training set and 3) a group for complex brands and formulations (i.e. tablets and pellets). All training and projected samples were included in the Hotelling T^2 ellipse at 95 % CL, except the replicate spectra for P11, 13-15 and 20 (ellipse not shown in Figure 7.8). The examination of these spectra using the influence plot (plot not shown) demonstrated that they are not well described by the calibration model (high Hotelling T^2) and high F-residuals indicating a high sample distance from the calibration model. These results were expected for P11 and 13-15 (coloured pellets with purity < 8 % by HPLC) and poor quality of Raman spectra (raised baselines and very poor S/N). However, these results were not fully understood for P20, which exhibited raised baselines and explained most of the variance along PC1. This is because it was relatively pure by HPLC (ca. 77 %) and the Raman spectra displayed defined peaks, which were consistent with the reference spectrum for N-Me-2-AI (Figure 7.10).

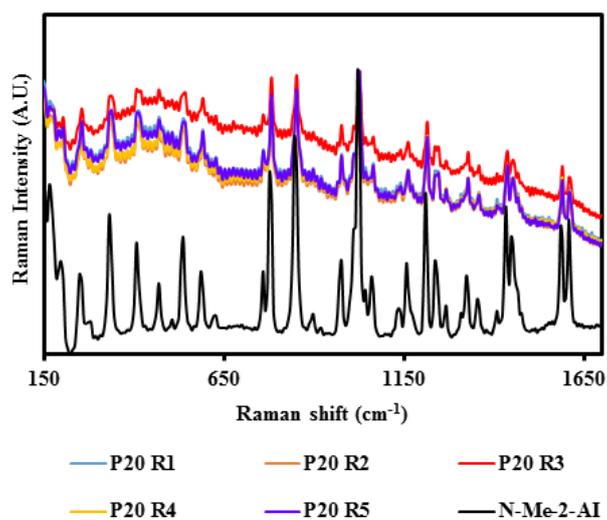


Figure 7. 10: Normalised replicate Raman spectra for P20 and the average reference Raman spectrum for N-Me-2-AI.

For group 1, results for PC1/PC2/PC3 (Figure 7.8) have shown that relatively pure samples (i.e. P1-3, 18 and 19) grouped in the same plot region with the NPS reference standard that correspond to the main NPS identified in the mixture (Chapter 6 Table 6.2). Both P1 and P2 grouped in close proximity to the 2-AI reference spectra. Similarly, P3 grouped with the AB-FUBINACA reference spectra, whereas P18 and P19 grouped with the MPA reference spectra. In such cases, the model developed using a ‘representative’ database and chemometrics in conjunction with handheld Raman spectroscopy equipped with a laser λ_{ex} of 1064 nm has

successfully aligned the main NPS contained in internet mixtures with their corresponding NPS reference spectra as anticipated. However, both P18 and P19 were also in close proximity to 5-EAPB, another arylalkylamine, showing an example of alignment with respect to their EMCDDA/EDND class analogue.

In contrast, P4, 5, 10 and 25 grouped with members of their expected superclusters. For example, both P4 and P5, ca. 50 % pure by HPLC, known to contain AB-PINACA (cluster 3.02), grouped in close proximity to the tryptamines including DPT (cluster 3.08). Furthermore, P10, a relatively pure sample by HPLC (> 95 %), known to contain the benzodiazepine flubromazepam (cluster 5.06) grouped in the same plot region as the cathinone 4F- α -PVP (cluster 5.03). P25, a relatively pure sample by HPLC (ca. 100 %), known to contain the synthetic cannabinoid THJ-018 (cluster 3.04) grouped in proximity to the synthetic cannabinoid AB-PINACA (cluster 3.02). In these cases, NPS internet products were aligned with members of their expected superclusters.

For group 2, results for PC1/PC2/PC3 (Figure 7.8) described the classification of NPS internet samples containing NPS not present in the training set. For example, the relatively pure samples P16 and P17 (\geq 94 % pure by HPLC), known to contain 5,6-MDAI, grouped in proximity to GHB. There is no structural similarity between these molecules, hence results represent a potential misclassification as suggested from the analysis of the line plots (Figure 7.11). Challenges faced in this case were two-fold. Firstly, there was no reference spectrum for 5,6-MDAI in the training set. Secondly, 5,6-MDAI was expected to cluster with members of supercluster 11, which is composed of one cluster only, with benzo[d][1,3]dioxole as a common substructure. These challenges may potentiate false positive alignment/ classification of new or 'unknown' samples.

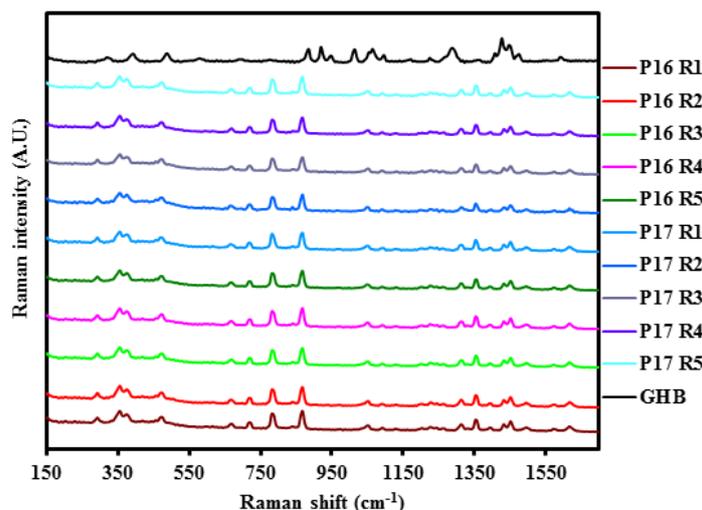


Figure 7. 11: Replicate Raman spectra for P16 and P17 and the average reference spectrum for GHB.

For group 3, results for PC1/PC2/PC3 (Figure 7.8) have shown that highly impure (< 10 % purity by HPLC), branded and formulated samples were scattered in the PCA projection scores plots and were delineated from most training samples. These include P7-9, 11 and 13-15. As described above, P11 and P13-15 were not well described by the calibration model. This could possibly be because of poor quality spectra with a handheld instruments and 1064 nm λ_{ex} (i.e. raised baselines, poor S/N and poor resolution). However, a few replicate spectra for P7 grouped in proximity to the flubromazolam reference spectra. Due to the low purity of these samples, each measurement may represent a different constituent in the mixtures. P7 is an example, where only a few spectra suggested the presence of an NPS in the mixture. This demonstrated that numerous replicate measurements are often needed in ‘street-like’ samples. In this study, five replicate measurements were made and a greater number of measurements e.g. 10 could possibly provide insight into further details regarding other constituents in NPS mixtures. A compromise is often needed between the numbers of measurements and enhanced rapid identification in the field.

In conclusion, based on the PCA projection of 19 NPS internet samples onto PCA1 derived from ‘53 representative’ NPS reference standards, 53 % (10/19 samples) of NPS internet ‘test’ samples were classified according to their EMCDDA/EDND classification and/or assigned to their anticipated superclusters. These include relatively pure samples. Challenges include the identification/ classification of NPS not present in the initial training set, highly impure, formulated and branded samples with very low amounts of NPS and high amounts of cutting

agents/ adulterants. Optimisation of pre-processing methods may improve the identification/ classification of NPS displaying raised baselines and poor S/N with the 1064 nm laser λ_{ex} .

7.3.5. Prediction of test NPS internet samples (test set 3) via PCA projection onto PCA2

In this Section, the Raman spectra of 19 NPS internet samples were projected onto a PCA model derived from the Raman spectra of NPS and cutting agents/ adulterants. Overlap/ closeness of projected test samples (i.e. 19 NPS internet products) to training samples (i.e. 53 NPS reference standards, 14 cutting agents and eight adulterants) was evaluated using PCA projection plots (Figure 7.12) and direct comparison to results obtained in Chapter 6. The aim of this analysis was to evaluate the transferability of methods developed using a benchtop instrument onto a handheld platform for the identification of claimed NPS in NPS products from a database containing ‘53 representative’ NPS and non-NPS reference standards.

The cumulative calibrated and validated explained variance by PCA1 (Figure 7.2a) for 20 PCs was 76 and 71 % respectively. The first three PCs explained approximately 33 % of the total variance (18/8/7 % for PC1/PC2/PC3). PCA projection of internet samples was 55 % (48/2/5 % for PC1/PC2/PC3) demonstrating improved performance of the model as compared to Section 7.3.4. (Figures 7.12a and 7.12b).

All training and projected samples were included in the Hotelling T^2 ellipse at 95 % CL, except the replicate spectra for P11, 13-15 and 20 (ellipse not shown in Figure 7.12). The examination of these spectra using the influence plot (plot not shown) demonstrated that they are not true outliers, except P20 R3, possibly due to raised baseline and poor S/N.

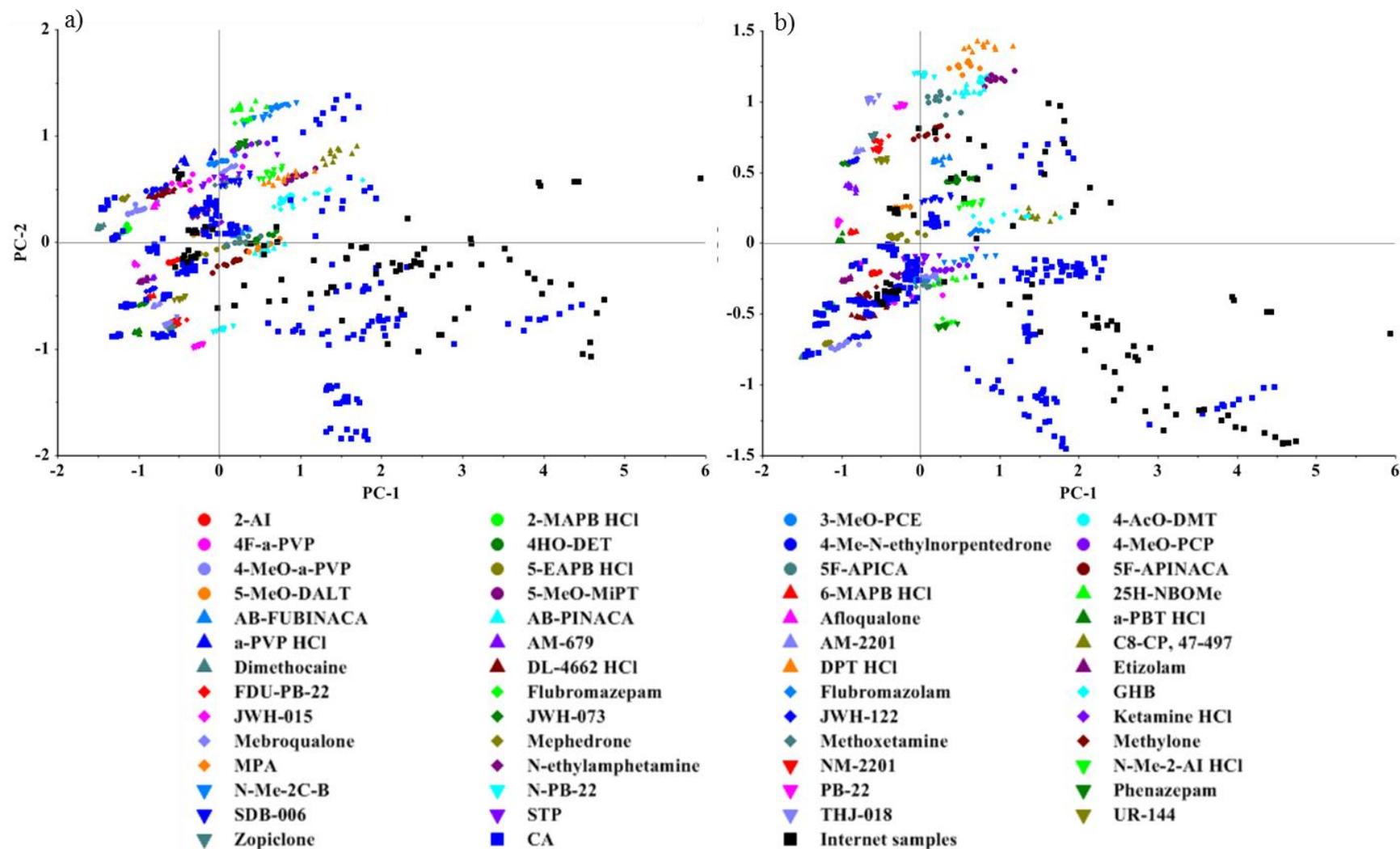


Figure 7. 12: a) 2D-projection PCA scores plot for PC1/PC2; b) 2D-projection PCA scores plot for PC1/PC3. a and b are pre-processed Raman spectra of 19 NPS internet products projected onto PCA2 generated from training samples ('53 representative' NPS and 22 cutting agents/adulterants reference standards).

PCA projection of 19 NPS internet samples (19 NPS internet samples x 5 measurements) showed relatively poor projection values ca. 54 % for PC1/PC2/PC3 (< 100 %) (Figure 7.12). This demonstrates that the ‘query’ samples are still not well described by the computed PCs of the original calibration model. However, there is an improvement achieved by increasing the diversity of the training set. It is postulated that this is due to data overfitting since the projection values for PC1 are significantly greater than both calibrated and validated values for PC1 (48 versus 17 % respectively). This was demonstrated through the projection explained variance plot, where the projection line (green) is diverging away from both calibration and validation lines (Figure 7.13). All lines are rising upwards, suggesting the potential improvement of the projection model if more PCs were computed. However, this may not improve the model since the projection line is diverging away from both calibration and validation lines.

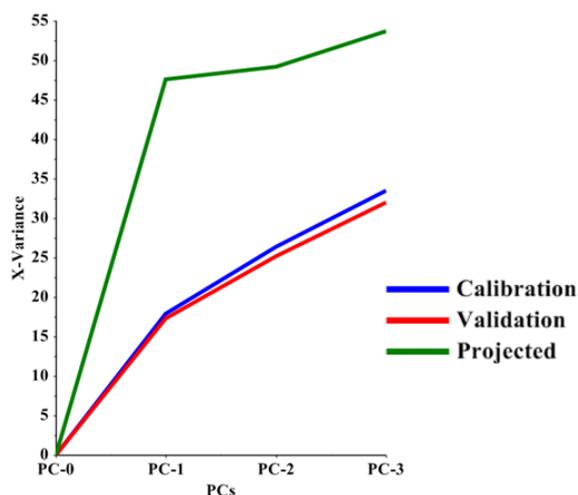


Figure 7. 13: PCA projection explained variance plot. This plot shows the calibration (blue) and validation (red) plots of the training samples (i.e. ‘53 representative’ NPS, 14 cutting agents and eight adulterants) and the projection plot (green) of projected ‘query’ samples (i.e. 19 NPS internet products).

By examining the 2D-PCA projection plots for PC1/PC2 (Figure 7.12a), the classification of NPS internet samples based on their relative purity was observed. By moving from the right to the left of the scores plot, a pattern can be observed where samples with relative low purity (e.g. P11 and 13-15) are located on the right, whereas samples with relative medium purity (e.g. P18 and 19) and located in the middle and samples with high purity (e.g. P1, 2 and 25) are located on the left of the scores plot. P20 is clearly an outlier to this pattern as discussed above. As discussed in Section 7.3.4., raised baselines and poor S/N for P11, 13-15 and 20 explained most of the variance on PC1 (Figures 7.12a and 7.12b).

Furthermore, the PCA projection plots for PC1/PC2/PC3 showed an overlap between training NPS samples with adulterants such as BEN, CAF, DIL, LID, PRO, PHE, PAR and THEO, and cutting agents such as CRE, L-TYR, TAU and MgS, in a similar pattern as discussed in Chapter 6 Section 6.3.2.3. and Table 6.3. Similarly, results for P1, 2, 7-15, 18-20 and 25 showed patterns akin to those discussed in Section 7.3.4. (Figures 7.12a and 7.12b). However, the relatively pure sample P3 (HPLC analysis 77 ± 3), known to contain AB-FUBINACA (cluster 3.01), grouped with the training sample class analogue AB-PINACA (cluster 3.02) as previously described in Figure 6.12. Furthermore, both P4 and P5, with medium purity (ca. 50 % by HPLC analysis), known to contain AB-PINACA, have clustered in close proximity to SUC (Figure 7.14).

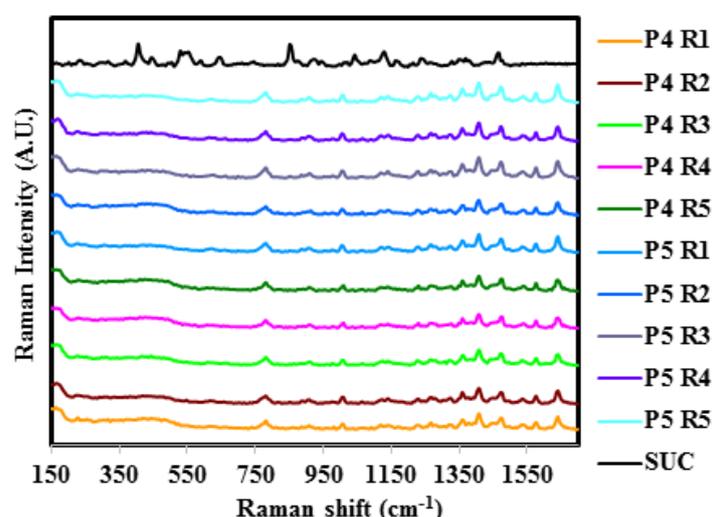


Figure 7. 14: Replicate Raman spectra for P4 and P5 and the average reference spectrum for SUC.

In this study, correct identification/ classification of ‘unknown’ NPS in NPS internet products has slightly declined, suggesting possible data overfitting and sub-optimal pre-processing methods. However, increasing the diversity of the training set by adding cutting agents and adulterants has given a successful indication of NPS sample purity but has contributed to increased interference from cutting agents/ adulterants. Optimisation of pre-processing methods and increasing the number of replicate measurements may refine the identification/ classification of NPS and other constituents in the mixtures and may reduce interference with cutting agents.

7.4. Conclusions

This Chapter provided a proof-of-concept for the transferability of the methods developed using a benchtop Raman instrument to a handheld platform. Results showed the potential of

identifying, suggesting the chemical scaffolds, classifying and/ or assigning ‘unknown’ NPS to their near neighbours, EMCDDA/EDND class analogues or superclusters/ clusters. Results revealed the challenges faced with complex ‘street-like’ NPS mixtures and the limitations of the chemometric techniques used including high dependence on minimum common substructures (MCSs) between training samples used to build the initial ‘hierarchical clustering’ model. Results also illustrated the compromise between reducing fluorescent backgrounds and obtaining adequate S/N, which may, subsequently enable or not the identification of ‘unknown’ samples. Unlike benchtop Raman instruments, where powder regions are carefully selected to optimise the signal with confocal microscopes, using a handheld platform, the quality of spectra are primarily influenced with the instrument features such as poor resolution.

In this Chapter, 89 % of both ‘test’ and ‘OOM’, projected onto a PCA model containing the ‘53 representative’ NPS reference standards, were aligned with their EMCDDA/EDND classes and/ or anticipated superclusters/ clusters. This clearly demonstrated the potential of the ‘NPS reference standards/handheld’ model to classify/ assign previously seen or unseen pure samples to their previously designated classifications.

By contrast, using the ‘NPS reference standards/handheld’ for the identification of internet samples, 53 % of the samples were aligned with their EMCDDA/EDND classes and/ or assigned to their anticipated superclusters. These include the relatively pure samples. Challenges include the identification/ classification of an ‘unknown’ sample containing an NPS (s) not represented in the training set ($n = 2/19$). In these cases, the ‘unknown’ samples were misclassified. Other challenges include the identification/ classification of highly impure branded formulated samples with minute amounts of NPS versus high amounts of cutting agents/ adulterants ($n = 4/19$). In total, five samples (26 %) were not described well by the calibration model and constitute the proportion of rejected samples by the model.

Finally, using the ‘NPS and non-NPS/handheld’ model for the identification of internet samples, only 42 % of NPS internet samples were aligned with members of their superclusters. The calibration model has not described well the highly impure branded formulated samples and interference with cutting agents was observed. This is possibly because of data overfitting, sub-optimal pre-processing methods as well as insufficient number of replicate spectra to represent mixture components. Optimisation of pre-processing methods may refine the identification/ classification of NPS displaying raised baselines and poor S/N with the 1064 nm

laser λ_{ex} . It may also improve the identification of other constituents in the mixtures and may reduce interference with cutting agents. Other chemometric techniques employing different parameters such as different molecular moieties and properties could improve prediction selectivity.

