5.5 Task 4: Profiling of Manufacturing Impurities by GC-FID

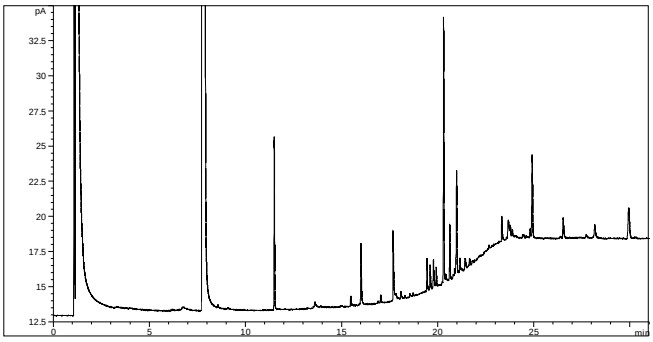
After quantifying the levels of the eight major components, the case samples were subjected to semi-quantitative analysis using GC-FID without the aid of chemical standards. This task is divided into four subtasks: 1) GC-FID optimization and method validation, 2) statistical validation of GC-FID data using a sample weight equivalent to 15 mg heroin base, 3) statistical validation of GC-FID data using a constant 650 mg sample weight, and 4) analysis and statistical classification of the case samples for sample-to-sample comparison at the production level using manufacturing impurities.

5.5.1 GC-FID Method Optimization and Validation

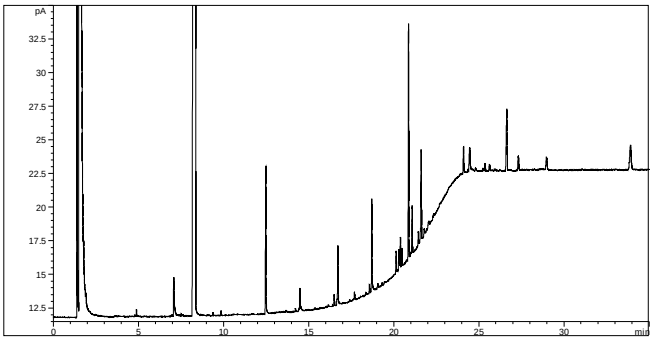
Several case samples were first tested on GC-MS and the results indicated that this technique has a relatively lower sensitivity than GC-FID where quantification of trace impurities is concerned. As only limited amounts of sample (maximum 650 – 700 mg per sample) were available for this task, it was decided to employ chiefly GC-FID for semi-quantitative analysis. Identification of the target peak was performed based on the relative retention time, RRT (retention time of the target peak relative to that of C40, the IS) while the concentration level of each analyte was estimated based on the peak area. In fact, the use of the combination of these two parameters in the profiling of manufacturing impurities was demonstrated by Strömberg *et al.* (2000). As the target impurities are not commercially available, three locally seized samples marked ‘A’, ‘B’ and ‘C’ containing the target impurities were used for method optimization. In addition, a novel control sample containing a series of n-alkanes was designed to check the system stability.

5.5.1.1 Choice of GC Capillary Column

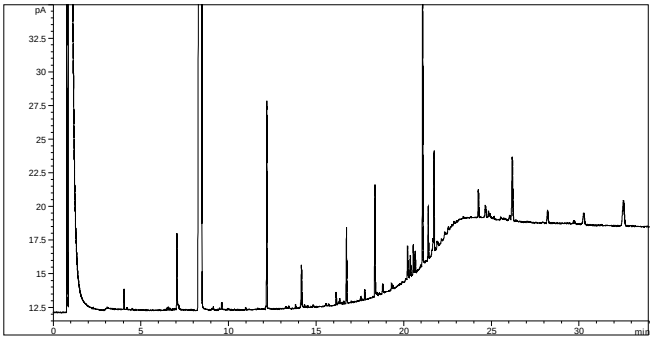
In trace impurity profiling, for an ideal column the retention times of the trace impurities must not be excessively long. It should also have a relatively constant/flat baseline along which a large number of sharp peaks could be obtained. Prior to use, four different capillary columns, namely J&W HP-5 (30 m x 250 µm x 0.25 µm), J&W DB-1 (30 m x 250 µm x 0.25 µm), J&W HP Ultra 2 (25 m x 200 µm x 0.33µm) and J&W HP Ultra 2 (25 m x 200 µm x 0.11µm) were conditioned overnight by heating the columns at 320 ^{o}C and washed with ten methanol blank injections. Their separation performance was grossly evaluated using a single extract of Sample B analyzed at 8 ^{o}C/min ramping rate from 140 ^{o}C to 320 ^{o}C (Figure 5.45). Except for the HP-Ultra 2 (film thickness 0.11µm) column, all the other columns showed unacceptable baselines. In particular, their sloping baselines significantly affected the peak shapes. As the low levels of analytes took a longer time to elute in the thicker films, some peaks became flatter and sometimes may become undetectable in these columns. In contrast, the thinner film of the HP-Ultra 2 column displayed enhanced sensitivity. The use of this column resulted in relatively sharper and higher peaks. It was also able to detect smaller peaks. This is attributed to the fact that as the trace impurities elute rapidly, sharper peaks can be obtained on the chromatogram and hence the improved sensitivity of the system. Therefore, this column was chosen for the current profiling work.



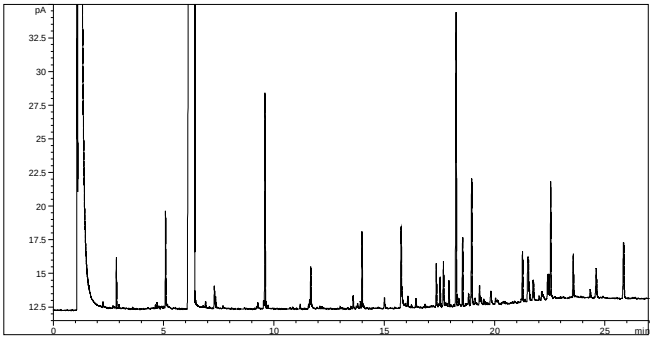
(a)



(b)



(c)

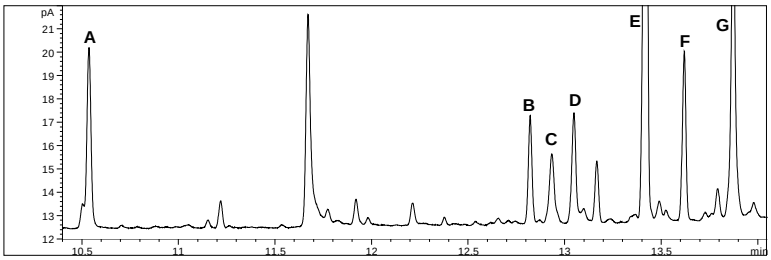


(d)

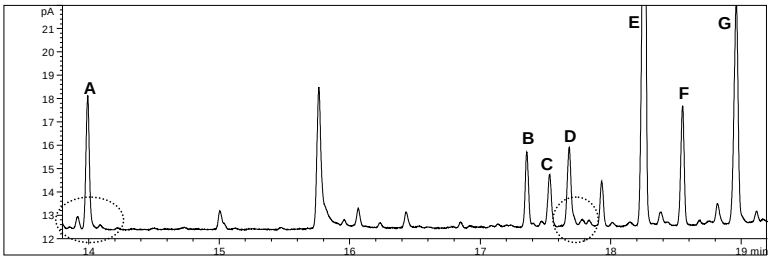
Figure 5.45: Chromatographic performance shown by four different GC columns using a single extract of Sample B (The ramping rate was set at 8 ^{o}C/min from 140 ^{o}C to 320 ^{o}C. (a) HP-5 column (30 m x 250 µm x 0.25 µm), (b) DB-1 column (30 m x 250 µm x 0.25 µm), (c) HP Ultra 2 (25 m x 200 µm x 0.33µm) and (d) HP Ultra 2 (25 m x 200 µm x 0.11µm))

5.5.1.2 Choice of Ramping Rate

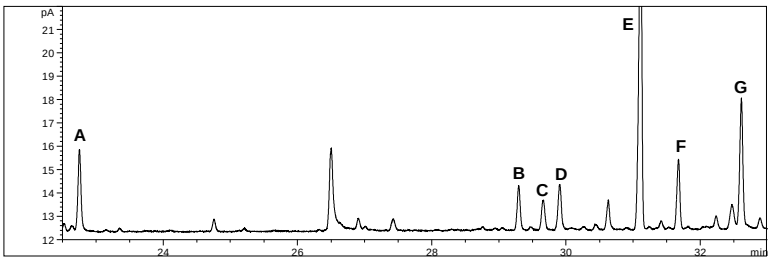
An optimum temperature programming rate should be able to demonstrate complete separation of peaks as well as good peak shapes. Generally, the peak shape is defined by the peak symmetry (PS) and this information can be automatically retrieved using the ChemStation software. With Sample B, six ramping rates were studied utilizing the chosen HP Ultra 2 (film thickness 0.11 µm) column with the initial and final oven temperatures respectively set at 140 ^{o}C and 320 ^{o}C. According to Figure 5.46, complete separation was not achieved with high ramping rates (10 ^{o}C/min and 12 ^{o}C/min); low ramping rates (2 ^{o}C/min and 4 ^{o}C/min) however required a longer analysis time and the peak heights are significantly lower due to peak broadening over a longer eluting time. In this study, it was found that ramping rates between 6 ^{o}C/min and 8 ^{o}C/min were suitable for the separation and elution of the target peaks. However, some of the impurity peaks were not symmetrical (e.g. PS = 0.469 and 1.761 whereby the perfect PS = 1) and hence resulted in poorer peak shapes. Finally, the method was optimized by starting the ramping at 8 ^{o}C/min from 145 ^{o}C. The oven temperature was held at this temperature for 0.4 min before it reached 320 ^{o}C at 6 ^{o}C/min. The average PS for all the target peaks eventually achieved 1.095 ± 0.182 under this optimized condition.



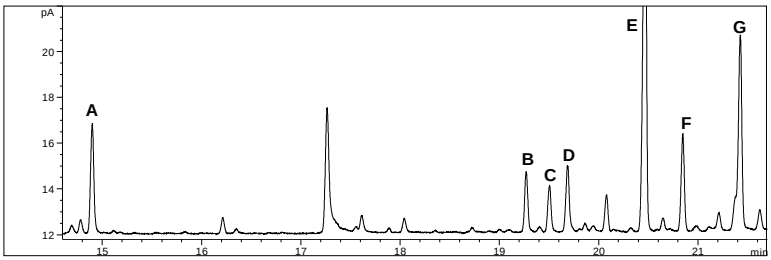
(a)



(b)



(c)



(d)

Figure 5.46: Partially reconstructed chromatograms of a heroin extract of Sample B with the selected ramping rate (The extract was run at (a) 12 ^{o}C/min, (b) 8 ^{o}C/min, (c) 4 ^{o}C/min from 140 ^{o}C and (d) a combination of 8 ^{o}C/min and 6 ^{o}C/min from 145 ^{o}C; labels A to G are the target peaks; circled areas indicate complete separation of unwanted peaks from the target peaks)

The four columns were again assessed using a control sample containing eight n-alkanes (including C40, the IS) under the optimized temperature program. As the most sensitive system will usually give the highest peak heights, therefore the peak

heights of the n-alkanes obtained with the four columns were again compared. According to Figure 5.47, the HP-Ultra 2 (0.11 µm) column showed the highest peak heights for all the n-alkanes (except C25) and hence the use of this column with the optimized temperature program was superior. Finally, this column was used for the subsequent investigation.

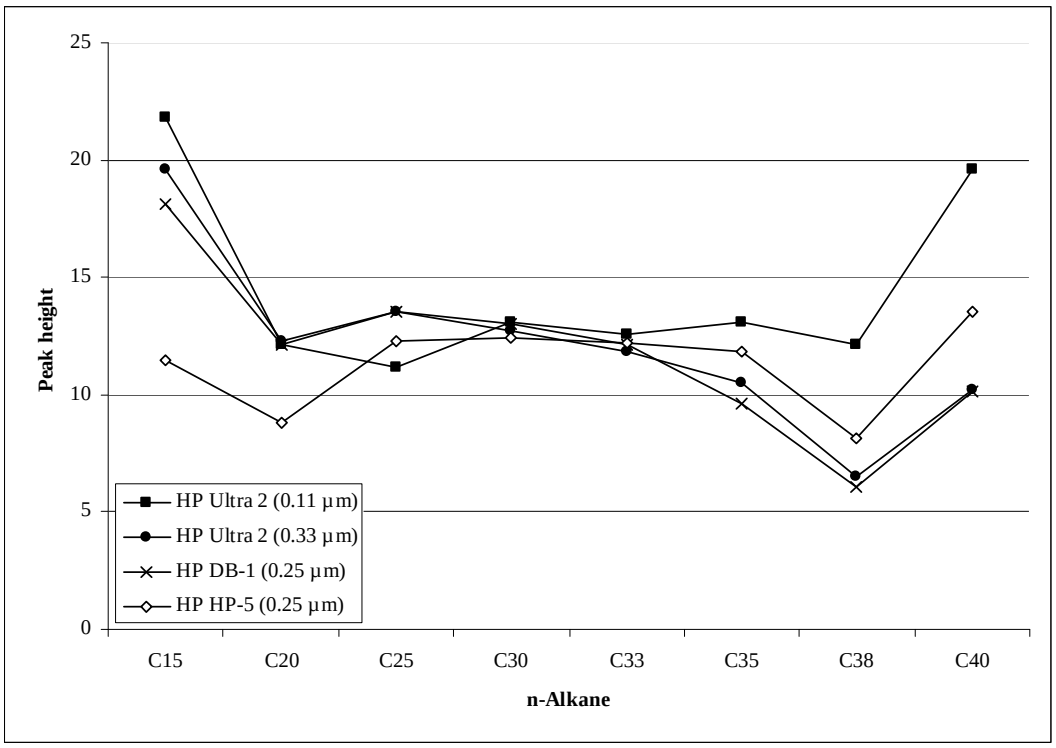


Figure 5.47: Peak heights of n-alkanes analyzed in four different columns

5.5.1.3 Peak Identification and Relative Retention Times

i) Target Manufacturing Impurities

A varying number of manufacturing impurities have been reported in the literature. From a particular street sample, it is possible to extract more than 60 peaks to collectively define a unique fingerprint for a sample. However, it is difficult to obtain all the manufacturing impurity peaks in this large number of peaks. For highly cut samples, extraction of some target impurities may pose a problem. Specifically,

extremely high amounts of adulterants often trapped these impurities in the aqueous portion. In practice, high levels of caffeine present in the case samples in this study resulted in poor extracts (or less impurity peaks). This corroborates the findings of Neumann and Gloger (1982) which showed that the Malaysian samples had a relatively lower number of manufacturing impurities. Despite the scarcity in peaks, 12 acidic impurities were successfully detected and identified by the GC-MS employing the chosen column and temperature program. As the conventional MS library is not usually furnished with the mass spectra of the target impurities, hence manual comparison between the obtained mass spectra and the reference mass spectra available from the works of Allen *et al.* (1984), Strömberg *et al.* (2000), Collins *et al.* (2006), Morello *et al.* (2010), is the alternative way to confirm the peak identities. Figures 5.48 and 5.49 and Table 5.32 respectively summarize the information of the 12 mass spectra of interest obtained from the local samples (Appendix 13).

