



# CLASSIFICATION AND IDENTIFICATION OF BLACK BALLPOINT PEN INKS BASED ON MULTIVARIATE ANALYSIS AND INFRARED SPECTRA

Loong Chuen LEE<sup>1,2</sup>, Mohamed Rozali OTHMAN<sup>1</sup>, Hiang PUA<sup>2</sup>, Ab Aziz ISHAK<sup>2</sup>, Seri Mirianti ISHAR<sup>2</sup>

<sup>1</sup> Faculty of Science & Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

<sup>2</sup> Faculty of Health Science, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

## Abstract

A non-destructive method for the classification and identification of black ballpoint pen inks on a questioned document was developed. Micro-ATR-FTIR spectroscopy allowed ink samples to be analysed *in situ* on a paper substrate. A total of 155 sample pens encompassing forty models of nine different brands were sampled from stationary shops at Subang Jaya, Malaysia. IR spectra data was processed by a number of multivariate analysis techniques, including cluster analysis (CA), principal component analysis (PCA) and discriminant analysis (DA). The proposed method gave satisfactory results relating to classification and identification of pen inks at the level of the brand as well as the model.

## Key words

Infrared spectrum; Ink analysis; Classification; Identification; Principal component analysis; Discriminant analysis.

*Received 27 March 2012; accepted 2 August 2012*

## 1. Introduction

Forensic ink analysis is one of the main focuses of questioned document examination. The forensic examination of writing inks usually involves various instrumental techniques to reveal the composition of inks, which assist in identification of the source of the questioned ink. The brand, model, manufacturer and year of production of the instrument that could have produced the ink specimen are the information needed in the identification process [16].

There are two main types of instrumental techniques. Most of the time, destructive techniques such as high performance thin layer chromatography (HPTLC) [16], high performance liquid chromatography (HPLC) [2], time-of-flight secondary ion mass spectrometry (TOF-SIMS) [9], and laser desorption ionization-mass spectrometry (LDI-MS) [11] could provide more information than non-destructive tech-

niques. However, these techniques disrupt the integrity of the sample. As a consequence, the evidence might not be intact for further testing or review at a later date. Therefore, non-destructive techniques like optical techniques are preferred.

Attenuated total reflectance (micro-ATR-FTIR) is a non-destructive sampling method for Fourier transform infrared (FT-IR) spectroscopy. It has been used to determine the sequence of intersecting lines [4] and to analyze blue ballpoint pen inks [13, 22] as well as red seal inks [10]. Micro-ATR-FTIR utilizes an attenuated total reflection (ATR) accessory, which operates by measuring the changes that occur in a totally internally reflected infrared beam when the beam comes into contact with a sample. An infrared beam is directed onto an optically dense crystal in contact with the sample. This evanescent wave protrudes only a few microns beyond the crystal surface and into the sample. FTIR-ATR has advantages over FTIR-KBr in that

sample preparation is quicker and less messy, user-to-user spectra variation is minimized and the impact of sample preparation on results is reduced [21].

Recently, multivariate techniques have been applied to various forensic analyses to assure objective interpretation of the results. Cluster analysis (CA), principal component analysis (PCA) and discriminant analysis (DA) are the most popular techniques being explored thoroughly by forensic scientists for discrimination of soils and ivories, determination of production batches of illegal heroin, and comparison of hair fibres as well as identification of pen inks, post-blast residues, firearms and also paper [2, 3, 5, 6, 13, 14, 17, 18, 19, 20]. Multivariate analysis is a powerful tool when dealing with multi-component systems and offers an interpretative methodology for discerning significant differences from complicated datasets, e.g. IR spectrum of ink [20].

This study attempted to test the capability of micro-FTIR in attenuated total reflectance (ATR) mode as a non-destructive tool to analyse black ballpoint pen inks. Additionally, multivariate analysis techniques: cluster analysis (CA) and principal component analysis (PCA) were combined with discriminant analysis (DA) to objectively classify and identify the ink samples at the level of brand and model sequentially.

## 2. Material and methods

### 2.1. Samples

A total of 155 black ballpoint pens were sampled from stationery shops in Malaysia. All of these were obtained in multiple packs of the same product to ensure all of them were from the same production batches, as the variation between batches was considered in this study. All pen brands and models were allocated a reference number for the purpose of this study. Detailed descriptions of collected samples used in this study are presented in Table I. Double A from Thailand was the white paper used throughout the study. Each of the 155 pens was used to draw three different small circles to ensure complete coverage of paper by the pens. Experimental work was carried out on the ink deposited on the white paper not later than one day after drawing to minimize changes due to ink aging.

### 2.2. Micro-ATR-FTIR analysis

All experimental spectroscopy was carried out on a Thermo Scientific Nicolet iN10 MX FT-IR microscope with mercury cadmium telluride (MCT) detec-

TABLE I. REFERENCE NUMBER FOR LIST OF STUDIED BLACK BALLPOINT PENS

Reference number	Brand	Reference number	Model	Quantity
B1	Bic	M1	BU3 0.8	4
		M2	BU2 0.7	3
		M3	RS 2	4
		M4	NS Med	3
		M5	Org NS fine	3
		M6	Round Stic	4
B2	Faster	M7	CX813	1
		M8	CX818	4
		M9	CX1006	4
		M10	CX838	4
		M11	CX608	3
		M12	828	5
B3	Faber-Castell	M13	Super click	4
		M14	Ballpen 1.0	4
		M15	Ballpen 0.7	4
		M16	Click ball	4
		M17	Needle ball	4
		M18	Ballpen 0.5	4
B4	G-soft	M19	gs 66	3
		M20	gs 77	3
		M21	r 100	4
B5	MGM	M22	BP 713	6
		M23	e-Rite	4
B6	PILOT	M24	Super GP	4
B7	Paper Mate	M25	KM 100 Med	4
		M26	KM 100 Fine	4
		M27	KM 100 S/F	4
		M28	KV 2 Med	4
		M29	KV 2 Fine	4
		M30	KV 2 S/F	4
		M31	KM Med PT	4
		M32	KM Fine PT	5
		M33	KM F	4
		B8	Stabilo	M34
M35	Liner 808 F			4
M36	Liner 308			3
M37	Marathon 318			2
M38	Galaxy 818 M			5
B9	UNicorn	M39	BP 318 GS	5
		M40	BP 333 DT	5
			Total	155

tor. A Ge crystal tip ATR objective was used as a micro-sampling accessory. The background spectrum was reacquired after every analysis to reduce variation in the spectra due to instrumental drift. Each spectrum was the result of an average 16 scans at  $4\text{ cm}^{-1}$  resolution over a spectrum range of 4000 to 675 wavenumber [ $\text{cm}^{-1}$ ]. Three spectra were collected from each of 155 pens. The IR spectra data was stored in a data matrix in Microsoft Excel spreadsheet.

