



THE POTENTIAL USE OF PRINCIPAL COMPONENT ANALYSIS AND MAHALANOBIS DISTANCE IN HANDWRITING COMPARISON

Loren WILLIAMS¹, Morgan MILLS², Carol RITTER¹, Lawrence QUARINO¹

¹ Forensic Science Program, Cedar Crest College, Allentown, PA, USA

² Bode Technology, Lorton, VA, USA

Abstract

The potential use of multivariate statistical analysis in handwriting comparison to provide objective criteria to an inherently subjective discipline was explored in this study. Twenty sets of exemplar writing in the form of checks from fifty participants were used as a primary dataset. Using 3-dimensional principal component analysis (PCA) plots of the first three principal components, writing from thirty-eight of the fifty participants could be differentiated. Applying both PCA and Mahalanobis distance to unknowns created from a closed set of exemplar writing from ten participants in the dataset resulted in a true positive rate of 70%. Conversely, the use of PCA and Mahalanobis distance in the comparison of seven true negatives to the entire database resulted in a false positive rate of only 1.1%. The method was unsuccessful in helping to determine the possible origin of both traced and simulated writing, although forged writing could be differentiated from exemplar writing. The method was also able to correctly determine that the writing on two checks from a group of six checks had a common origin even when a professional forensic handwriting examiner erroneously concluded that no common origin existed. The paper does not advocate for a replacement of traditional handwriting examination with objective criteria such as the methodology described in the study. However, the results of the study support the incorporation of mathematical models into standard practice of handwriting comparison as a means of verification and as a way to ensure consistency of conclusions.

Keywords

Handwriting comparison; Principal Component Analysis; Mahalanobis distance.

Received 10 March 2021; accepted 2 July 2021

1. Introduction

In the post-Daubert era, many court challenges in the United States have been raised as to the reliability of handwriting comparisons. Although most decisions support acceptance and reliability, examples do exist where reservations about the reliability of testimony were expressed by a court (*United States v. Prime*, 220 F. Supp. 2d 1203, 2002) and testimony regarding handwriting comparison was completely excluded in a few cases (*United States v. Fujii*, 152 F. Supp. 2d

939, 2000; *United States v. Saelee*, 162 F. Supp. 2d 1097, 2001). The individuality of handwriting has been questioned as some believe handwriting comparisons are overly susceptible to contextual bias due to the subjective nature of traditional handwriting comparative analysis (Sulner, 2018). Furthermore, researchers have found that handwriting comparisons are subject to high error rates (Kam, Fielding, Conn, 1997; Sita, Found, Rogers, 2002). Although the 2009 National Academy of the Sciences Report on strengthening forensic science in the United States stated that there is

value in handwriting comparisons and concluded that a scientific basis for handwriting comparisons likely exists, research in proving the scientific basis is limited (National Research Council, 2009).

Over the last few decades, models that can determine a degree of confidence or probability have been sought to provide a more objective basis for handwriting comparisons. Interestingly, however, the idea is not new. Albert Osborn, a pioneer in the area of forensic handwriting examination, suggested application of the Newcomb rule of probability to provide a statistical framework for handwriting samples concluded to be from the same source (Osborn, 1929).

In an attempt to improve the reliability of handwriting comparisons, several machine learning and computer methods have been developed since the 1990s and have been shown to have discriminating potential between sources of writing. These include FISH (Hecker, 1993), Wanda (Franke et al., 2003), CEDAR-FOX (Srihari, Srinivasan, Desai, 2007), and Fourier analysis (Marquis, Schmittbuhl, Mazzella, Taroni, 2005). In addition, Bayesian models and calculations of likelihood ratios have demonstrated not only the ability to differentiate handwriting for comparative examinations (Marquis, Bozza, Schmittbuhl, Taroni, 2011) but for investigative purposes (such as inference of gender and handedness) as well (Taroni et al., 2012, 2014).