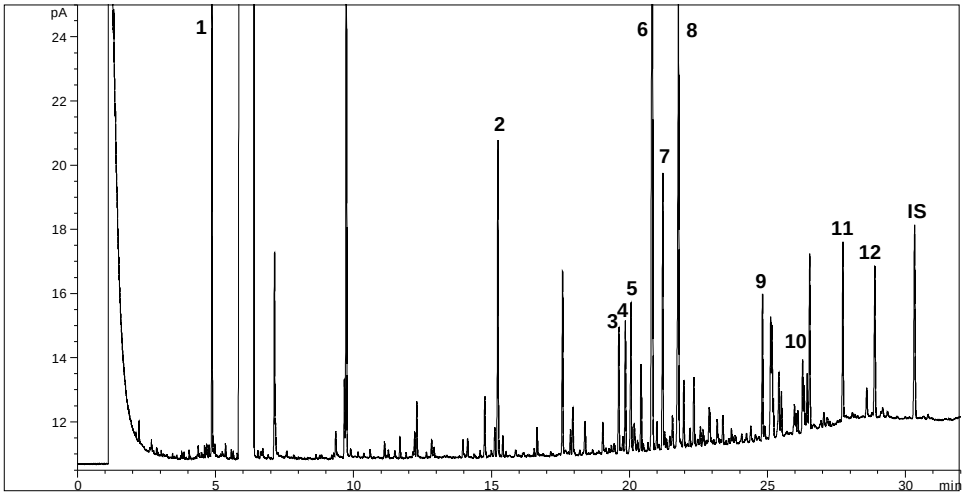


Figure 5.48: A chromatogram showing the positions of 12 target impurity peaks and the IS in a validation sample

Table 5.32: Tentative identities of 12 significant impurity peaks

Peak no.	Tentative compound name	Molecular weight	Base peak	RRT[a]
1	Meconine	194	165	0.157
2	4-O-Acetylthebaol	296	254	0.499
3	Unknown-270		270	0.645
4	6-O,N-Diacetylnorcodeine	369	87	0.652
5	Unknown-254		254	0.658
6	4-Acetoxy-3,6-dimethoxy-5-[2-(N-methyl-acetamido)]ethylphenanthrene	395	265	0.684
7	3-O,6-O,N-Triacetylnormorphine	397	87	0.697
8	N-Acetylnorlaudanosine	385	234	0.716
9	Unknown-151		151	0.817
10	N-Acetylnornarcotine	441	248	0.864
11	(E)-N-Acetylanhydronornarceine	455	193	0.913
12	(Z)-N-Acetylanhydronornarceine	455	193	0.951

[a]RRT = Relative retention time is the retention time of each peak relative to that of the IS.

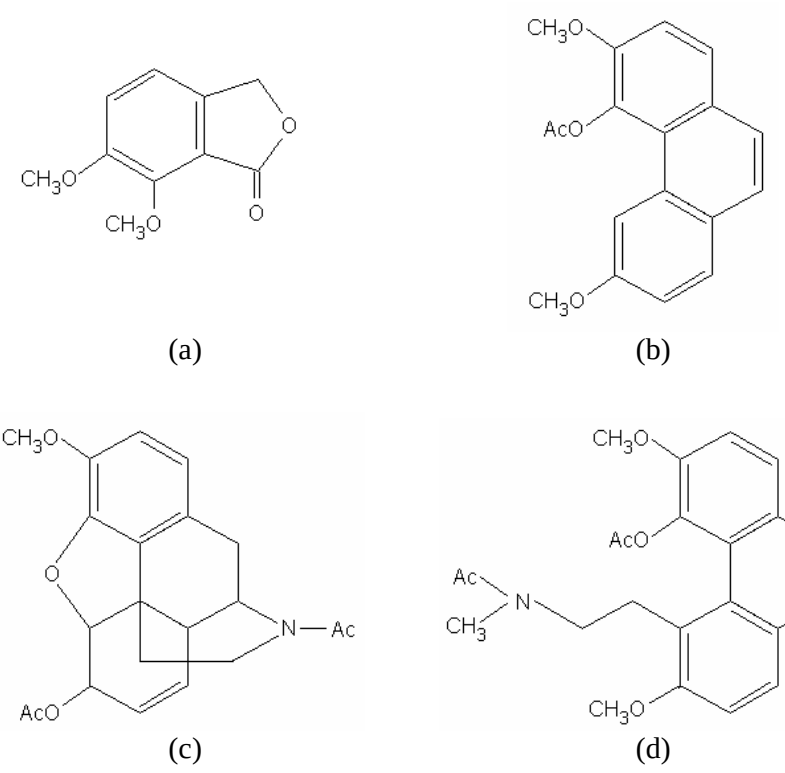
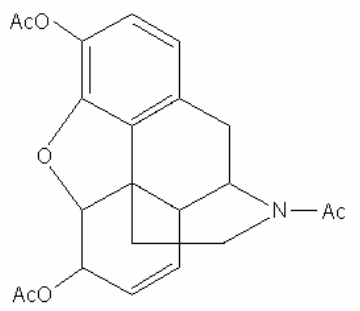
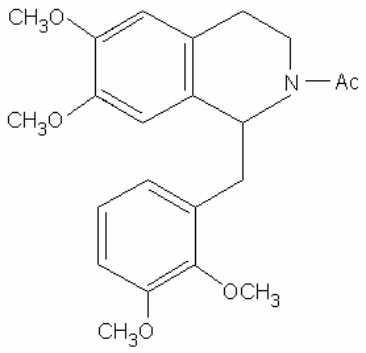


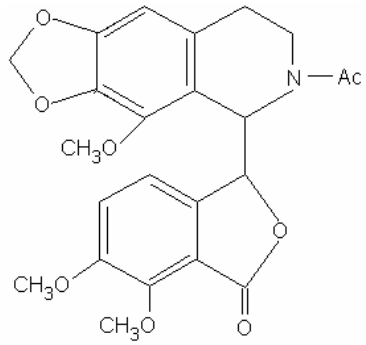
Figure 5.49: Names and structures of 12 impurity compounds (They are (a) Peak 1: Meconine, (b) Peak 2: 4-O-Acetylthebaol, (c) Peak 4: 6-O,N-Diacetylnorcodeine, (d) Peak 6: 4-Acetoxy-3,6-dimethoxy-5-[2-(N-methyl-acetamido)]ethylphenanthrene (e) Peak 7: 3-O,6-O,N-Triacetylnormorphine, (f) Peak 8: N-Acetylnorlaudanosine, (g) Peak 10: N-Acetylnornarcotine, (h) Peak 11: (E)-N-Acetylanhydronornarceine and (i) Peak 12: (Z)-N-Acetylanhydronornarceine)



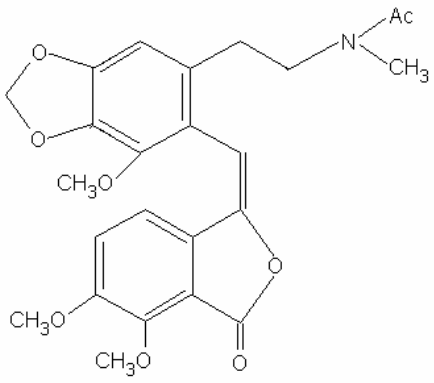
(e)



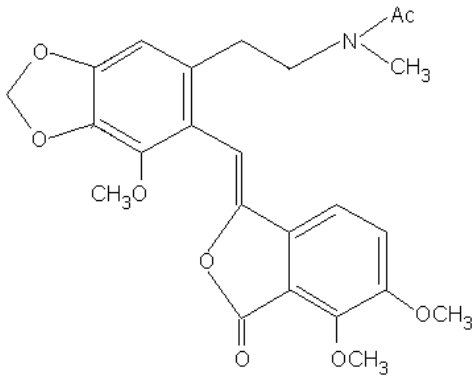
(f)



(g)



(h)



(i)

Figure 5.49: Continued

These 12 peaks were chosen for subsequent investigation using GC-FID based on the RRT. These peaks were selected because they were frequently extractable from the heroin samples and showed relatively well-defined mass spectra. Other unreported peaks could be adulterants (such as caffeine and dextromethorphan) and other unidentified compounds.

ii) n-Alkanes in Control Sample

To validate the performance of the GC system, the control sample was used throughout the course of the analysis. Since n-alkanes have been commonly chosen as the IS of choice in most heroin profiling work (Neumann & Gloger, 1982; Allen *et al.*, 1984; Strömberg *et al.*, 2000), they are thus ideally suited to serve as a control mixture in this study. According to Figure 5.50, more n-alkanes (C30 to C38) are included in the RRT range from 0.0665 – 0.931 because a relatively higher number of impurities elute within that range using the chosen column and temperature program. With the aid of this control mixture, it served to check the system stability when the sample matrix was totally absent.

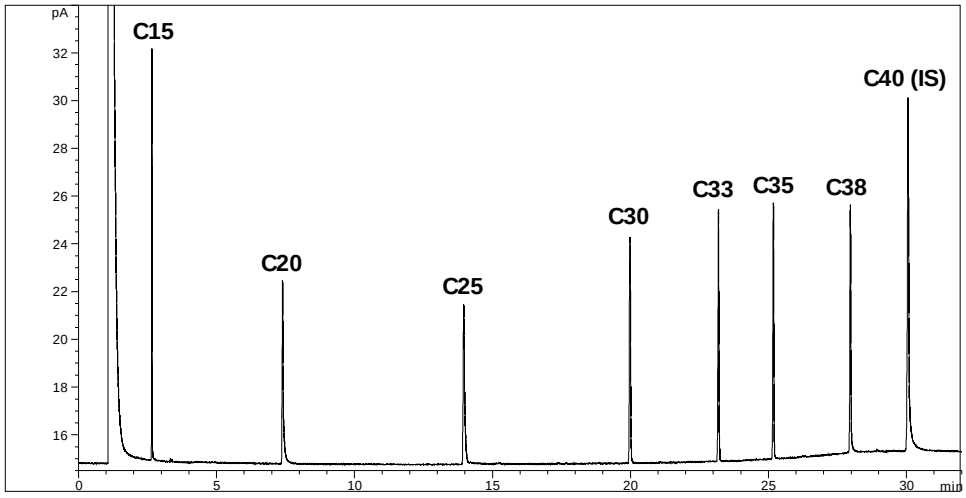


Figure 5.50: A chromatogram showing the positions of seven n-alkanes and the IS in a control sample

5.5.1.4 Injection Volume

As the Malaysian street samples are highly cut, the diluents often hinder maximum recovery of the impurities. The low amounts of impurities resulting from ineffective recovery will render the peaks undetectable unless a suitable injection volume is employed. Consequently, an injection volume that allows for repeatable readings was determined using Samples B and C and the control sample, each injected

at six target volumes. Table 5.33 summarizes the RSD of the area ratio (peak relative to the IS) for each peak calculated from four consecutive injections at each level. From the results obtained in Table 5.33, it was found that 0.2 µL showed the worst performance for the validation and control samples. Some impurity peaks were not detected at this low injection volume. Higher RSD values (> 10% for individual peaks) were observed in the validation samples when the injection volumes were set at 1.0 µL and 2.0 µL. This was largely attributed to the poor peak shapes when the amounts of compounds were insufficient for quantification. A better performance (RSD < 8% for individual peaks) was obtained when injection volumes ≥ 3.0 µL of the sample extracts were used. Based on the results, the best injection volume could be 5.0 µL but this volume will introduce excessive amounts of unwanted compounds into the column. In addition, high amounts of caffeine and other involatile compounds will also be deposited in the split liner and therefore more frequent maintenance may be required. In some cases, this will also lead to sample carry-over in the subsequent analysis. As a result, the use of 3.0 µL as the injection volume of choice for sample introduction was decided for this task.