### 2.3. Software

Data collected was processed and analysed using statistical package SPSS (Statistical Package for the Social Sciences, Windows version 15.0, SPSS Inc., Chicago, USA).

## 3. Results & discussion

### 3.1. Sampling

In brief, there are three different sources of variations present in pen ink samples: a – variation between groups; b – variation within a group (differences between different individual pens of a single group) and c – variation between replicates within a single individual pen (errors inherent to the method of analysis). Two groups can be considered different if variation (a) is significantly bigger than variation (b) and (c). On the other hand, the aforementioned sources of variations could be present at four different levels including batches of production, brands, models and ballpoint sizes. Most studies in the literature draw only one sample to represent a particular group and seldom take into account variation (b) and (c). This study considered all three sources of variations simultaneously. In this way, we could evaluate the method more thoroughly and explicitly distinguish the differences between each group in more detail. Moreover, the resulting conclusion can be applied to the population rather than to the sub-population. In this study, the variations between groups at the level of brand (inter-brand variability) as well as model (inter-model variability) were determined systematically. For determination of variation (b) and (c), each brand/model was represented by more than one individual pen, and each pen was subjected to triplicate analysis.

### 3.2. Infrared spectrum

Careful examinations of all infrared spectra obtained revealed that the regions from 2000–675  $\text{cm}^{-1}$  were the most informative part. IR absorption peaks

due to water vapour and carbon dioxide from the atmosphere were present in the region between 4000 and 2000  $\text{cm}^{-1}$  [13]. Figure 1 shows representative IR spectra for each of forty different models according to nine different brands. All of the IR spectra exhibit two prominent peaks at approximately 1584 and 1360  $\text{cm}^{-1}$  but with differences in the form of relative peak height and shape. They are characteristic peaks of ballpoint pen inks, and have previously been assigned to a skeletal vibration of triarylmethane dye [22]. Besides that, IR spectra of model M14-16 and M34-38 show a minute peak at around 1730  $\text{cm}^{-1}$  whereas IR spectra of model M22-23 and M39-40 show a minute peak at around 1250  $\text{cm}^{-1}$ . According to Wang et al. [22], both peaks support the presence of polyester and phthalic ester, both of which are well known constituents of resins of pen inks [13]. Peaks present in the region between wavenumbers 1100–1000  $\text{cm}^{-1}$  are mostly from paper substrate [7]. The vibrational assignment for other peaks has been explained elsewhere [22].

### 3.3. Statistical analysis

All three spectra of each sample were included when assessing material reproducibility and homogeneity, thus ensuring representativeness of each sample (this is necessary due to the minimal size of inks contacted by the ATR Ge tip and the potential heterogeneity of the inks) [5]. Triplicate analysis of 155 pens gave 464 spectra, as one replica spectrum of sample 11 was lost. The full data set comprising 464 spectra, consisting of absorbance values as a function of wavenumber, were normalized prior to carrying out the multivariate analysis to reduce variation in the data due to different thickness of the pen inks deposited on paper samples. The absorbance values for each spectrum were divided by the total absorbance resulting from each spectrum across wavenumbers 2000–675  $\text{cm}^{-1}$ .

As mentioned in the “Introduction”, this study aimed to analyse IR spectra of pen inks at the level of brand and model. To this end, all 464 normalised spectra were rearranged into a few different files. On the one hand, a file comprising all 464 spectra was used to classify IR spectra of pen inks according to their brands. On the other hand, those 464 spectra were grouped into different files based on their brands for determination of differences between IR spectra of pen inks at model-level within a single brand.

Subsequently, each file was processed by CA and PCA sequentially to produce new sets of variables for that particular file. Firstly, CA was performed by using Ward’s method and squared Euclidean distance to form eight different clusters composed of similar data