8. Conclusions

The work presented in this thesis serves as a proof-of-concept for the identification of existing or newly emerging NPS. It has demonstrated that the use of Raman spectroscopy in conjunction with a ‘representative’ NPS database and chemometrics has the potential for rapidly and non-destructively classifying ‘unknown’ NPS according to their chemical structure and purity in ‘street-like’ mixtures.

In this thesis, five models were developed, and three of these provided an insight into the identification and classification of NPS depending on their purity. These are: the ‘NPS and non-NPS/benchtop’ model, the ‘NPS reference standards/handheld’ model and the ‘NPS and non-NPS/handheld’ model. In the ‘NPS and non-NPS/benchtop’ model (laser $\lambda_{\text{ex}} = 785 \text{ nm}$), NPS internet samples were projected onto a PCA model derived from a Raman database comprising ‘representative’ NPSs and cutting agent/ adulterant reference standards. This proved the most successful in suggesting the likely chemical scaffolds for NPS present in samples bought from the internet. It does this by grouping the relatively pure internet samples in close proximity to their corresponding reference standards where they exist. If the NPS claimed in the internet samples were not represented in the training set for the model, the test samples were grouped in close proximity to their NPS class or structurally similar analogues (supercluster/ cluster). For branded and formulated products with high impurity profiles, a few of the spectra obtained from each product were grouped next to their corresponding reference standard, or other products previously confirmed to incorporate the NPS in the sample. Of the spectra, 23% displaying high fluorescent backgrounds and very low purity profiles were not classified. It has been demonstrated that this model performed the best of the three investigated for suggesting the chemical scaffolds of NPS internet samples (Chapter 6).

The ‘NPS reference standards/handheld’ model successfully identified 89 % of the ‘query’ NPS reference standards (Chapter 7). In this model, NPS reference standards were projected onto a PCA model derived from a Raman database comprising the ‘representative’ NPS reference standards collected with a handheld instrument ($\lambda_{\text{ex}} = 1064 \text{ nm}$). This model outperformed results obtained with the benchtop instrument (76 %), because of reduced fluorescent backgrounds and, subsequent improved identification and classification of NPS reference standards.

The ‘NPS and non-NPS/handheld’ model successfully classified 95 % (18/19 samples) of the NPS internet samples with respect to their purity on the PCA projection scores plot. In this model, NPS internet samples were projected onto a PCA model derived from a Raman database comprising the ‘representative’ NPS and cutting agent/ adulterant reference standards collected with a handheld instrument ($\lambda_{\text{ex}} = 1064 \text{ nm}$) (Chapter 7).

These models can be considered as useful tools for the early screening and classification of existing and newly emerging NPS. The systematic approach underpinned by the group’s *in-silico* clustering analysis [204], has shown great promise with its ability to detect the similarities and differences between ‘unknown’ samples and references contained in the spectral database based on a) chemical similarity, b) Raman spectral features over designated spectral ranges and c) high loading values along calculated principal components.

The potential of the proposed models for the task at hand was established by successfully classifying/suggesting the chemical scaffolds of NPS reference standards and internet samples using benchtop and handheld Raman platforms with and without the presence of matrix interferences such as adulterants and cutting agents (Chapters 6 and 7). Confirmation is still challenging for internet samples due to numerous interferences from cutting agents and adulterants. Nevertheless, these results are valuable in reducing false negatives since they may serve as a means to propose further assessment of internet samples using other lab-based techniques. The use of spectral subtraction is a possible tool that could be applied to mixtures that may reduce these types of false negatives i.e. it could be used to improve the identification of NPS with low content in the presence of large amounts of excipients with masking Raman signals.

The models developed demonstrated practicability, rapidity, the need for minimal sample manipulation *via* analysis through glass vials, and significant reduction in false negatives/positives subsequent to reduction in fluorescence, all of which are vital requirements for in-field Raman applications with respect to NPS. This project suggested, for the first time, the possibility of identifying NPS or classifying them with their near-neighbours based on a dual mechanism i.e. classifying or aligning ‘unknown’ NPS with members of their anticipated superclusters/ clusters or EMCDDA/EDND class, whether or not they were analogous to existing training NPS used to build the model.

To achieve these results, a number of studies were conducted. Initially, there was a need to identify a suitable technique specifically for the identification of NPS. Thus, a pilot study was conducted [128] to investigate the feasibility in using different handheld techniques (i.e. FTIR, NIR and Raman) for the identification of NPS in simulated binary mixtures and internet mixtures. This study illustrated that handheld Raman outperformed both NIR and FT-IR to give a greater discrimination between NPS and excipients i.e. cutting agents and adulterants in internet products. However, the study revealed that the use of on-board libraries and library matching in-built algorithms was limited in identifying NPS in NPS mixtures due to swamping by fluorescence signals. Therefore, a study was conducted to propose a handheld Raman instrument for the identification of NPS (Chapter 2).

Three different handheld Raman instruments i.e. TruscanTM GP ($\lambda_{\text{ex}} = 785 \text{ nm}$), FirstGuardTM ($\lambda_{\text{ex}} = 1064 \text{ nm}$) and ProgenyTM ($\lambda_{\text{ex}} = 1064 \text{ nm}$) were evaluated for the identification of NPS in more complex simulated mixtures (i.e. binary, ternary and random mixtures) and internet mixtures. Results showed that ProgenyTM operating with a 1064 nm laser source and multiple algorithms was the optimal instrument since it has demonstrated to be more suitable for the identification of NPS in ‘street-like’ NPS products. In this respect, ProgenyTM showed better selectivity and successfully identified 100 % of pure substances, the ‘drug’ constituent in model binary mixtures within a minimum range of 10 – 95 % m/m without any false positive correlations for model binary mixtures, random ternary mixtures and NPS internet products. The ProgenyTM instrument identified the NPS/ adulterant content in 83 % of the NPS internet products in consistence with GC-EI-MS confirmatory analysis (Chapter 2).

This pilot study has highlighted important factors that influence the identification of NPS mixtures using handheld Raman spectroscopy that will now be discussed in turn.

First there are the instrumentation factors related to their specification. A critical one is the laser λ_{ex} (i.e. 785 versus 1064 nm), where an adequate Raman spectrum could not be obtained for a few cutting agents and NPS products using a high energy short laser λ_{ex} of 785 nm, but identification of constituents in NPS internet products was improved with a longer laser λ_{ex} of 1064 nm, possibly because of reduced fluorescence. Also, the design and weight of the instruments made them difficult to hold steady and so present a stable sample for analysis. This may have had a great impact on collecting Raman spectra for samples placed on aluminium plates as opposed to those in glass vials. The content and quality of library spectra had an impact on the number of positive correlations using the instruments’ in-built algorithms.

Instruments which have fixed non-adjustable parameters, designed for non-experts, may cause sample degradation (burning/ melting) which will hinder NPS mixture identification.

Secondly, there are problems related to the in-built algorithms within the handheld instruments. These are: the presence of a threshold for positive correlations; and the impact of poor S/N on correlation values and algorithm calculation which will also affect accurate identification of NPS in mixtures.

Thirdly, there are factors related to the nature of the sample. These are: sample heterogeneity; colour; the number of adulterants in a single mixture; Raman scattering activity of constituents in mixtures; and the presence of fluorescing impurities. All these were identified as important factors that influence the identification of NPS.

As a result of the instrumentation issues, a study was conducted to investigate the impact of the laser λ_{ex} on the identification of a wide range of NPS. This was done using two handheld Raman instruments only differing in the laser λ_{ex} , Xantus-1TM ($\lambda_{\text{ex}} = 785 \text{ nm}$) and FirstGuardTM ($\lambda_{\text{ex}} = 1064 \text{ nm}$) (Chapter 3) [12]. Results demonstrated that a 1064 nm laser source significantly reduced background fluorescence for a wide range of NPS products compared to a 785 nm laser source and subsequently improved identification using an in-built ‘first pass’ matching algorithm in 48 % of the samples. An increase in the laser excitation wavelength from 785 to 1064 nm improved positive NPS identification mainly due to reduced fluorescence, most likely originating from cutting agents and coloured constituents in the products. Correlations between the internet products with the NPS signatures that were confirmed with GC-MS, ranged from 57.0 to 84.1 % using the 1064 nm source, demonstrating that reduced % HQI thresholds may be required when monitoring NPS products in the field. Results also showed that a higher number of false positives and false negative were observed when using the 785 nm source, again resulting mainly from the fluorescent background produced by these samples. False negatives observed for both wavelength sources were also attributable to a variety of factors: low NPS concentration; absence of appropriate reference spectra from the instrument’s library; and the high chemical complexity of the product. Chemically complex samples, such as some of the ‘branded products’, did not correlate to an NPS signature but did show marked improvement in the Raman spectra and characteristic Raman bands upon using the 1064 nm source.

To overcome the limitations of the in-built matching algorithm and to enable the maximum extraction of hidden patterns and chemical information from the Raman multivariate data, further work was needed to optimise spectral pre-processing, with the aim of improving the identification and classification of NPS.

Various pre-processing strategies for Raman data of NPS-related substances were evaluated. An iterative step-wise pre-processing of Raman spectra was carried out to evaluate the contribution of each stage to enhancing the accurate classification of NPS. Based on the analysis, a universal protocol was developed and was recommended prior to conducting exploratory PCA and prediction *via* PCA projection (Chapter 4). The protocol included the visual inspection of Raman spectra, rejection of the spectra with errors, removal of cosmic spikes, interpolation of the spectra if required, examination and evaluation of spectra using descriptive statistics using e.g. the Unscrambler® software, and, finally, the comparison and assessment of the effect of step-wise pre-processing on spectral data using PCA. Interestingly, this study also suggested that PCA of pre-processed Raman data could be employed to identify substance peak identifiers and improve the classification of drug mixtures using a discrete number of Raman bands. This method had the potential of enhancing and optimising samples delineation using PCA. Optimised spectral pre-processing made it possible to explore and visualise the important differences in the dataset based on chemical differences and/ or patterns in the mixtures.

In order to overcome the limitation of the on-board libraries in field instruments and to address the dramatic increase in the numbers of emerging NPS, there has been an interest in developing predictive models that could enable law enforcement agents to classify ‘unknown’ NPS by exploiting chemical information related to existing NPS. Based on the ‘hierarchical clustering analysis’ performed by Zloh et al. [204], a ‘representative’ Raman database was therefore developed and evaluated for its ability to suggest the chemical scaffolds of ‘unknown’ NPS. The projection of pure NPS reference standards against the ‘representative’ Raman database highlighted the success of the clustering analysis performed by Zloh et al. [204] and the robustness of the pre-processing methods developed in this thesis. It also demonstrated the feasibility of predicting the chemical scaffolds of ‘unknown’ NPS by examining proximity on the PCA scores plots to their near-neighbours (Chapters 5 – 7).

Accordingly, a number of models were developed, where ‘query’ NPS reference standards and NPS internet samples were projected onto PCA models derived from training sets including

the ‘representative’ NPS database with and without cutting agents/ adulterants. Raman spectra of ‘query’ samples, which displayed intense fluorescent background were not successfully aligned with their anticipated supercluster/ cluster/ class even after pre-processing. Fluorescent background may have hindered the accurate correlations of pre-processed spectra with the spectra of training sample members from the same supercluster, possibly because of shot noise and poor S/N. Other ‘query’ samples were not successful as anticipated as a consequence of poor structural similarity to the training set molecules in the dataset used to generate the models.

The identification of the likely chemical scaffolds of ‘out-of-model’ samples was more challenging because they exhibited structural similarity to a number of superclusters/ clusters, which increases the likelihood of a misclassification. Additionally, the overlap between NPS and non-NPS may result in false negatives, where NPS samples contain large amounts of these cutting agents and adulterants. Greatest overlaps were observed between training NPS and adulterants of the same chemical class, such as anaesthetics. This is understandable given that their structural similarities would give rise to similar Raman spectra.

By comparing the general limitations of the models developed in this thesis with the marketed library and mixtures algorithm deployed on the ProgenyTM platform, the following could be observed: with a very limited NPS library (n = 2) corresponding to claimed NPS on the products’ labels, and by using the in-built mixtures algorithm, 100 % of the NPS reference standards correlated to themselves in the library (Table 2.12). However, the NPS content in only 6 out of 7 NPS products were identified (86 %) (Table 2.15). This is because of the absence of the 7th unclaimed NPS from the instrument’s library. Furthermore, by populating the library with reference spectra of representative NPSs, test and ‘OOM’ NPS (n = 74), the NPS content in only 10 out of 25 NPS products were identified (40 %) (Table 6.2).

In contrast, using the models developed *via* a combination of similarity studies, chemometrics and a training set comprising the reference spectra of representative NPSs, test and ‘OOM’ NPS (n=74), 89 % of the test and ‘OOM’ NPS reference standards were identified and 53 % of the NPS products were aligned with their EMCDDA/EDND classes and/ or assigned to their anticipated superclusters. When the reference spectra of 22 cutting agents and adulterants were added to the training set, only 42 % of NPS internet products were aligned with members of their superclusters because samples with low amounts of NPS correlated to cutting agents.

From the above, it could be observed, that the developed models resulted in similar findings to that of the proprietary mixtures algorithm deployed on Progeny™. However, the advantage of the developed models is their ability to suggest the chemical scaffolds of newly emerging NPS even if they are absent from the library.

Overall results from the PCA modelling revealed the challenges faced with identifying NPS in complex ‘street-like’ mixtures and the limitations of the chemometric techniques used including a high dependence on the minimum common substructures between training samples used to build the initial clustering model by Zloh et al. [204]. These challenges were exacerbated by the difficulty of obtaining good quality Raman spectra for NPS internet samples using a handheld platform, which may have added to problems for complex mixtures. This was compounded by the poor sensitivity of the 1064 nm laser. However, despite the limitations, the technique used enabled the extraction of Raman signals *via* data reduction and identification of important patterns for classification of NPS out of noisy and corrupt data.

The overall aims and objectives of this work were met. This work has potentially significant practical implications with respect to improving the rapid, non-contact, non-destructive identification of NPS in solid mixtures in the field, in particular by first responders, front-line staff in emergency settings, customs and border control, forensic officers and researchers. This work has highlighted the successes and challenges in the identification of these mixtures using Raman spectroscopy including the strengths and limitations of bench-top versus hand-held instruments and different in-built detection and classification algorithms.

Implications for practical use

The proposed step-wise pre-processing protocols (Chapter 4) is designed to assist stakeholders in extracting Raman information from noisy complex Raman data. Automated protocols including pre-processing of Raman spectra in combination with a ‘representative’ NPS Raman database would be of paramount importance for first responders as it may potentially assist in the indication of the chemical scaffolds of an ‘unknown’ NPS. This may, in turn, be very useful for front-line healthcare professionals in informing decision making in emergency situations. It can also be useful for manufacturers of these instruments as they may financially benefit from it if the pre-processing method and the ‘representative’ Raman database were transferred onto/ deployed on a handheld platform equipped with a 1064 nm laser λ_{ex} .

Also, NPS have been shown to be incorporated in illicit drug products and unintentionally consumed [38]. Therefore, it is indispensable to be able to screen these products for NPS, adulterant and cutting agent content in music festivals and nightclubs in order gain the understanding of the current patterns and trends of drug misuse, changes to the drug market, inform treatment decision-making if needed and inform law enforcement on potential supply chains. It is crucially important for clinicians to understand prevalent trends particularly in those who are resisting or not responding to abstinence treatment.

NPS are also often undetected using common drug testing kits, and, therefore are preferred for users who undergo regular drug check such as for those among the military personnel, individuals in jail or those in mental health units [64]. Therefore, the proposed method can be used as a prototype screening protocol to overcome the limitations of the current drug testing kits if a suspect sample is seized. There is no doubt that the wide chemical diversity, patterns of use and toxicity profiles of NPS, samples' complexity, and the current limitations of detection kits require varied interventions and harm reductions techniques that are complex to implement. Therefore, preliminary classification of emerging NPS can be considered as an invaluable contribution to assist these interventions.

Information presented herein (i.e. spectra and PCA plots) provide an important source of identification data for NPS reference standards as well as for adulterants and cutting agents commonly incorporated in street samples which could be exploited by the Home Office, police officers, investigators, personnel in the criminal justice system, forensic laboratories and stakeholders involved in the routine analysis of seized NPS, toxicologists, medical examiners and the training and education of healthcare professionals. The 'representative' Raman database accompanied with proposed supercluster/ cluster membership and EMCDDA/EDND classification could subsequently be deployed on handheld Raman platforms to create a searchable library for newly emerging NPS.

Despite all the advances in analytical techniques used by forensic and toxicology scientists in order to enable the identification of NPS, the 'unpredictability factor' makes it very challenging to detect an 'unknown' substance [2]. This makes the developed models a significant contribution in this field for their ability to suggest the chemical scaffold of an 'unknown' molecule and its potential alignment with supercluster/ cluster or class of known NPS. This method may accelerate the screening of newly emerging NPS by assigning them to a supercluster/ cluster or class. This, in turn could assist in informing law enforcement and

treatment decision-making. Ideally, this would suggest the need for a full characterisation of the suspected sample in a forensic lab.

Future Work

In view of the misclassification of NPS internet mixtures and false positives engendered by the complex nature of these mixtures, the poor S/N and intense fluorescent background, future work should focus on the optimisation of the pre-processing of Raman data obtained with poor resolution handheld Raman instruments. Future work should also focus on the use of PCA in conjunction with machine learning techniques such as SVM and ANN, which may, in turn improve the identification of newly emerging NPS. Linear SVM could be applied to screen whether a confiscated mixture (powder or tablet) contains an NPS or not. This could be optimised by testing the ‘unknown’ mixture against the ‘representative’ NPS Raman database. The use of ANN may be more successful if the input neurones were trained using the ‘representative’ NPS Raman database. Since ANN works by identifying patterns in noisy corrupt data, it may perform well with noisy corrupt Raman spectra of internet complex mixtures. Both SVM and PCR could be investigated to give numerical accuracy to NPS classification. Independent component analysis (ICA), known as a ‘blind source separation’ is another method that could be used to filter off the patterns related to a specific NPS in a complex mixtures [292].

A current limitation of the supervised PCA methods is their limited ability of classifying diverse datasets comprising a large number of classes. Hence, there is a need to apply classification algorithms such as SIMCA analysis in order to assign class membership of clusters, define class boundaries and distinguish between closely related classes (e.g. crystallising and corresponding non-crystallising substances [165]), enable the assignment of cluster numbers and concentration limitations [293]. This in turn will enable the theoretical contribution of this work to be developed further for in-field detection of NPS. SIMCA allows an optimum data reduction by calculating a PCA for each group and producing a model from the combined PCs [178]. SIMCA also allows the identification of outliers at certain confidence limits and the evaluation of the distribution of data. Therefore, future work will include SIMCA analysis in order to measure the ‘Mahalanobis distance’ between ‘unknown’ samples and the multivariate centroid of the variables using the scores of the retained PCs and assign ‘unknown’ samples to pre-defined classes [184, 194, 195].

Quantitative methods using PCR and partial least squares regression (PLSR) may be employed to determine the NPS amounts in complex mixtures. However, ‘design of experiments (DoE)’ may be employed to optimise the efficiency of the experiments to be conducted. In this thesis, the ‘representative’ database was based on selecting ‘representative NPS with maximum structural similarity to others in their cluster (medoids), but maximum dissimilarity to members of other clusters. Other chemometric techniques employing different parameters such as different molecular moieties and properties could be used to improve prediction selectivity. Also, further work needs to be done to improve in-built algorithms on handheld Raman instruments to match the rapid explosion of NPS.

In this work, limitations due to the instruments employed in the studies i.e. laser spot sizes, laser penetration depth, etc. may have had a direct impact on the power of the statistical sampling. In Chapters 5 and 6, the number of measurements taken per sample coupled with the random selection of sampling was evaluated and optimised such that the analysis time was considered and the robustness of the PCA models was not compromised. In future work, the power of the statistical sampling should be considered in order to ensure that the combined results for all measurements per street sample are representative of the constituents of the sample.

Future work may also include optimising the use of exploratory PCA for its ability to discriminate between NPS, adulterants and cutting agents and also between NPS salts by individually evaluating each class or supercluster previously described in this thesis. This could be beneficial for law enforcement and forensic examiners to inform them on supply chains, common origins of NPS batches and new patterns/ trends in the NPS market. Further work needs to be done to improve the model on a handheld platform with reduced number of variables and low detector sensitivity.

The last decade has witnessed dramatic changes to the NPS market at a local and international level. Therefore, information sharing and dissemination of findings with our global partners, researchers, forensic scientists, law enforcement agents, customs and border control agents, personnel of the criminal justice system and emergency departments is crucial.