A possible simpler approach that has been used in the examination of other pattern-based evidence is combining principal component analysis (PCA) and Mahalanobis distance. PCA is an unsupervised, data reduction technique often used to better understand large datasets that are uncorrelated. It is often used prior to applying more discriminating non-linear classification algorithms or independent component analysis but can be used to separate different data sets (Caprihan, Pearlson, Calhoun, 2008). Mahalanobis distance is a multivariate distance measurement between a data point and the mean (centroid) of multivariate data generated from a reference distribution. The technique compares a single data point to a known mean and produces a distance describing the relationship between samples. While some promising work has been reported using PCA (Al Haddad, White, Cowell, 2011) and Mahalanobis distance (Marquis et al., 2005) separately in handwriting comparisons, other pattern disciplines have utilized one or both with successful results. For instance, accidental characteristics on footwear were analyzed using PCA and demonstrated clustering of datapoints for individual samples (Petraço, Gambino, Kubic, Olivio, Petraço, 2010). In addition, PCA has been shown to produce clearer images of fingerprints

by separating substrate datapoints (Sonnex, Almond, Bond, 2016). Furthermore, PCA has been applied to digital scans of faces with resulting Mahalanobis distances used to compare and identify an unknown facial scan to a known facial scan (Aeria, Claes, Vandermeulen, Clement, 2010).

The purpose of the current study is to determine if PCA and Mahalanobis distance can be successfully applied to forensic handwriting examination in hopes of adding a simple, yet objective step into traditional handwriting comparative practice. The study will incorporate quantitative-based features, such as height and width measurements of letters, into a multivariate statistical model.

2. Methods

The study consisted of fifty participants who were 18–60 years of age, including males and females irrespective of left or right-handed dominancy, and had no known illnesses impacting handwriting. Institutional Review Board approval was obtained for the procedures used to collect writing samples. All participants were informed of the collection process and were notified that they could remove themselves from the study at any given time. The participants were provided twenty mock bank checks downloaded from an open access internet website (creativetemplate.net). Using names from a disputed nineteenth century United States presidential election, cursive and print handwriting was used to make entries onto each check (not historically accurate) over a period of several weeks (Fig. 1). To ensure a natural flow and less rigidity, signature information was practiced by participants prior to making entries onto the checks. Common ballpoint pens were used since the ink dries quickly to minimize the chances of smudging. Alias information was placed on the back of each check to protect the identity of participants and to assist with comparisons.

After checks were collected from participants, the paper copies were digitized using a scanner (Ricoh MP4054, Tokyo, Japan) and uploaded to Adobe® Photoshop® (Adobe® Photoshop® 2015, San Jose, USA) for measurements. Adobe measurement tools were used to measure the characteristics outlined in Table 1. Before measurements were completed on samples, the measurement process was verified using an extra check collected from a participant to ensure the digital measurements were to scale and to ensure that measurements were consistent from check to check. The extra check was hand measured and then digitally measured and compared. Data from measurements

were used to perform principal component analysis (BioVinci 2.0.0, San Diego, CA, USA) and calculate Mahalanobis distance (RStudio 1.1.38, Boston, MA, USA) where appropriate.

The study was completed in five parts. Part 1 examined the variability in all fifty participants' writing in 3-dimensional principal component analysis plots. In addition, a Mahalanobis distance was generated for each participant. Mahalanobis distance was

calculated by randomly removing a single check from each participant's set of twenty checks with the use of a random number generator and then comparing the Euclidean distance from the single check to the centroid of the remaining nineteen checks in the set. Part 2 was designed to show if the methodology used in the study was effective in finding true positives. In Part 2, PCA and Mahalanobis distance were used to compare one additional check from ten of the study's