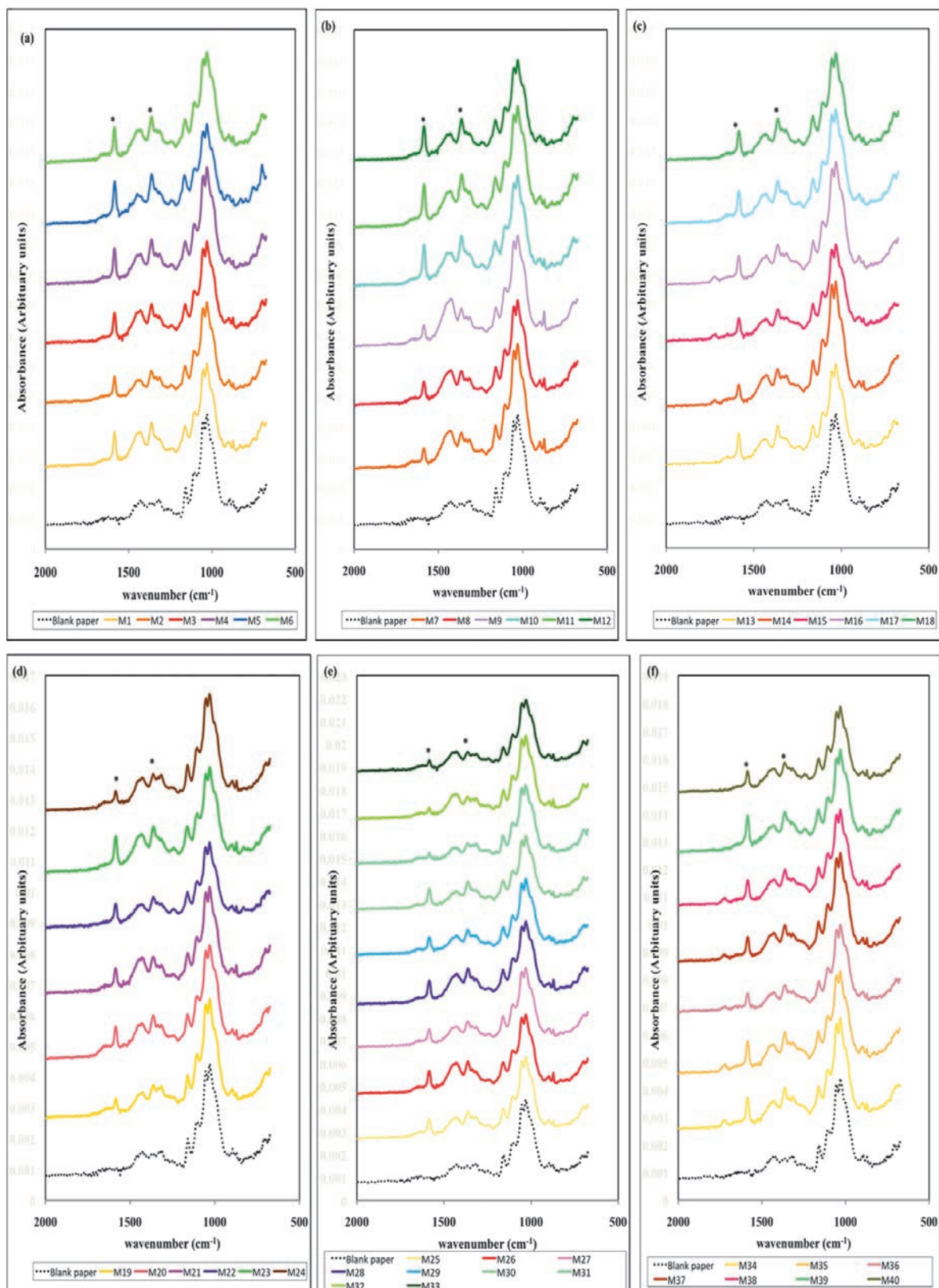


Fig. 1. Representative IR spectra of black ballpoint pen inks (–) according to brands and blank paper (...) analysed by micro-ATR-FTIR spectroscopy in the region of 2000–675  $\text{cm}^{-1}$ : a – Bic; b – Faster; c – Faber-castell; d – G-soft, MGM and Pilot; e – Paper-mate; f – Stabilo and UNIcorn.

points on the IR spectrum. Secondly, PCA was performed on each of the eight clusters, separately. The rationale behind this approach is that PCA running on a smaller number of data points would result in a more stable matrix than PCA running on all 688 data points all at once. The number of components to be extracted was decided based on the Kaiser criterion [12]. The new set of variables comprised all selected principal components, which were labelled accordingly, e.g. PCA 1\_2 refers to the first component produced by PCA conducted on the second cluster of data points. Transformation of a high number of raw data into a small number of new sets of variables is vital to ensure accuracy and efficiency of DA [15]. On the other hand, PCA scores have been proven to be more effective than raw variables for DA [19, 20]. In addition, by using extracted components as new latent variables, the problem of redundancy is eliminated as the components are orthogonal and uncorrelated. Finally, stepwise DA was performed by using the Mahalanobis distance method and probability- $F$  as the criterion. Prior probabilities were equal for all groups (brands) as the representativeness of the population was unknown [8].

### 3.4. Classification

All 464 IR spectra were classified according to their brands as well as models sequentially. Firstly, all 464 IR spectra were classified according to brand (regardless of the pen models) by discriminant analysis (DA). Subsequently, pen models within a single brand were classified by conducting another DA. Table II summarises the results of classification of 464 IR spectra at the level of brand and model by DA.

Based on Table II, all nine different pen brands were divided into three separate groups according to the linear discriminant function LDF 1 and LDF 2 scores. The first, second and third clusters are composed of B5 and B9; B3 and B8; B1, B2, B4, B6 and B7, respectively. Pen brands that were grouped together might exhibit similar mid-IR characteristics. On the one hand, group 2 (G2), composed of B3 and B8, shows a unique similarity, which is a small peak present at around  $1730\text{ cm}^{-1}$  on IR spectra of pen inks from both B3 (Figure 1c) and B8 (Figure 1f). This peak does not present at IR spectra of other brands. On the other hand, IR spectra of pen inks from both B5 and B9 that were clustered into group 3 (G3) give a minute peak at around  $1245\text{ cm}^{-1}$ .

Nevertheless, examinations of all spectra (Figure 1) did not show any obvious differences among IR spectra of pen inks from different brands. In other

TABLE II. CLASSIFICATION OF 464 IR SPECTRA OF BLACK BALLPOINT PEN INKS ACCORDING TO BRANDS AS WELL AS MODELS WITHIN A SINGLE BRAND

Brand	Models within a single brand
G1	B1 M1, M3, M4    M2, M5    M6
	B2 M7    M8, M9, M11    M10, M12
	B4 M19, M20, M21
	B6 M24
	B7 M25, M26, M27    M28, M29, M30    M31, M32, M33
G2	B3 M13, M18, M17    M14, M15, M16
	B8 M34, M35, M36, M3    M37
G3	B5 M22, M23
	B9 M39, M40

words, the minor differences of a particular ink relative to other inks which could not be noticed easily by visual inspection, were successfully assessed quantitatively by multivariate analysis. For those pen brands that could not be discriminated from each other, other techniques including Raman spectroscopy or mass spectrometry could be conducted to provide more information about the ink's composition.

Subsequently, each model was further classified by another DA after being grouped according to brand. All forty models can be divided into at least two different groups within a particular brand, except for models from B4, B5, B6 and B9. For example, all six models (M7-12) of Faster (B2) black ballpoint pens have been grouped into three different clusters by linear discriminant function LDF 1 and LDF 2 as shown by Figure 2. In brief, results indicated a reasonable level of differentiation does exist among studied pen brands/models.

### 3.5. Identification of pen inks based on IR spectra

As mentioned in 3.3., all 464 normalised spectra were rearranged into a few different files with the aim of analysing IR spectra of pen inks at the level of brand and model sequentially. For each file, a flow chart was designed to identify the pen brand/models of studied inks objectively based on the absence or presence of significant statistical difference of selected principal components in comparison with a pre-chosen reference spectrum. For that purpose, one-way ANOVA at 95% confidence limits together with a suitable *post-hoc* test was conducted to compare each pair of pen inks using the new set of variables. For instance, Figure 3 shows the flow chart for identifying six models (M7-12) of