9. References

1. C.M. Goodair, J. Corkery, H. Claridge, Legal highs: a problem of definitions?, *Lancet*. 383 (2014) 1715. DOI: 10.1016/S0140-6736(14)60822-9.
2. S.D. Brandt, L.A. King, M. Evans-Brown, The new drug phenomenon, *Drug Test. Anal.* 6 (2014) 587-597. DOI: 10.1002/dta.1686. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.1686/full>. (accessed 02.02.17).
3. Advisory Council on the Misuse of Drugs (ACMD). Consideration of the Novel Psychoactive Substances ('Legal Highs'). Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/119139/acmdnps2011.pdf (accessed 21.09.15).
4. L. Iversen, Psychoactive Substances Bill, Advisory Council on the Misuse of Drugs (ACMD): London, Home Office, UK, 2015. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470421/ACMD_definitions_advice_final-23-October-2015.pdf (accessed 02.02.17).
5. F. Schifano, L. Orsolini, G.D. Papanti, J.M. Corkery, Novel psychoactive substances of interest for psychiatry, *World Psychiatry*. 14 (2015) 15-26. DOI: 10.1002/wps.20174. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/wps.20174/full> (accessed 16.01.17)
6. Psychoactive Substances Act 2016, The Stationery Office Limited, 2016. Available at: <http://services.parliament.uk/bills/2015-16/psychoactivesubstances.html> (accessed 21.09.15).
7. R. Ralphs, L. Williams, R. Askew, A. Norton, Adding spice to the porridge: the development of a synthetic cannabinoid market in an English prison, *Int. J. Drug Policy*. 40 (2016) 57-69. DOI:10.1016/j.drugpo.2016.10.003. Available at: http://ac.els-cdn.com/S0955395916303073/1-s2.0-S0955395916303073-main.pdf?_tid=2456c7d4-e958-11e6-9542-00000aabb0f02&acdnat=1486047725_700dfca6dbf64c9b2dbae5e979165971 (accessed 02.02.17).
8. G. Henderson, Designer drugs: past history and future prospects, *J Forensic Sci.* 43 (1988) 569-575. Available at: https://www.astm.org/DIGITAL_LIBRARY/JOURNALS/FORENSIC/PAGES/JFS11976J.htm (accessed 02.02.17).
9. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), New psychoactive substances in Europe: An update from the EU Early Warning System, Luxembourg, EMCDDA. Available at: http://www.emcdda.europa.eu/attachements.cfm/att_235958_EN_TD0415135ENN.pdf (accessed 21.09.15).
10. E. Lahaie, M. Martinez, A. Cadet-Tairou, New psychoactive substances and the internet: current situations and issues, *Tendances*. 84 (2013) 1-8. Available at: https://www.researchgate.net/publication/285900715_New_Psychoactive_Substances_and_the_Internet (accessed 02.02.17).
11. T. Shimane, K. Wada, Y. Hidaka, M. Funada, Prevalence and patterns of the use of novel psychoactive substances, "Kiken drugs", among younger adults at dance parties in Japan, *Drug Alcohol Depend.* 156 (2015) e202-e203. Available at: [http://www.drugandalcoholdependence.com/article/S0376-8716\(15\)00916-3/abstract](http://www.drugandalcoholdependence.com/article/S0376-8716(15)00916-3/abstract) (accessed 02.02.17).
12. A. Guirguis, S. Giroto, B. Berti, J.L. Stair, Identification of new psychoactive substances (NPS) using handheld Raman Spectroscopy employing both 785 and 1064 nm laser sources, *J. Forensic Sci. Int.* 273 (2017) 113-123. DOI: 10.1016/j.forsciint.2017.01.027.

13. F. Schifano, G.D. Papanti, L. Orsolini, J.M. Corkery, The consequences of drug misuse on post-marketing surveillance, *Expert Rev. Clin. Pharmacol.* 9 (2016) 867-871. DOI: 10.1080/17512433.2016.1178571. Available at: <http://www.tandfonline.com/doi/pdf/10.1080/17512433.2016.1178571?needAccess=true> (accessed 02.02.17).
14. S. Beharry, S. Gibbons, An overview of emerging and new psychoactive substances in the United Kingdom, *Forensic Sci. Int.* 267 (2016) 25-34. DOI: 10.1111/add.12287. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/add.12287/full> (accessed 02.02.17).
15. T. Staton, Pfizer sets painkiller marketing code with Chicago's opioid police. FiercePharma, Available at: <http://www.fiercepharma.com/marketing/pfizer-sets-painkiller-promo-code-chicago-s-opioid-marketing-police> (accessed 17.01.17).
16. P. Griffiths, M. Evans-Brown, R. Sedefov, Getting up to speed with the public health and regulatory challenges posed by new psychoactive substances in the information age, *Addiction.* 108 (2013) 1700-1703. DOI: 10.1111/add.12287. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/add.12287/full> (accessed 02.02.17).
17. United Nations Office on Drug Control, A century of international drug control in: 2008 World Drug Report, Vienna, UNODC. Available at: https://www.unodc.org/documents/wdr/WDR_2008/WDR_2008_eng_web.pdf (accessed 21.09.15).
18. A. Shulgin, A. Shulgin, PIHKAL: a chemical love story, Berkeley, CA. Transform Press, 1991.
19. A. Shulgin, A. Shulgin, TIHKAL: The continuation, Berkeley, CA. Transform Press, 1997.
20. D.S. Dolliver, J.B. Kuhns, The Presence of New Psychoactive Substances in a Tor Network Marketplace Environment, *J. Psychoactive Drugs.* 48 (2016) 321-329. DOI:10.1080/02791072.2016.1229877. Available at: <http://www.tandfonline.com/doi/abs/10.1080/02791072.2016.1229877> (accessed 02.02.17).
21. D. Rhumorbarbe, L. Staehli, J. Broseus, Q. Rossy, P. Esseiva, Buying drugs on a Darknet market: A better deal? Studying the online illicit drug market through the analysis of digital, physical and chemical data, *Forensic Sci. Int.* 267 (2016) 173-182. DOI:10.1016/j.forsciint.2016.08.032. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073816303620> (accessed 02.02.17).
22. A.J.M. Forsyth, Virtually a drug scare: Mephedrone and the impact of the Internet on drug news transmission. *Int. J. Drug Policy.* 23 (2012) 198-209. DOI:10.1016/j.drugpo.2011.12.003. Available at: <http://www.sciencedirect.com/science/article/pii/S0955395911002398> (accessed 02.02.17).
23. The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), The internet and drug markets, Luxembourg, EMCDDA. Available at: http://www.emcdda.europa.eu/system/files/publications/2155/TDXD16001ENN_FINAL.pdf (accessed 21.09.15).
24. A.T. Shulgin, M. Tania, P.F. Daley, The Shulgin Index: Psychedelic phenethylamines and related compounds, volume one, Berkely, CA. Transform Press, 2011.
25. O. Corazza, G. Valeriani, F.S. Bersani, J. Corkery, G. Martinotti, G. Bersani, F. Schifano, "Spice," "Kryptonite," "Black Mamba": An Overview of Brand Names and Marketing Strategies of Novel Psychoactive Substances on the Web, *J. Psychoactive Drugs.* 46 (2014) 287-294. DOI:10.1080/02791072.2014.944291. Available at: <http://www.tandfonline.com/doi/abs/10.1080/02791072.2014.944291> (accessed 02.02.17).
26. S. Levissianos, The UNODC Early Warning Advisory (EWA) on New Psychoactive Substances (NPS) – New Trends. In *Novel Psychoactive Substances.* 2014. Rome, United Nations Office on Drugs and Crime.

27. EMCDDA-EDND, European Monitoring Centre on Drugs and Drug Addiction - EDND European Information System on New Drugs (access restricted to registered users only). 2017.
28. Anonymous, Esters of morphine, *Bulletin on Narcotics*, 4 (1953) 36.
29. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), *New drugs in Europe, 2012. EMCDDA–Europol 2015 Annual Report on the implementation of Council Decision 2005/387/JHA*, Luxembourg, EMCDDA. Available at: <http://www.emcdda.europa.eu/system/files/publications/2880/TDAS16001ENN.pdf> (accessed 21.09.15).
30. A. Gallegos, New psychoactive substances and the EU Early Warning System: chemistry matters. In *Novel Psychoactive Substances*. 2014. Rome: European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), Third International Conference on Novel Psychoactive Substances.
31. D. Abdulrahim, O. Jones, Novel Psychoactive Treatment UK Network NEPTUNE, *Guidance on the Management of Acute and Chronic Harms of Club Drugs and Novel Psychoactive Substances*, Available at: <http://neptune-clinical-guidance.co.uk/wp-content/uploads/2015/03/NEPTUNE-Guidance-March-2015.pdf> (accessed 21.09.15).
32. Adley, M., *The drug wheel*, Available at: http://www.thedrugswheel.com/downloads/TheDrugsWheel_2_0_2_colour.pdf (accessed 02.02.17).
33. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), *Perspectives on drugs: Injection of synthetic cathinones*, Available at: http://www.emcdda.europa.eu/attachements.cfm/att_228233_EN_POD2014_Injection%20of%20synthetic%20cathinones.pdf (accessed 21.09.15).
34. P. Heikman, M. Sundstrom, A. Pelander, I. Ojanpera, New psychoactive substances as part of polydrug abuse within opioid maintenance treatment revealed by comprehensive high-resolution mass spectrometric urine drug screening, *Hum. Psychopharmacol. Clin. Exp.* 31 (2016) 44-52. DOI: 10.1002/hup.2512. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/hup.2512/full> (accessed 02.02.17).
35. A. Batisse, M. Gregoire, M. Marillier, M. Fortias, S. Djeddar, Usage de cathinones à Paris [Cathinones use in Paris], *L'Encephale*, 42 (2016) 354-360. DOI: 10.1016/j.encep.2015.09.002. Available at: <http://europepmc.org/abstract/med/26847479> (accessed 02.02.17).
36. K.D. Wagner, R.F. Armenta, A.M. Roth, J.C. Maxwell, J. Cuevas-Mota, R.S. Garfein, Use of synthetic cathinones and cannabimimetics among injection drug users in San Diego, California, *Drug Alcohol Depend.* 141 (2014) 99-106. DOI:10.1016/j.drugalcdep.2014.05.007. Available at: <http://www.sciencedirect.com/science/article/pii/S037687161400876X> (accessed 07.05.14).
37. F. Schifano, L. Orsolini, D. Pappanti, J Corkery, NPS: medical consequences associated with their intake, *Curr. Top. Behav. Neurosci.* 2016 1-30. DOI: 10.1007/7854_2016_15. Available at: https://www.researchgate.net/profile/Laura_Orsolini/publication/303822786_NPS_Medical_Consequences_Associated_with_Their_Intake/links/578550b308ae3949cf538ce2.pdf (accessed 02.02.17).
38. A. Guirguis, J. Corkery, J.L. Stair, M. Zloh, S.B. Kirton, F. Schifano, Intended and unintended use of cathinone mixtures, *Hum. Psychopharmacol.* DOI: 10.1002/hup.2598. March 2017. (In Press).
39. M.R. Meyer, New psychoactive substances: an overview of recent publications on their toxicodynamics and toxicokinetics, *Arch. Toxicol.* 90 (2016) 2421-2444. DOI:

- 10.1007/s00204-016-1812-x. Available at: <http://link.springer.com/article/10.1007/s00204-016-1812-x> (accessed 02.02.17).
40. S.L. Hill, S.H. Thomas, Clinical toxicology of newer recreational drugs, *Clin. Toxicol.* 49 (2011) 705-719. DOI:10.3109/15563650.2011.615318. Available at: <http://www.tandfonline.com/doi/abs/10.3109/15563650.2011.615318> (accessed 02.02.17).
41. J. Mounteney, I. Giraudon, G. Denisson, P. Griffiths, Fentanyls: are we missing the signs? Highly potent and on the rise in Europe, *Int. J. Drug Policy.* 26 (2015) 626-631. DOI:10.1016/j.drugpo.2015.04.003. Available at: <http://www.sciencedirect.com/science/article/pii/S0955395915000973> (accessed 02.02.17).
42. L. Morrison, Notice of meeting and agenda. Health, social care and housing committee. 2015, Edinburgh: The City of Edinburgh Council.
43. D.M. Wood, S. Davies, M. Puchnarewicz, J. Button, R. Archer, H. Ovaska, J. Ramsey, T. Lee, D.W. Holt, P.I. Dargan, Recreational use of mephedrone (4-methylmethcathinone, 4-MMC) with associated sympathomimetic toxicity, *J. Med. Toxicol.* 6 (2010) 327-330. DOI: 10.1007/s13181-010-0018-5. Available at: <http://link.springer.com/article/10.1007/s13181-010-0018-5> (accessed 02.02.17).
44. B.J. Warrick, J. Wilson, M. Hedge, S. Freeman, K. Leonard, C. Aaron, Lethal serotonin syndrome after methylone and butylone ingestion, *J. Med. Toxicol.* 8 (2012) 65-68. DOI: 10.1007/s13181-011-0199-6. Available at: <http://link.springer.com/article/10.1007/s13181-011-0199-6> (accessed 02.02.17).
45. A. Kamour, D. James, D.J. Lupton, G. Cooper, M. Eddleston, A. Vale, J.P. Thompson, R. Thanacoody, S.L. Hill, S.H.L. Thomas, Patterns of presentation and clinical features of toxicity after reported use of ([2-aminopropyl]-2, 3-dihydrobenzofurans), the 'benzofuran' compounds. A report from the United Kingdom National Poisons Information Service, *Clin. Toxicol.* 52 (2014) 1025-1031. DOI:10.3109/15563650.2014.973115. Available at: <http://www.tandfonline.com/doi/abs/10.3109/15563650.2014.973115> (accessed 02.02.17).
46. D.M. Wood, P.I. Dargan, Novel psychoactive substances: How to understand the acute toxicity associated with the use of these substances, *Ther. Drug Monit.* 34 (2012) 363-367. DOI: 10.1097/FTD.0b013e31825b954b. Available at: http://journals.lww.com/drugmonitoring/Abstract/2012/08000/Novel_Psychoactive_Substances_How_to_Understand.2.aspx (accessed 02.02.17).
47. P.J. Nicholson, M.J. Quinn, J.D. Dodd, Headshop heartache: acute mephedrone 'meow' myocarditis, *Heart.* 96 (2010) 2051-2052. DOI:10.1136/hrt.2010.209338. Available at: <http://heart.bmj.com/content/96/24/2051.short> (accessed 02.02.17).
48. S. Fröhlich, E. Lambe, J. O'Dea, Acute liver failure following recreational use of psychotropic "head shop" compounds, *Irish J. Med. Sci.* 180 (2011) 263-264. DOI: 10.1007/s11845-010-0636-6. Available at: <http://link.springer.com/article/10.1007/s11845-010-0636-6> (accessed 02.02.17).
49. M.J. Valente, P.G. de Pinho, M. de Lourdes Bastos, F. Carvalho, M. Carvalho, Khat and synthetic cathinones: a review, *Arch. Toxicol.* 88 (2014) 15-45. DOI: 10.1007/s00204-013-1163-9. Available at: <http://link.springer.com/article/10.1007/s00204-013-1163-9> (accessed 02.02.17).
50. H. Regunath, V.K. Ariyamuthu, P. Dalal, M. Misra, Bath salt intoxication causing acute kidney injury requiring hemodialysis, *Hemodialysis International.* 16 (2012) S47-S49. DOI: 10.1111/j.1542-4758.2012.00750.x. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1542-4758.2012.00750.x/full> (accessed 02.02.17).

51. S.L. Thornton, J. Lo, R.F. Clark, A.H.B. Wu, R.R. Gerona, Simultaneous detection of multiple designer drugs in serum, urine, and CSF in a patient with prolonged psychosis, *Clin. Toxicol.* 50 (2012) 1165-1168. DOI:10.3109/15563650.2012.744996 <http://www.tandfonline.com/doi/abs/10.3109/15563650.2012.744996> (accessed 02.02.17).
52. J.M. Striebel, J.M. Pierre, Acute psychotic sequelae of “bath salts”, *Schizophr. Res.* 133 (2011) 259-260. DOI:10.1016/j.schres.2011.09.001. Available at: http://ac.els-cdn.com/S0920996411004762/1-s2.0-S0920996411004762-main.pdf?_tid=8ef0b10a-e966-11e6-b93c-00000aacb361&acdnat=1486053917_cc33c595fd21b2cac9c94a5e6696a6b8 (accessed 02.02.17).
53. G.L. Kashyap, A.G. Patel, Unusual presentation of a patient with GBL withdrawal: a case report. *Psychiatr Danub.* 23 (2011) S32-34. Available at: http://www.hdbp.org/psychiatria_danubina/pdf/dnb_vol23_sup/dnb_vol23_sup_32.pdf (accessed 02.02.17).
54. J. Corkery, H. Claridge, B. Loi, C. Goodair, F. Schifano, National Programme on Substance Abuse Deaths, Annual Report 2013. 2014, National Programme on Substance Abuse Deaths (NPSAD). International Centre for Drug Policy (ICDP). St George’s, University of London, UK. Available at: <http://www.yumpu.com/en/document/view/23352804/national-programme-on-substance-abuse-deaths-annual-report-2013-on-drug-related-deaths-in-the-uk-january-december-2012-pdf> (accessed 02.02.17).
55. J.M. Corkery, F. Schifano, A.H. Ghodse, Mephedrone-Related Fatalities in the United Kingdom: Contextual, Clinical and Practical Issues, Chapter 17, pp. 355-380, 2012, in *Pharmacology*. Rijeka, Croatia: InTech - Open Access Publisher. ISBN 979-953-307-482-4. Edited by Dr. Luca Gallelli Department of Experimental and Clinical Medicine, School of Medicine, University of Catanzaro, Clinical Pharmacology Unit, Mater Domini Hospital, Italy. Published 14 March 2012. DOI: 10.5772/32935. Available at: <http://www.intechopen.com/books/pharmacology/mephedrone-related-fatalities-in-the-united-kingdom-contextual-clinical-and-practical-issues> (accessed 02.02.17).
56. F. Schifano, J. Corkery, A.H. Ghodse, Suspected and confirmed fatalities associated with mephedrone (4-methylmethcathinone; ‘meow meow’) in the United Kingdom, *J. Clin. Psychopharm.* 32 (2012) 710-714. DOI: 10.1097/JCP.0b013e318266c70c. Available at: http://journals.lww.com/psychopharmacology/Abstract/2012/10000/Suspected_and_Confirmed_Fatalities_Associated_With.21.aspx (accessed 16.01.17).
57. S.H. Cosbey, K.L. Peters, A. Quinn, A. Bentley, Mephedrone (Methylmethcathinone) in Toxicology Casework: A Northern Ireland Perspective, *J. Anal. Toxicol.* 37 (2013) 74-82. DOI: 10.1093/jat/bks094. Available at: <https://academic.oup.com/jat/article/37/2/74/857955/Mephedrone-Methylmethcathinone-in-Toxicology> (accessed 02.02.17).
58. L.J. Marinetti, H.M. Antonides, Analysis of Synthetic Cathinones Commonly Found in Bath Salts in Human Performance and Postmortem Toxicology: Method Development, Drug Distribution and Interpretation of Results, *J. Anal. Toxicol.* 37 (2013) 135-146. DOI:https://doi.org/10.1093/jat/bks136. Available at: <https://academic.oup.com/jat/article/37/3/135/819450/Analysis-of-Synthetic-Cathinones-Commonly-Found-in> (accessed 16.01.17).
59. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), EMCDDA–Europol, EMCDDA–Europol Joint Report on a new psychoactive substance: MDPV (3,4-methylenedioxypropylvalerone). Luxembourg, EMCDDA. Available at: http://www.emcdda.europa.eu/system/files/publications/819/TDAS14001ENN_466653.pdf (accessed 21.09.15).