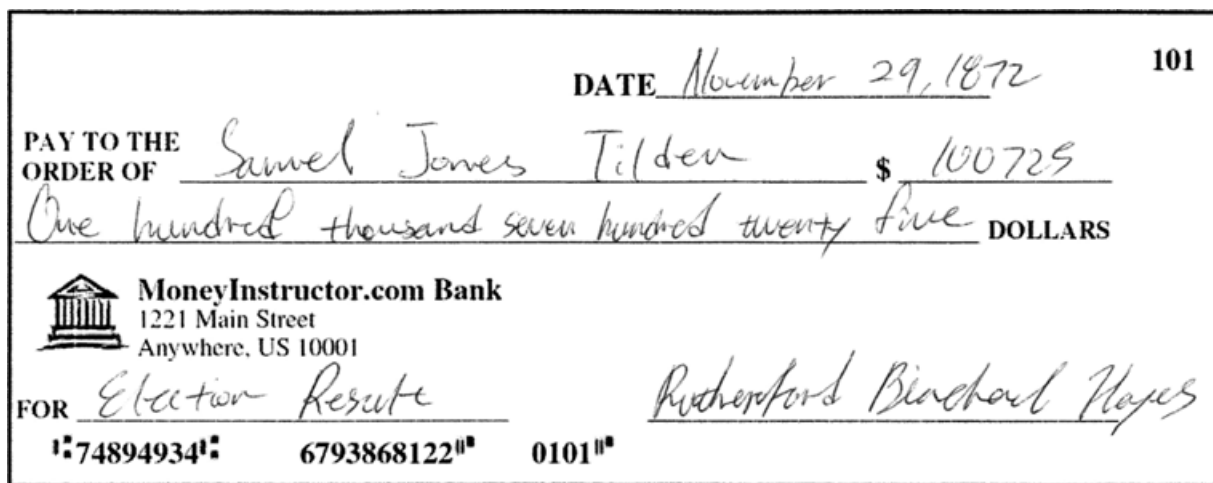


Fig. 1. A completed mock bank check with alias information in script and print handwriting.

Table 1
52 writing characteristics measured for each check collected

Line 1 – “DATE”	Line 2 – “PAY TO THE ORDER OF”	Line 3 – Dollar Amount in Cursive	Line 4 – “FOR” and Signature
Height of “N” Height of “o” Spacing in front of comma Spacing behind comma	Height of “s” Height of “a” Distance between “I” and “j” Height of “o” Distance between “s” and “T” Spacing of “I” dot Spacing in front of comma Spacing behind comma	Height of “o” Height of first “n” Height of second “n” Height of first “d” Height of first “e” Height of second “d” Length of first “t” cross Height of third “n” Height of third “d” Height of second “e” Height of fourth “d” Length of second “t” cross Length of third “t” cross Spacing of “I” dot	Height of first “e” Height of first “I” Height of second “e” Length of “t” cross Spacing of “I” dot Height of first “r” Height of “u” Height of second “r” Height of “f” Height of second “r” Height of first “d” Distance between “d” and “B” Height of “b” Height of second “I” Height of “c” Height of first “h” Height of third “r” Height of second “d” Distance between “d” and “H” Height of second “h” Height of “a” Height of “y”

participants to the data collected from those same ten participants in Part 1. The ten collected checks were treated as unknowns and the researcher performing the measurements was blind to their origin. Part 3 demonstrated if the methodology used in the study was susceptible to false positives. Handwriting samples were generated from seven individuals not in the original dataset. Using PCA and Mahalanobis distance, one additional check from each of these seven individuals was compared to the dataset collected in Part 1. Part 4 attempted to determine the effect of disguised writing on the methodology proposed in this study by comparing traced and simulated handwriting from four of the fifty participants in the dataset created in Part 1. Using transmitted light, traced samples were produced by placing a blank check on top of a check from the dataset. Four traced sets (10 checks in each set) were collected in total, two from females and two from males. The simulated writing samples were created by a participant trying to imitate the handwriting of another individual. Four simulated sets (10 checks in each set) were again collected, two from females and two from males. For both studies, PCA plots were produced containing data from the simulated and traced samples along with the datasets of relevant exemplars from Part 1. The average Mahalanobis distance from exemplar sets was determined for simulated and traced checks from each participant. The separation of male and female participants in this study is due to past research showing that gender may be inferred from handwriting using Bayesian statistics for investigative purposes (Taroni et al., 2012, 2014). The authors of this study make no claim as to differences in handwriting based on gender. Part 5 compared Mahalanobis distance data to the results from a forensic handwriting examiner with over 30 years of experience working in a public sector laboratory. The practitioner was asked to provide conclusions to two scenarios. First, the practitioner was given six checks used in the dataset collected in Part 1 and asked if any of them had a common origin. Second, the practitioner was asked to compare an unknown check to four checks taken from different participants in the study.