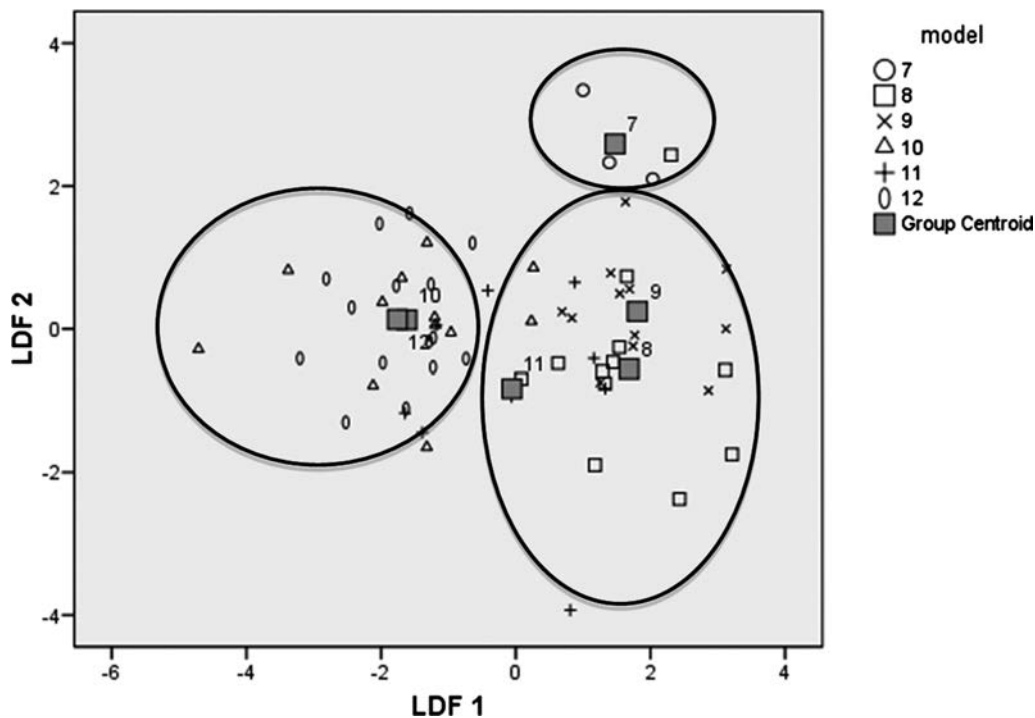


Fig. 2. Scores plot of first two canonical discriminant functions showing the separation of group centroids for six different models of Faster pen inks.

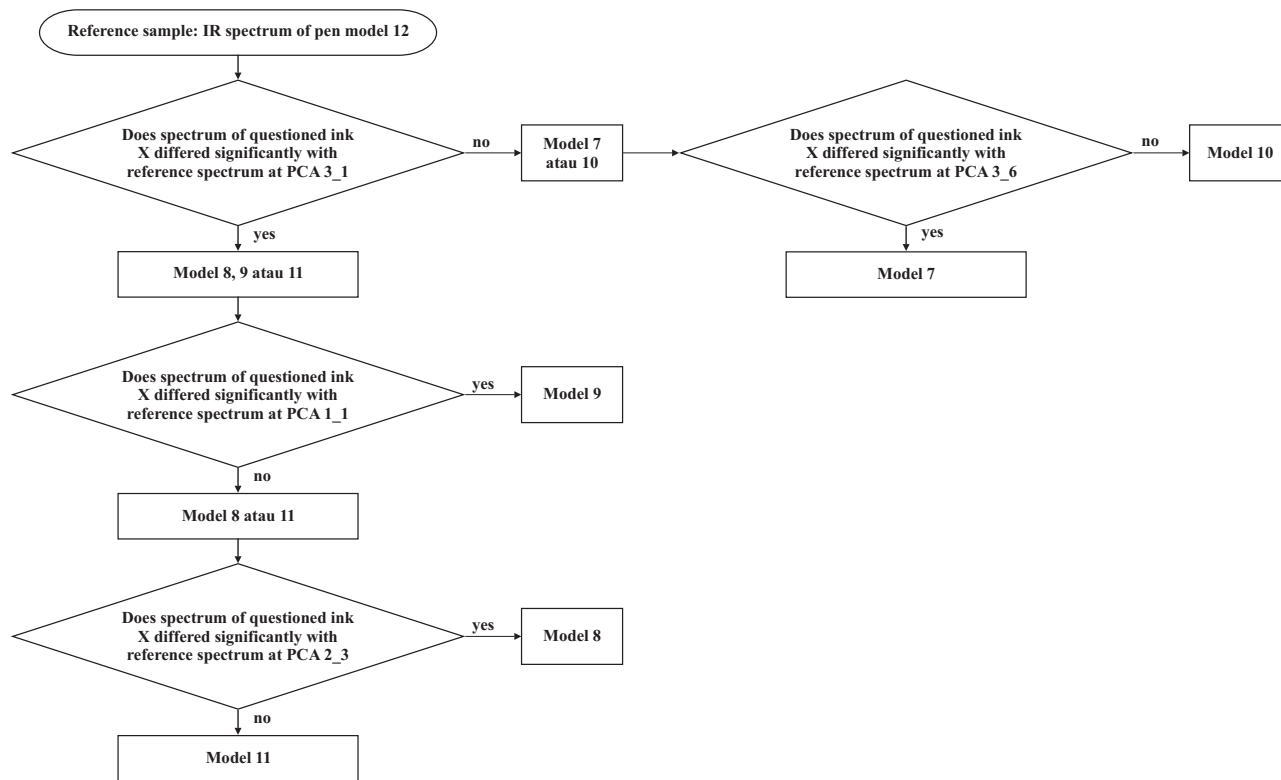


Fig. 3. Flow chart showing identification of Faster pen ink models based on IR spectrum.

Faster (B2) black ballpoint pens. After examining the output of post-hoc tests, it was decided that the IR spectrum of M12 would act as a reference spectrum. Firstly, questioned ink X is compared against the reference sample (the IR spectrum of pen model 12) at PCA 3\_1. If the score value of both at PCA 3\_1 is significantly different at 95% confidence interval, then sample X could be from model 8, 9 or 11. On the other hand, an insignificant result indicates sample X could likely be model 7 or 10. The identification process may continue until the model of sample X is eventually identified.

The idea was to first determine the brand of questioned pen inks using a flow chart established at the level of brand (not shown here). After that, the model of the pen might be determined using a flow chart established at the level of model for a particular brand. By determining the brand/model of the inks, we would be able to narrow down our scope which would then lead to a high chance of identification of the questioned ink. However, the proposed flow charts at this stage are only applicable to the studied brands/models. For future application, it is necessary to acquire more samples from different manufacturers for database building. Meanwhile, more analyses need to be conducted to identify those brands/models that cannot be discriminated by the IR technique. At the same time, the proposed flow charts need to be evaluated by the likelihood ratio approach as explained by [1] prior to application in a real case.

#### 4. Conclusions

In conclusion, IR spectra of pen inks could be discriminated better at brand-level than model-level. Although full discrimination between pens based on brands and models was not achieved, the obtained results showed that a combination of micro-ATR-FTIR and multivariate analysis was able to act as a fast and easy discrimination tool in differentiating pen inks according to selected brands and models.

#### Acknowledgements

We are grateful for assistance in the acquisition of infrared data from Mr. Mohd Fizol and Mrs. Rusikah bt. Minhad from Jabatan Kimia Malaysia. We wish to thank Prof. Dr. Aziz Jermain for a useful discussion on multivariate analysis. We would also like to extend our gratitude to all the staff of the forensic science program from UKM (Universiti Kebangsaan Malaysia), as well as Mrs. Hartinah Annuar. Also thanks to UKM for research grant UKM-DLP-2012-024.