60. J. Krueger, H. Sachs, F. Musshoff, T. Dame, J. Schaeper, M. Schwerer, M. Graw, G. Roeder, First detection of ethylphenidate in human fatalities after ethylphenidate intake, *Forensic Sci. Int.* 243 (2014) 126-129. DOI: 10.1016/j.forsciint.2014.07.017 <http://www.sciencedirect.com/science/article/pii/S0379073814003028> (accessed 21.09.15).
61. J.M. Corkery, F. Schifano, A.H. Ghodse, Phenazepam abuse in the UK: an emerging problem causing serious adverse health problems, including death, *Hum. Psychopharm.* 27 (2012) 254-261. DOI: 10.1002/hup.2222. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/hup.2222/full> (accessed 21.09.15).
62. C. O'Neill, Novel Psychoactive Substances: risks and harms, *Community Pract.* 87 (2014) 45-47. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/25226709> (accessed 02.02.17).
63. Home Office, New psychoactive substances review: Report of the expert panel. London, Home Office. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/368583/NPExpertReviewPanelReport.pdf (accessed 21.09.15).
64. B.L. Miller, J.M. Stogner, J.M. Miller, M.I. Fernandez, The Arrest and Synthetic Novel Psychoactive Drug Relationship: Observations From a Young Adult Population, *J. Drug Issues.* 47 (2017) 91-103. DOI: 10.1177/0022042616678611. Available at: <http://journals.sagepub.com/doi/pdf/10.1177/0022042616678611> (accessed 02.02.17).
65. D. Ladder, Drug misuse: Findings from the 2014/15 crime survey for England and Wales (second edition). London, Home Office. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/462885/drug-misuse-1415.pdf (accessed 21.09.15). ISBN: 978-1-78246-828-8.
66. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), EMCDDA-Europol 2013 Annual Report on the implementation of Council Decision 2005/387/JHA. Luxembourg, EMCDDA. Available at: http://www.emcdda.europa.eu/system/files/publications/814/TDAN14001ENN_475519.pdf (accessed 21.09.15).
67. M. Evans-Brown, A. Kimergård, J. McVeigh, M. Chandler, S.D. Brandt, Is the breast cancer drug tamoxifen being sold as a bodybuilding dietary supplement?, *BMJ.* 348 (2014) g1476. DOI: 10.1136/bmj.g1476. Available at: https://www.researchgate.net/profile/Jim_Mcveigh/publication/260194844_BODYBUILDING_SUPPLEMENTS_Is_the_breast_cancer_drug_tamoxifen_being_sold_as_a_bodybuilding_dietary_supplement/links/0deec52fdf77d9685e000000.pdf (accessed 21.09.15).
68. A. Pourmand, P. Armstrong, M. Mazer-Amirshahi, The evolving high new designer drugs of abuse, *Hum. Exp. Toxicol.* 33 (2014) 993-999. DOI: 10.1177/0960327113514100. Available at: <http://journals.sagepub.com/doi/pdf/10.1177/0960327113514100> (accessed 21.09.15).
69. A. Boys, M. Farrell, P. Bebbington, T. Brugha, J. Coid, R. Jenkins, G. Lewis, J. Marsden, H. Meltzer, N. Singleton, C. Taylor, Drug use and initiation in prison: results from a national prison survey in England and Wales, *Addiction.* 97 (2002) 1551-1560. DOI: 10.1046/j.1360-0443.2002.00229.x. Available at: <http://onlinelibrary.wiley.com/doi/10.1046/j.1360-0443.2002.00229.x/full> (accessed 21.09.15).
70. RAPt, RAPt research and policy briefing series No. 4 - tackling the issue of new psychoactive substances in prisons. 2015, Rehabilitation for Addicted Prisoners Trust. Available at: <http://www.rapt.org.uk/sites/default/files/16/RAPt%20Research%20and%20Policy%20Briefing%20Number%204%20v10%20AW%20edit%20-%20%281.9.2015%29.pdf> (accessed 02.02.17).

71. T.M. Brunt, A. Poortman, R.J.M. Niesink, W. van den Brink, Instability of the ecstasy market and a new kid on the block: mephedrone, *J. of Psychopharmacol.* 25 (2011) 1543-1547. Available at: <http://journals.sagepub.com/doi/abs/10.1177/0269881110378370> (accessed 21.09.15).
72. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), New drugs in Europe, 2012. EMCDDA–Europol 2012 Annual Report on the implementation of Council Decision 2005/387/JHA. Luxembourg, EMCDDA. Available at: http://www.emcdda.europa.eu/system/files/publications/734/EMCDDA-Europol_2012_Annual_Report_final_439477.pdf (accessed 21.09.15).
73. B. Miserez, O. Ayrton, J. Ramsey, Analysis of purity and cutting agents in street mephedrone samples from South Wales, *Forensic Toxicol.* 32 (2014) 305-310. DOI: 10.1007/s11419-014-0232-y. Available at: <http://link.springer.com/article/10.1007/s11419-014-0232-y> (accessed 21.09.15).
74. A.M. Leffler, P.B. Smith, A. de Armas, F.L. Dorman, The analytical investigation of synthetic street drugs containing cathinone analogs, *Forensic Sci. Int.* 234 (2014) 50-56. DOI:10.1016/j.forsciint.2013.08.021. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073813003988> (accessed 21.09.15).
75. Home Office, Annual Report on the Home Office Forensic Early Warning System (FEWS): A system to identify new psychoactive substances in the UK. London, Home Office. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/461333/1280_EL_FEWS_Annual_Report_2015_WEB.pdf (accessed 21.09.15).
76. Home Office, Annual Report on the Home Office Forensic Early Warning System (FEWS): A system to identify New Psychoactive Substances in the UK. London, Home Office. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/225737/FEWS_Annual_Report_WEB.pdf (accessed 21.09.15).
77. Home Office, Annual Report on the Home Office Forensic Early Warning System (FEWS): A system to identify new psychoactive substances in the UK. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/344551/2014-08-12_-_FEWS_Annual_Report_Aug_2014_-_Final_2_.pdf (accessed 21.09.15).
78. T.J. Weatherholt, Not for human consumption: how inept legislative policy proliferates the synthetic drug problem. *KY Law J.* 859 (2015) 257-4747. Available at: <http://www.kentuckylawjournal.org/index.php/2015/02/22/not-for-human-consumption/> (accessed 21.09.15).
79. S.D. Brandt, S. Freeman, H.R. Sumnall, F. Measham, J. Cole, Analysis of NRG ‘legal highs’ in the UK: identification and formation of novel cathinones. *Drug Test. Anal.* 3 (2011) 569-575. DOI: 10.1002/dta.204. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.204/full> (accessed 21.09.15).
80. H. Torrance, G. Cooper, The detection of mephedrone (4-methylmethcathinone) in 4 fatalities in Scotland, *Forensic Sci. Int.* 202 (2010) e62-e63. DOI:10.1016/j.forsciint.2010.07.014. Available at: http://ac.els-cdn.com/S037907381000352X/1-s2.0-S037907381000352X-main.pdf?_tid=3a2c8cf2-e970-11e6-b7a4-00000aacb362&acdnat=1486058070_e5bc1d6e5aa8c90673344c08f3554bb4 (accessed 21.09.15).
81. M.D. Swortwood, D. Boland, A. de Caprio, Determination of 32 cathinone derivatives and other designer drugs in serum by comprehensive LC-QQQ-MS/MS analysis, *Anal. Bioanal. Chem.* 405 (2013) 1383-1397. DOI: 10.1007/s00216-012-6548-8. Available at: <http://link.springer.com/article/10.1007/s00216-012-6548-8> (accessed 21.09.15).

82. E.Y. Santali, A.K. Cadogan, N.N. Daeid, K.A. Savage, O.B. Sutcliffe, Synthesis, full chemical characterisation and development of validated methods for the quantification of (\pm)-4'-methylmethcathinone (mephedrone): A new "legal high", *J. Pharmaceut. Biomed.* 56 (2011) 246-255. DOI:10.1016/j.jpba.2011.05.022. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708511002810> (accessed 21.09.15).
83. P. Jankovicsa, A. Váradi, L. Tölgyesi, S. Lohner, J. Németh-Palotás, H. Kőszegi-Szalai, Identification and characterization of the new designer drug 4'-methylethcathinone (4-MEC) and elaboration of a novel liquid chromatography–tandem mass spectrometry (LC–MS/MS) screening method for seven different methcathinone analogs, *Forensic Sci. Int.* 210 (2011) 213-220. DOI:10.1016/j.forsciint.2011.03.019. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073811001460> (accessed 21.09.15).
84. S. Strano-Rossi, L. Anzillotti, E. Castrignanò, F.S. Romolo, M. Chiarotti, Ultra high performance liquid chromatography–electrospray ionization–tandem mass spectrometry screening method for direct analysis of designer drugs, "spice" and stimulants in oral fluid. *J. Chromatogr. A.* 1258 (2012) 37-42. DOI:10.1016/j.chroma.2012.07.098. Available at: <http://www.sciencedirect.com/science/article/pii/S0021967312012198> (accessed 21.09.15).
85. M. Martin, J.F. Muller, K. Turner, M. Duez, V. Cirimele, Evidence of mephedrone chronic abuse through hair analysis using GC/MS, *Forensic Sci. Int.* 218 (2012) 44-48. DOI:10.1016/j.forsciint.2011.10.016. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073811004919> (accessed 21.09.15).
86. A.D. Lesiak, R.A. Musah, R.B. Cody, M.A. Domin, A.J. Dane, J.R.E. Shepard, Direct analysis in real time mass spectrometry (DART-MS) of "bath salt" cathinone drug mixtures, *Analyst* 138 (2013) 3424-3432. DOI: 10.1039/C3AN00360D. Available at: <http://pubs.rsc.org/en/content/articlelanding/2013/an/c3an00360d/unauth#!divAbstract> (accessed 21.09.15).
87. M. Baron, M. Elie, L. Elie, An analysis of legal highs—do they contain what it says on the tin?, *Drug Test. Anal.* 3 (2011) 576-581. DOI: 10.1002/dta.274. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.274/full> (accessed 21.09.15).
88. M.P. Elie, L.E. Elie, M.G. Baron, Keeping pace with NPS releases: fast GC-MS screening of legal high products, *Drug Test. Anal.* 5 (2013) 281-290. DOI: 10.1002/dta.1434. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.1434/full> (accessed 21.09.15).
89. L. Glicksberg, K. Bryand, S. Kerrigan, Identification and quantification of synthetic cathinones in blood and urine using liquid chromatography–quadrupole/time of flight (LC-Q/TOF) mass spectrometry, *J. Chromatogr. B.* 1035 (2016) 91-103. DOI: 10.1016/j.jchromb.2016.09.027. Available at: <http://www.sciencedirect.com/science/article/pii/S1570023216308868> (accessed 21.09.15).
90. C. Margalho, A. Castanheira, F.C. Real, E. Gallardo, M. López-Rivadulla, Determination of "new psychoactive substances" in postmortem matrices using microwave derivatization and gas chromatography–mass spectrometry, *J. Chromatogr. B.* 1020 (2016) 14-23. DOI:10.1016/j.jchromb.2016.03.001. Available at: <http://www.sciencedirect.com/science/article/pii/S1570023216301325> (accessed 02.02.17).
91. F. Vaiano, F.P. Busardò, D. Palumbo, C. Kyriakou, A. Fioravanti, V. Catalani, F. Mari, E. Bertol, A novel screening method for 64 new psychoactive substances and 5 amphetamines in blood by LC–MS/MS and application to real cases. *Anal. Bioanal. Chem.* 129 (2016) 441-449. DOI:10.1016/j.jpba.2016.07.009. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708516303818> (accessed 02.02.17).
92. P.M. Geyer, M.C. Hulme, J.P.B. Irving, P.D. Thompson, R.N. Ashton, R.J. Lee, L.J. Johnson, J. Marron, C.E. Banks, O.B. Sutcliffe, Guilty by dissociation - development of gas chromatography–mass spectrometry (GC-MS) and other rapid screening methods for the analysis of 13 diphenidine-derived new psychoactive substances (NPSs), *Anal. Bioanal.*

- Chem. 408 (2016) 8467-8481. DOI: 10.1007/s00216-016-9969-y. Available at: <http://link.springer.com/article/10.1007/s00216-016-9969-y> (accessed 02.02.17).
93. J. Welter-Luedeke, H.H. Maurer, New psychoactive substances: chemistry, pharmacology, metabolism, and detectability of amphetamine derivatives with modified ring systems, *Ther. Drug Monit.* 38 (2016) 4-11. DOI: 10.1097/FTD.0000000000000240. Available at: http://journals.lww.com/drug-monitoring/Abstract/2016/02000/New_Psychoactive_Substances_Chemistry..2.aspx (accessed 02.02.17).
94. United Nation Office on Drugs and Crime (UNODC), The challenge of new psychoactive substances, Global SMART Programme, Vienna, 2013. https://www.unodc.org/documents/scientific/NPS_2013_SMART.pdf (accessed 21.09.15).
95. K.Y. Noonan, L.A. Tonge, O.S. Fenton, D.B. Damiano, K.A. Frederick, Rapid classification of simulated street drug mixtures using Raman spectroscopy and principal component analysis, *Appl. Spectrosc.* 63 (2009) 742-747. Available at: <https://www.osapublishing.org/as/abstract.cfm?uri=as-63-7-742> (accessed 02.02.17).
96. S. Bell, *Forensic Chemistry*, *Annu. Rev. Anal. Chem.* 2 (2009) 297-319. Available at: <http://annualreviews.org/doi/abs/10.1146/annurev-anchem-060908-155251> (accessed 02.02.17).
97. P.I. Dargan, D.M. Wood, *Novel Psychoactive Substances: Classification, Pharmacology and Toxicology*, London, Elsevier Inc., 2013. Available at: <https://www.elsevier.com/books/novel-psychoactive-substances/dargan/978-0-12-415816-0> (accessed 02.02.17). ISBN: 9780124159112.
98. A. Grigoryev, A. Melnik, S. Savchuk, A. Simonov, V. Rozhanets, Gas and liquid chromatography–mass spectrometry studies on the metabolism of the synthetic phenylacetylindole cannabimimetic JWH-250, the psychoactive component of smoking mixtures, *J. Chromatogr. B*, 879 (2011) 2519-2526. DOI:10.1016/j.jchromb.2011.07.004. Available at: <http://www.sciencedirect.com/science/article/pii/S1570023211004557> (accessed 02.02.17).
99. R.B. Cody, J.A. Laramée, H.D. Durst, Versatile new ion source for the analysis of materials in open air under ambient conditions, *Anal. Chem.* 77 (2005) 2297-2302. DOI: 10.1021/ac050162j. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac050162j> (accessed 02.02.17).
100. E. Cuyppers, A.J. Bonneure, J. Tytgat, The use of presumptive color tests for new psychoactive substances. *Drug Test. Anal.* 8 (2016) 136-140. DOI: 10.1021/ac050162j. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.1847/abstract> (accessed 02.02.17).
101. C.A.F. de O. Penido, M.T. T. Pacheco, I.K. Lednev, L. Silveira Jr., Raman spectroscopy in forensic analysis: identification of cocaine and other illegal drugs of abuse, *J. Raman Spectrosc.* 47 (2015) 28-38. DOI: 10.1002/jrs.4864. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.4864/full> (accessed 02.02.17).
102. United Nations Office on Drugs and Crime (UNODC), Rapid testing methods of drugs of abuse. 1995, Vienna, United Nations Office on Drugs and Crime. Available at: https://www.unodc.org/documents/scientific/Rapid_Testing_Methods_of_Drugs_of_Abuse_E.pdf accessed 02.02.17).
103. M. Philp, R. Shimmon, M. Tahtouh, S. Fu, Development and validation of a presumptive colour spot test method for the detection of synthetic cathinones in seized illicit materials, *Forensic Chem.* 1 (2016) 39-50. DOI: 10.1016/j.forc.2016.06.001. Available at: <http://www.sciencedirect.com/science/article/pii/S2468170916300236> (accessed 02.02.17).
104. J.P. Smith, J.P. Metters, C. Irving, O.B. Sutcliffe, C.E. Banks, *Forensic electrochemistry: the electroanalytical sensing of synthetic cathinone-derivatives and their*

- accompanying adulterants in "legal high" products, *Analyst*. 139 (2014) 389-400. DOI: 10.1039/C3AN01985C. Available at: <http://pubs.rsc.org/is/content/articlelanding/2014/an/c3an01985c#!divAbstract> (accessed 02.02.17).
105. K. Kellett, J.H. Broome, M. Zloh, S.B. Kirton, S. Fergus, U. Gerhard, J.L. Stair, K.J. Wallace, Small molecule recognition of mephedrone using an anthracene molecular clip. *Chem. Commun.* 52 (2016) 7474-7477. DOI: 10.1039/C6CC03404G. Available at: <http://pubs.rsc.org/-/content/articlelanding/2016/cc/c6cc03404g/unauth#!divAbstract>.
106. M.J. Swortwood, W.L. Hearn, A.P. de Caprio, Cross-reactivity of designer drugs, including cathinone derivatives, in commercial enzyme-linked immunosorbent assays. *Drug Test. Anal.* 6 (2013) 716-727. DOI: 10.1002/dta.1489. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.1489/full> (accessed 02.02.17).
107. K.N. Ellefsen, S. Anizan, M.S. Castaneto, N.A. Desposiers, T.M. Martin, K.L. Klette, M.A. Huestis, Validation of the only commercially available immunoassay for synthetic cathinones in urine: Randox Drugs of Abuse V Biochip Array Technology, *Drug Test. Anal.* 6 (2014) 728-738. DOI: 10.1002/dta.1633. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/24659527> (accessed 02.02.17).
108. A. de Castro, E. Lendoiro, H. Fernandez-Vega, S. Steinmeyer, M. Lopez-Rivadulla, A. Cruz, Liquid chromatography tandem mass spectrometry determination of selected synthetic cathinones and two piperazines in oral fluid. Cross reactivity study with an on-site immunoassay device, *J. Chromatogr.A.* 1374 (2014) 93-101. DOI: 10.1016/j.chroma.2014.11.024. Available at: <http://www.sciencedirect.com/science/article/pii/S0021967314017828> (accessed 02.02.17).
109. M. Paillet-Loilier, A. Cesbron, R. Le Boisselier, J. Bourguine, D. Debruyne, Emerging drugs of abuse: current perspectives on substituted cathinones, *Subst. Abuse Rehabil.* 5 (2014) 37. Available at: <http://pubmedcentralcanada.ca/pmcc/articles/PMC4043811/> (accessed 02.02.17).
110. Ojanperä, I.A., P.K. Heikman, and I.J. Rasanen, Urine analysis of 3, 4-methylenedioxypropylvalerone in opioid-dependent patients by gas chromatography–mass spectrometry. *Ther. Drug Monit.* 33 (2011) 257-263. Available at: http://journals.lww.com/drugmonitoring/Abstract/2011/04000/Urine_Analysis_of_3,4_Methylenedioxypropylvalerone.18.aspx (accessed 02.02.17).
111. L.E. Register, J.D. Chmiel, J.M. Holler, S.P. Vorce, B. Levine, T.Z. Bosy, Determination of designer drug cross-reactivity on five commercial immunoassay screening kits, *J. Anal. Toxicol.* 39 (2015) 144-151. DOI: 10.1093/jat/bku133. Available at: <https://academic.oup.com/jat/article/39/2/144/762843/Determination-of-Designer-Drug-Cross-Reactivity-on> (accessed 02.02.17).
112. K. Tsujikawa, T. Yamamuro, K. Kuwayama, T. Kanamori, Y.T. Iwata, K. Miyamoto, F. Kasuya, H. Inoue, Application of a portable near infrared spectrometer for presumptive identification of psychoactive drugs, *Forensic Sci. Int.* 242 (2014) 162-171. DOI: 10.1016/j.forsciint.2014.05.020. Available at: <http://www.sciencedirect.com/science/article/pii/S037907381400231X> (accessed 02.02.17).
113. S. Mabbott, E. Correa, D.P. Cowcher, J.W. Allwood, R. Goodacre, Optimisation of parameters for the quantitative surface-enhanced raman scattering detection of mephedrone using a fractional factorial design and a portable Raman spectrometer, *Anal. Chem.* 85 (2012) 923-931. DOI: 10.1021/ac302542r. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac302542r> (accessed 02.02.17).
114. S. Mabbott, A. Eckmann, C. Casiraqhi, R. Goodacre, 2p or not 2p: tuppence-based SERS for the detection of illicit materials, *Analyst*, 138 (2013) 118. DOI: 10.1039/C2AN35974J. Available at:

<http://pubs.rsc.org/en/content/articlelanding/2012/an/c2an35974j/unauth#!divAbstract>
(accessed 02.02.17).

115. Z. Han, H. Liu, J. Meng, L. Yang, J. Liu, J. Liu, Portable kit for identification and detection of drugs in human urine using surface-enhanced Raman spectroscopy, *Anal. Chem.* 87 (2015) 9500-9506. DOI: 10.1021/acs.analchem.5b02899. Available at:

<http://pubs.acs.org/doi/abs/10.1021/acs.analchem.5b02899> (accessed 12/06/2017).

116. J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves, Vibrational spectroscopy sampling techniques in: J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves (Eds.) *Infrared and Raman Spectroscopy in Forensic Science*, UK, John Wiley & Sons. Chapter 2. pp 10-44, 2012. ISBN: 978-0-470-74906-7.

117. G.S. Bumbrah, R.M. Sharma, Raman spectroscopy - Basic principle, instrumentation and selected applications for the characterisation of drugs of abuse. *Egypt. J. Forensic Sci.* 6 (2016) 209-215. DOI: 10.1016/j.ejfs.2015.06.001. Available at:

<http://www.sciencedirect.com/science/article/pii/S2090536X15000477> (accessed 02.02.17).