3. Results

PCA was performed on each set of 20 checks from all 50 participants in Part 1 of the study. Using PCA, the writing of 38 of the 50 participants could be differentiated visually. Fig. 2 displays 3-dimensional PCA plots for the first three principal components shown from two different views for participants 31–40 (for clarity

purposes, data points from a representative number of participants are presented). Although overlap between participants does exist in some cases, clusters of data points from each participant are observed. The difference in intra-writer variability in handwriting for each participant is observed by the calculated Mahalanobis distance for each set (Table 2). The Mahalanobis distance ranged from the lowest variability of 0.03833 (set 4) to the highest variability of 19.95 (set 8) with a mean of 2.893.

Table 2
Mahalanobis distance data generated in Part 1 for each set of samples (four significant figures)

Set #	Mahalanobis distance	Set #	Mahalanobis distance
1	1.361	26	8.064
2	0.2431	27	1.996
3	0.09052	28	1.783
4	0.03833	29	3.605
5	0.1702	30	0.8821
6	9.403	31	1.372
7	2.228	32	13.45
8	19.95	33	1.223
9	1.769	34	5.084
10	0.3802	35	0.3886
11	1.959	36	1.676
12	11.32	37	0.4276
13	0.2329	38	0.09054
14	9.873	39	0.3356
15	0.5850	40	3.157
16	0.3638	41	0.5529
17	0.8028	42	2.171
18	1.752	43	1.207
19	0.2817	44	9.117
20	5.871	45	0.3562
21	0.5244	46	2.028
22	2.920	47	3.428
23	0.9579	48	2.486
24	4.016	49	0.1903
25	1.662	50	0.4819

In Part 2, a single check was collected from ten participants (participants 1, 4, 6, 7, 9, 10, 12, 14, 19, and 22) who had participated in Part 1. The 10 collected checks were treated as unknowns and compared to the set of 20 checks collected in Part 1 from the same participants. Using PCA, nine out of the ten unknown

checks were visually clustered with the correct participant set (Fig. 3); for clarity purposes, PCA plots were divided into two with each showing the results from five unknowns. However, calculation of Mahalanobis distance only correctly classified seven of the ten unknowns (true positive rate of 70%). Unknowns 1, 6, 7, 10, 12, 14, and 19 provided the lowest Mahalanobis distance with its correct known set. Of those

classified incorrectly, two produced the second lowest Mahalanobis distance (unknowns 9 and 22) with its correct set and one (unknown 4) produced the third lowest Mahalanobis distance when compared with its correct set (Table 3).

When examining the false positive rate (Part 3), Mahalanobis distances were calculated for each of the seven checks from individuals not in the original