#### References

1. Aitken C. G. G., Statistics in forensic science. Part II. An aid to the evaluation of evidence, *Problems of Forensic Sciences* 2006, 65, 68–81.
2. Banas K., Banas A., Moser H. O. [et al.], Multivariate analysis techniques in the forensic investigation of the postblast residues by means of Fourier-transform infrared spectroscopy, *Analytical Chemistry* 2010, 82, 3038–3044.
3. Barton P. M. J., A forensic investigation of single human hair fibres using FTIR-ATR spectroscopy and chemometrics, Queensland University of Technology 2011 [PhD thesis].
4. Bojko K., Roux C., Reedy B. J., An examination of the sequence of intersecting lines using attenuated total reflectance-Fourier transform infrared spectra imaging, *Journal of Forensic Sciences* 2008, 53, 1458–1467.
5. Brody R. H., Edwards H. G. M., Pollard A. M., Chemometric method applied to the differentiation of Fourier-transform Raman spectra of ivories, *Analytica Chimica Acta* 2001, 427, 223–232.
6. Brooks E., Comber B., Robertson J., Digital imaging and image analysis applied to numerical application in forensic hair examination, *Science & Justice* 2011, 51, 28–37.
7. Causin V., Marega C., Marigo A. [et al.], Forensic differentiation of paper by X-ray diffraction and infrared spectroscopy, *Forensic Science International* 2010, 197, 70–74.
8. Chan Y. H., Biostatistics 303: Discriminant analysis, *Singapore Medical Journal* 2005, 46, 54–61.
9. Denman J. A., Skinner W. M., Kirkbride K. P. [et al.], Organic and inorganic discrimination of ballpoint pen inks by TOF-SIMS and multivariate statistics, *Applied Surface Science* 2010, 256, 2155–2163.
10. Dirwono W., Park J. S., Agustin-Camacho M. R. [et al.], Application of micro-attenuated total reflectance FTIR spectroscopy in the forensic study of questioned documents involving red seal inks, *Forensic Science International* 2010, 199, 6–8.
11. Gallidabino M., Weyermann C., Marquis R., Differentiation of blue ballpoint pen inks by positive and negative mode LDI-MS, *Forensic Science International* 2011, 204, 169–177.
12. Kaiser H. F., The application of electronic computers to factor analysis, *Applied Psychological Measurement* 1960, 20, 141–151.
13. Kher A., Mulholland M., Green E. [et al.], Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis, *Vibrational Spectroscopy* 2006, 40, 270–277.
14. Klemenc S., In common batch searching of illicit heroin samples-evaluation of data by chemometrics methods, *Forensic Science International* 2001, 115, 43–52.

15. Mallet Y., Coomans D., de Vel O., Recent developments in discriminant analysis on high dimensional spectra data, *Chemometrics and Intelligent Laboratory Systems* 1996, 35, 157–173.
16. Neumann C., Ramotowski R., Genessay T., Forensic examination of ink by high-performance thin layer chromatography – the United States Secret Service Digital Ink Library, *Journal of Chromatography A* 2011, 1218, 2793–2811.
17. Nor A., Md G., Liong C. Y. [etal], Analysis of geometric moments as features for firearm identification, *Forensic Science International* 2010, 198, 143–149.
18. Panayiotou H., Vibrational spectroscopy of keratin fibers. A forensic approach, Queensland University of Technology 2004 [PhD thesis].
19. Thanasoulis N. C., Parisi N. A., Evmiridis N. P., Multivariate chemometrics for the forensic discrimination of blue ball-point pen inks based on their Vis spectra, *Forensic Science International* 2003, 138, 75–84.
20. Thanasoulis N. C., Piliouris E. T., Kotti M-S. E. [et al.], Application of multivariate chemometrics in forensic soil discrimination based on the UV-Vis spectrum of the acid fraction of humus, *Forensic Science International* 2002, 130, 73–82.
21. Thompson T. J. U., Gauthier M., Islam M., The application of a new method of Fourier Transform Infrared Spectroscopy to the analysis of burned bone, *Journal of Archaeological Science* 2009, 36, 910–914.
22. Wang J., Lou G., Sun S. [et al.], Systematic analysis of bulk blue ballpoint pen ink by FTIR spectrometry, *Journal of Forensic Sciences* 2001, 46, 1093–1097.

---

**Corresponding author**

L. C. Lee  
Program of Forensic Science  
Faculty of Health Science, UKM  
Jalan Raja Muda Abdul Aziz  
50300 Kuala Lumpur, Malaysia  
e-mail: dkcindy@gmail.com

---



# KLASYFIKACJA I IDENTYFIKACJA CZARNYCH PAST DŁUGOPISOWYCH NA PODSTAWIE ANALIZY WIDM W PODCZERWIENI Z WYKORZYSTANIEM METOD WIELOWYMIAROWEJ ANALIZY DANYCH

## 1. Wprowadzenie

Analiza past długopisowych dokonywana na potrzeby wymiaru sprawiedliwości jest jednym z głównych problemów, na jakim skupiają się biegli zajmujący się analizą autentyczności dokumentów. Tego typu badania wymagają zwykle zastosowania różnych metod analizy instrumentalnej w celu ustalenia składu pasty długopisowej, co z kolei pozwala ustalić ich źródło pochodzenia. Informacje uzyskane w procesie identyfikacji narzędzia pisarskiego, które mogło pozostawić próbkę pasty długopisowej na badanym dokumencie, dotyczą zwykle marki, modelu, producenta czy też roku produkcji narzędzia pisarskiego [16].

Metody analizy instrumentalnej stosowane w badaniach past długopisowych można podzielić na dwie grupy. Do najczęściej stosowanych należą metody niszczące, takie jak wysokosprawna chromatografia cienkoinwarstwowa (HPTLC) [16], wysokosprawna chromatografia cieczowa (HPLC) [2], spektrometria jonów wtórnych z wykorzystaniem analizatora czasu przelotu (TOF-SIMS) [9] czy spektrometria mas z jonizacją przez desorpcję laserową (LDI-MS) [11]. Metody te z reguły dostarczają więcej informacji niż techniki nieniszczące. Jednakże techniki te ze swej natury powodują zniszczenie analizowanej próbki (dowodu) i w konsekwencji mogą uniemożliwić jej dalsze badanie np. w późniejszym czasie. Dlatego też w badaniach sądowych preferowane są techniki nieniszczące, takie jak np. badania optyczne.

Technika osłabionego całkowitego wewnętrznego odbicia (mikro-ATR) należy do nieniszczących metod analizy stosowanych w fourierowskiej spektroskopii w podczerwieni (FT-IR). Wykorzystywano ją m.in. w ustalaniu kolejności nakładania krzyżujących się linii [4], w analizie niebieskich past długopisowych [13, 22], jak również w badaniu czerwonych atramentów do pieczęci [10]. Metoda mikro-ATR-FTIR wykorzystuje przystawkę do osłabionego całkowitego odbicia (ATR), która działa na zasadzie pomiaru zmian intensywności promieniowania w zakresie podczerwieni, które pojawiają się w wyniku całkowitego wewnętrznego odbicia wiązki podczerwonej podczas oddziaływania tej wiązki na próbkę. Wiązka promieniowania podczerwonego kierowana jest na kryształ o dużej gęstości optycznej oraz na próbkę, która pozostaje w kontakcie z tym kryształem. Wnika ona na głębokość rzędu kilku mikronów w głąb kryształu oraz badanej próbki. Zaletą metody FTIR-ATR w sto-

sunku do metody FTIR-KBr jest szybsze i mniej żmudne przygotowanie próbki, ograniczenie wpływu operatora na zmienność uzyskiwanych widm, a także ograniczenie wpływu przygotowania próbki na wyniki analiz [21].