118. D. Mainali, J. Seelenbinder, Automated fast screening method for cocaine identification in seized drug samples using a portable Fourier transform infrared (FT-IR) instrument. *Appl. Spectrosc.* 70 (2016) 916-922. DOI: 10.1177/0003702816638305.

Available at: <http://journals.sagepub.com/doi/pdf/10.1177/0003702816638305> (accessed 02.02.17).

119. Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) recommendations. 2016: USA, SWGDRUG. Available at:

<http://www.swgdrug.org/approved.htm> (accessed 02.02.17).

120. J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves, *Infrared and Raman spectroscopy in forensic science*, John Wiley & Sons, London, UK, 2012. ISBN: 978-0-470-74906-7.

121. P. Kavanagh, J. O'Brien, J. Fox, C. O'Donnell, R. Christie, J.D. Power, S.D. McDermott, The analysis of substituted cathinones. Part 3. Synthesis and characterisation of 2,3-methylenedioxy substituted cathinones. *Forensic Sci. Int.* 216 (2012) 19-28. DOI: 10.1016/j.forsciint.2011.08.011. Available at:

<http://www.sciencedirect.com/science/article/pii/S037907381100404X> (accessed 02.02.17).

122. S.D. Brandt, P.F. Daley, N.V. Cozzi, Analytical characterisation of three trifluoromethyl-substituted methcathinone isomers. *Drug Test. Anal.* 4 (2012) 525-529. DOI: 10.1002/dta.382. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.382/full> (accessed 02.02.17).

123. C.R. Maheux, C.R. Copeland, Chemical analysis of two new designer drugs: buphedrone and penthedrone. *Drug Test. Anal.* 4 (2012) 17-23. DOI: 10.1002/dta.385.

<http://onlinelibrary.wiley.com/doi/10.1002/dta.385/full> (accessed 02.02.17).

124. M.D. Hargreaves, *Drugs of abuse: application of handheld FT-IR and Raman spectrometers* in: J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves (Eds.) *Infrared and Raman spectroscopy in forensic science*, UK, John Wiley & Sons. Chapter 6, pp 339-350, 2012. ISBN 978-0-470-74906-7.

125. P. Larkin, *Infrared and Raman spectroscopy principles and spectral interpretation*, Elsevier, USA, 2011. Available at: <https://www.elsevier.com/books/infrared-and-raman-spectroscopy/larkin/978-0-12-386984-5>. ISBN: 9780123870186.

126. P. Thomas, *Learning QC: Lonza rolls out Raman for materials ID*, *Pharmaceutical Manufacturing*, (2010) 20-21.

127. S. Christesen, B. Maciver, L. Procell, D. Sorrick, M. Carrabba, J. Bello, Non-intrusive analysis of chemical agent identification sets using a portable fiber-optic Raman spectrometer. *Appl. Spectrosc.* 53 (1999) 850-855. Available at:

<http://journals.sagepub.com/doi/abs/10.1366/0003702991947432> (accessed 02.02.17).

128. S. Assi, A. Guirguis, S. Halsey, S. Fergus, J.L. Stair, Analysis of 'legal high' substances and common adulterants using handheld spectroscopic techniques, *Anal. Methods* 7 (2015) 736-746. DOI: 10.1039/C4AY02169J. Available at: <http://pubs.rsc.org/en/content/articlelanding/2014/ay/c4ay02169j#!divAbstract> (accessed 02.02.17).
129. S. Coticone, B. Morgan, N. Ludwigsen, Forensic Applications of vibrational spectroscopy techniques to identify prescription drugs and mixtures. *Internet J. Forensic Sci.* 4 (2011) 1-5. DOI: 10.1016/j.sbsr.2014.11.002. Available at: <http://www.sciencedirect.com/science/article/pii/S2214180414000415> (accessed 16.01.17).
130. B. Diehl, C. Chen, B. Grout, J. Hernandez, S. O'Neill, C. McSweeney, J.M. Alvarado, M. Smith, non-destructive material identification: in-depth focus, an implementation perspective on handheld Raman spectrometers for the verification of material identity, *Eu. Pharm. Rev.* Available at: <http://www.europeanpharmaceuticalreview.com/wp-content/uploads/Raman-Supplement-2012.pdf> (accessed 21.09.15).
131. K. Tsujikawa, K. Kuwayama, H. Miyaguchi, T. Kanamori, Y.T. Iwata, T. Yoshida, H. Inoue, Development of an on-site screening system for amphetamine-type stimulant tablets with a portable attenuated total reflection Fourier transform infrared spectrometer, *Anal. Chim. Acta.* 608 (2008) 95-103. DOI: 10.1016/j.aca.2007.12.002. Available at: <http://www.sciencedirect.com/science/article/pii/S0003267007019654> (accessed 02.02.17).
132. S.P. Stewart, S.E.J. Bell, N.C. Fletcher, S. Bouazzaoui, Y.C. Ho, S.J. Speers, K.L. Peters, Raman spectroscopy for forensic examination of β -ketophenethylamine "legal highs": Reference and seized samples of cathinone derivatives, *Anal. Chim. Acta.* 711 (2012) 1-6. DOI: 10.1016/j.aca.2011.10.018. Available at: <http://www.sciencedirect.com/science/article/pii/S0003267011013286> (accessed 02.02.17).
133. C.R. Maheux, C.R. Copeland, M.M. Pollard, Characterisation of three methcathinone analogs: 4-methylmethcathinone, methylone, and bk-MBDB, *Microgram J.* 7 (2010) 42-49. Available at: https://www.dea.gov/pr/microgram-journals/2010/mj7-2_42-49.pdf (accessed 02.02.17).
134. S.E.J. Bell, D.T. Burns, A.C. Dennis, L.J. Matchett, J.S. Speers, Composition profiling of seized ecstasy tablets by Raman spectroscopy, *Analyst.* 125 (2000) 1811-1815. DOI: 10.1039/B005662F. Available at: <http://pubs.rsc.org/en/content/articlelanding/2000/an/b005662f#!divAbstract> (accessed 02.02.17).
135. A.G. Ryder, G.M. O'Connor, T.J. Glynn, Quantitative analysis of cocaine in solid mixtures using Raman spectroscopy and chemometric methods, *J. Raman Spectrosc.* 31 (2000) 221-227. DOI: 10.1002/(SICI)1097-4555(200003)31:3<221::AID-JRS518>3.0.CO;2-5. Available at: <https://aran.library.nuigalway.ie/handle/10379/2927> (accessed 02.02.17).
136. A.G. Ryder, G.M. O'Connor, T.J. Glynn, Identification and quantitative measurements of narcotics in solid mixtures using near-IR Raman spectroscopy and multivariate analysis, *J. Forensic Sci.* 44 (1999) 1013-1019. DOI: 10.1520/JFS12031J. Available at: https://www.astm.org/DIGITAL_LIBRARY/JOURNALS/FORENSIC/PAGES/JFS12031J.htm (accessed 02.02.17).
137. M.D. Hargreaves, K. Page, T. Munshi, R. Tomsett, G. Lynch, H.G.M. Edwards, Analysis of seized drugs using portable Raman spectroscopy in an airport environment - a proof of principle study, *J. Raman Spectrosc.* 39 (2008) 873-880. DOI: 10.1002/jrs.1926. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.1926/abstract> (accessed 02.02.17).

138. S. Assi, M.D. Osselton, B. Wallis, The evaluation of dual laser handheld Raman spectroscopy for identifying novel psychoactive substances. *Am. Pharm. Rev.* 2016: p. 1-8. Available at: http://eprints.bournemouth.ac.uk/20960/1/Assi_EPR%20Issue%205%202013.pdf (accessed 02.02.17).
139. L. Elie, M. Elie, G. Cave, M. Vetter, R. Croxton, M. Baron, Microcrystalline testing used in combination with Raman micro-spectroscopy for absolute identification of novel psychoactive substances, *J. Raman Spectrosc.* 47 (2016) 1343-1350. DOI: 10.1002/jrs.4957. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.4957/full> (accessed 02.02.17).
140. L. E. Jones, A. Stewart, K. L. Peters, M. McNaul, S. J. Speers, N. C. Fletcher, S. E. J. Bell, Infrared and Raman screening of seized novel psychoactive substances: a large scale study of >200 samples, *Analyst.* 141 (2016) 902-909. DOI: 10.1039/C5AN02326B. Available at: <http://pubs.rsc.org/-/content/articlehtml/2016/an/c5an02326b> (accessed 02.02.17).
141. J. Calvo-Castro, M. Warzecha, A.J. McLean, C. McHugh, Impact of substituent effects on the Raman spectra of structurally related N-substituted diketopyrrolopyrroles. *Vib Spectrosc.* 83(2015) 8-16. DOI: 10.1016/j.vibspec.2015.12.004. Available at: <http://www.sciencedirect.com/science/article/pii/S0924203115300382> (accessed 12.06.2017).
142. M. Donahue, E. Botonjic-Sehic, D. Wells, C.W. Brown, Understanding infrared and Raman spectra of pharmaceutical polymorphs. *American Pharmaceutical Review.* 14(2011). Available at: <http://www.americanpharmaceuticalreview.com/Featured-Articles/37183-Understanding-Infrared-and-Raman-Spectra-of-Pharmaceutical-Polymorphs/> (accessed 12.06.2017).
143. J.N. Jr Willis, R.B. Cook, R. Jankow, Raman spectrometry of some common barbiturates. *Anal Chem.* 44(1972) 1228-34. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac60315a032> (accessed 12.06.2017).
144. S.A. Schönlichler, L.K.H. Bittner, A.K.H. Weiss, U.J. Griesser, J.D. Pallua, C.W. Huck, Comparison of NIR chemical imaging with conventional NIR, Raman and ATR-IR spectroscopy for quantification of furosemide crystal polymorphs in ternary powder mixtures. *Eur J Pharm Biopharm.* 84(2013) 616-625. DOI: 10.1016/j.ejpb.2013.01.006. Available at: <http://www.sciencedirect.com/science/article/pii/S0939641113000301> (accessed 12.06.2017).
145. Home Office, Circular 013/2014: Home Office approved kits for testing controlled substances, Available at: <https://www.gov.uk/government/publications/circular-0132014-the-testing-of-substances-suspected-to-be-drugs> (accessed 21.09.15).
146. N. Reandeau, Legislative solutions to the synthetic drug problem. 2015, Gordon Thomas Honeywell Governmental Affairs: Tacoma, USA.
147. E. Smith, G. Dent, *Modern Raman spectroscopy - A practical approach*, London, John Wiley & Sons, UK, 2005. ISBN 0-471-49668-5.
148. S.J. Fraser, K.C. Gordon, Simultaneous qualitative and quantitative analysis of counterfeit and unregistered medicines using Raman spectroscopy, *J. Raman Spectrosc.* 44 (2013) 1172-1180. DOI: 10.1002/jrs.4344. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.4344/full> (accessed 02.02.2017).
149. F.A. Settle, *Handbook of instrumental techniques for analytical chemistry.* 1997: Prentice Hall PTR. Upper Saddle River, NJ. Available at: <http://ceplecture.yolasite.com/resources/Handbook%20Of%20Instrumental%20Techniques%20For%20Analytical%20Chemistry%20-%20Fran%20A.Settle.pdf> ISBN 0-13-177338-0 (accessed 02.02.2017).
150. A.C. Moffat, A.C., *Molecular excitation using Raman scattering. Medicines complete. Introduction and Theory.* Clarke's Analysis of Drugs and Poisons, 2012, London, Pharmaceutical Press.

151. C.R. Maheux, C.R. Copeland, Chemical analysis of two new designer drugs: buphedrone and penthedrone. *Drug Test. Anal.* 4 (2012) 17-23. DOI: 10.1002/dta.385. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.385/full> (accessed 05.02.17).
152. R. Christie, E. Horan, J. Fox, C. O'Donnell, H.J. Byrne, S. McDermott, J. Power, P. Kavanagh, Discrimination of cathinone regioisomers, sold as 'legal highs', by Raman spectroscopy. *Drug Test. Anal.* 6 (2013) 651-657. DOI: 10.1002/dta.1518. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.1518/full> (accessed 05.02.17).
153. 149. S.E.J. Bell, S.P. Stewart, and S.J. Speers, Raman spectroscopy of drugs of abuse in: J.M. Chalmers, H.G.M. Edwards, M.D. Hargreaves (Eds.) *Infrared and Raman spectroscopy in forensic science*, John Wiley & Sons, London, UK, 2012. Chapter 6, pp 317-338, 2012. ISBN: 978-0-470-74906-7.
154. M.D. Hargreaves, A.D. Burnett, T. Munshi, J.E. Cunningham, E.H. Linfield, A.G. Davies, H.G.M. Edwards, Comparison of near infrared laser excitation wavelengths and its influence on the interrogation of seized drugs-of-abuse by Raman spectroscopy, *J. Raman Spectrosc.* 40 (2009) 1974-1983. DOI: 10.1002/jrs.2352. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.2352/abstract> (accessed 05.02.17).
155. R. Jiménez-Pérez, J.M. Sevilla, T. Pineda, M. Blazquez, J. Gonzalez-Rodriguez, Comparative study of γ -hydroxybutyric acid (GHB) and other derivative compounds by spectroelectrochemistry Raman (SERS) on platinum surface, *Electrochim. Acta.* 193 (2016) 154-159. DOI: 10.1016/j.electacta.2016.02.041. Available at: <http://www.sciencedirect.com/science/article/pii/S001346861630306> (accessed 05.02.17).
156. W.W.Y. Lee, V.A.D. Silverson, L.E. Jones, Y.C. Ho, N.C. Fletcher, M. McNaul, K.L. Peters, S.J. Speers, S.E.J. Bell, Surface-enhanced Raman spectroscopy of novel psychoactive substances using polymer-stabilized Ag nanoparticle aggregates, *Chem. Commun.* 52 (2016) 493-496. DOI: 10.1039/C5CC06745F. Available at: <http://pubs.rsc.org/is/content/articlelanding/2016/cc/c5cc06745f/unauth#!divAbstract> (accessed 05.02.17).
157. J.D. Jaganathan, G.P. Singh, Vibrational spectroscopic studies on drugs, *World J. Pharm. Pharm. Sci.* 2 (2013) 3467-3479. Available at: file:///C:/Users/ag10aan/Downloads/article_wjpps_1382332665.pdf (accessed 05.02.17).
158. B.L. Volodin, S. Dolgy, C. Lieber, H. Wu, W. Yang, Quantitative and qualitative analysis of fluorescent substances and binary mixtures by use of shifted excitation Raman difference spectroscopy, *Proc. SPIE.* 8572 (2013) 1-9. DOI:10.1117/12.2008241. Available at: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1671939> (accessed 02.02.17).
159. W. Yang, H. Wu, J. Qian, L. Chandler, C.A. Lieber, C. Dentinger, Multi-wavelength excitation Raman spectrometers and microscopes for measurements of real-world samples, *Proc. SPIE.* 8546 (2012). DOI:10.1117/12.970567. Available at: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1388162> (accessed 02.02.17).
160. E. Ali, H.G. Edwards, R. Cox, Forensic and security applications of a long-wavelength dispersive Raman system. *J. of Raman Spectrosc.* 46 (2015) 322-326. DOI: 10.1002/jrs.4632. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.4632/full> (accessed 05.02.17).
161. S. Harkai, M. Putz, Comparison of rapid detecting optical techniques for the identification of New Psychoactive Substances in 'Legal High' preparations, *Toxicchem Krimtech.* 82 (2015) 229-239. Available at: https://www.gtfc.org/cms/images/stories/media/tb/tb2015/Harkai_Putz_2015.pdf (accessed 05.02.17).

162. R.L. Green, M.D. Hargreaves, C.M. Gardner, Performance characterization of a combined material identification and screening algorithm, *Proc. SPIE*. 8726 (2013). DOI:10.1117/12.2015953. Available at: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1692493> (accessed 05.02.17).
163. R.L. Green, C.D. Brown, Raw-material authentication using a handheld Raman spectrometer, *Pharm. Technol.* 32 (2008) 1-6. Available at: <http://www.pharmtech.com/raw-material-authentication-using-handheld-raman-spectrometer>.
164. K.A. Bakeev, R.V. Chimenti, Pros and cons of using correlation versus multivariate algorithms for material identification via handheld spectroscopy, *Eu. Pharm. Rev.* (2013) 1-5. Available at: http://www.europeanpharmaceuticalreview.com/wp-content/uploads/Bakeev-Web-Article_proof.pdf (accessed 02.02.17).
165. C.M. Gryniewicz-Ruzicka, J.D. Rodriguez, S. Arzhantsev, L.F. Buhse, J.F. Kauffman, Libraries, classifiers, and quantifiers: a comparison of chemometric methods for the analysis of Raman spectra of contaminated pharmaceutical materials. *J Pharm Biomed Anal.* 61(2012) 191-8. DOI: 10.1016/j.jpba.2011.12.002. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/22206890> (accessed 02.02.17).
166. R.L. McCreery, A.J. Horn, J. Spencer, E. Jefferson, Non-invasive identification of materials inside USP vials with Raman spectroscopy and a Raman spectral library, *J. Pharm. Sci.* 87 (1998) 1-8. DOI: 10.1021/js970330q. Available at: <http://onlinelibrary.wiley.com/doi/10.1021/js970330q/abstract> (accessed 02.02.17).
167. K. Carron, R. Cox, Qualitative analysis and the answer box: a perspective on portable Raman spectroscopy, *Anal. Chem.* 82 (2010) 3419-3425. DOI: 10.1021/ac901951b. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac901951b> (accessed 02.02.17).
168. H. Inoue, Y. Maeno, M. Iwasa, R. Matoba, M. Nagao, Screening and determination of benzodiazepines in whole blood using solid-phase extraction and gas-chromatography / mass spectrometry, *Forensic Sci. Int.* 113 (2000) 367-373. DOI: 10.1016/S0379-0738(00)00226-7. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073800002267> (accessed 02.02.17).
169. B. Moosmann, S. Kneisel, U. Girreser, V. Brecht, F. Westphal, V. Auwarter, Separation and structural characterization of the synthetic cannabinoids JWH-412 and 1-[(5-fluoropentyl)-1H-indol-3yl]-(4-methylnaphthalen-1-yl) methanone using GC-MS, NMR analysis and a flash chromatography system, *Forensic Sci. Int.* 220 (2012) e17-e22. DOI: 10.1016/j.forsciint.2011.12.010. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073811006025> (accessed 02.02.17).
170. B&W Tek, The detector, in *Spectrometer knowledge*, 2017, Japan, B.W. Tek.
171. E.G. Bartick, Technical Evaluation: Rigaku Analytical Devices Progeny™ ResQ™. Application: Illegal Drug Screening. 2015, E.G. Bartick: Woodbridge, Suffolk.
172. B. Buttingsrud, *Multivariate data analysis - Level 1*. 2015, Camo: Oslo, Norway.
173. K.Y. Noonan, M. Beshire, J. Darnell, K.A. Fredericl, Qualitative and quantitative analysis of illicit drug mixtures on paper currency using Raman microspectroscopy, *Appl. Spectrosc.* 59 (2005) 1493-1497. Available at: <https://www.osapublishing.org/as/abstract.cfm?uri=as-59-12-1493> (accessed 02.02.17).
174. L.S. Taylor, F.W. Langkilde, Evaluation of solid-state forms present in tablets by Raman spectroscopy, *J. Pharm. Sci.* 89 (2000) 1342-1353. DOI: 10.1002/1520-6017(200010)89:10<1342::AID-JPS12>3.0.CO;2-X. Available at: [http://onlinelibrary.wiley.com/doi/10.1002/1520-6017\(200010\)89:10%3C1342::AID-JPS12%3E3.0.CO;2-X/full](http://onlinelibrary.wiley.com/doi/10.1002/1520-6017(200010)89:10%3C1342::AID-JPS12%3E3.0.CO;2-X/full) (accessed 02.02.17).