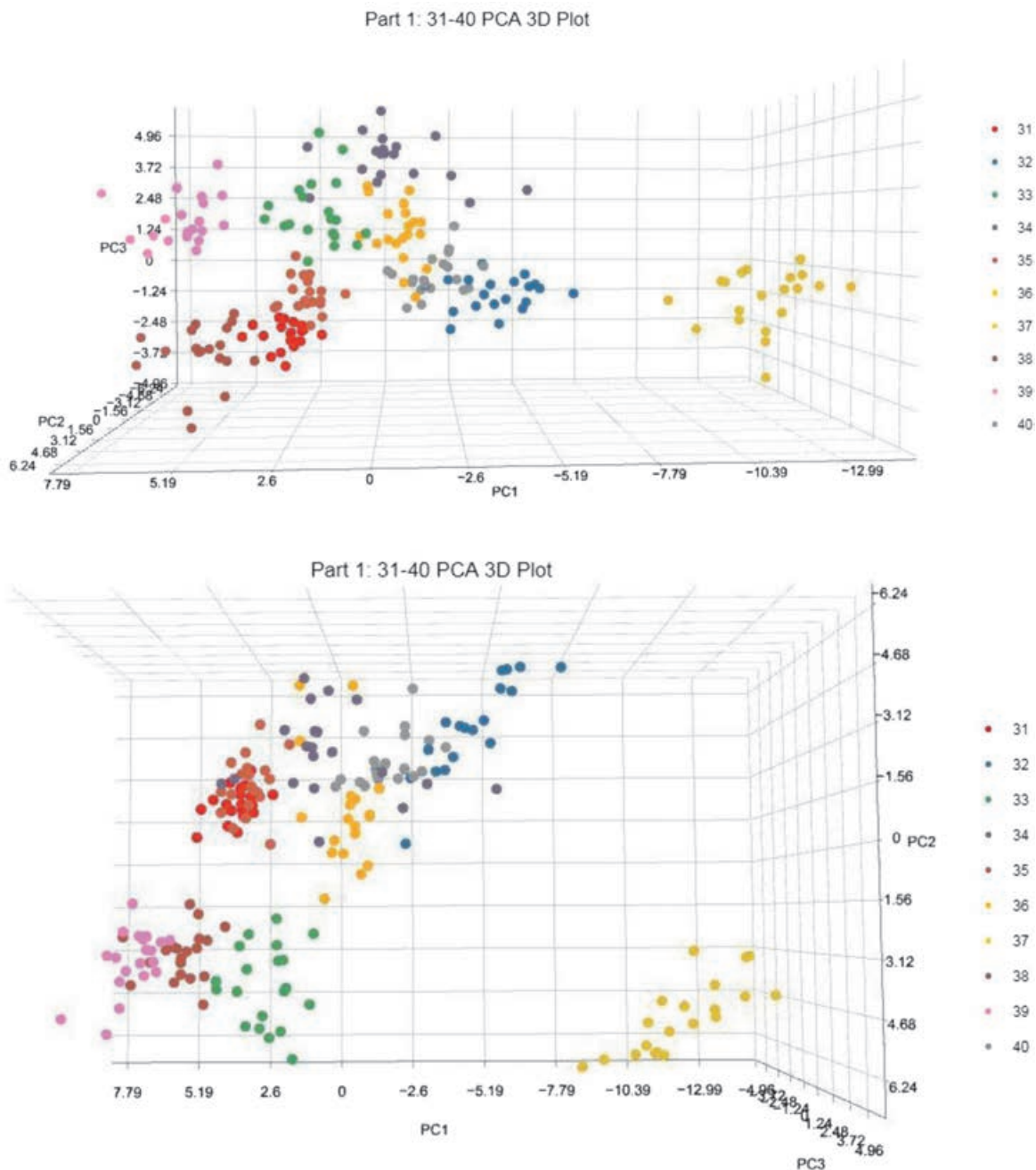


Fig. 2. The side view and top view of a generated 3-dimensional PCA plot representing samples 31–40 (three principal components account for 83.64% of the variance). The mustard yellow cluster (samples 37) is visually separated from other clusters.