Table 5.33: RSD (%) of area ratios (peak relative to IS) for 12 impurities found in Samples B and C and a control sample analyzed at six injection volumes (n = 4)

Peak no./n-alkane	0.2 µl	1.0 µl	2.0 µl	3.0 µl	4.0 µl	5.0 µl
			Sample B			
1	3.68	3.17	2.05	1.75	1.09	1.63
2	N.D.	3.30	2.00	1.05	0.89	0.40
3	29.36	4.06	3.95	0.90	1.86	1.16
4	13.73	12.17	2.29	3.81	3.10	0.83
5	10.25	6.13	2.66	1.99	1.31	0.83
6	8.35	1.21	0.74	0.77	0.19	0.26
7	12.38	6.66	2.72	1.58	0.60	0.81
8	1.69	3.75	1.97	1.19	0.64	0.54
9	N.D.	9.08	7.32	4.53	1.88	3.28
10	N.D.	9.05	5.72	3.79	1.91	1.63
11	25.27	7.41	6.68	2.79	3.06	1.64
12	N.D.	10.84	10.81	2.99	2.05	2.83
Average	**13.09**	**6.40**	**4.08**	**2.26**	**1.55**	**1.32**

N.D. = Not detected

Table 5.33: Continued

Peak no./n-alkane	0.2 µl	1.0 µl	2.0 µl	3.0 µl	4.0 µl	5.0 µl
			Sample C			
1	7.10	2.04	1.93	1.19	1.09	2.78
2	24.16	2.06	1.71	0.74	0.74	2.44
3	16.96	7.57	10.54	5.65	2.68	1.67
4	N.D.	11.60	3.56	7.54	3.59	2.60
5	N.D.	14.53	3.79	1.67	0.98	1.75
6	4.65	2.28	1.13	0.83	0.60	0.23
7	39.39	8.27	0.87	2.26	1.05	1.59
8	13.30	14.00	1.92	2.71	1.44	2.41
9	N.D.	14.24	10.15	5.33	2.51	1.10
10	N.D.	8.15	7.29	3.34	2.43	1.48
11	N.D.	8.00	2.64	2.05	2.09	0.77
12	N.D.	23.27	7.84	6.63	3.81	1.32
Average	**17.59**	**9.67**	**4.45**	**3.33**	**1.92**	**1.68**
			Control sample			
C15	3.43	0.29	0.73	0.49	0.63	0.57
C20	40.68	0.82	1.42	0.76	0.56	0.94
C25	24.54	0.90	1.04	0.66	0.52	1.56
C30	10.97	1.51	1.53	0.61	0.34	1.36
C33	11.82	0.88	0.89	0.57	0.42	0.85
C35	10.43	1.00	1.28	0.50	0.18	0.53
C38	7.98	0.77	1.18	0.30	0.09	0.10
Average	**15.69**	**0.88**	**1.15**	**0.56**	**0.39**	**0.84**

N.D. = Not detected

5.5.1.5 Injector Temperature

An ideal injector temperature should be able to offer sufficient heat to consistently volatilize all the target compounds in a repeatable manner. This was assessed based on the RSD calculated from four consecutive injections performed at each of the four chosen inlet temperatures using the above-mentioned optimized parameters with Samples A, B and C. According to Table 5.34, the system performance generally improved with increased injector temperature. This is because the higher heat content around the inlet is able to produce more volatiles, leading to sharper peaks, and hence better RSD values. Besides, higher temperatures were found to be particularly crucial for compounds with long retention times. These compounds with relatively larger molecular weights are usually less volatile and therefore more heat is required. In this study, 320 ^{o}C was chosen as the preferred inlet temperature since it proved to be the ideal value for the local samples. At this injector temperature, 19 peaks (out of 35 peaks

from the three validation samples, excluding peak 3 which was absent in Sample A) showed an RSD < 1.5% compared to their corresponding peaks at lower inlet temperatures.

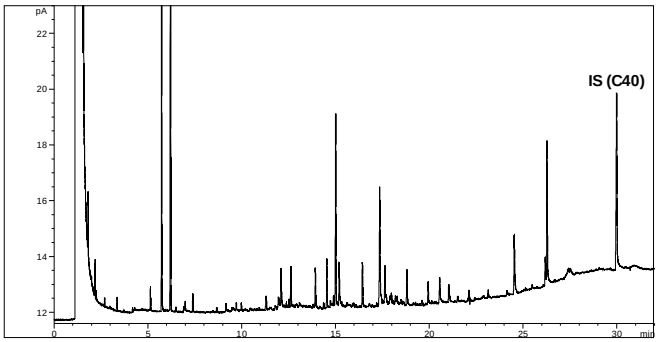
Table 5.34: RSD (%) of area ratios (peak relative to IS) for 12 impurities found in Samples A, B and C analyzed at four injector temperatures (n = 4)

Peak no.	Sample	Injector Temperature (°C)			
		260	280	300	320
1	A	1.45	0.71	0.62	0.27
	B	4.95	1.52	0.84	1.84
	C	1.77	0.66	1.19	1.16
2	A	2.11	1.27	1.23	0.73
	B	2.78	1.10	0.74	0.50
	C	1.25	0.62	0.86	0.91
3	A	N.D.	N.D.	N.D.	N.D.
	B	5.20	0.58	0.83	1.37
	C	2.42	3.33	2.43	2.87
4	A	1.41	0.83	0.64	0.54
	B	4.81	1.26	0.50	0.64
	C	2.18	1.30	0.43	1.82
5	A	1.52	1.30	1.70	0.87
	B	5.64	1.00	1.47	0.19
	C	1.31	0.60	0.36	2.19
6	A	0.91	0.61	0.79	0.58
	B	2.02	1.53	0.62	0.32
	C	1.22	0.29	0.35	0.22
7	A	0.74	1.10	0.61	0.85
	B	3.13	3.63	1.62	1.45
	C	2.18	2.41	4.65	3.43
8	A	12.09	7.07	6.77	7.12
	B	3.11	1.34	0.96	0.19
	C	1.72	2.63	0.43	0.34
9	A	5.44	1.60	2.79	2.50
	B	4.91	0.93	0.93	0.73
	C	3.83	2.03	0.23	1.81
10	A	1.89	1.01	3.38	2.98
	B	0.84	0.70	1.25	0.56
	C	3.08	1.12	1.16	0.74
11	A	5.98	1.05	0.67	0.54
	B	9.44	1.96	0.59	0.34
	C	2.37	1.36	0.99	0.18
12	A	4.95	1.52	0.84	1.84
	B	7.98	0.96	1.29	1.20
	C	5.29	0.81	1.26	0.47
Grand average	**A+B+C**	**3.36**	**1.47**	**1.31**	**1.23**

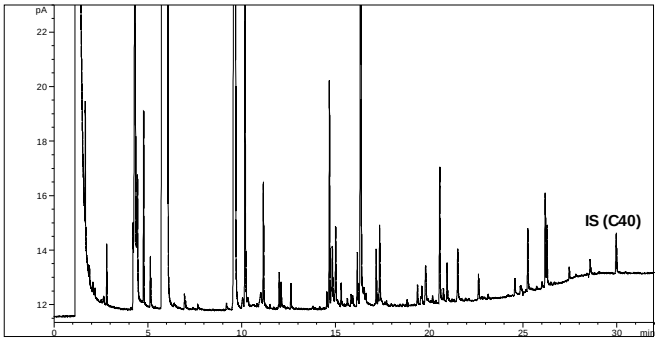
N.D. = Not detected.

5.5.1.6 Choice of Extraction Solvent

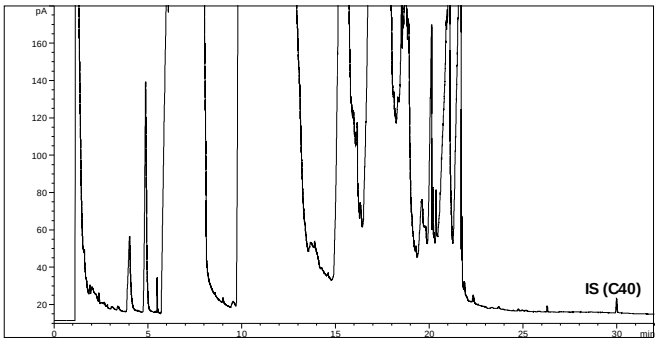
As far as heroin impurities are concerned, toluene has been the widely accepted extraction solvent used in liquid-liquid extraction (LLE) for more than a decade. Four commonly used solvents, namely n-hexane, ethyl acetate, chloroform and toluene were employed in this study. Each solvent was used to extract the impurities found in the local heroin sample and the chromatographic outcomes are illustrated in Figure 5.51. The resulting chromatograms showed that n-hexane and ethyl acetate were not able to extract much of the impurities although they were capable of excluding large amounts of unwanted peaks such as caffeine. On the other hand, chloroform showed excessive extraction capability by extracting too much of caffeine and dextromethorphan (respectively at RT = 6 min and 9 min). Overall, the toluene extract showed an optimal extraction power as indicated by the relatively higher number of impurities (especially after RT = 10 min) and with reasonably low amounts of caffeine and dextromethorphan as shown in the chromatogram. This solvent therefore remained as the ideal extraction solvent for this task.



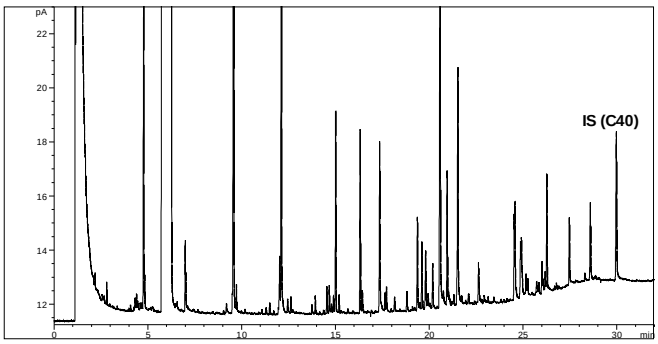
(a)



(b)



(c)



(d)

Figure 5.51: Chromatograms of heroin extracts in (a) n-hexane, (b) ethyl acetate, (c) chloroform and (d) toluene

5.5.1.7 Extraction pH

As only neutral and acidic manufacturing impurities are emphasized in this task, an acidic aqueous medium is employed to facilitate the extraction of acidic analytes. In terms of normality, 0.5 N and 2 N sulfuric acid have been widely employed in heroin impurity studies. A more acidic medium is not favorable as extreme pH will lead to compound degradation. Therefore, the performance of four normality levels in combination with toluene as the extracting solvent was investigated via recovery studies. At each level, the sample was extracted consecutively for three times. For each peak, the sum of the compounds (on the basis of area ratio or peak relative to the IS) from the three extracts was considered to be 100% (van Deursen, Lock & Poortman-van der Meer, 2006). In each extract, the percentage recovery of each peak was calculated. Finally, the capability of the acid was evaluated based on the percentage recovery in the first extract.

Example: The GC readings (peak area relative to IS) for Peak 1 of Sample C is 2.72 in the first extract, 0.47 in the second extract and 0.20 in the third extract at 2 N are obtained.

% Recovery for Peak 1 in the first extract = 2.72 / (2.72 + 0.47 + 0.20) X 100%

= 80.2%

The detailed results of the recovery study for the first extract obtained from each sample at each acid strength are presented in Appendix 14. At each level, the mean recovery value for the first extract was calculated from the three validation samples for each peak. According to Table 5.35, it was found that the strength of the acid did not have a significant impact on the recovery. From the normality range, the 2 N acid

strength resulted in the best extraction capability (mean recovery = 85.7%), and thus it was used for LLE.

Table 5.35: Comparison of the mean recoveries (%) between four different normality strengths of sulfuric acid calculated from three validation sample extracts

Peak	0.5 N	1 N	1.5 N	2 N
1	79.7	83.5	86.1	82.9
2	95.4	96.5	96.1	96.1
3	56.2	52.0	54.2	55.4
4	87.6	89.1	90.2	91.2
5	57.6	55.5	43.1	57.8
6	95.7	95.9	95.7	95.9
7	85.2	84.6	84.9	85.7
8	92.3	91.5	91.4	91.7
9	92.3	90.9	91.0	90.4
10	90.8	92.7	91.8	92.7
11	95.1	94.7	95.3	94.8
12	95.6	94.8	94.6	94.3
Mean	**85.3**	**85.1**	**84.5**	**85.7**

Note 1: Three successive extractions were performed and the sum of all extracts was considered to be 100%.
Note 2: All findings are reported as peak relative to the IS.

Table 5.36: Recovery (%) in the first and second extracts of Sample B with 2 N sulfuric acid

Peak	First Extract	Second Extract
1	89.2	8.3
2	95.7	3.8
3	55.2	25.6
4	93.6	4.9
5	57.9	18.0
6	96.2	3.5
7	83.2	11.2
8	92.7	6.0
9	98.1	1.9
10	95.6	3.9
11	94.0	5.0
12	94.4	4.3
Mean	**87.2**	**8.0**

Note: All findings are reported as peak relative to the IS.

5.5.1.8 Extraction Vessels

Plastic centrifuge tubes should not be used for organic analysis as they contain many extractable organic matters such as additives on the inner wall. Figure 5.52 illustrates the degree to which artifacts could be extracted when both the plastic and glass centrifuge tubes were employed for blank extraction. High amounts of unwanted peaks were present in the blank extract obtained from the plastic centrifuge tube. The noticeable peak heights were exacerbated by the reconstitution of the organic matters after the organic solvent was evaporated off. To ensure a clean target peak will be obtained, the glass centrifuge tube must be used for heroin profiling.

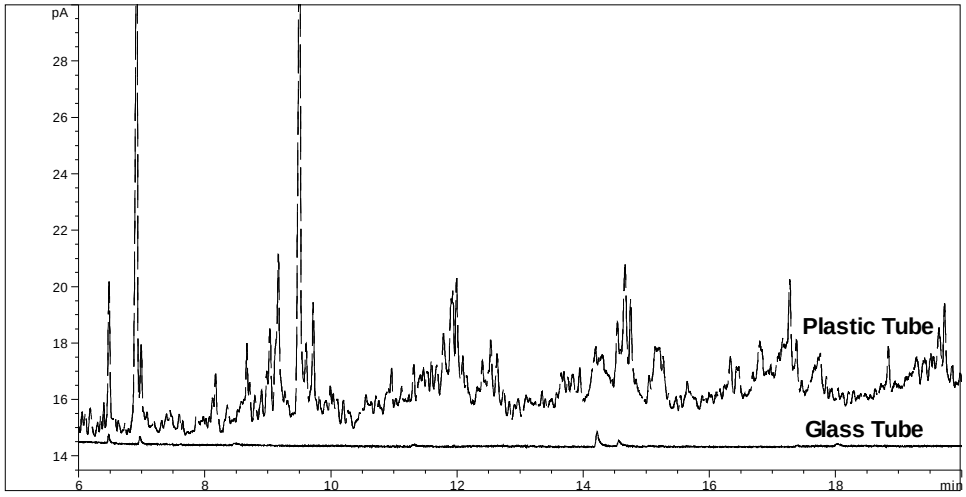


Figure 5.52: Comparison of blank extracts obtained with a plastic tube and a glass centrifuge tube.