175. C. Eliasson, N.A. Macleod, L.C. Jayes, F.C. Clarke, S.V. Hammond, M.R. Smith, P. Matousek, Non-invasive quantitative assessment of the content of pharmaceutical capsules using transmission Raman spectroscopy, *J. Pharmaceut. Biomed.* 47 (2008) 221-229. DOI:10.1016/j.jpba.2008.01.013. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708508000447> (accessed 02.02.17).
176. M.L. O'Connell, T. Howley, A.G. Ryder, M.N. Leger, M.G. Madden, Classification of a target analyte in solid mixtures using principal component analysis, support vector machines, and Raman spectroscopy, *Proc. SPIE.* 5826 (2005). DOI:10.1117/12.605156. Available at: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=865137> (accessed 02.02.17).
177. S.E.J. Bell, J.R. Beattie, J.J. McGarvey, K.L. Peters, N.M.S. Sirimuthu, S.J. Speers, Development of sampling methods for Raman analysis of solid dosage forms of therapeutic and illicit drugs, *J. Raman Spectrosc.* 35 (2004) 409-417. DOI: 10.1002/jrs.1160. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.1160/full> (accessed 02.02.17).
178. K. Varmuza, P. Filzmoser, Introduction to multivariate statistical analysis in chemometrics, 2009, London, Taylor & Francis Group. [ISBN] Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.409.3893> (accessed 02.02.17). ISBN 9781420059472.
179. C. Muehlethaler, G. Massonnet, P. Esseiva, The application of chemometrics on Infrared and Raman spectra as a tool for the forensic analysis of paints, *Forensic Sci. Int.* 209 (2011) 173-182. DOI: 10.1016/j.forsciint.2011.01.025. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073811000387?np=y> (accessed 02.02.17).
180. F.D.S.L. Borba, R.S. Honorato, A. de Juan, Use of Raman spectroscopy and chemometrics to distinguish blue ballpoint pen inks, *Forensic Sci. Int.* 249 (2015) 73-82. DOI: 10.1016/j.forsciint.2016.03.049. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073816301360> (accessed 02.02.17).
181. N. Scoutaris, The use of chemometrics in extracting information from Raman data, *Eu. Pharm. Rev.* 20 (2015) 39-42. Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0.084923213603&partnerID=40&md5=5fed0013a47150a53a4ec3129987b577> (accessed 02.02.17).
182. S.J. Fraser, K.C. Gordon, Raman spectroscopy in the study of pharmaceuticals: the problems and solutions to sub-sampling and data analysis, *Eu. Pharm. Rev.* 19 (2014) 3-7. Available at: https://www.researchgate.net/publication/285964972_Raman_spectroscopy_in_the_study_of_pharmaceuticals_the_problems_and_solutions_to_sub-sampling_and_data_analysis (accessed 02.02.17).
183. T. Howley, M.G. Madden, M.L. O'Connell, A.G. Ryder, The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data, *Knowl-Based Syst.* 19 (2006) 363-370. DOI: 10.1016/j.knosys.2005.11.014. Available at: <http://www.sciencedirect.com/science/article/pii/S0950705106000141>.
184. M.L. O'Connell, A.G. Ryder, M.N. Leger, T. Howler, Qualitative analysis using Raman spectroscopy and chemometrics: a comprehensive model system for narcotics analysis. *Appl. Spectrosc.* 64 (2010) 1109-1121. Available at: <http://journals.sagepub.com/doi/abs/10.1366/000370210792973541> (accessed 02.02.17).
185. A.C. Moffat, M.D. Osselton, B. Widdop, Clarke's Analysis of Drugs and Poisons, 4th ed., EGC Clarke, 2011, London, Pharmaceutical Press. ISBN 978 0 85369 711 4.
186. N.V.S. Rodrigues, E.M. Cardoso, M.V.O. Andare, C.L. Donnici, M.M. Sena,

- Analysis of seized cocaine samples by using chemometric methods and FTIR spectroscopy, *J. Braz. Chem. Soc.* 24 (2013) 507-517. DOI: 10.5935/0103-5053.20130066. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-50532013000300019 (accessed 02.02.17).
187. R. Risoluti, S. Materazzi, A. Gregori L. Ripani, Early detection of emerging street drugs by near infrared spectroscopy and chemometrics, *Talanta*. 153 (2016) 407-413. DOI: 10.1016/j.talanta.2016.02.044. Available at: <http://www.sciencedirect.com/science/article/pii/S0039914016301126> (accessed 02.02.17).
188. D.E. Bugay, R.C. Brush, Chemical identity testing by remote-based dispersive Raman spectroscopy, *Appl. Spectrosc.* 64 (2010) 467-475. Available at: <http://journals.sagepub.com/doi/abs/10.1366/000370210791211745> (accessed 02.02.17).
189. D. Zhang, D. Ben-Amotz, Enhanced chemical classification of Raman images in the presence of strong fluorescence interference, *Appl. Spectrosc.* 54 (2000) 1379-1383. Available at: <http://journals.sagepub.com/doi/abs/10.1366/0003702001951066> (accessed 02.02.17).
190. M.N. Leger, A.G. Ryder, Comparison of Derivative Pre-processing and Automated Polynomial Baseline Correction Method for Classification and Quantification of Narcotics in Solid Mixtures, *Appl. Spectrosc.* 60 (2006) 182-193. Available at: <https://www.osapublishing.org/as/abstract.cfm?uri=as-60-2-182> (accessed 02.02.17).
191. M.G. Madden, A.G. Ryder, Machine learning methods for quantitative analysis of Raman spectroscopy data, *Proc. SPIE*. 4876 (2003). DOI:10.1117/12.464039. Available at: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=879148> (accessed 16.01.17).
192. A.G. Ryder, Classification of narcotics in solid mixtures using principal component analysis and Raman spectroscopy, *J. Forensic Sci.* 47 (2002) 275-284. DOI: 10.1520/JFS15244J. Available at: https://www.astm.org/DIGITAL_LIBRARY/JOURNALS/FORENSIC/PAGES/JFS15244J.htm (accessed 02.02.17).
193. P. Vandenabeel, A. Hardy, H.G.M. Edwards, L. Moens, Evaluation of a principal components-based searching algorithm for Raman spectroscopic identification of organic pigments in 20th century artwork, *Appl. Spectrosc.* 55 (2001) 525-533. Available at: <http://journals.sagepub.com/doi/pdf/10.1366/0003702011952307> (accessed 02.02.17).
194. W. Krzanowski, *Principles of multivariate analysis*, 2nd ed., Oxford Statistical Science Series, Oxford University Press, Oxford, 2000. ISBN 0-19-850708-9.
195. Seber, G.A., *Multivariate observations*, 2009, London, John Wiley & Sons. ISBN: 978-0-471-69121-1.
196. N.A. Northrop, B.K. Yamamoto, Methamphetamine effects on blood-brain barrier structure and function. *Frontiers Neurosci.* 9 (2015) 69. DOI: 10.3389/fnins.2015.00069. Available at: <http://journal.frontiersin.org/article/10.3389/fnins.2015.00069/full> (accessed 02.02.17).
197. K. Varmuza, M. Karlovits, W. Demuth, Spectral similarity versus structural similarity: infrared spectroscopy, *Anal. Chim. Acta* 409 (2003) 313-324. DOI: 10.1016/S0003-2670(03)00668-8. Available at: <http://www.sciencedirect.com/science/article/pii/S0003267003006688> (accessed 02.02.17).
198. D.H. Rouvray, Similarity studies. The necessity of analogies in the development of science, *J. Chem. Inf. Comp. Sci.* 34 (1994) 446-452. DOI: 10.1021/ci00018a036. Available at: <http://pubs.acs.org/doi/abs/10.1021/ci00018a036> (accessed 02.02.17).
199. A. Bourne, D. Reid, F. Hickson, S. Torres-Rueda, P. Weatherburn, Illicit drug use in sexual settings ('chemsex') and HIV/STI transmission risk behaviour among gay men in South London: findings from a qualitative study, *Sex. Transm. Infect.* 91 (2015) 564-568.

200. V. Schoonjans, F. Questier, A.P. Borosy, B. Walczak, D.L. Massart, B.D. Hudson, Use of mass spectrometry for assessing similarity/diversity of natural products with unknown chemical structures, *J. Pharm. Biomed.* 21 (2000) 1197-1214. DOI: 10.1016/S0731-7085(99)00236-8. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708599002368> (accessed 02.02.17).
201. G.M. Downs, P. Willett, W. Fisanick, Similarity Searching and Clustering of Chemical-Structure Databases Using Molecular Property Data, *J. Chem. Inf. Comput. Sci.* 34 (1994) 1094-1102. DOI: 10.1021/ci00021a011. Available at: <http://pubs.acs.org/doi/abs/10.1021/ci00021a011> (accessed 02.02.17).
202. V. Schoonjans, D.L. Massart, Combining spectroscopic data (MS, IR): exploratory chemometric analysis for characterising similarity/diversity of chemical structures, *J. Pharm. Biomed.* 26 (2001) 225-239. DOI: 10.1016/S0731-7085(01)00427-7. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708501004277> (accessed 02.02.17).
203. D.H. Drewry, S. Stanley Young, Approaches to the design of combinatorial libraries, *Chemometr. Intell. Lab.* 48 (1999) 1-20. DOI: 10.1016/S0169-7439(99)00010-6. Available at: <http://www.sciencedirect.com/science/article/pii/S0169743999000106> (accessed 02.02.17).
204. M. Zloh, E.G. Samaras, J. Calvo-Castro, A. Guirguis, J.L. Stair, S.B. Kirton, Drowning in diversity? A systematic way of clustering and selecting a representative set of New Psychoactive Substances. *Sci. Rep.* (Under review), January 2017.
205. V. Schoonjans, F. Questier, Q. Guo, Y. Van der Heyden, D.L. Massart, Assessing molecular similarity/diversity of chemical structures by FT-IR spectroscopy, *J. Pharm. Biomed.* 24 (2001) 613-627. DOI: 10.1016/S0731-7085(00)00437-4. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708500004374> (accessed 02.02.17).
206. V.L. Brewster, H.G.M. Edwards, M.D. Hargreaves, T. Munshi, Identification of the date-rape drug GHB and its precursor GBL by Raman spectroscopy, *Drug Test. Anal.* 1 (2009) 25-31. DOI: 10.1002/dta.11. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.11/full> (accessed 02.02.17).
207. H. Markert, J. Ring, N. Campbell, K. Grates, Comparison of four commercially available portable Raman spectrometers, National Criminal Justice Reference Service (NCJRS), 2011, National Forensic Science Technology Centre (NFSTC): Florida. p. 1-8.
208. S. Daily, Class A - National drugs intelligence bulletin. Q1 2013, S. Daily, Editor. 2013, Teddington, UK, LGC Limited.
209. S. Daily, Class A - National drugs intelligence bulletin. Q1 2014, S. Daily, Editor. 2014, Teddington, UK, LGC Limited.
210. K.G. Ray, R.L. McCreery, Simplified calibration of instrument response function for Raman spectrometers based on luminescent intensity, *Appl. Spectrosc.* 51 (1997) 108-116. Available at: <https://www.osapublishing.org/as/abstract.cfm?uri=as-51-1-108> (accessed 02.02.17).
211. J.L. Stair, A. Guirguis, J. Calvo-Castro, E.G. Samaras, M. Zloh, S.B. Kirton, Practical Considerations for the Detection of New Psychoactive Substances (NPS) using Raman Spectroscopy. 2016, University of Hertfordshire, UK. p. 1-2. Available at: http://www.eumadness.eu/wp-content/uploads/2016/06/TechnicalNote_NPSRamanSpectroscopy-310516.pdf (accessed 05.02.17).
212. Rigaku, Search algorithm comparison, in Technical Guide. 2015, Rigaku Analytical Devices, USA.
213. S. Daily, Class A - National drugs intelligence bulletin. Q1 2016, 2016, S. Daily, Teddington, UK, LGC Limited.

214. P. Sainsbury, A. Kicman, R. Archer, L. King and R. Braithwaite, Aminoindanes—the next wave of ‘legal highs’? *Drug Test Anal.* 3 (2011) 479-482. DOI: 10.1002/dta.318. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.318/full> (accessed 02.02.17).
215. T. Biermann, B. Schwarze, B. Zedler, P. Betz, On-site testing of illicit drugs: the use of the drug-testing device “Toxiquick®”, *Forensic Sci. Int.* 143 (2004) 21-25. DOI: 10.1016/j.forsciint.2004.01.013. Available at: <http://www.sciencedirect.com/science/article/pii/S037907380400057X> (accessed 02.02.17).
216. T.M. Woo, J.R. Hanley, “How High Do They Look?”: Identification and Treatment of Common Ingestions in Adolescents, *Journal of Paediatric Health Care* 27 (2013) 135-144. DOI: 10.1016/j.pedhc.2012.12.002. Available at: <http://www.sciencedirect.com/science/article/pii/S089152451200291X>.
217. O.I. Khreit, C. Irving, E. Schmidt, J.A. Parkinson, N.N. Daeid, O.B. Sutcliffe, Synthesis, full chemical characterisation and development of validated methods for the quantification of the components found in the evolved “legal high” NRG-2, *J. Pharmaceut. Biomed.* 61 (2012) 122-135. DOI: 10.1016/j.jpba.2011.11.004. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708511006315> (accessed 02.02.17).
218. Home Office, Introduction of new powers to allow law enforcement agencies to seize, detain and destroy chemical substances suspected of being used as drug cutting agents: Consultation response. 2014. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/298900/Cutting_agents_consultation_response.pdf (accessed 21.09.15).
219. M. Takahashi, M. Nagashima, J. Suzuki, T. Seto, I. Yasuda, T. Yoshida, Creation and application of psychoactive designer drugs data library using liquid chromatography with photodiode array spectrophotometry detector and gas chromatography–mass spectrometry, *Talanta.* 77 (2009) 1245-1272. DOI: 10.1016/j.talanta.2008.07.062. <http://www.sciencedirect.com/science/article/pii/S0039914008005808?np=y> (accessed 02.02.17).
220. T. Nogami, F. Muta, Pharmaceutical raw material inspection with handheld Raman spectrometer, *Rigaku Journal.* 29 (2013) 16-18. Available at: http://rigaku.com/downloads/journal/RJ29-2/RigakuJournal29-2_16-18.pdf (accessed 02.02.17).
221. P. Vandenabeele, H.G. Edwards, J. Jehlicka, The role of mobile instrumentation in novel applications of Raman spectroscopy: archaeometry, geosciences, and forensics, *Chem. Soc. Rev.* 43 (2014) 2628-2649. DOI: 10.1039/C3CS60263J. Available at: <http://pubs.rsc.org/en/content/articlelanding/2014/cs/c3cs60263j/unauth#!divAbstract> (accessed 02.02.17).
222. J.W. Daly, W.L. Padgett, M.T. Shamim, Analogues of Caffeine and Theophylline: Effect of Structural Alterations on Affinity at Adenosine Receptors, *Am. Chem. Soc.* 29 (1986) 1305-1306. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/3806581> (accessed 02.02.17).
223. J.D. Rodriguez, B.J. Westenberger, L.F. Buhse, J.F. Kauffman, Quantitative evaluation of the sensitivity of library-based Raman spectral correlation methods. *Anal. Chem.* 83 (2011), 4061-4067. DOI: 10.1021/ac200040b. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac200040b> (accessed 02.02.17).
224. A.M. Araújo, M.J. Valente, M. Carvalho, D.D. da Silva, H. Gaspar, F. Carvalho, M. de L. Bastos, P.G. de Pinho, Raising awareness of new psychoactive substances: chemical analysis and in vitro toxicity screening of ‘legal high’ packages containing synthetic cathinones, *Arch. Toxicol.* 89 (2014) 775-771. DOI: 10.1007/s00204-014-1278-7. Available at: <http://link.springer.com/article/10.1007/s00204-014-1278-7> (accessed 02.02.17).

225. D. W. Hill, B. T. McSharry, L. S. Trzuppek, Quantitative analysis by isotopic dilution using mass spectroscopy: The determination of caffeine by GC-MS, *J. Chem. Educ.* 65 (1988) 907-910. DOI: 10.1021/ed065p907. Available at: <http://pubs.acs.org/doi/abs/10.1021/ed065p907?journalCode=jceda8> (accessed 02.02.17).
226. J.F. Casale, P.A. Hays, Characterisation of the “methylenedioxy-2-aminoindans”, *Microgram. J.* 8 (2011) 43. Available at: <http://forendex.southernforensic.org/uploads/references/MicrogramJournal/8.2.43.52.pdf> (accessed 02.02.17).
227. Y.J. Wu, Y.Y. Cheng, S. Zeng, M.M. Ma, Determination of dextromethorphan and its metabolite dextromethorphan in human urine by capillary gas chromatography without derivatization, *J. Chromatogr. B.* 784 (2003) 219-224. DOI: 10.1016/S1570-0232(02)00797-3. Available at: <http://www.sciencedirect.com/science/article/pii/S1570023202007973?np=y> (accessed 02.02.17).
228. H. M. Cartwright, M. Sztandera, *Soft Computing Approaches in Chemistry*, Springer-Verlag Berlin Heidelberg, Germany, 2003. ISBN 978-3-642-53507-9.
229. S.S. Khan, M.G. Madden, New similarity metrics for Raman spectroscopy. *Chemom. Intell. Lab. Syst.* 114 (2012) 99-108. DOI: 10.1016/j.chemolab.2012.03.007. Available at: <http://www.sciencedirect.com/science/article/pii/S0169743912000524> (accessed 02.02.17).
230. P. Adamowicz, D. Zuba, B. Byrska, Fatal intoxication with 3-methyl-N-methylcathinone (3-MMC) and 5-(2-aminopropyl) benzofuran (5-APB). *Forensic Sci. Int.* 245 (2014) 126-132. DOI: 10.1016/j.forsciint.2014.10.016. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073814004228> (accessed 02.02.17).
231. W.L. Chan, D.M. Wood, S. Hudson, P.I. Dargan, Acute psychosis associated with recreational use of benzofuran 6-(2-aminopropyl) benzofuran (6-APB) and cannabis, *J. Med. Toxicol.* 9 (2013) 278-281. DOI: 10.1007/s13181-013-0306-y. Available at: <http://link.springer.com/article/10.1007/s13181-013-0306-y> (accessed 02.02.17).
232. D.E. Nichols, M.P. Johnson, R. Oberlender, 5-Iodo-2-aminoindan, a nonneurotoxic analogue of p-iodoamphetamine, *Pharmacol. Biochem. Behav.* 38 (1991) 135-139. DOI: 10.1016/0091-3057(91)90601-W. Available at: <http://www.sciencedirect.com/science/article/pii/009130579190601W> (accessed 02.02.17).
233. S. Anne, R. Tse, A.D. Cala, A fatal case of isolated methiopropamine (1-(thiophen-2-yl)-2-methylaminopropane) toxicity: a case report. *Am. J. Forensic Med. Pathol.* 36 (2015) 205-206. DOI: 10.1097/PAF.0000000000000170. Available at: http://journals.lww.com/amjforensicmedicine/Abstract/2015/09000/A_Fatal_Case_of_Isolate_d_Methiopropamine.19.aspx (accessed 02.02.17).
234. G.P. Bailey, J.H. Ho, S. Hudson, A. Dines, J.Rh. Archer, P.I. Dargan, D.M. Wood, Nopaine no gain: recreational ethylphenidate toxicity, *Clin. Toxicol.* 53 (2015) 498-499. DOI: 10.3109/15563650.2015.1033062. Available at: <http://www.tandfonline.com/doi/pdf/10.3109/15563650.2015.1033062?needAccess=true> (accessed 02.02.17).
235. K.E. Kellett, Development of chemical sensors for rapid identification of amphetamine-related new psychoactive substances, Unpublished Ph.D. thesis, University of Hertfordshire. UK, 2016.
236. K. Kellett, J.D. Owen, J.L. Stair, Determination of Molecular and Elemental Constituents in Aminoindane ‘legal high’ Products, (in preparation).
237. J.F. Casale, P.A. Hays, The characterisation of 4-and 5-iodo-2-aminoindan, *Microgram. J.* 9 (2012) 18. Available at: https://www.dea.gov/pr/microgram-journals/2012/mj9-1_18-26.pdf (accessed 02.02.17).
238. J. F. Casale, P. A. Hays, Methiopropamine: An analytical profile, *Microgram Journal* 8 (2011) 53-57. Available at:

- https://www.researchgate.net/profile/John_Casale/publication/258111648_Methiopropamine_An_Analytical_Profile/links/0046352836f2c60f65000000.pdf (accessed 02.02.17).
239. DrugWatch, Drug watch information sheet - MPA, Drug Watch, UK, 2014 1-6. Available at:
https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiVweuS8PjRAhUkLMAKHTWkB7cQFggcMAA&url=http%3A%2F%2Fwww.sdf.org.uk%2Findex.php%2Fdownload_file%2Fview%2F670%2F106%2F&usg=AFQjCNGxnXJ30W29W-TiICEBAsOHsNfcw&cad=rja accessed 05.02.17).
240. J. Casale, P. Hays, The Characterization of 5-and 6-(2-Aminopropyl) 2, 3-dihydrobenzofuran, *Microgram J.* 8 (2011) 62. Available at:
https://www.researchgate.net/profile/John_Casale/publication/258111355_The_Characterization_of_6-2-Aminopropylbenzofuran_and_Differentiation_from_its_4-5- and 7-Positional_Analogues/links/00b49528295a7788fa000000.pdf (accessed 16.01.17).
241. J.F. Casale, P.A. Hays, The Characterisation of α -Pyrrolidinopentiophenone, *Microgram J.* 9 (2012) 33-38. Available at:
https://www.dea.gov/pr/microgram-journals/2012/mj9-1_33-38.pdf (accessed 16.01.17).
242. D. McQuade, S. Hudson, P. I. Dargan, D. M. Wood, First European case of convulsions related to analytically confirmed use of the synthetic cannabinoid receptor agonist AM-2201, *Eur. J. Clin. Pharmacol.* 69 (2013) 373. DOI: 10.1007/s00228-012-1379-2. Available at:
<http://link.springer.com/article/10.1007/s00228-012-1379-2> (accessed 16.01.17).
243. C.A. Lieber, A. Mahadevan-Jansen, Automated method for subtraction of fluorescence from biological Raman spectra. *Appl. Spectrosc.* 57 (2003) 1363-1367. Available at: <https://www.osapublishing.org/as/abstract.cfm?uri=as-57-11-1363> (accessed 16.01.17).
244. H.B. Klinke, I.B. Müller, S. Steffenrud, R. Dahl-Sørensen, Two cases of lysergamide intoxication by ingestion of seeds from Hawaiian Baby Woodrose, *Forensic Sci. Int.* 197 (2010) e1-e5. DOI: 10.1016/j.forsciint.2009.11.017. Available at:
<http://www.sciencedirect.com/science/article/pii/S0379073809004745?np=y> (accessed 16.01.17).
245. United Nations Office on Drugs and Crimes (UNODC), *Global Synthetic Drugs Assessment: Amphetamine-type stimulants and new psychoactive substances.* 2014, Vienna, UNODC. Available at:
https://www.unodc.org/documents/scientific/2014_Global_Synthetic_Drugs_Assessment_web.pdf (accessed 21.09.15).
246. Advisory Council on the Misuse of Drugs (ACMD). Letter to Home Secretary. Home Office: London. Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/119087/methoxetamine2012.pdf (accessed 21.09.15).
247. D. Zuba, B. Byrska, Prevalence and co-existence of active components of 'legal highs', *Drug Test. Anal.* 5 (2013) 420-429. DOI: 10.1002/dta.1365. Available at:
<http://onlinelibrary.wiley.com/doi/10.1002/dta.1365/full> (accessed 16.01.17).
248. J. Corkery, *Novel Psychoactive Substances*, Personal communication with A Guirguis. 2014.
249. S.D. Brandt, H.R. Sumnall, F. Measham, J. Cole, Analyses of second-generation 'legal highs' in the UK: Initial findings, *Drug Test. Anal.* 2 (2010) 377-382. DOI: 10.1002/dta.155. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/dta.155/full> (accessed 16.01.17).