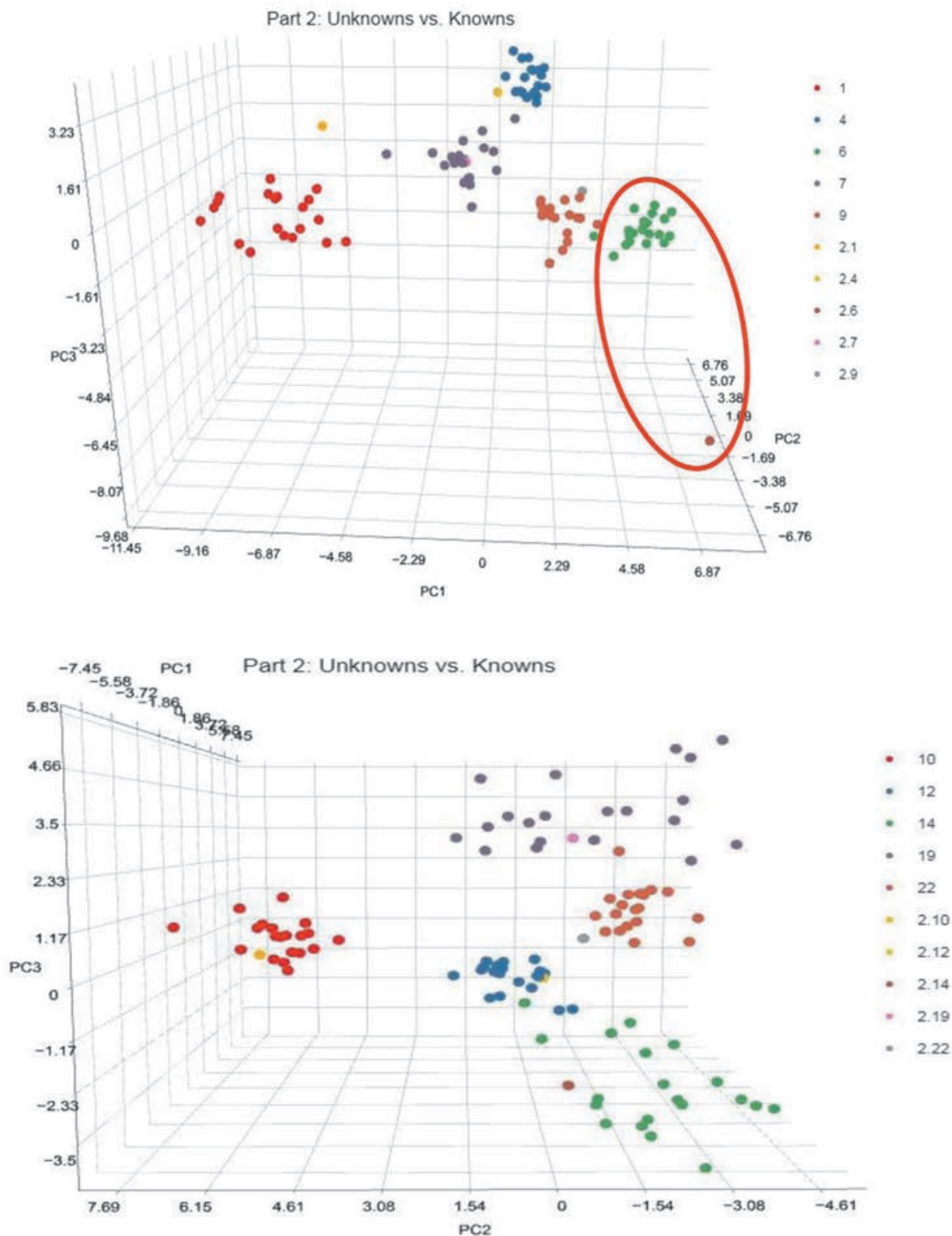


Fig. 3. Two generated 3-dimensional PCA plots comparing the first five samples and second five samples used as unknowns to ten knowns from each participant in Part 1 (three principal components accounted for 74.11% of the variance). Nine of the ten unknowns clustered with the correct set. However, participant six’s handwriting samples from Part 1 (the green cluster) and their “unknown” from Part 2 (single brown dot) indicate that the “unknown” is dissimilar to the original handwriting samples and, therefore, did not classify with the correct group (known set and unknown are circled, top plot). Unknowns are classified with the numeral “2” followed by a decimal and the correct classification set number; each participant set is shown by the set number.