5.5.1.9 Additional Optimization Aspects

Several other aspects of the method optimization are not discussed in detail but their rationale of use is outlined in this section. The reconstitution solvent was fixed at 100 µL to ensure that the impurities are not too diluted. The split ratio of 1:25 is necessary to eliminate a significant amount of background interference. The 0.6 µg/mL IS was used for highly cut samples because this IS concentration has relatively the same

peak height as that of the target compounds after reconstitution. This practically serves to ensure that the IS behaves in the same manner as the other analytes.

The precision of the retention time of each target impurity was not performed because the chosen column is easily worn out (Strömberg *et al.*, 2000). After a sequence of 30 – 50 runs, the retention time will usually shift by - 0.5 min, suggesting the short life span of the column. Alternatively, the retention time was re-evaluated/re-confirmed using Sample B when the retention time showed a shift of - 0.5 min, otherwise it was usually performed after the weekly maintenance.

5.5.1.10 Repeatability, Reproducibility and Linearity Checked by Validation Samples

With the optimized conditions, system stability and suitability was concurrently checked with the validation samples and the control sample although the results are presented separately. The use of the case samples (validation samples A, B and C) for validation confers the advantage of assessing the possible matrix interference with the system.

Intra-day precision (repeatability) provides a good measure for not only assessing the system stability but also the stability of the analytes during analysis. At the chosen inlet temperature (320 ^{o}C), the precision of the peak area relative to the IS was examined and the findings are presented as the RSD in Table 5.37. All the compounds in the three different matrices achieved an RSD < 5% (except for peak 8 in Sample A). As there were no significant outliers observed in this study, the compounds are thus demonstrated to be stable in the system.

Reproducibility is a better measure to confirm the system stability. In order to ensure that inconsistent results are not due to the issue of sample instability, the reproducibility was assessed in terms of the inter-hour precision. Calculations for the

reproducibility were done similar to that for the repeatability. According to Table 5.37, all the target compounds maintained an RSD < 10% (except for peak 12 in Samples B and C) for the inter-hour precision. A relatively higher RSD was probably due to the interference from the sample matrices. The overall performance showed that the instrument was sufficiently stable for the analytes quantified within 48 hours or two days.

The FID response was evaluated to determine the extent to which matrix effects will affect the system. As the peaks varied in their heights, the lowest concentration level included for this study should render all the peaks detectable. A graph showing the regression line for each analyte was constructed as the area ratio (peak relative to the IS) versus the amount of extract spiked. Based on the linearity curves, all the peaks reached an $r^2 > 0.968$ (except for peak 9 in Sample A), suggesting insignificant matrix effects on the successful compounds. The system basically showed a satisfactory linear response with the target analyte.

The deviations observed in the repeatability, reproducibility and linearity could be ascribed to the matrix interference. As far as trace impurity profiling is concerned, this level of deviation is still acceptable. Hence, the GC system is generally stable for impurity profiling.

Table 5.37: Repeatability and reproducibility in RSD (%) and r^2 value for the linearity obtained from Samples A, B and C

Peak no.	Sample	Repeatability (n=10)	Reproducibility[a] (n=17)	Correlation coefficient, r^2
	A	1.41	1.16	0.9971
1	**B**	2.11	1.98	0.9982
	C	1.80	0.81	0.9903
	A	0.56	3.37	0.9958
2	**B**	0.92	1.16	0.9974
	C	1.66	1.30	0.9886
	A	N.D.	N.D.	N.D.
3	**B**	3.38	0.91	0.9963
	C	2.25	1.58	0.9870
	A	1.41	2.94	0.9975
4	**B**	0.96	0.98	0.9971
	C	1.94	0.94	0.9876
	A	1.09	2.45	0.9984
5	**B**	1.61	1.72	0.9969
	C	1.00	1.53	0.9915
	A	0.70	0.89	0.9969
6	**B**	0.59	0.59	0.9972
	C	2.84	0.78	0.9910
	A	0.75	1.34	0.9973
7	**B**	1.92	0.52	0.9907
	C	4.37	0.82	0.9734
	A	11.21	6.81	0.9914
8	**B**	0.65	1.08	0.9971
	C	0.69	2.56	0.9912
	A	4.09	6.95	0.6873
9	**B**	1.06	2.83	0.9981
	C	4.04	1.96	0.9853
	A	1.59	2.17	0.9853
10	**B**	1.48	1.40	0.9953
	C	1.60	1.13	0.9909
	A	2.71	9.79	0.9686
11	**B**	1.61	7.82	0.9814
	C	3.58	7.43	0.9980
	A	3.55	7.22	0.9711
12	**B**	3.18	11.58	0.9949
	C	5.87	13.01	0.9690
Grand average	**A+B+C**	**2.31**	**3.19**	**-**

[a]Reproducibility: Due to the concern of sample stability, the GC system was checked by inter-hour reproducibility. Each validation sample was programmed to inject once every 3 hours from 0-hour until the last injection at the 48th-hour.
N.D. = Not detected.
Note: All findings are reported as peak relative to the IS.

5.5.1.11 Repeatability, Reproducibility, Linear Range, LOD, LOQ and Linearity Checked by the Control Sample

System validation was further carried out using the control sample. To evaluate the system repeatability (intra-day precision) and the inter-day precision (reproducibility), the RSD of the area ratio (peak relative to the IS) was calculated for each n-alkane. According to Table 5.38, all the n-alkanes achieved RSD < 2% and RSD < 5% respectively for the intra-day and inter-day precision.

The linear response of the FID was investigated at the concentration range from 0.5 – 100 µg/mL within which all the target n-alkanes could be detected by the instrument. According to Table 5.38, all the n-alkanes obtained an $r^2 > 0.997$ within that range. As this measure cannot represent the linearity for the impurities, the r^2 values achieved by the n-alkanes only infer that the FID functions optimally well for general profiling purposes.

Due to the unavailability of the standards for the impurities, the sensitivity of the system represented by the LOD and LOQ was grossly determined from the control sample with 0.01 – 0.5 µg/mL n-alkanes. The LOD was theoretically determined using the same procedure described in Appendix 10. On this basis, the LODs within 0.09 – 0.41 µg/mL based on 3 S/N and the LOQs within 0.17 – 1.36 µg/mL based on 10 S/N for the range of the n-alkanes in the control sample were obtained. As the maximum value of RSD = 10% (based on the obtained maximum inter-hour precision determined from the validation samples) is regarded as acceptable in this profiling program, the practical LOQ was determined based on the lowest concentration level at which the RSD < 10% from the six consecutive injections was achieved. However in the validation, better LOQs ranging from 0.05 – 0.50 µg/mL for the range of the n-alkanes in the control sample were obtained in this study. Hence, the GC system is considered to be generally stable and reasonably sensitive for impurity analysis.

Table 5.38: Repeatability and reproducibility in RSD (%) and linear range, equation, r^2 value for the linearity, LOD, LOQ and practical LOQ obtained from a control sample

n-Alkane	RRT[a]	Intra-day precision (n = 10)	Inter-day precision (n = 10)	Concentration range covered (µg/mL)	Linearity function	Correlation coefficient, r^2	LOD (µg/mL)	LOQ (µg/mL)	Practical LOQ (µg/mL)
C15	0.088	1.87	3.96	0.5 – 100	y = 0.0350x + 0.0342	0.9975	0.05	0.17	0.05
C20	0.246	1.75	4.28	0.5 – 100	y = 0.0345x + 0.0081	0.9977	0.38	1.27	0.50
C25	0.464	1.30	3.86	0.5 – 100	y = 0.0412x + 0.0203	0.9978	0.41	1.36	0.50
C30	0.665	1.64	3.53	0.5 – 100	y = 0.0408x + 0.0273	0.9974	0.19	0.63	0.50
C33	0.771	0.49	2.58	0.5 – 100	y = 0.0384x + 0.0249	0.9976	0.16	0.54	0.10
C35	0.838	0.49	2.18	0.5 – 100	y = 0.0383x + 0.0398	0.9972	0.10	0.35	0.10
C38	0.931	1.01	1.80	0.5 – 100	y = 0.0396x + 0.0382	0.9976	0.09	0.29	0.05

[a]RRT = Relative retention time is the retention time of each n-alkane relative to that of the IS.
Note: All findings are reported as peak relative to the IS.

5.5.1.12 Extraction Reproducibility

With the 2 N sulfuric acid and toluene, six individual extractions were performed for each validation sample. As the peak relative to the sum of all peaks (PRS) is more reliable than the peak relative to the IS (or area ratio) whereby the former can minimize analytical errors to a larger extent than the latter, so the extraction reproducibility was assessed in terms of RSD for each target peak by using the PRS approach. Table 5.39 shows that the RSD values are relatively high in all the samples. High RSDs for peaks 1 and 5 could be due to some unknown factor in the sample. Undesirable RSDs for peaks 3, 9 and 12 were largely due to their low peak areas. On average, the extraction reproducibility was achieved with RSD $\leq$ 11%.

Despite the poor extraction precision, it is important to know whether the six related extracts of each validation sample could be associated using the impurity peaks. This was done by computing the Pearson correlation coefficient, r^2 achieved by the related samples when they were compared against one another. In general, a mean $r^2 > 0.97$ was obtained, indicating that high correlational relationships existed between the related extracts. In other words, the close relationships of the related samples can still be established using the peaks of interest. Hence, the method is sufficiently good to facilitate impurity profiling for sample comparison.

Table 5.39: Extraction reproducibility (n = 6) in RSD (%) for Samples A, B and C

Peak	Sample A	Sample B	Sample C
1	19.27	4.70	20.18
2	5.31	7.15	11.91
3	N.D.	16.04	10.52
4	2.35	2.40	9.68
5	12.85	11.18	13.79
6	4.03	2.22	5.59
7	3.71	5.06	8.97
8	5.43	3.95	5.51
9	24.54	10.01	16.89
10	6.19	8.62	8.30
11	7.45	10.47	8.19
12	11.23	12.15	12.81
Average	**9.31**	**7.83**	**11.03**
Mean r^2	**0.9838 ± 0.0166**	**0.9959 ± 0.0024**	**0.9787 ± 0.0155**

N.D. = Not detected.
Note: All findings are reported as peak area relative to the sum of all peak areas.

5.5.1.13 Summary

A GC-FID method was optimized using three validation samples and a control sample for impurity profiling of street doses of illicit heroin. The validation samples containing the 12 target impurities were crucial to evaluate the method performance under the influence of sample matrices. The n-alkanes in the control sample were important for the validation of the instrument. Using the compounds in the validation and control samples, the parameters in Table 5.40 were finally achieved based on the ideal optimization results obtained using these samples. The method was also found to be sufficiently good and precise. It also showed sufficient capability for sample classification despite the poor extraction precision.

Table 5.40: GC-FID parameters and liquid-liquid extraction for semi-quantitative determination of 12 target impurities

Condition	Setting
GC-FID	
Column:	J&W HP Ultra 2
Dimensions:	Length: 25 m I.D.: 200 µm Film thickness: 0.11 µm
Carrier gas:	Helium
Injection volume:	3 µL
Split ratio:	25 : 1
Flow rate:	1.2 mL/min
Injector temp.:	320 ^{o}C
Temp. programming:	145 ^{o}C to 190 ^{o}C at 8 ^{o}C/min and hold for 0.4 min, then to 320 ^{o}C at 6 ^{o}C/min and hold for 5-7 min.
Detector temp.:	330 ^{o}C
H_2 flow:	30 mL/min
Air flow:	400 mL/min
He makeup flow:	30 mL/min
Total run time:	< 35 min
Liquid-liquid Extraction	
Acid strength:	2 N sulfuric acid
Extraction solvent:	Toluene
Extraction vessel:	Glass centrifuge tube
Reconstitution volume:	100 µL

5.5.2 Statistical Evaluation Using Simulated Heroin Links

Conventionally, it is recommended to use a sample weight equivalent to 15 mg heroin base or 10 – 45 mg total morphine content for impurity profiling (Neumann & Gloger, 1982; Allen *et al.*, 1984; Strömberg *et al.*, 2000; Morello *et al.*, 2010). This approach can be conveniently used for samples with relatively higher purity or low cut samples. In cases where the sample is of lower purity, a larger sample weight is required for profiling. For example, approximately 3 g material is required for profiling a heroin sample of 0.5% purity to achieve an equivalent weight of 15 mg heroin base. In relation to this, Dufey *et al.* (2007) has reported that such a requirement is not always practical for many laboratories. Due to this constraint, some of the samples with < 2.3% purity in this study could not be profiled following the conventional approach as it requires a minimum of 650 mg substance for analysis. As only 1 g of homogenized heroin substance was collected from each case, a sample weight in excess of 650 mg is therefore not recommended since the collected sample amount was marginally sufficient to accomplish three major tasks of chemical profiling. Furthermore, a larger sample amount in the tube can also saturate the extracting aqueous phase. To rectify the above-mentioned issues, two sample weight approaches were employed for case sample analysis, namely the equivalent to 15 mg heroin base sample weight approach and the constant 650 mg sample weight approach. The latter approach was used for samples with the purity level < 2.3%. Both of these approaches were statistically validated with two separate sets of five simulated links prepared from five unrelated heroin case seizures (M, P, K, T and Z). All the samples were analyzed with the optimized GC-FID method targeting the 12 manufacturing impurities.

5.5.2.1 Variation Associated with LLE and Sample Weight Difference

A set of 5 simulated links each containing eleven related samples were analyzed following the first approach using a sample weight equivalent to 15 mg heroin base. Another set of 5 simulated links each containing five related samples were analyzed following the second approach using a constant 650 mg sample weight. In each dataset, the variation of each peak expressed as RSD was calculated from the total related samples available for each link based on two forms of data, 1) area ratio or peak relative to IS (denoted as AR) and 2) peak area relative to the sum of peak areas (denoted as PRS) and the results are summarized in Table 5.41. On the basis of AR, the RSD values obtained with the first approach were better than the corresponding RSD values of the second approach. This could be due to the consistent amounts of impurities available in the aliquots prepared with the first approach. As the amounts of impurities prepared for each link/batch were highly variable in the second approach, the mean values were greatly affected, and hence the poor RSD values.