250. M. McGraw, L. McGraw, Bath salts: not as harmless as they sound, *Journal of Emergency Nursing*, 38 (2012) 582-588. DOI: 10.1016/j.jen.2012.07.025. Available at: [http://www.intljourtranur.com/article/S0099-1767\(12\)00394-7/abstract](http://www.intljourtranur.com/article/S0099-1767(12)00394-7/abstract) (accessed 16.01.17).
251. J.B. Zawilska, J. Wojcieszak, Designer cathinones - An emerging class of novel recreational drugs, *Forensic Sci. Int.* 231 (2013) 42-53. DOI: 10.1016/j.forsciint.2013.04.015. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073813002387> (accessed 16.01.17).
252. L.R. Watterson, L. Hood, K. Sewalia, S.E. Tomek, S. Yahn, C.T. Jognson, S. Wegner, B.E. Blough, J.A. Marusich, M.F. Olive, The Reinforcing and Rewarding Effects of Methylone, a Synthetic Cathinone Commonly Found in "Bath Salts", *J. Addict. Res. Ther. S* 9 (2012) 1-18. DOI: 10.4172/2155-6105.S9-002. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3828752/> (accessed 16.01.17).
253. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), EMCDDA - Europol, EMCDDA - Europol Joint Report on a new psychoactive substance: MDPV (3,4-methylenedioxypropylvalerone). Luxembourg, EMCDDA. Available at: http://www.emcdda.europa.eu/system/files/publications/819/TDAS14001ENN_466653.pdf (accessed 21.09.15).
254. B.M. Murphy, S.W. Prescott, I. Larson, Measurement of lactose crystallinity using Raman spectroscopy, *J. Pharmaceut. Biomed.* 38 (2005) 186-190. DOI: 10.1016/j.jpba.2004.12.013. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708504006867?np=y> (accessed 16.01.17).
255. J. Gorgol, Formulation, microscopy and moisture, Surface Measurement Systems Ltd. UK, Hull. Available from: <http://iformulate.biz/news-and-views/formulation-microscopy-moisture/> (accessed: 20/02/2015).
256. U.B. Cappel, I.M. Bell, L.K. Pickard, Removing cosmic ray features from Raman map data by a refined nearest neighbour comparison method as a precursor for chemometric analysis, *Appl. Spectrosc.* 64 (2010) 195-200. Available at: <http://journals.sagepub.com/doi/abs/10.1366/000370210790619528> (accessed: 20/02/2015).
257. M.C.M. Grimbergen, C.F.P. Van Swol, C. Kendall, R.M. Verdaasdonk, N. Stone, J.L.H.R. Bosch, Signal-to-noise contribution of principal component loads in reconstructed near-infrared Raman tissue spectra, *Appl. Spectrosc.* 64 (2010) 8-14. Available at: <http://journals.sagepub.com/doi/abs/10.1366/000370210790572052> (accessed: 20/02/2015).
258. P. Heraud, B.R. Wood, J. Beardall, D. McNaughton, Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells, *J. Chemometr.* 20 (2006) 193-197. DOI: 10.1002/cem.990. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/cem.990/full> (accessed: 20/02/2015).
259. B. Li, A. Calvet, Y. Casamayou-Boucau, C. Morris, A.G. Ryder, Low-Content Quantification in Powders Using Raman Spectroscopy: A Facile Chemometric Approach to Sub 0.1% Limits of Detection, *Anal. Chem.* 87 (2015) 3419-3428. DOI: 10.1021/ac504776m. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac504776m> (accessed: 20/02/2015).
260. British Pharmacopoeia, Appendix II H. Raman Spectrometry, 2014, UK, London, The Stationery Office.
261. N.K. Afseth, V.H. Segtnan, J.P. Wold, Raman spectra of biological samples: A study of pre-processing methods, *Appl. Spectrosc.* 60 (2006) 1358-1367. Available at: <http://journals.sagepub.com/doi/abs/10.1366/000370206779321454> (accessed: 20/02/2015).
262. T. Bocklitz, A. Walter, K. Hartmann, P. Rosch, J. Popp, How to pre-process Raman spectra for reliable and stable models?, *Anal. Chim. Acta* 704 (2011) 47-56. <http://dx.doi.org/10.1016/j.aca.2011.06.043>

- <http://www.sciencedirect.com/science/article/pii/S0003267011008749> (accessed: 20/02/2015).
263. S. Romero-Torres, J.D. Perez-Ramos, K.R. Morris, E.R. Grant, Raman spectroscopy for tablet coating thickness quantification and coating characterization in the presence of strong fluorescent interference, *J. Pharmaceut. Biomed.* 41 (2006) 811-819. DOI:10.1016/j.jpba.2006.01.033. Available at: <http://www.sciencedirect.com/science/article/pii/S0731708506000926> (accessed: 20/02/2015).
264. Anonymous, Advanced Pre-processing: Noise, Offset, and Baseline Filtering, Eigenvector Research, Inc. Manson, WA. Available at: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Noise,_Offset,_and_Baseline_Filtering (accessed 29/07/2015).
265. H. Mark, J. Workman, Derivatives in spectroscopy, *Spectroscopy*, 18 (2003) 32-37. Available at: http://images.alfresco.advanstar.com/alfresco_images/pharma/2014/08/22/4f351328-85b3-485c-9cac-1a011970f60c/article-69304.pdf (accessed 02.02.17).
266. V. Mazet, C. Carteret, D. Brie, J. Idier, B. Humbert, Background removal from spectra by designing and minimising a non-quadratic cost function, *Chemometr. Intell. Lab. 76* (2005) 121-133. DOI: 10.1016/j.chemolab.2004.10.003. Available at: <http://www.sciencedirect.com/science/article/pii/S0169743904002412?np=y> (accessed 02.02.17).
267. P. Vandenabeele, J. Jehlicka, P. Vitek, H.G.M. Edwards, On the definition of Raman spectroscopic detection limits for the analysis of biomarkers in solid matrices, *Planet. Space Sci.* 62 (2012) 48-54. DOI:10.1016/j.pss.2011.12.006. <http://www.sciencedirect.com/science/article/pii/S0032063311003485?np=y> (accessed 02.02.17).
268. Camo, The Unscrambler X 10.3, Camo: Oslo, Norway, 2015.
269. J. Johansson, S. Pettersson, S. Folestad, Characterization of different laser irradiation methods for quantitative Raman tablet assessment, *J. Pharmaceut. Biomed.* 39 (2005) 510-516. DOI: 10.1016/j.jpba.2005.04.029. Available at: <http://www.sciencedirect.com/science/article/pii/S073170850500292X> (accessed 02.02.17).
270. K.H. Liland, T. Almøy, B.-H. Mevik, Optimal choice of baseline correction for multivariate calibration of spectra, *Appl. Spectrosc.* 64 (2010) 1007-1016. Available at: <http://journals.sagepub.com/doi/abs/10.1366/000370210792434350> (accessed 02.02.17).
271. E. Podstawka, M. Swiatlowska, E. Borowiec, L.M. Proniewicz, Food additives characterization by infrared, Raman, and surface-enhanced Raman spectroscopies, *J. Raman Spectrosc.* 38(2007) 356-363. DOI: 10.1002/jrs.1653. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jrs.1653/full> (accessed 02.02.17).
272. A.R. Leach, V.J. Gillet, An introduction to chemoinformatics, 2nd ed., Springer Science & Business Media, UK, 2007. ISBN 978-1-4020-6291-9.
273. EMCDDA-EDND. European Monitoring centre on Drugs and Drug Addiction - EDND European Information System on New Drugs (access restricted to registered users only). 2015.
274. K. Varmuza, P.N. Penchev, H. Scsibrany, Maximum common substructures of organic compounds exhibiting similar infrared spectra, *J. Chem. Inf. Comput. Sci.* 38 (1998) 420-427. DOI: 10.1021/ci9700889. Available at: <http://pubs.acs.org/doi/abs/10.1021/ci9700889> (accessed 02.02.17).
275. S. Odoardi, F.S. Romolo, S. Strano-Rossi, A snapshot on NPS in Italy: Distribution of drugs in seized materials analysed in an Italian forensic laboratory in the period 2013–2015, *Forensic Sci. Int.* 265 (2016) 116-120. DOI: 10.1016/j.forsciint.2016.01.037.

- <http://www.sciencedirect.com/science/article/pii/S0379073816300111> (accessed 02.02.17).
276. E. Bertol, F. Mari, R.B. Berto, G. Mannaioni, F. Vaiano, D. Favretto, A mixed MDPV and benzodiazepine intoxication in a chronic drug abuser: determination of MDPV metabolites by LC–HRMS and discussion of the case, *Forensic Sci. Int.* 243 (2014) 149–155. DOI: 10.1016/j.forsciint.2014.08.002.
- <http://www.sciencedirect.com/science/article/pii/S0379073814003168> (accessed 02.02.17).
277. Z.-P. Chen, L.-M. Li, J.-W. Jin, A. Nordon, D. Littlejohn, J. Yang, J. Zhang, R.-Q. Yu, Quantitative Analysis of Powder Mixtures by Raman Spectrometry: the influence of particle size and its correction, *Anal. Chem.* 84 (2012) 4088–4094. DOI: 10.1021/ac300189p. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac300189p> (accessed 02.02.17).
278. G. Socrates, *Infrared and Raman characteristics group frequencies: Tables and charts*, 3rd ed., 2015, Middlesex, UK, Wiley. ISBN: 978-0-470-09307-8.
279. M. Joshi, B. Cetroni, A. Camacho, C. Krueger, A.J. Midey, Analysis of synthetic cathinones and associated psychoactive substances by ion mobility spectrometry, *Forensic Sci. Int.* 244 (2014) 196–206. DOI: 10.1016/j.forsciint.2014.08.033. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073814003600> (accessed 02.02.17).
280. Advisory Council on the Misuse of Drugs (ACMD): Consideration of the cathinones. 2010, London, Home Office. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/119173/acmd-cathinodes-report-2010.pdf (accessed 21.09.15).
281. S. Kumar, S.B. Rai, Spectroscopic studies of L-arginine molecule, *Indian J. Pure Ap. Phy.* 48 (2010) 251–255. Available at: <http://nopr.niscair.res.in/handle/123456789/7643> (accessed 16.01.17)
282. C. Miliano, G. Serpelloni, C. Rimondo, M. Mereu, M. Marti, M.A. De Luca, Neuropharmacology of new psychoactive substances (NPS): focus on the rewarding and reinforcing properties of cannabimimetics and amphetamine-like stimulants, *Front. Neurosci.* 10 (2016) 153. DOI: 10.3389/fnins.2016.00153. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4835722/> (accessed 02.02.17).
283. V.M.R. Zancajo, J. Brito, M.P. Carrasco, M.R. Bronze, R. Moreira, A. Lopes, Analytical profiles of “legal highs” containing cathinones available in the area of Lisbon, Portugal, *Forensic Sci. Int.* 244 (2014) 102–110. DOI: 10.1016/j.forsciint.2014.08.010. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073814003247> (accessed 02.02.17).
284. J. Klavž, M. Gorenjak, M. Marinšek, Suicide attempt with a mix of synthetic cannabinoids and synthetic cathinones: Case report of non-fatal intoxication with AB-CHMINACA, AB-FUBINACA, alpha-PHP, alpha-PVP and 4-CMC, *Forensic Sci. Int.* 265 (2016) 121–124. DOI: 10.1016/j.forsciint.2016.01.018. Available at: <http://www.sciencedirect.com/science/article/pii/S0379073816000372> (accessed 02.02.17).
285. S. Daily, Class A - National drugs intelligence bulletin. Q1 2015, S. Daily, Editor. 2015, Teddington, UK, LGC Limited.
286. C. Cole, CUT, A guide to adulterants, bulking agents and other contaminants found in illicit drugs. Available at: <http://www.cph.org.uk/wp-content/uploads/2012/08/cut-a-guide-to-the-adulterants-bulking-agents-and-other-contaminants-found-in-illicit-drugs.pdf> (accessed 21.09.15).
287. M. Coppola, R. Mondola, Synthetic cathinones: Chemistry, pharmacology and toxicology of a new class of designer drugs of abuse marketed as “bath salts” or “plant food”, *Toxicology Letters.* 211(2012) 144–149. DOI: 10.1016/j.toxlet.2012.03.009. Available at: <http://www.sciencedirect.com/science/article/pii/S0378427412001117?np=y> (accessed 02.02.17).

288. emc, Metoclopramide 10 mg tablets. 2016, Actavis UK Ltd, Devon, UK. Available at: <http://www.medicines.org.uk/emc/medicine/24123> (accessed 02.02.17).
289. emc, Diltiazem hydrochloride tablets 60 mg. 2016, Actavis UK Ltd, Devon: UK. Available at: <http://www.medicines.org.uk/emc/medicine/23945> (accessed 02.02.17).
290. E. Tyrkko, M. Andersson, R. Kronstrand, The toxicology of new psychoactive substances: synthetic cathinones and phenethylamines, *Ther. Drug Monit.* 38 (2016) 190-216. DOI: 10.1097/FTD.0000000000000263. Available at: http://journals.lww.com/drug-monitoring/Abstract/2016/04000/The_Toxicology_of_New_Psychoactive_Substances_.3.aspx (accessed 02.02.17).
291. C. Davidson, Developing treatments for stimulant abuse: A brief overview, *East Asian Arch. Psychiatry.* 26 (2016) 59. Available at: <https://search.informit.com.au/documentSummary;dn=170508764209385;res=IELHEA> (accessed 02.02.17).
292. D.N., Rutledge, D.J. Bouveresse, Independent components analysis with the JADE algorithm, *TrAC Trends Anal. Chem.* 50 (2013) 22-32. DOI: 10.1016/j.trac.2013.03.013. Available at: <http://www.sciencedirect.com/science/article/pii/S0165993613001222> (accessed 02.02.17).
293. Y. Qiu, D. Rajagopalan, S.C. Connor, D. Damian, L. Zhu, A. Handzel, G. hu, A. Amanullah, S. Bao, N. Woody, D. MacLean, K. Lee, D. Vanderwall, T. Ryan, Multivariate classification analysis of metabolomic data for candidate biomarker discovery in type 2 diabetes mellitus. *Metabolomics*, 4(2008) 337-346.

10. Appendices

10.1. Appendix A – Chapter 2

Table A2.1: Cutting agents and adulterants analysed in this study

Name	Supplier	Batch Number
Adulterants		
Benzocaine (BEN)	Sigma-Aldrich (Dorset, UK)	SLBB1067V
Caffeine anhydrous (CAF)	Fluka Analytical (Dorset, UK)	1428211V
Diltiazem HCl (DIL)	Medicines Testing Lab (UK)	N/A
Lidocaine HCl (LID)	Sigma-Aldrich (Dorset, UK)	050M0098V
Paracetamol (PAR)	Sigma-Aldrich (Dorset, UK)	SLBB2780V
Phenacetin (PHE)	Sigma-Aldrich (Dorset, UK)	STBB2177V
Procaine HCl (PRO)	Sigma-Aldrich (Dorset, UK)	STBB9655V
Teophylline HBr (THEO)	Sigma-Aldrich (Dorset, UK)	120M0211V
Cutting agents		
α -lactose monohydrate (LAC)	Sigma-Aldrich (Dorset, UK)	SLBJ4979V
α -D-glucose, anhydrous 96 % (GLU)	Sigma-Aldrich (Dorset, UK)	STBC9352V
Alginic acid sodium salt (ALG)	Sigma-Aldrich (Dorset, UK)	MKBP7317 V
Calcium carbonate (CaCO ₃)	Sigma-Aldrich (Dorset, UK)	071M0073V
Creatine monohydrate (CRE)	Sigma-Aldrich (Dorset, UK)	SLBH1411 V
Dextrose monohydrate (DEX)	Sigma-Aldrich (Dorset, UK)	011M0083V
D-Mannitol (MAN)	Sigma-Aldrich (Dorset, UK)	WXBC0281 V
L (+) - Glutamic acid monosodium salt monohydrate (GLUT)	Acros Organics (Geel, Belgium)	A0363200
L-Tyrosine (L-TYR)	Sigma-Aldrich (Dorset, UK)	BCBF4244 V
Magnesium stearate (MgS)	Sigma-Aldrich (Dorset, UK)	-
Maize Starch EP (STA)	J. M. Loveridge Ltd (Andover, UK)	BN K363
Microcrystalline cellulose (MCC)	Sigma-Aldrich (Dorset, UK)	MKBH4403 V
Niacinamide (Nicotinamide) (NIA)	Sigma-Aldrich (Dorset, UK)	031M0198V
Sucrose (SUC)	Sigma-Aldrich (Dorset, UK)	SLBF7618V
Talc (Hydrous magnesium silicate) (TAL)	Sigma-Aldrich (Dorset, UK)	MKBS2607 V
Taurine (TAU)	Sigma-Aldrich (Dorset, UK)	1419568V

10.2. Appendix B – Chapter 3

Table A3.1: Details of websites accessed to purchase the NPS products

Website code	Websites
1	www.brc-finechemicalc.com
2	www.buyanychem.com/rearch-chemicals
3	www.benzo_fury.me.uk
4	www.chem_shop.co.uk
5	www.rc-lab.co.uk
6	www.chemicalwire.com
7	www.buyresearchchemicals.co.uk/buy-mdai.html
8	www.buckledbonzi.co.uk
9	www.herbalhighs.co.uk
10	www.highstore.net
11	www.acechem.com
12	www.elegalhighs.com

10.3. Appendix C – Chapter 4

Raman spectra of BEN, CAF, CRE and LAC (WiRE 3.4).