Ostatnio, aby zapewnić większy obiektywizm w procesie interpretacji wyników uzyskanych w różnych badaniach prowadzonych dla wymiaru sprawiedliwości, zaczęto stosować chemometryczne metody wielowymiarowej analizy danych. Do najbardziej popularnych i gruntownie poznanych przez biegłych sądowych metod należy analiza skupień (CA), analiza głównych składowych (PCA) oraz analiza dyskryminacyjna (DA). Metody te zastosowano m.in. w różnicowaniu próbek ziemi i kości słoniowej, w różnicowaniu partii produkcji nielegalnej heroiny, w badaniach porównawczych włosów, jak również w identyfikacji past długopisowych, śladów powybuchowych, broni palnej czy papieru [2, 3, 5, 6, 13, 14, 17, 18, 19, 20]. Wielowymiarowa analiza danych jest znakomitym narzędziem wykorzystywanym w interpretacji danych zawartych w bazach danych zawierających próbki opisane przez kilkadziesiąt zmiennych, do jakich przynależą widma past długopisowych uzyskane w zakresie promieniowania w podczerwieni [20].

W prezentowanej pracy w badaniu czarnych past długopisowych zastosowano metodę mikro-FTIR z przystawką do osłabionego całkowitego odbicia (ATR), która jest techniką nieniszczącą. Dodatkowo w celu zapewnienia obiektywizacji procesu identyfikacji i klasyfikacji próbek past długopisowych (dotyczącej marki lub modelu długopisu) zastosowano wielowymiarową analizę danych, a konkretnie analizę skupień (CA) oraz analizę głównych składowych w połączeniu z analizą dyskryminacyjną (DA).

## 2. Materiały i metody

### 2.1. Przygotowanie próbek

Przedmiotem badań było 155 czarnych past długopisowych nabytych w sklepie papierniczym w Maleszji. Długopisy różnych marek i modeli zakupiono w liczbie kilku sztuk, aby mieć pewność, że pochodzą z tej samej partii produkcji, jako że zmienność między partiami nie była przedmiotem badań. Wszystkim markom i modelom długopisów przyporządkowano odpowiedni numer, a szczegółowy opis zgromadzonych próbek zamiesz-

czono w tabeli I. Próbkę past długopisowych nanoszono na biały papier firmy Double A wyprodukowany w Tajlandii. Każdym z długopisów rysowano trzy niewielkie koła tak, aby pasta długopisowa całkowicie pokryła powierzchnię papieru. Próbkę past długopisowych analizowano nie później niż jeden dzień od naniesienia na papier, aby zminimalizować wpływ ewentualnych zmian powstałych na skutek ich starzenia się.

## 2.2. Analiza techniką mikro-ATR-FTIR

Wszystkie pomiary spektroskopowe wykonano przy użyciu spektrometru Thermo Scientific Nicolet iN10 MX FT-IR z detektorem rtęciowo-kadmowo-telurowym (MCT) wyposażonym również w przystawkę ATR z kryształem germanowym. Aby zmniejszyć wpływ aparatury na otrzymywane wyniki, widmo podłoża rejestrowano ponownie po każdym pomiarze próbki. Wszystkie rejestrowane widma stanowiły średnią z 16 pomiarów wykonanych przy rozdzielczości  $4\text{ cm}^{-1}$ . Zakres pomiarowy ustalono między  $4000\text{ a }675\text{ cm}^{-1}$ . Dla każdego ze 155 atramentów rejestrowano trzy widma. Widma podczerwieni zapisano w formie macierzy danych w arkuszu kalkulacyjnym programu Microsoft Excel.

## 2.3. Oprogramowanie

Uzyskane wyniki przetwarzano i analizowano, wykorzystując pakiet statystyczny SPSS (Statistical Package for the Social Sciences, Windows wersja 15.0, SPSS Inc., Chicago, Stany Zjednoczone).

# 3. Dyskusja wyników

## 3.1. Pobieranie próbek

W próbkach past długopisowych można wyróżnić trzy źródła zmienności. Należą do nich: a – zmienność międzyobiektoowa; b – zmienność wewnątrzobiektoowa (różnice między poszczególnymi długopisami należącymi do tej samej grupy) oraz c – zmienność pomiędzy powtórzeniami w obrębie wybranego długopisu (błąd związany z pojedynczym pomiarem). Wyżej wymienione źródła zmienności mogą być związane z czterema różnymi poziomami, wliczając w to: partię produkcji, markę, model czy rozmiar długopisu. Dwie porównywane grupy obiektów można traktować jako różne, jeżeli zmienność (a) jest znacząco większa niż zmienność (b) i (c). Większość badań prezentowanych w literaturze przedmiotu opiera się na pomiarze pojedynczej próbki, która reprezentuje daną grupę obiektów, a zmienność (b) i (c) jest rzadko brana pod uwagę. W prezentowanej pracy wszystkie trzy źródła zmienności zostały jednocześnie uwzględnione. Stosując takie podejście, autorzy byli w stanie w spo-

sób bardziej gruntowny ocenić użyteczność stosowanej metody analizy past długopisowych oraz wyraźniej wskazać różnice pomiędzy poszczególnymi grupami. Ponadto wnioski końcowe sformułowano w odniesieniu do populacji, a nie tylko subpopulacji. W prezentowanej pracy została ustalona zmienność pomiędzy różnymi markami długopisów (tj. zmienność wyników uzyskiwanych wewnątrz danej marki długopisów), jak również na poziomie modelu długopisu (zmienność uzyskiwanych wyników w obrębie danego modelu długopisu). W celu określenia zmienności (b) i (c) użyto więcej niż jednego długopisu każdej marki lub modelu, a każdy z długopisów był analizowany trzykrotnie.