In order to minimize the sample weight effect, it is more reliable to assess the RSD based on the PRS. Most of the RSD values calculated from the AR for the second approach have been improved by the PRS. This is also true in the first approach, however, the RSD values calculated from the PRS for the second approach were still unsatisfactory in links M, K and T as compared to the corresponding RSD values of the PRS for the first approach. This indicates that data normalization through PRS could not effectively minimize the intra-sample/intra-batch variation arising from the sample weight difference. The large variations observed in the first approach are most likely due to the presence of cutting agents. Inconsistent extraction efficiency was caused by the large amounts of cutting agents which often trapped the target impurities in the solution. However for the second approach, the poor RSD values could be the errors imparted from the poor extraction efficiency and the weight difference.

Table 5.41: Variation in RSD (%) encountered in the simulated datasets analyzed by GC-FID using the sample weight equivalent to 15 mg heroin base sample weight approach and the 650 mg constant weight approach

Peak			1	2	3	4	5	6	7	8	9	10	11	12
M	15	AR	31.75	7.57	30.78	19.83	41.74	21.07	22.59	15.58	70.61	6.58	13.16	17.60
		PRS	26.32	10.08	23.21	14.03	32.92	8.99	13.38	9.48	60.41	13.14	6.80	10.81
	650	AR	36.20	30.64	94.67	45.14	92.34	49.03	52.35	44.04	N.A.	73.13	37.64	N.A.
		PRS	18.54	20.17	115.50	30.97	92.38	16.97	21.60	10.66	N.A.	106.20	18.75	N.A.
P	15	AR	40.37	8.07	29.49	17.12	35.37	15.79	21.13	10.56	21.30	11.62	14.07	18.54
		PRS	35.96	16.22	24.83	5.93	28.55	5.06	11.84	8.85	13.84	13.28	5.14	11.71
	650	AR	34.70	25.69	23.77	27.10	31.50	32.71	37.64	26.45	43.96	27.06	32.74	34.15
		PRS	28.92	6.74	17.74	3.97	19.17	6.28	13.57	5.58	31.21	7.83	8.40	11.52
K	15	AR	72.17	N.A.	44.70	29.25	33.67	30.37	31.76	34.59	331.66	124.49	57.67	N.A.
		PRS	55.38	N.A.	24.00	3.62	9.65	3.48	1.82	41.86	331.66	106.56	54.36	N.A.
	650	AR	58.74	N.A.	31.58	66.55	92.52	82.50	79.49	95.16	N.A.	N.A.	60.32	N.A.
		PRS	50.40	N.A.	58.54	29.59	58.13	16.74	16.15	137.32	N.A.	N.A.	118.02	N.A.
T	15	AR	19.73	11.60	46.49	14.76	20.01	23.04	33.43	12.93	22.00	11.67	14.64	13.62
		PRS	13.28	8.33	36.86	6.40	15.43	13.54	29.02	12.12	11.50	5.49	8.51	7.35
	650	AR	49.08	57.93	31.37	43.61	78.46	82.53	115.76	45.54	67.08	67.48	62.87	75.53
		PRS	18.50	10.57	85.38	23.97	29.77	32.22	98.13	18.78	56.06	9.46	3.63	39.78
Z	15	AR	36.39	9.04	N.A.	18.23	30.31	21.58	22.86	19.71	39.18	12.59	19.25	20.95
		PRS	38.21	20.90	N.A.	9.33	19.08	10.34	13.05	10.74	32.77	11.36	9.79	12.76
	650	AR	43.74	25.53	N.A.	46.64	77.56	27.61	23.44	18.66	N.A.	50.47	25.70	69.37
		PRS	26.66	15.85	N.A.	47.92	79.08	10.00	5.70	10.49	N.A.	32.91	22.03	60.13

N.A. = Not applicable.
15 = A sample weight equivalent to 15 mg heroin base
650 = A constant sample weight at 650 mg
AR = Area ratio or peak relative to IS
PRS = Peak area relative to the sum of peak areas

In general, the PRS shows that the intra-batch variations were still measurable in both approaches. The next step is to find the best pretreatment method and linkage-distance combinations to minimize these variations.

5.5.2.2 Sample Weight Equivalent to 15 mg heroin base Sample Weight Approach

i) Evaluation of Pretreatment Methods

Using the sample weight equivalent to 15 mg heroin base, 5 unrelated links each containing eleven linked samples were analyzed and the GC-FID data were reported as peak areas. Thereafter, 55 data points derived from the individual aliquots analyzed were subjected to a number of pretreatment methods according to Table 5.42.

Table 5.42: Pretreatment methods for GC-FID impurity data

Pretreatment	Abbreviation	Description
Normalization or PRS	N	Each peak area is divided by the sum of peak areas. This is similar to PRS.
Standardization	S	Each peak area is divided by the standard deviation calculated from that peak variable
Fourth root	4R	Application of fourth root to each peak area
Logarithm	L	Application of logarithm to each peak area
Normalization + standardization	N + S	Each normalized peak is divided by the standard deviation of that normalized peak variable
Normalization + fourth root	N + 4R	Application of fourth root to each normalized peak variable
Normalization + logarithm	N + L	Application of logarithm to each normalized peak variable

Each of the pretreated dataset was screened by PCA in the covariance mode to verify for the best pretreatment method for clustering. A successful method will display

five distinct groups representing the unrelated batches. Besides, it must also show that the linked samples are closely packed within their groups. Pretreatment methods such as S, 4R and L were not successful and the sample units in the groups were too widely distributed. This was largely due to the uncorrected experimental errors such as split ratio, injection volume and extraction efficiency associated with the original data recorded in the peak areas. Therefore, data normalization is expected to give a better outcome. However, the N data also met with failure and only one group was successfully clustered. This could be due to the presence of large peaks or extreme values in the dataset. Subsequent application of fourth root, logarithm and standardization to the normalized data plays a significant role in minimizing the influence of the large peaks. Finally, N + S was found to be the most successful method to group all the linked samples into five distinct groups when they were decomposed into three components by PCA (Figure 5.53). The success lies with the ability of the standardization to equalize the weightage of all the peaks in the dataset. In particular, the use of logarithm is not applicable to zero-values. When the untreated zero-values indicated as missing values were replaced with zeroes after the pretreatment, the dataset would have therefore become less reliable for sample classification.

Each peak has its contribution to the component. When the N + S data were used to define the loadings of the peaks in the first three principal components, Table 5.43 suggests that the contributions of peaks 2, 11 and 12 having loadings > 0.3 were found to be associated with the first component. With loadings > 0.4, peaks 1 and 10 were associated with the second component. Peaks 5, 9 – 12 had higher contributions to the third component with loadings > 0.3. Besides, the first three components accounted for 94.2% of the total variability of the data.

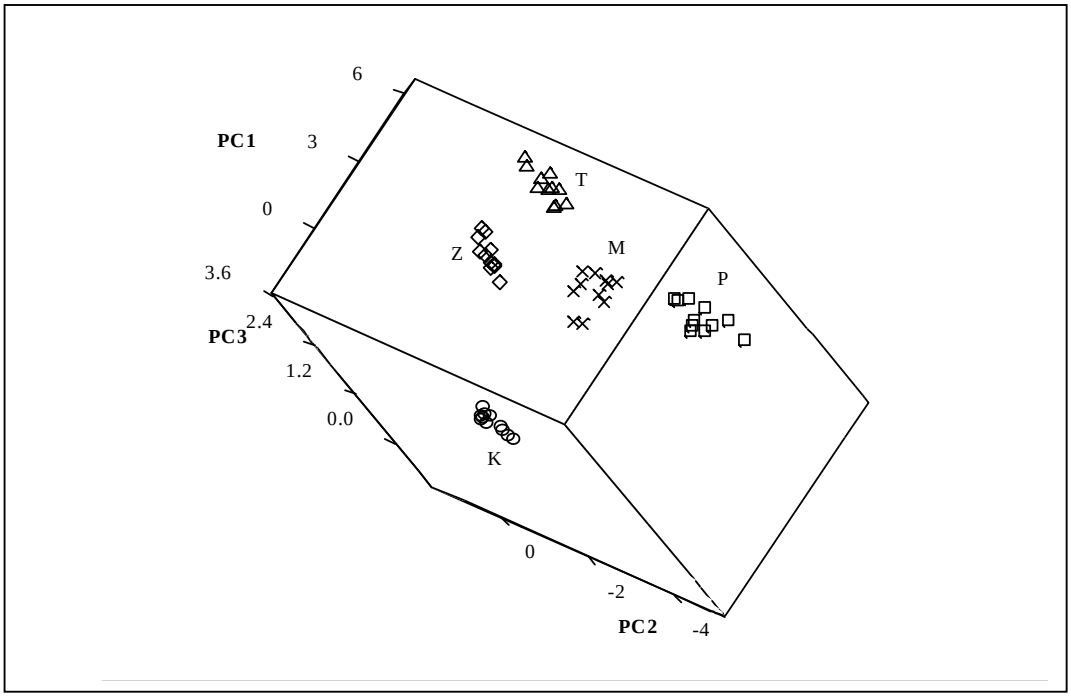


Figure 5.53: A score plot representing 12 N + S pretreated impurity peaks of 55 data points decomposed by PCA in covariance mode into three dimensions, $\%V_1$ = 71.2%, $\%V_2$ = 16.7% and $\%V_3$ = 6.3% (The distribution shows distinct groups)

Table 5.43: Loadings of the first three principal components of 12 N + S data of 55 simulated samples

N + S of peak no.	PC1	PC2	PC3
1	0.217	0.424	-0.484
2	0.328	0.089	-0.091
3	0.018	-0.670	-0.247
4	-0.333	-0.064	0.120
5	-0.313	-0.085	0.366
6	0.297	-0.183	-0.076
7	-0.337	0.027	0.092
8	0.294	-0.309	-0.204
9	0.263	-0.348	0.301
10	0.279	0.312	0.330
11	0.323	-0.039	0.307
12	0.309	0.046	0.445

The suitability of the N + S data was further confirmed by DA. Of the 55 samples, 35 samples were randomly assigned as a training set which essentially provides the source characteritics to the DA so that it can classify the remaining 20

blind test samples based on the characteristics. Table 5.44 shows that all the training and blind test samples were correctly grouped under their respective links. In other words, the DA achieved 100% correctnes using the N + S data for sample classification.

Table 5.44: Summary of classification with cross-validation for 55 simulated samples

Put into Group	K	T	P	M	Z
1	**7** (4)				
2		**7** (4)			
3			**7** (4)		
4				**7** (4)	
5					**7** (4)
Total N	**7**	**7**	**7**	**7**	**7**
N correct	**7**	**7**	**7**	**7**	**7**
Proportion	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
N = 35	**N Correct = 35**		**Proportion Correct = 1.000**		

Note 1: Linear discriminant function was used.
Note 2: Figure in bold = Figure for training set.
Note 3: Parenthesis = Number of blind test samples assigned.
Note 4: Only 11 peaks were used in DA as peak 12 is highly correlated with others, hence it was excluded.

ii) Evaluation of Linkages and Distance Measures

Seven distance measures and five linkage methods working in combination were assessed using the 55 samples of known relationships. The data were initially pretreated with N + S. The number of mistakes and the d_m values for each combination are presented in Table 5.45. As predicted, the pair of Euclidean and Pearson as well as the pair of Squared Euclidean and Squared Pearson achieved the same performance in the clustering. The use of one can supplant the use of the other in its pair. Based on this preliminary investigation, 24 out of 35 combinations were able to give zero mistakes when the linked sample units were repeatedly clustered. Manhattan is considered as the best distance measure since it consistently showed zero mistakes with up to 6 linkages (except for a Single linkage). In addition, linkages including Average, McQuitty and

Ward also showed the best performance with any of the distance measures by having null mistakes in the sample classification. As a smaller d_m value indicating minimum intra-batch variability and maximum inter-batch variability (whereby batch in this section refers to each link) is desirable, the Ward linkage apparently displays significantly lower d_m values with every distance measure compared to its corresponding linkages in Table 5.45. Similarly, Squared Euclidean and Squared Pearson also gave lower d_m values with all the linkages. With these simulated links, Ward-Squared Euclidean and Ward-Squared Pearson were found to have zero mistakes and the lowest d_m value and hence the best linkage-distance combination for sample classification using HCA. As a result, one of these combinations has been chosen for the classification of case samples in the later section.

Table 5.45: Number of samples erroneously clustered and the d_m value in parenthesis obtained with 55 simulated samples analyzed by HCA

Method		Distance measure				
		Euclidean	**Manhattan**	**Pearson**	**Squared Euclidean**	**Squared Pearson**
Linkage method	**Average**	0 (24.17)	0 (18.73)	0 (24.17)	0 (6.24)	0 (6.24)
	Centroid	2 (21.46)	0 (17.65)	2 (21.46)	0 (5.35)	0 (5.35)
	Complete	6 (22.83)	0 (20.96)	6 (22.83)	6 (5.45)	6 (5.45)
	McQuitty	0 (24.46)	0 (18.76)	0 (24.46)	0 (6.46)	0 (6.46)
	Median	2 (21.53)	0 (17.23)	2 (21.53)	2 (5.09)	2 (5.09)
	Single	0 (22.43)	1 (15.53)	0 (22.43)	0 (5.26)	0 (5.26)
	Ward	0 (4.17)	0 (2.94)	0 (4.17)	0 (1.16)	0 (1.16)

5.5.2.3 The Constant Sample Weight (650 mg) Approach

As previously stated, the use of a constant weight for highly cut samples results in insufficient amounts of impurities for extraction. Figure 5.54 shows a chromatogram of the 12 impurities extracted from a 650 mg case sample (1.18% heroin base). All the peaks were detected but with significantly lower levels. For example, peaks 3 to 9 are

relatively small in Figure 5.54. This means that the constant weight approach to some extent is still useful for impurity profiling. However, this approach was only employed for samples containing < 2.3% heroin base.