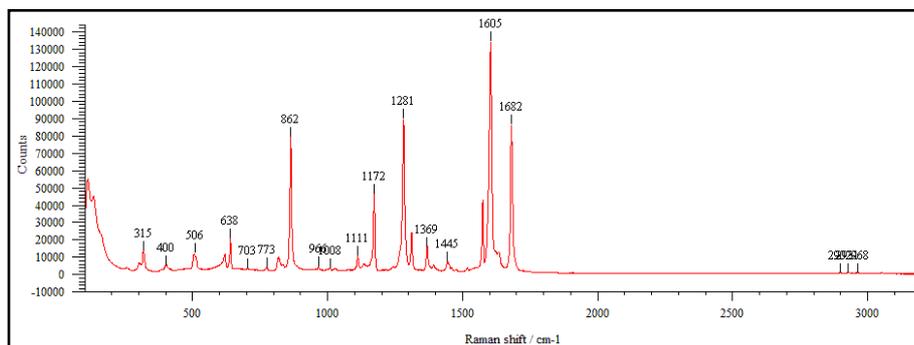


Figure A4.1: An example of a BEN Raman spectrum (WiRE 3.4)

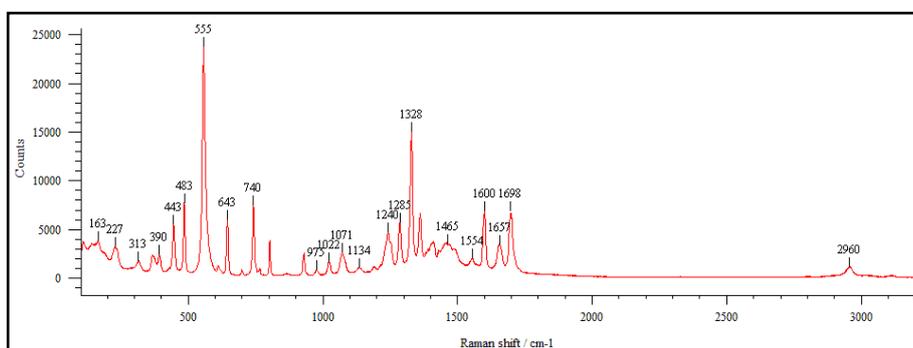


Figure A4.2: An example of a CAF Raman spectrum (WiRE 3.4)

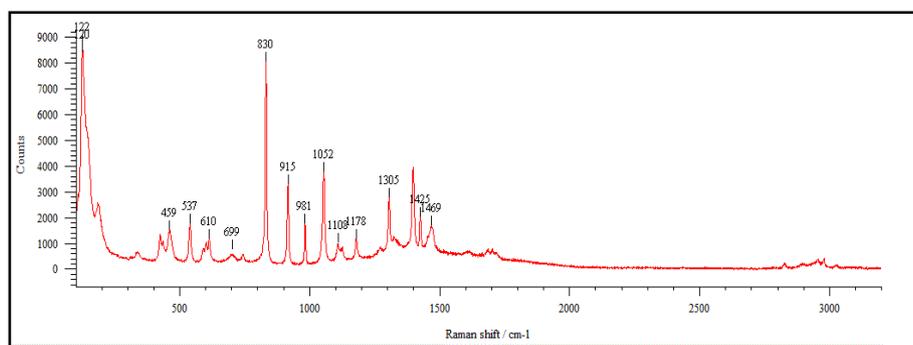


Figure A4.3: An example of a CRE Raman spectrum (WiRE 3.4)

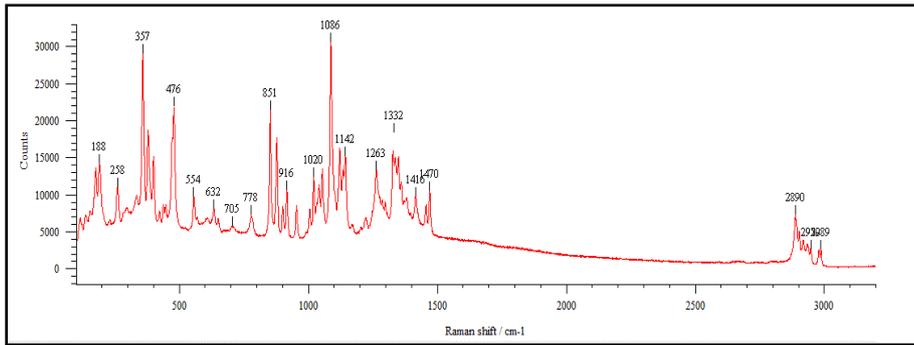


Figure A4.4: An example of a LAC Raman spectrum (WiRE 3.4)

10.4. Appendix D – Chapter 6

Table A6.1: The chemical structures of cutting agents/ adulterants analysed in this Chapter

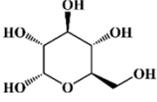
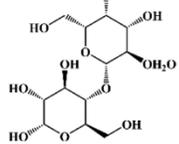
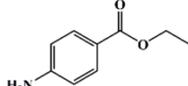
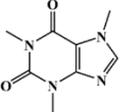
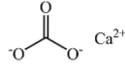
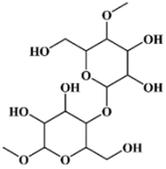
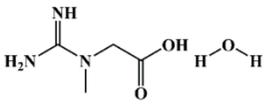
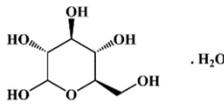
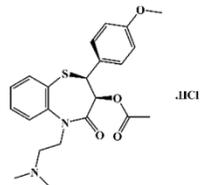
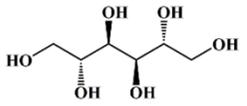
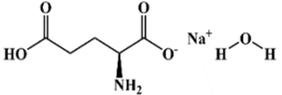
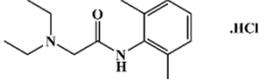
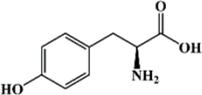
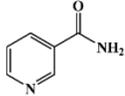
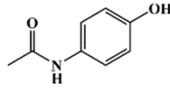
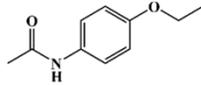
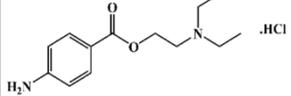
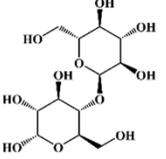
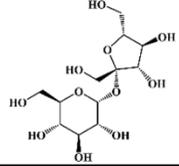
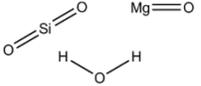
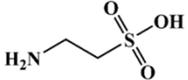
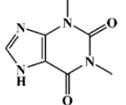
α -D-glucose anhydrous	α -lactose monohydrate	Benzocaine	Caffeine (anhydrous)	Calcium carbonate
				
Cellulose	Creatine monohydrate	Dextrose monohydrate	Diltiazem HCl	D-mannitol
				
Glutamic acid monosodium salt monohydrate	Lidocaine HCl	L-tyrosine	Niacinamide	Paracetamol
				
Phenacetine	Procaine HCl	Starch	Sucrose	Talc
				
Taurine	Theophylline			
				

Table A6.2: Summary of GC-MS method parameters used in the qualitative analysis of NPS internet products¹

Sample ID		GC-MS Method Parameters						
Product No.	Sample name	Injector		Column Temperature Programme				
		Injector T°	Injection mode	Flow (ml min ⁻¹)	Temperature (°C)	Rate (°C/min)	Hold (min)	Total (min)
P1	2-AI	200	Splitless	1.5	70	0	1	1
P2	2-AI				280	15	2	17
					310	20	2	20.5
P3	AB-FUBINACA	250	Splitless	1	160	0	1	1
P4	AB-PINACA				280	15	2	11
P5	AB-PINACA				310	20	2	14.5
					320	20	3	18
P6	Etizolam	275	1 in 100	1				
P7	Etizolam				50	0	2	2
P8	Etizolam				300	15	5	28.67
P9	Etizolam							
P10	Flubromazepam	220	Splitless	1.5				
P11	Flubromazepam				150	0	3	3
P12	Flubromazepam				250	15	1	10.67
P13	Flubromazepam				310	20	4	17.67
P14	Flubromazepam							
P15	Flubromazepam							
P16	MDAI	200	Splitless	1.2	90	0	1	1
P17	MDAI				250	15	2	13.67
					310	25	3	19.07
P18	MPA	275	1 in 100	1	50	0	2	2
P19	MPA				300	15	5	28.67
P20	N-Me-2-AI	200	Splitless	1.2	90	0	1	1
					250	15	2	13.67
					310	25	3	19.07
P21	Pink Champagnes	275	1 in 100	1				
P22	Pink Champagnes				50	0	2	2
P23	Pink Champagnes				300	15	5	28.67
P24	Magic Beans							
P25	THJ-018	250	Splitless	1	160	0	1	1
					280	15	2	11
					310	20	2	14.5
					320	20	3	18

¹GC-MS experiments were conducted under my supervision by Miss V. Guarino, a visiting Erasmus student

Table A6.3: Summary of HPLC method parameters used for NPS internet products

Sample ID		HPLC method parameters						
Product No.	Sample name	Absorption wavelength (nm)	Sample solution concentration ($\mu\text{g mL}^{-1}$)	Flow rate (mL min^{-1})	Injection volume (μL)	Mobile Phase A (%)	Mobile Phase B (%)	LC run time (min.)
P1	2-AI	210	40	1.0	15	OAB ² (90)	ACN (10)	3
P2	2-AI	210	40	1.0	15	OAB (90)	ACN (10)	3
P3	AB-FUBINACA	210	40	1.2	10	OAB (40)	ACN (60)	4
P4	AB-PINACA	210	40	1.2	5	OAB (40)	ACN (60)	4
P5	AB-PINACA	210	40	1.2	5	OAB (40)	ACN (60)	4
P6	Etizolam	205	40	1.0	5	OAB (60)	ACN (40)	5
P7	Etizolam	205	40	1.0	5	OAB (60)	ACN (40)	5
P8	Etizolam	205	40	1.0	5	OAB (60)	ACN (40)	5
P9	Etizolam	205	40	1.0	5	OAB (60)	ACN (40)	5
P10	Flubromazepam	227	40	1.0	5	OAB (40)	ACN (60)	4
P11	Flubromazepam	227	40	1.0	5	OAB (40)	ACN (60)	4
P12	Flubromazepam	227	40	1.0	5	OAB (40)	ACN (60)	4
P13	Flubromazepam	227	40	1.0	5	OAB (40)	ACN (60)	4
P14	Flubromazepam	227	40	1.0	5	OAB (40)	ACN (60)	4
P15	Flubromazepam	227	40	1.0	5	OAB (40)	ACN (60)	4
P16	MDAI	210	40	1.5	15	OAB (90)	ACN (10)	5
P17	MDAI	210	40	1.5	15	OAB (90)	ACN (10)	5
P18	MPA	233	40	1.5	15	OAB (90)	ACN (10)	4
P19	MPA	233	40	1.5	15	OAB (90)	ACN (10)	4
P20	N-Me-2-AI	210	40	1.5	5	OAB (90)	ACN (10)	3
P21	Pink Champagnes	210	80	1.5	5	OAB (90)	ACN (10)	3
P22	Pink Champagnes	210	80	1.5	15	OAB (90)	ACN (10)	3
P23	Pink Champagnes	210	80	1.5	15	OAB (90)	ACN (10)	3
P24	Magic beans	210	80	1.5	15	OAB (90)	ACN (10)	3
P25	THJ-018	216	40	1.5	5	OAB (10)	ACN (90)	3

¹HPLC experiments were conducted under my supervision by Miss V. Guarino, a visiting Erasmus student; ²OAB: Orthophosphoric aqueous buffer

10.5. Appendix E - Published work included in this thesis

Publications

1. A. Guirguis,* J.M. Corkery, J.L. Stair, M. Zloh, S.B. Kirton, F. Schifano, Intended and unintended use of cathinone mixtures. Dec 2016. **Journal of Human Psychopharmacology: Clinical and Experimental**. DOI: 10.1002/hup.2598. March 2017. (In Press).
2. A. Guirguis,* R. Gittins, J.M. Corkery, F. Schifano, Substance use and misuse. Second edition 2012. A CPPE open learning programme. Update September 2016. Legal highs. Section 1.8. **Centre for Pharmacy Post Graduate Education (CPPE)**. Accessed to registered users only.
3. A. Guirguis, S. Giroto, B. Berti, J.L. Stair*, Evaluation of two laser sources, 785 and 1064 nm, for the identification of new psychoactive substances using handheld Raman spectroscopy, **Forensic Science International**. 273 (2017) 113-123. DOI: [10.1016/j.forsciint.2017.01.027](https://doi.org/10.1016/j.forsciint.2017.01.027). Available at: [http://www.fsijournal.org/article/S0379-0738\(17\)30041-5/abstract](http://www.fsijournal.org/article/S0379-0738(17)30041-5/abstract)
4. A. Guirguis*, Legal highs and lows, **The Analytical Scientist**. 17 (2016) 34-40. Available at: <https://theanalyticalscientist.com/issues/0216/legal-highs-lows/>
5. A. Guirguis*, J. Corkery, J.L. Stair, S. Kirton, M. Zloh, C. Goodair, F. Schifano, C. Davidson, Survey of knowledge of legal highs (novel psychoactive substances) amongst London pharmacists, **Drugs and Alcohol Today**. 15(2015) 93-99. Available at: <http://www.emeraldinsight.com/doi/abs/10.1108/DAT-03-2015-0012>. ISSN: 1745-9265.
6. A. Guirguis*, Discovering how to identify new psychoactive substances through PhD research (Online headline). The designer drug detective (Print headline). **The Pharmaceutical Journal**. 294 (2015). URI: 20067578. Available at: <http://www.pharmaceutical-journal.com/careers/career-qa/discovering-how-to-identify-new-psychoactive-substances-through-phd-research/20067578.article>
7. S. [Assi](#), A. [Guirguis](#), S. [Halsey](#), S. [Fergus](#), J.L. Stair*, Analysis of ‘legal high’ substances and common adulterants using handheld spectroscopic techniques. **Analytical Methods**. 7 (2015) 736-746, DOI: 10.1039/C4AY02169J. Available at: <http://pubs.rsc.org/-/content/articlelanding/2015/ay/c4ay02169j#!divAbstract>

Short papers

8. A. Guirguis, S.B. Kirton, M. Zloh, E. Samaras, S. Giroto, B. Berti, J.L. Stair, Identification of new psychoactive substances using Raman spectroscopy: handheld and computational approaches. 2016. Short paper In: **Research and Advances in Psychiatry**. 2(2016) 94. Available at:
<http://www.rapjournal.eu/common/php/portiere.php?ID=add5a67fbc7c0f2d562360838e1ae633>
9. J. Calvo-Castro, A. Guirguis, N. Ijaz, S.B. Kirton, M. Zloh, J.L. Stair, Chemical evaluation of internet products containing new psychoactive substances to investigate supply/distribution chains. Short paper In: **Research and Advances in Psychiatry**. 2(2016) 95. Available at:
<http://www.rapjournal.eu/common/php/portiere.php?ID=add5a67fbc7c0f2d562360838e1ae633> (accessed 30.06.16).

Parliamentary Submissions

1. A. Guirguis and C. Heading, Written evidence to the psychoactive substances inquiry. Sep 2015, Accessed from:
<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/home-affairs-committee/psychoactive-substances/written/19888.html>
2. The British Pharmacological Society recommendations regarding education on new psychoactive substances and IV users (points 2 and 4 made by both Guirguis, A. and Heading, C.). Accessed from
<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/home-affairs-committee/psychoactive-substances/written/20228.html>

Technical Notes

1. J.L. Stair, A. Guirguis, J. Calvo-Castro, E. Samaras, S. Patel, S.B. Kirton, M. Zloh, Practical considerations for the identification of new psychoactive substances (NPS) using handheld Raman spectroscopy. Printed and electronic versions available to forensic laboratories. Submitted to the **EU Commission** (Jan 2016).
2. J.L. Stair, M. Zloh, S.B. Kirton, J. Calvo-Castro, A. Guirguis, S. Patel, Chemical analysis and risk assessment of new psychoactive substances (NPS). Submitted to the **Interpol** (Criminology analysis) (22 Jan 2016).

Publications In-progress

1. M. Zloh, E.G. Samaras, J. Calvo-Castro, A. Guirguis, Stair, J.L., Kirton, S. B.* Drowning in diversity? A systematic way of clustering and selecting a representative set of new psychoactive substances. Jan 2017. **Sci. Rep. (under review)**.
2. J. Calvo-Castro, §* A. Guirguis, § E.G. Samaras, M. Zloh, S.B. Kirton, J.L. Stair,* Hide and Seek: On the Quest to Detect Newly Appearing Psychoactive Substances Using Raman Spectroscopy and Chemometrics. Feb 2017. **Journal of the American Chemical Society (JACS) (in preparation)**.

Book Sections

3. J.L. Stair, M. Zloh, J. Calvo-Castro, A. Guirguis, S.B. Kirton, Raman and light. In: Stair, J. L., Zloh, M. (editors) Raman spectroscopy, Royal Society of Chemistry, Cambridge, UK, January 2017 (**under review**).

Databases

4. J.L. Stair, M. Zloh, J. Calvo-Castro, A. Guirguis, S.B. Kirton, NPS monographs. In: Clarke's Drug Analysis for Drugs and Poisons. Medicines Complete. Moffat, A. (editor), January 2017 (**in preparation**).

Poster presentations

1. A. Guirguis, E.G. Samaras, J. Calvo-Castro, S.B. Kirton, M. Zloh, S. Girotto, B. Berti, J.L. Stair, Poster: Identification of new psychoactive substances using Raman spectroscopy: handheld and computational approaches. **Life and Medical Science conference 2016**, Hatfield 05 April 2016.
2. A. Guirguis, S. Girotto, B. Berti, J.L. Stair, Poster: Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **The Royal Pharmaceutical Society 2015**, London 10 November 2015.
3. A. Guirguis, S. Girotto, B. Berti, J.L. Stair, Poster: Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **The Analytical Research Forum 2015**, London 3 July 2015.
4. A. Guirguis, S. Girotto, B. Berti, J.L. Stair, Poster: Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **EU MADNESS conference 2015**, London 10 March 2015.
5. A. Guirguis, S. Girotto, B. Berti, J.L. Stair, Poster: Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **Life and Medical Science conference 2015**, Hatfield 14 April 2015.
6. A. Guirguis, S.B. Kirton, S. Fergus, M. Zloh, J.L. Stair, Poster: Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **Life and Medical Science conference 2014**, Hatfield 08 September 2014.
7. A. Guirguis, S.B. Kirton, S. Fergus, M. Zloh, J.L. Stair, Poster: Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **The Analytical Research Forum 2014**, London 7 July 2014.

Oral Presentations

1. A. Guirguis, Identification of new psychoactive substances using handheld Raman spectroscopy coupled with chemometrics. In **Bright Spark Symposium**, University of Warwick, Monday 12th September 2016, **invited** presentation.
2. A. Guirguis, The challenge of new psychoactive substance. In **Royal Pharmaceutical Society Conference**, Birmingham, UK, Sunday 4th September 2016, **invited** presentation.
3. A. Guirguis, Identification of new psychoactive substances using Raman spectroscopy: handheld and computational approaches. In **IV International Conference on Novel Psychoactive Substances**, Budapest, Hungary, May 2016.
4. A. Guirguis, New psychoactive substances: Implications for the pharmacy profession in the UK. In **NPS EU Meeting**, Rome, Italy, Tuesday 16th Feb 2016, **invited** presentation.
5. A. Guirguis, Evaluation of two wavelengths, 785 and 1064 nm, for the identification of new psychoactive substances using handheld Raman Spectroscopy. In **SciX/ FACSS 2015 (The Great Scientific Exchange/ Federation of Analytical Chemistry and Spectroscopy Societies)**, Rhode Island, USA, Monday 27th September 2015.
6. A. Guirguis, New psychoactive substances: The legal high phenomenon. In **Hillingdon Council**, London, Tuesday 30th June 2015, **invited** presentation.
7. A. Guirguis, Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. In **Emerging Analytical professionals (EAP) 2015. Lighting the Way for analytical science (Bristol)**. (8th - 11th May), **invited** presentation.
8. A. Guirguis, Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **The LMS conference**. University of Hertfordshire, Hatfield, 14th April 2015.
9. A. Guirguis, Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. **Analytical Symposium**. Manchester University, Monday 20th April 2015, **invited** presentation.
10. A. Guirguis, Investigation of ‘Legal High’ Substances, Common Cutting Agents and Adulterants using Portable Raman Spectroscopy. In **APS UK PharmSci**, University of Hertfordshire, Hatfield, 2014.

Awards

1. The JPAG's Geoffrey Phillips Analytical Science Award (2014).
2. Joint RPS national award for the Research, Evaluation and Audits award for three projects carried out over the period 2013 – 2014 by the LNWLPF: Lung cancer and dementia audits and the EU MADNESS study (2014).
3. The Life and Medical Science conference presentation prize (2015).
4. Highly commended for The Life and Medical Sciences Dean's Award for the outstanding engagement in extra curricula activities related to my PhD research (2016).

Other

1. Nominated to speak about my Ph.D. project to the Royal Party led by the Duke of Edinburgh at the official opening of the Science Building – University of Hertfordshire (3rd of Nov, 2016).
2. Novel psychoactive substances: what are they? And what implications can they have for pharmacists. Robinson, J. **The Pharmaceutical Journal**, PJ September 2016 online, | DOI: 10.1211/PJ.2016.20201674. Article written about a talk delivered by A. Guirguis at the Royal Pharmaceutical Society Annual Conference 2016.