## 3.2. Analiza widm w podczerwieni

Na podstawie analizy widm w podczerwieni ustalono, że najwięcej informacji widmowych dostarcza zakres  $2000\text{--}675\text{ cm}^{-1}$ . Pasma absorbcyjne pojawiające się w zakresie między  $4000\text{ a }2000\text{ cm}^{-1}$  pochodzą od pary wodnej i dwutlenku węgla pochłanianego z powietrza [13]. Reprezentatywne widma IR past długopisowych uzyskanych z czterdziestu różnych modeli długopisów zestawionych zgodnie z 9 różnymi markami przedstawiono na rycinie 1. Na wszystkich prezentowanych widmach widoczne są dwa znaczące pasma położone około  $1584\text{ i }1360\text{ cm}^{-1}$ , lecz różniące się względną wysokością oraz kształtem. Są to charakterystyczne pasma składnika atramentów długopisowych utożsamiane wcześniej z drganiami szkieletowymi barwnika trifenylometanowego [22]. Ponadto widma IR modelu M14-16 i M34-38 charakteryzuje niewielkie pasmo około  $1730\text{ cm}^{-1}$ , podczas gdy w widmach IR modeli M22-23 i M39-40 widoczne jest pasmo o niewielkiej intensywności położone około  $1250\text{ cm}^{-1}$ . Według Wang i innych [22], oba pasma świadczą o obecności poliestru i estru ftalowego, z których oba związki są dobrze znanymi składnikami żywności wchodzącej w skład past długopisowych [13]. Pasma obecne w zakresie pomiędzy  $1100\text{--}1000\text{ cm}^{-1}$  pochodzą głównie od składników papieru [7]. Przypisanie drgań innym pasmom zostało omówione gdzie indziej [22].

## 3.3. Analiza chemometryczna i statystyczna

Niejednorodność past długopisowych oraz fakt, że tylko niewielka ich ilość pozostaje w kontakcie z germanowym kryształem przystawki ATR, może wpływać na reprezentatywność próbki [5]. Stąd do oceny jednorodności materiału (powtarzalności wyników uzyskiwanych dla konkretnej próbki pasty długopisowej) wykorzystano wszystkie trzy widma dla niej zarejestrowane, wykonując po trzy pomiary dla past długopisowych pochodzących z każdego ze 155 długopisów. Pełen zestaw danych zawierał 464 widma (jedna kopia widma próbki nr 11 została utracona) składające się z wartości absor-

bancji w funkcji liczby falowej. Przed wykonaniem wielowymiarowej analizy danych wszystkie widma znormalizowano, aby zredukować zmienność występującą w danych, a wynikającą z różnic w grubości pasty długopisowej osadzającej się na powierzchni papieru. W tym celu dla każdego widma wartości absorbancji podzielono przez absorbancję całkowitą uzyskaną dla każdego widma w zakresie liczb falowych 2000–675  $\text{cm}^{-1}$ .

Jak już wcześniej wspomniano, celem badań była analiza widm IR past długopisowych w obrębie marki i modelu długopisu. Wszystkie 464 znormalizowane widma skatalogowano w kilku różnych plikach. Z jednej strony plik złożony ze wszystkich 464 widm wykorzystano do klasyfikacji atramentów według marek długopisów, z drugiej zaś strony widma pogrupowano, przyjmując jako kryterium podziału markę długopisu. Następnie badano różnice między widmami past długopisowych uzyskanymi dla różnych modeli długopisów w obrębie konkretnej marki (zmienność na poziomie modelu długopisu).

Dane analityczne utworzone według procedury opisanej powyżej poddano w obrębie grup analizie CA i PCA w celu uzyskania nowego zestawu zmiennych. Analizę CA wykonywano w pierwszej kolejności, wykorzystując kwadrat odległości euklidesowej jako miarę odległości oraz metodę Warda jako metodę tworzenia skupień. W ten sposób otrzymano osiem różnych skupień, z których każde zawierało obiekty mające podobne widma IR. Następnie na podstawie wyników uzyskanych dla próbek wchodzących w skład każdego z ośmiu skupień przeprowadzono analizę PCA. Za zastosowaniem takiego podejścia przemawiał fakt, iż analiza PCA przeprowadzana na mniejszej liczbie danych skutkuje otrzymaniem stabilniejszej macierzy, niż gdyby analizę prowadzono jednocześnie dla wszystkich 688 wyników. Decyzję o wyborze liczby istotnych głównych składowych podjęto w oparciu o kryterium Kaisera [12]. Nowy zestaw zmiennych zawierał wszystkie wybrane główne składowe, które odpowiednio oznaczono. Przykładowo składowa PCA 1\_2 odnosiła się do pierwszej składowej uzyskanej w wyniku analizy PCA przeprowadzonej na podstawie wyników uzyskanych dla próbek wchodzących w skład drugiego skupienia. Przekształcenie dużej liczby oryginalnych danych w niewielką liczbę nowych zmiennych (tj. głównych składowych) jest istotne w celu zapewnienia odpowiedniej dokładności i skuteczności metody DA [15]. Z drugiej strony udowodniono, że analiza DA przeprowadzona na podstawie wyników PCA jest skuteczniejsza niż dla danych nieprzetworzonych [19, 20]. Ponadto stosowanie uzyskanych głównych składowych jako nowych tzw. utajonych zmiennych eliminuje problem redundancji zmiennych (tj. problem występowania zmiennych zawierających zbędną informację), jako że główne składowe są ortogonalne i tym samym nieskorelowane. Dlatego też zastosowano analizę PCA

przed wykonaniem analizy dyskryminacyjnej. Na koniec zmienne poddano krokowej analizie DA, stosując odległości Mahalanobisa oraz wartość parametru  $F$  jako kryterium klasyfikacji. Założono również, że prawdopodobieństwa *a priori* dla wszystkich grup marek długopisów (analizowanych w badaniach) są równe, ponieważ reprezentatywność badanej populacji past długopisowych nie była znana [8].

#### 3.4. Klasyfikacja

Wszystkie 464 widma IR sklasyfikowano kolejno według marki oraz modelu długopisu. W pierwszej kolejności przy zastosowaniu analizy dyskryminacyjnej (DA) sklasyfikowano wszystkie 464 widma IR, biorąc pod uwagę jedynie markę długopisu. W dalszej kolejności w obrębie wybranej marki długopisów przeprowadzono inną analizę DA, grupując modele długopisów. Tabela II podsumowuje wyniki klasyfikacji 464 widm IR na poziomie marki i modelu.

W oparciu o wyniki liniowej funkcji dyskryminacyjnej LDF 1 oraz LDF 2 dziewięć marek długopisów podzielono na trzy oddzielne grupy. Pierwszą, drugą i trzecią grupę stanowiły marki oznaczone odpowiednio: B5 i B9; B3 i B8; B1, B2, B4, B6 oraz B7. Marki długopisów połączone w jedną grupę mogły wykazywać zbliżoną charakterystykę w zakresie średniej widma w podczerwieni. Przykładowo grupę drugą (G2) stanowiły marki B3 i B8, których widma atramentów wykazywały wyjątkowe wzajemne podobieństwo w postaci niewielkiego pasma położonego około 1730  $\text{cm}^{-1}$  widocznego w widmach obu past długopisowych B3 i B8 (rycina 1c dla próbki B3 oraz rycina 1f dla próbki B8). Ponadto pasmo to nie występowało w widmach IR żadnych innych marek. Natomiast widma IR past długopisowych obu marek: B5 i B9 tworzących grupę trzecią (G3) charakteryzowały się pasmem o niewielkiej intensywności położonym około 1245  $\text{cm}^{-1}$ .