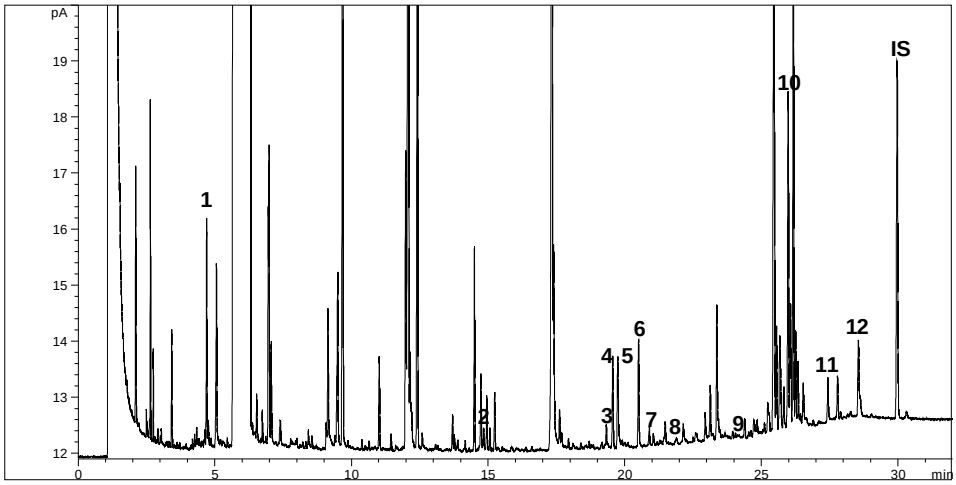


Figure 5.54: Positions of 12 impurity peaks on an enlarged chromatogram for a 650 mg case sample (1.18 % heroin base)

As the N + S pretreatment proved promising with the samples prepared with the conventional approach, this data pretreatment method was also employed for the 650 mg constant sample weight approach. In this subtask, PCA and DA were not used for screening purposes. Instead, four statistical techniques were deliberately investigated in order to find the best model that is able to classify all the linked samples under their respective links. Two unsupervised pattern recognitions (PCA and HCA) and two supervised pattern recognitions (K-means cluster, K-MC and DA) were applied to the 25 post-cut samples prepared from the five unrelated heroin seizures (M, P, K, T and Z). For sample clustering, the former technique does not require the knowledge of sample origin whereas the latter technique is usually trained with samples of known origins.

i) Principal Component Analysis

The post-cut samples were decomposed by PCA in covariance mode into three components and the distribution of the 25 post-cut samples is presented in Figure 5.55. From the score plot, the five known groups were not distinctively segregated from one another. The linked samples in each group are widely related. Groups P and T as well as M and Z were clustered in close proximity and the distinction between the groups is not apparent. The inability to distinguish the groups is probably due to the presence of zero values (absence of peaks) in samples highly-cut with caffeine in addition to the wide inter-extraction variation (equivalent to intra-batch variation). Based upon the score plot, PCA is not an ideal statistical technique for clustering since it is not able to minimize the variation of the related samples and maximize the variation of the unrelated samples.

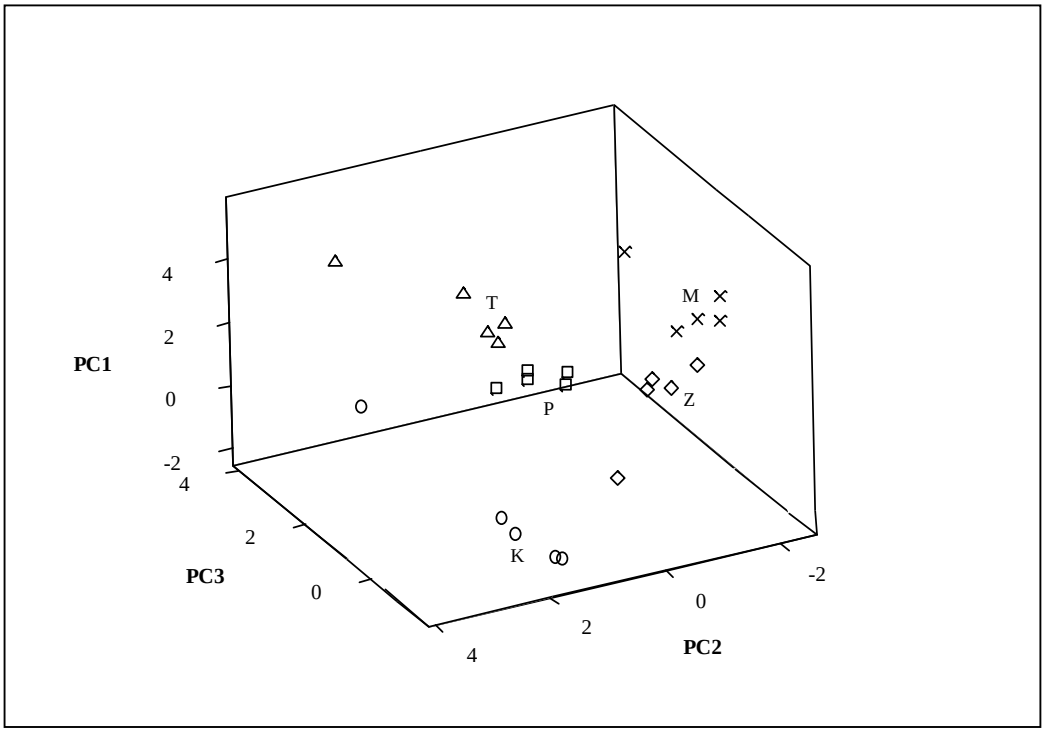


Figure 5.55: A score plot representing 12 N + S pretreated impurity peaks of 25 data point decomposed by PCA in covariance mode into three dimensions, $\%V_1$ = 48.0%, $\%V_2$ = 23.3% and $\%V_3$ = 12.0% (The distribution hardly shows distinct groups)

Table 5.46: Loadings of the first three principal components of 12 N + S data of 25 simulated samples

N + S of peak no.	PC1	PC2	PC3
1	0.186	-0.385	0.387
2	0.323	-0.322	0.072
3	-0.091	0.368	0.445
4	-0.354	0.141	0.314
5	-0.160	0.331	-0.520
6	0.202	-0.268	-0.435
7	-0.401	0.071	-0.059
8	0.379	0.163	-0.044
9	0.262	0.390	0.089
10	0.334	0.143	0.003
11	0.314	0.238	0.230
12	0.279	0.389	-0.151

ii) Hierarchical Cluster Analysis

The post-cut samples were analyzed by HCA with five distance measures and seven linkage methods to evaluate the mistaken units and d_m value. Table 5.47 summarizes the number of mistaken units determined from the linkage methods and distance measures working in combination.

Table 5.47: Number of samples erroneously clustered and the d_m value in parenthesis obtained with 25 simulated samples analyzed by HCA

Method		Distance measure				
		Euclidean	**Manhattan**	**Pearson**	**Squared Euclidean**	**Squared Pearson**
Linkage method	**Average**	5 (26.61)	3 (25.66)	5 (26.61)	5 (7.11)	5 (7.11)
	Centroid	6 (24.39)	4 (43.55)	6 (24.39)	4 (9.03)	4 (9.03)
	Complete	2 (38.92)	2 (28.20)	2 (38.92)	2 (18.32)	2 (18.32)
	McQuitty	6 (21.58)	4 (17.77)	6 (21.58)	5 (6.30)	5 (6.30)
	Median	6 (27.47)	4 (28.01)	6 (27.47)	5 (9.19)	5 (9.19)
	Single	4 (35.07)	4 (35.12)	4 (35.07)	4 (15.24)	4 (15.24)
	Ward	2 (10.77)	0 (12.04)	2 (10.77)	0 (8.42)	0 (8.42)

Among all combinations, three combinations, namely Ward-Manhattan, Ward-Squared Euclidean and Ward-Squared Pearson showed promising results in the sample

classification with zero mistakes. Ward-Squared Euclidean and Ward-Squared Pearson are the best combinations as both showed the smallest d_m value among the three successful combinations.

iii) K-means clustering

K-MC is useful for clustering unknown samples by employing samples of known or unknown origins as initial knowledge. If the origin is unknown, certain samples can be chosen to represent the identities of the sources. In this study, K-MC operates according to MacQueen's algorithm (MacQueen, 1967). Basically, K-MC (also called the statistical machine) will acquire some knowledge about the characteristics of each cluster from the specified samples which serve as the starting initials. The starting initials serve to train the K-MC machine so that it recognizes how each individual cluster should be defined and on this basis other samples will be grouped accordingly.

In this study, each source (pre-cut sample) was represented by an individual link containing five samples of different purity levels, ranging from relatively low purity to high purity when a varying amount of caffeine was added to make up the 650 mg constant sample weight. The effectiveness of K-MC was assessed by training the machine with samples of high, medium and low purity levels. In the first instance, only the highest purity sample from each group was used to train the machine before it was allowed to cluster all other samples accordingly. This step was repeated for medium and low purity levels. The effectiveness was assessed based on the number of samples grouped. Table 5.48 shows the performance of the three incidents. From the results as tabulated in Table 5.48, the K-MC worked satisfactorily well and was able to group the five related samples according to their respective links/clusters when the medium purity level was used to train the machine. Training with the high and low purity levels however gave rise to errors. One sample of Cluster T was erroneously grouped into

Cluster Z when the high purity was employed as a training set. According to Table 5.48 for the low purity, the four related samples were not successfully grouped into Cluster K (except the one used for training). Similarly, only one sample (other than that used for training) was grouped under Cluster M. These samples belonging to Clusters M and K were erroneously grouped into Clusters P and T. Based on their performance, K-MC in this case is only adequate when training is done using samples of medium purity level. In general, the test infers that correct assignment of the training set is very important to ensure K-MC can perform well with the target data. Unfortunately, it is rather difficult to estimate whether the chosen training set is the best model for K-MC, and hence the weakness of this technique.

Table 5.48: Number of samples grouped according to the cluster

	High	**Medium**	**Low**
Cluster M	5	5	2
Cluster P	5	5	8
Cluster K	5	5	1
Cluster T	4	5	9
Cluster Z	6	5	5

iv) Discriminant analysis

The effectiveness of DA was examined by having all the 25 post-cut samples randomly assigned into training and validation sets. The DA acquires the source characteristics from the training and validation sets before new samples are subsequently grouped. In this subtask, the effectiveness was assessed based on how well the DA machine learned the characteristics. Blind test samples were not used because the dataset was too small to serve this purpose. A good statistical tool must be able to classify all the related samples according to their respective groups. This is indicated in the proportion of correct grouping (Table 5.49). As the machine ‘knows’ about the

sample origins (or called the true group) and based on this knowledge; DA was able to achieve 96% correctness with only one sample from Group T being erroneously placed in Group M. With this level of correctness, DA is assumed to be relatively more robust than K-MC as the correctness of the latter technique strictly depends on the correct assignment of the training set.

Table 5.49: Summary of classification with cross-validation for 25 simulated samples

Put into group	**True Group**				
	K	**M**	**P**	**T**	**Z**
K	5				
M		5		1	
P			5		
T				4	
Z					5
Total N	5	5	5	5	5
Proportion	1.000	1.000	1.000	1.000	1.000
N = 25	**N Correct = 24**			**Proportion Correct = 0.960**	

Note 1: Linear discriminant function was used.
Note 2: Only 11 peaks were used in DA as peak 12 is highly correlated with others, hence it was excluded.

5.5.2.4 Harmonized Statistical Model for Both Approaches

Based on the overall statistical validation, it was decided to use HCA for unsupervised pattern recognition since this technique showed excellent clustering potential with both the weight approaches. As two different approaches have been involved for case sample analysis in the later section, it is also important to determine whether both separate sets of data could be statistically analyzed using a harmonized statistical model simultaneously. In the previous section, Ward-Squared Euclidean and Ward-Squared Pearson showed promising results for the two approaches. In this section, both combinations (both have similar performance) were also employed to simultaneously examine the 55 + 25 = 80 N + S data points obtained from the two

approaches using HCA. During the pretreatment, all the data in the original readings (peak areas) were prepared on a single datasheet. The N + S pretreatment was performed on the data of 80 samples as a single dataset. Figure 5.56 shows that all the related samples were successfully clustered under their respective links with the d_m value = 4.15 using either of the combinations. Hence, it was decided to use either one of these combinations as a harmonized statistical model for the case samples.

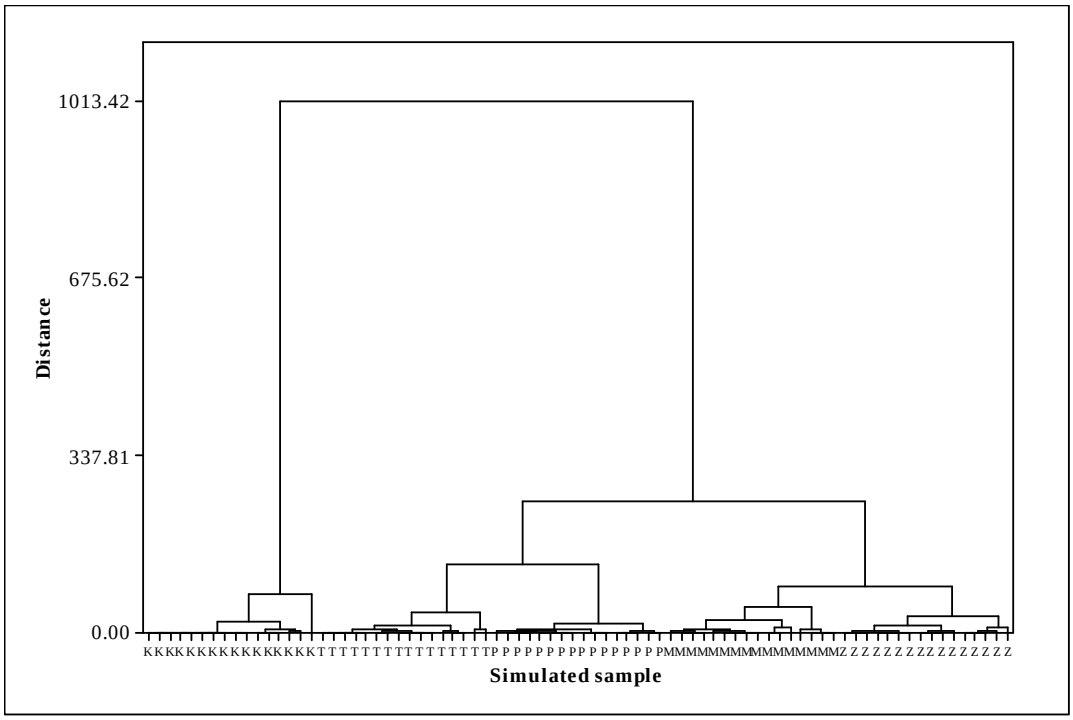


Figure 5.56: A dendogram expressed in distance showing the distance relationships between 80 simulated samples using Ward-Squared Euclidean or Ward-Squared Pearson.

5.5.2.5 Summary

The N + S pretreatment method was found to be ideal for clustering the related samples in the simulated dataset using data of 12 manufacturing impurities. When the samples were analyzed following the conventional 15 mg heroin base approach, PCA, DA and HCA were able to show correct sample grouping. However in HCA, although 24 combinations of linkage methods and distance measures were able to give zero

mistakes, only Ward-Squared Euclidean and Ward-Squared Pearson have the least d_m values and hence the most ideal combinations/models.

The limited amount of sample when analyzed via a constant 650 mg sample weight approach can still offer useful information. Four chemometric procedures were tested using 25 samples of known relationships that were analyzed at 650 mg each. HCA proved more powerful than PCA for unsupervised pattern recognition. When the assignment of known samples was used for supervised pattern recognition, DA was found to be more robust than K-MC as the latter requires correct assignment of the training samples.

The overall statistical validation showed that HCA operating with Ward-Squared Euclidean and Ward-Squared Pearson were superior for both approaches. Hence, this technique was used to simultaneously classify 80 samples analyzed via the two approaches. It showed zero mistake and reasonably low d_m value for sample grouping.

5.5.3 Analysis of Heroin Case Samples and Sample Classification by Trace Manufacturing Impurities

With the optimized GC-FID method, 298 samples (the remaining 11 samples from 309 genuine heroin samples determined in Task 3 were not selected and the reasons are given in the later section) were analyzed using either of the two sample weight approaches depending on the purity level. All data were obtained in peak areas. Peak identification was achieved solely by the RRT verified on a weekly basis (refer to Section 5.5.1.9 Additional Optimization Aspects). Subsequently, the proposed statistical models were applied to these samples. A maximum of 10 aliquots were analyzed on each day and therefore a control sample was not used. The instrument was cleaned on a weekly basis to ensure that the split liner was free of clogged material. After the maintenance, Validation Sample B was used to calibrate the retention times.

5.5.3.1 Sample Weight Equivalent to 15 mg heroin base Approach

252 illicit heroin cases were analyzed using the conventional sample weight equivalent to 15 mg heroin base approach as they contained > 2.3% heroin base. To display the frequency and concentration levels of the target peaks, each peak was first normalized to the sum of peak areas in order to eliminate any analytical error. The statistics of the concentrations of the target analytes are presented in Figure 5.57 and Table 5.50. Peaks 1, 4, 5 and 7 were detected in all the case samples. Peak 12 was not frequently detected owing to its low level. According to Figure 5.57, peaks 1, 4, 5 and 7 and to some extent peak 6 were often present in relatively higher concentrations while other impurities were only present in trace concentrations. For sample classification, these large peaks require standardization to minimize their dominant influence in the dataset.

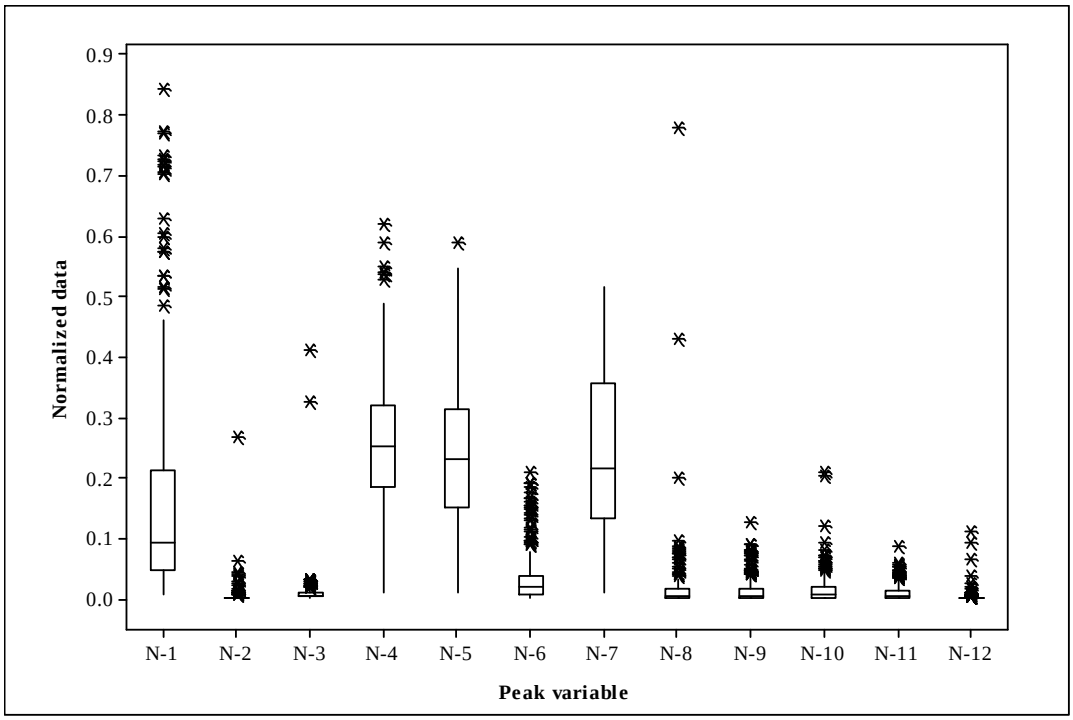


Figure 5.57: Boxplots showing the concentrations of 12 target impurity peaks in 252 illicit heroin case samples including zero values (absence) (The plots represent the normalized data)

Table 5.50: Statistical parameters for 12 impurity peaks found in 252 heroin case samples, excluding zero values (absence)

Peak	Frequency	Mean ± SD	Median	Range
1	252 (100%)	0.166 ± 0.178	0.095	0.0061 – 0.8431
2	163 (65%)	0.006 ± 0.022	0.001	0.0001 – 0.2679
3	235 (93%)	0.011 ± 0.034	0.005	0.0009 – 0.4128
4	252 (100%)	0.256 ± 0.116	0.253	0.0109 – 0.6193
5	252 (100%)	0.238 ± 0.109	0.231	0.0103 – 0.5904
6	245 (97%)	0.035 ± 0.042	0.020	0.0015 – 0.2085
7	252 (100%)	0.233 ± 0.133	0.215	0.0101 – 0.5160
8	245 (97%)	0.018 ± 0.059	0.006	0.0004 – 0.7776
9	232 (92%)	0.015 ± 0.020	0.007	0.0001 – 0.1281
10	211 (84%)	0.018 ± 0.026	0.010	0.0001 – 0.2105
11	215 (85%)	0.012 ± 0.014	0.006	0.0001 – 0.0862
12	88 (35%)	0.007 ± 0.017	0.002	0.0002 – 0.1125

Note: The figures represent the normalized data

The use of impurity data for sample classification can provide inferences about the relationships between samples at the production/manufacturing level. As both PCA

and HCA functioned well with this sample weight approach during statistical validation, so both techniques were also employed to analyze the case samples. Prior to PCA and HCA (Ward-Squared Euclidian or Ward-Squared Pearson), all the GC data reported in peak areas were subjected to the N + S pretreatment. Figures 5.58 and 5.59 show the relationships between the 252 case samples seized from four geographical locations of interest. According to the score plot, a large proportion of the samples were clustered in a dense area and they were likely to have come from the same manufacturing batch. It was also found that the majority of the samples from PH were grouped within this large cluster. Furthermore, the samples seized from SL illustrate that they could have come from more than one manufacturing batch. The outstanding outliers (Samples 23, 30, 32, 138 and 147 seized from SL) are less likely to be related to the main cluster and these samples could be from other origins. Based upon the distribution pattern on the score plot, it is likely that the street doses of heroin considered in this study were most probably from more than one manufacturing batch. The dendogram on the other hand indicates that many samples are closely related at the similarity level approaching 100%. These related samples are assumed to have come from similar sources. At the similarity level = -34.47, three major groups are obtained. At this level, Samples 23, 138 and 147 form a minor group. Before converging at the similarity level = -303.41, a major group and a minor group are finally obtained.

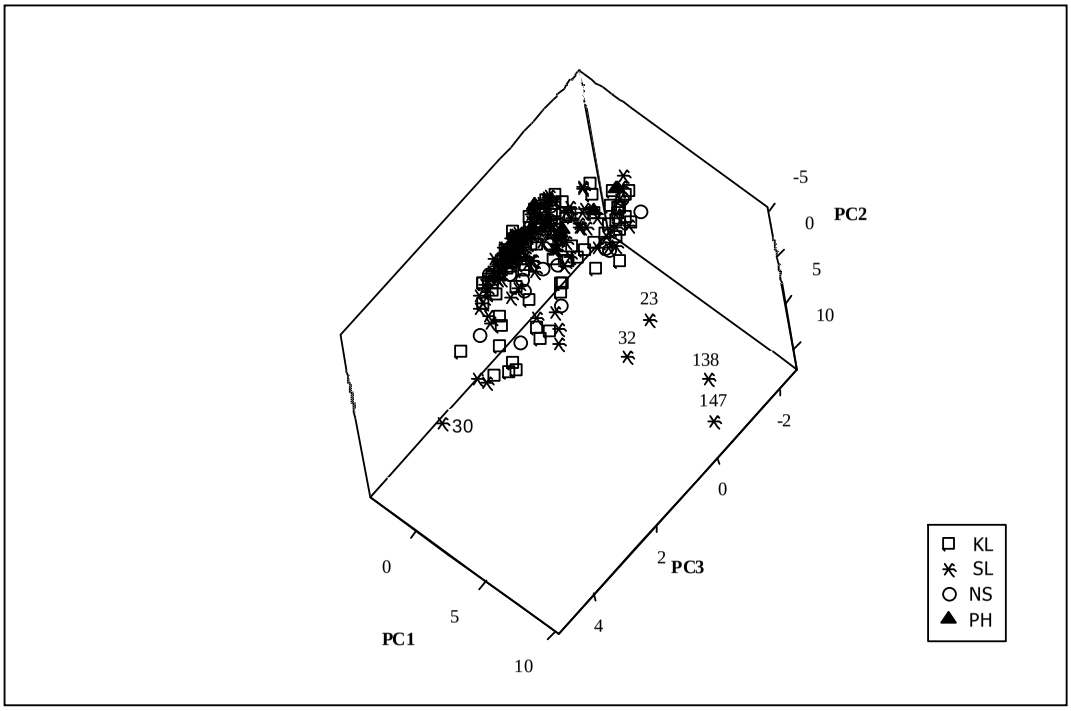


Figure 5.58: A score plot representing N + S pretreated impurity data obtained from 252 cases analyzed by PCA in covariance mode, $\%V_1 = 33.0$ %, $\%V_2 = 13.1\%$ and $\%V_3 = 9.7\%$

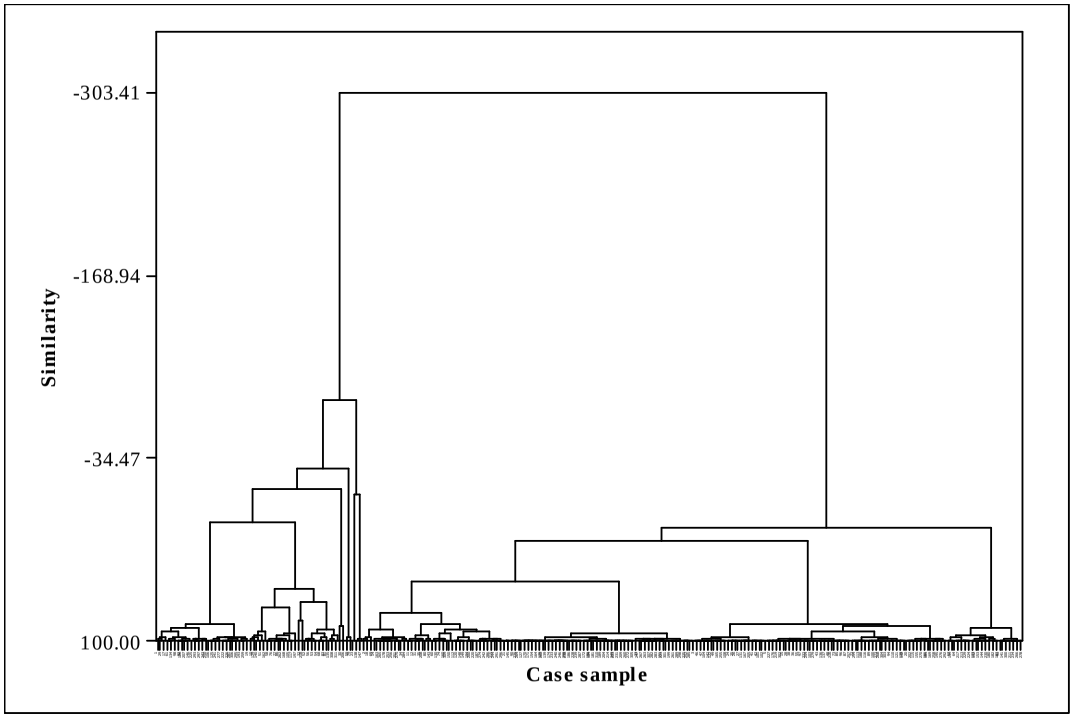


Figure 5.59: A dendogram expressed in similarity showing the relationships between 252 case samples using Ward-Squared Euclidian or Ward-Squared Pearson

5.5.3.2 The Constant Sample Weight (650 mg) Approach

A total of 46 case samples showing purity level < 2.3% were quantified using the 650 mg constant sample weight approach. The GC data are presented as N data in Figure 5.60 and Table 5.51. The statistics show that peaks 1, 4, 5 and 8 were still detectable in 100% of the samples by the GC-FID although only 650 mg of the sample was used for each case. However, peaks 3 and 9 became less detectable due to their trace amounts in the highly cut samples. The dominant peaks in this dataset were peaks 1, 4 and 5 as they were present in high concentrations. Prior to sample classification, standardization of the normalized data was performed to equalize the weightage of all the peaks in the dataset.