Wizualna analiza wszystkich widm w podczerwieni (rycina 1) nie wykazała jednak żadnych oczywistych różnic między pastami długopisowymi pochodzącymi z długopisów różnych marek. Innymi słowy, drobne różnice między poszczególnymi atramentami, których nie dało się stwierdzić podczas wizualnego porównania tych widm, zostały pomyślnie oszacowane w wyniku przeprowadzonej wielowymiarowej analizy danych. Dla tych marek długopisów, których nie można było rozróżnić od siebie, należałoby zastosować inne techniki analizy instrumentalnej, wliczając w to spektroskopię Ramana czy spektrometrię mas, aby uzyskać więcej informacji dotyczących składu atramentu.

W dalszej kolejności każdy model długopisu, który uprzednio sklasyfikowano według marki, poddano dalszemu grupowaniu przy zastosowaniu analizy DA. Wszystkie czterdzieści modeli należących do poszcze-

gólnych marek długopisów można było podzielić przy najmniej na dwie grupy. Wyjątek stanowiły modele w obrębie marek oznaczonych jako B4, B5, B6 i B9. Przykładowo, stosując liniową funkcję dyskryminacyjną LDF 1 oraz LDF 2, wszystkie sześć modeli czarnych past długopisowych (M7-12) marki Foster (B2) zgrupowało się w trzy różne skupiska (rycina 2). Tym samym uzyskane wyniki wskazują, że istnieją różnice pozwalające na rozróżnianie analizowanych marek oraz modeli długopisów.

### 3.5. Identyfikacja past długopisowych na podstawie widm IR

Jak wspomniano w rozdziale 3.3, aby przeprowadzić analizę widm IR past długopisowych dotyczącą zarówno marki, jak i modelu długopisu, wszystkie 464 znormalizowane widma podzielono na pięć grup. Dla każdej grupy zaprojektowano schemat blokowy. Jego celem było zaproponowanie możliwie obiektywnej procedury identyfikacji marki lub modelu długopisu, z którego pochodziła analizowana pasta długopisowa. Identyfikacja ta oparta została na określeniu obecności lub braku statystycznie istotnych różnic wartości głównych składowych uzyskanych w trakcie analizy widma klasyfikowanej próbki pasty długopisowej i widma odniesienia uzyskanego dla wytypowanej wcześniej próbki odniesienia. Podczas badań wykorzystano jednoczynnikową analizę wariancji (ANOVA – założono poziom ufności 95%) w połączeniu z odpowiednim testem *post-hoc*. Podczas analizy porównywano każdą parę past długopisowych, wykorzystując za każdym razem nowy zestaw zmiennych. Przykładowy schemat blokowy dotyczący identyfikacji sześciu modeli (M7-12) czarnych past długopisowych marki Foster (B2) przedstawiono na rycinie 3. W wyniku wykonanej analizy danych zdecydowano na podstawie wyników testów *post-hoc*, że widmem wzorcowym dla tej grupy past długopisowych będzie widmo atramentu M12. Dlatego też analizę atramentu X rozpoczyna się od jego porównania z próbką wzorca (widmo IR atramentu pochodzącego z modelu długopisu oznaczonego jako 12) w PCA 3\_1. Jeżeli w wyniku analizy PCA 3\_1 jej wartości dla porównywanych próbek są znacząco różne w przedziale ufności 95%, wtedy próbka atramentu X może pochodzić z modelu długopisu oznaczonego jako 8, 9 czy 11. Z drugiej strony stwierdzenie nieistotnych różnic wskazuje na to, że próbka X prawdopodobnie może być utożsamiana z modelem długopisu oznaczonym jako 7 lub 10. Taki proces identyfikacji należy prowadzić dopóty, dopóki model długopisu, z którego pochodzi próbka X, nie zostanie w końcu zidentyfikowany.

Celem prezentowanych badań było ustalenie w pierwszej kolejności marki długopisu, z którego pochodzi kwestionowana pasta długopisowa, przy wykorzystaniu schematu blokowego (nieopisany w tym artykule) utwo-

żonego dla danej marki. Model długopisu można natomiast ustalić w oparciu o schemat blokowy opracowany dla każdej z przebadanych marek długopisów. Poprzez ustalenie modelu lub marki pasty długopisowej zakres poszukiwań zostaje zawężony, co w konsekwencji pozwala zwiększyć szanse identyfikacji pasty długopisowej. Jednakże na tym etapie zaproponowany schemat blokowy znajduje zastosowanie jedynie do analizy marek bądź modeli długopisów. Do dalszych zastosowań konieczne jest uzyskanie większej liczby próbek pochodzących od różnych producentów w celu utworzenia odpowiednio reprezentatywnej bazy danych. Równocześnie należałoby wykonać analizy z wykorzystaniem metod analizy instrumentalnej innych niż spektrometria w podczerwieni tak, aby można było zidentyfikować te marki (modele) długopisów, których nie można rozróżnić przy wykorzystaniu metody IR. Ponadto przed ewentualnym wykorzystaniem w praktyce zaproponowanych schematów blokowych należałoby ocenić ich poprawność, np. porównując z wynikami, jakie można by uzyskać, stosując podejście oparte na obliczaniu ilorazu wiarygodności [1].

## 4. Wnioski

Podsumowując, widma IR past długopisowych są lepiej rozróżnialne dla marki niż modelu długopisu. Choć pełne rozróżnienie atramentów na podstawie marki i modelu długopisu nie zostało osiągnięte, uzyskane rezultaty pokazują, że metoda mikro-ATR-FTIR dostarcza danych, które przy zastosowaniu wielowymiarowej analizy danych mogą z powodzeniem być wykorzystywane i stanowią szybkie oraz sprawne narzędzie w różnicowaniu past długopisowych według wybranej marki lub modelu.

### Podziękowania

Autorzy dziękują Panu Mohd Fizol oraz Pani Rusikah bt. Minhad z Jabatan Kimia Malaysia za pomoc w uzyskaniu wyników metodą spektrometrii w podczerwieni, a Prof. dr Aziz Jermain za dyskusję na temat wielowymiarowej analizy danych. Autorzy pragną również wyrazić swoją wdzięczność wszystkim pracownikom biorącym udział w programie nauk sądowych realizowanym na UKM (Universiti Kebangsaan Malaysia), jak również Pani Hartinah Annuar. Podziękowania także dla UKM za grant UKM-DLP-2012-024.