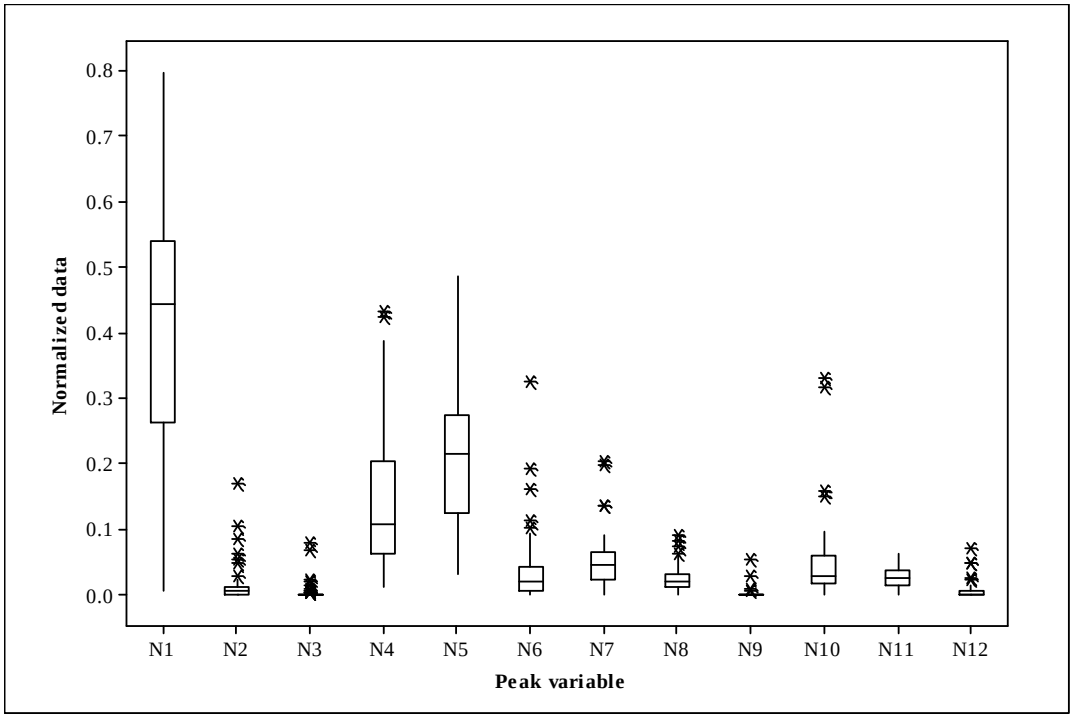


Figure 5.60: Boxplots showing the concentrations of 12 target impurity peaks in 46 illicit heroin case samples including zero values (absence) (The plots represent the normalized data)

Table 5.51: Statistical parameters for 12 impurity peaks found in 46 heroin case samples, excluding zero values (absence)

Peak	Frequency	Mean ± SD	Median	Range
1	46 (100%)	0.408 ± 0.199	0.442	0.00710 – 0.79710
2	32 (70%)	0.023 ± 0.037	0.008	0.00004 – 0.17040
3	11 (24%)	0.021 ± 0.027	0.008	0.00227 – 0.07828
4	46 (100%)	0.150 ± 0.116	0.109	0.01130 – 0.43170
5	46 (100%)	0.216 ± 0.113	0.216	0.03000 – 0.48610
6	40 (87%)	0.046 ± 0.063	0.022	0.00318 – 0.32597
7	45 (98%)	0.056 ± 0.042	0.047	0.00760 – 0.20234
8	46 (100%)	0.025 ± 0.022	0.020	0.00003 – 0.09144
9	4 (9%)	0.024 ± 0.022	0.019	0.00570 – 0.05320
10	45 (98%)	0.052 ± 0.068	0.029	0.00070 – 0.33030
11	44 (96%)	0.028 ± 0.016	0.027	0.00021 – 0.06190
12	17 (30%)	0.017 ± 0.020	0.008	0.00116 – 0.07055

Note: The figures represent the normalized data

As shown in Section 5.5.2.3 The Constant Sample Weight (650 mg) Approach, PCA demonstrated a relatively weak discriminative power for samples analyzed at the 650 mg constant sample weight. Therefore, the N + S data of the 46 case samples were subjected to HCA (Ward-Squared Euclidian or Ward-Squared Pearson) for unsupervised pattern recognition. The relationships between these samples are presented in Figure 5.61. At the similarity level approaching 100%, many case samples on the left side of the dendogram are closely related, whereas the samples on the right side are not closely packed. The closely related samples in their respective clusters most likely belonged to the same manufacturing batches. At the similarity level = 0.31, four groups are obtained with the pair of Samples 99 and 271 being the smallest group. One major group and one minor group are finally clustered at the similarity level = -199.07.

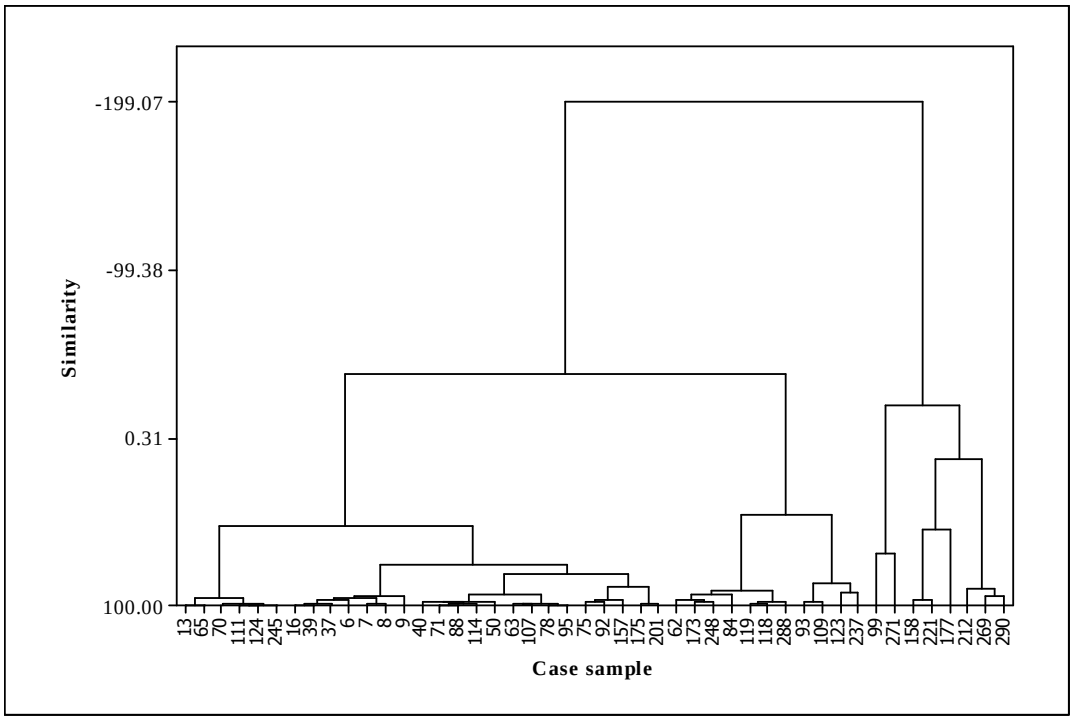


Figure 5.61: A dendogram expressed in similarity showing the relationships between 46 case samples using Ward-Squared Euclidian or Ward-Squared Pearson.

5.5.3.3 Classification of Heroin Case Samples Using Harmonized Statistical Model

The N + S impurity data of 298 case samples were subjected to HCA using Ward-Squared Euclidian or Ward-Squared Pearson for sample classification. The hierarchical relationships between these samples based on 12 impurity peaks are presented in Figure 5.62. At the similarity level = -54.10, three groups are obtained. Of the three groups, the middle group is the smallest cluster which contains only ten samples. At the similarity level = -362.30, two final groups are joined on the dendogram. In these two final clusters, the case samples in the right cluster are more closely related than the case samples in the left cluster. The right cluster is completely formed before it reaches a quarter of the similarity exhibited by the whole dataset. These samples were most likely processed from a similar production line. The results from the dendogram however are not in accord with those obtained with the major components in Figure 5.38. This could be due to the presence of zero-values in the

impurity profiling. Therefore, the dendograms obtained in this study can best serve as a general overview of the sample relationships rather than their true origins.

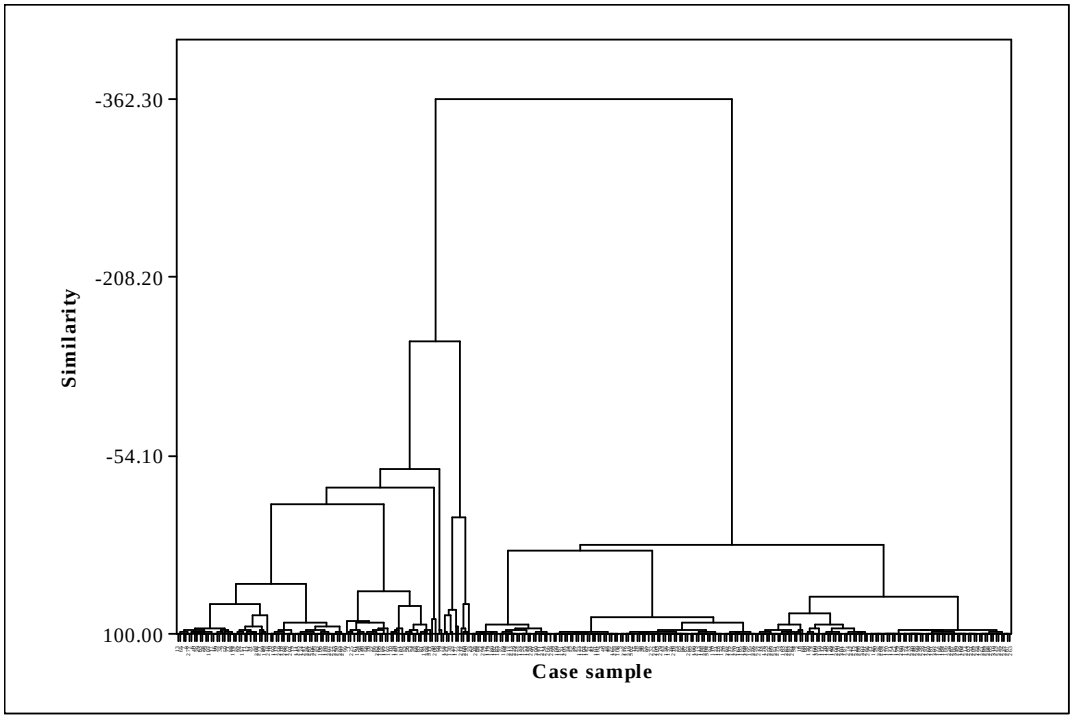


Figure 5.62: A dendogram expressed in similarity showing the relationships between 298 case samples using Ward-Squared Euclidian or Ward-Squared Pearson

5.5.3.4 Limitations

The overall analytical and statistical methods functioned satisfactorily for the 298 case samples. The analytical method however did not work well for the remaining 11 genuine heroin case samples as they contained interferents that reacted with sulfuric acid, rendering unreliable conditions for profiling. Besides, they also showed unusual profiles in the chromatograms (Figure 5.63). Hence, these samples were excluded for impurity profiling.

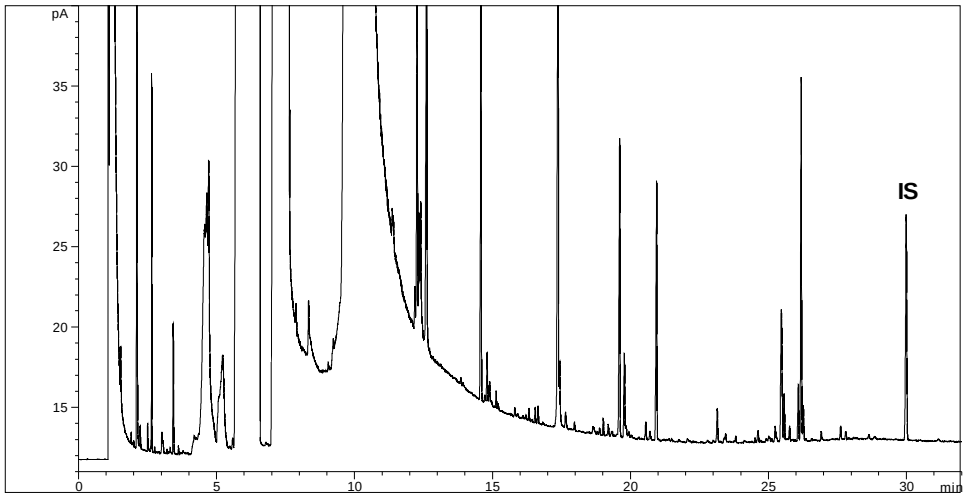


Figure 5.63: An unusual profile shown by Case 66

5.5.3.5 Summary

From the analysis of the 298 case samples, peaks 1, 4 and 5 were frequently detected. Peak 12 was not usually present in the samples. The 252 samples analyzed using the sample weight equivalent to 15 mg heroin base approach showed a major cluster on the score plot with a few exceptional cases. Hence, it is deduced that more than one manufacturing source were responsible for the samples. Besides, 46 samples analyzed using the 650 mg constant sample approach suggested that these samples could be grouped in a minor group and a major group by HCA (Ward-Squared Euclidean or Ward-Squared Pearson) as far as the final two clusters are concerned. The 298 samples were finally analyzed by HCA (Ward-Squared Euclidean or Ward-Squared Pearson) and it suggested that many samples are closely related at high similarity levels. The 11 case samples containing reactant interferents were not included for profiling.