Table 3

Mahalanobis distance data (four significant figures) generated in Part 2 that compared “unknown” checks to sets of checks from ten participants in Part 1. The yellow highlighted numbers indicate the unknown and the set which originated from the same person. The red highlighted numbers indicate unknowns that were more similar to handwriting sets they did not belong to

Unknown #	Set 1	Set 4	Set 6	Set 7	Set 9	Set 10	Set 12	Set 14	Set 19	Set 22
1	1.533	305.6	295.2	127.5	388.5	344.2	277.4	42.94	25.08	165.1
4	25.45	17.29	58.25	35.46	122.4	73.91	43.15	25.13	1.628	4.369
6	331.5	1081	260.6	422.9	370.6	322.2	475.8	341.7	121.9	1026.1
7	46.22	173.9	140.2	0.3318	50.92	43.65	219.3	19.81	9.385	69.17
9	89.02	178.8	42.31	46.54	7.335	2.027	118.6	46.15	13.52	153.8
10	80.84	152.2	34.02	43.06	10.72	0.3464	99.91	38.81	10.43	130.2
12	54.04	28.85	43.69	96.30	270.8	74.66	2.663	106.3	13.05	15.68
14	23.02	93.97	105.7	11.59	77.08	69.99	135.5	0.6051	1.827	27.23
19	28.35	54.49	67.08	15.98	64.31	50.01	84.53	5.722	0.1715	19.05
22	28.52	10.13	51.33	41.58	135.5	71.92	30.83	33.92	2.459	2.837

dataset with all fifty sets from Part 1. Of 350 comparisons, 327 (93.43%) produced a Mahalanobis distance above the mean value of 2.893 generated from the data in Table 2. Based on Mahalanobis distances alone (Table 4), six of the checks could be falsely assigned to a range of two to five test sets (totaling 23) from Part 1. PCA of each unknown with their included sets, however, showed greater differentiation. Of the six unknowns that could be falsely assigned, three (unknowns 3, 4, and 6) were shown not to be grouped with any set. For the other three, two showed overlap with one group (unknown 1 with set 25; unknown 7

with set 17) and one showed overlap with two groups (unknown 5 with sets 25 and 19). Representative PCA plots showing both types of examples (unknowns 4 and 7) are shown in Fig. 4.

Principal component analysis and calculation of Mahalanobis distance proved to be of limited utility when examining traced and simulated writing. PCA indicated that all four traced sets (two male and two female) did not group with either the writer’s exemplar set or the set that the trace samples were trying to duplicate (Fig. 5a). However, in all four traced sets, the Mahalanobis distance provided a lower value with the set being traced than the writer’s exemplar set (Table 5) although the values were well above the mean Mahalanobis value (2.893) determined from the original set in Part 1. Similar Mahalanobis values were seen with the simulated writing samples with the exception of one female whose writing produced a lower value with the set being imitated. PCA plots did show clustering between the simulated writing of the two males and the set attempting to be imitated but not with their own exemplar writing (Fig. 5b). The two female simulated writing samples showed no clustering with either the set attempting to be duplicated or their own set of exemplars (Fig. 5b).

Finally, results from a forensic handwriting examiner in Part 5 were in part divergent from the results of the method described in this study. In the first examination, the practitioner erroneously concluded that none of the six checks examined came from a common source. Two of the checks were in fact from a common source which was supported by Mahalanobis distance

Table 4

Mahalanobis distance data (four significant figures) for all test sets yielding a value below 2.893 when compared to true negatives

Unknown 1	Set 3	Set 14	Set 25	Set 34	Set 36
	1.868	0.5077	1.893	1.127	0.05921
Unknown 3	Set 29	Set 32			
	0.1798	0.4001			
Unknown 4	Set 3	Set 20	Set 31	Set 43	
	1.083	2.602	0.4509	0.5458	
Unknown 5	Set 3	Set 14	Set 19	Set 20	Set 25
	1.803	2.433	1.593	0.8646	0.007101
Unknown 6	Set 5	Set 19	Set 44		
	0.7091	2.034	0.3203		
Unknown 7	Set 5	Set 17	Set 37	Set 45	
	2.826	0.3811	1.574	1.266	

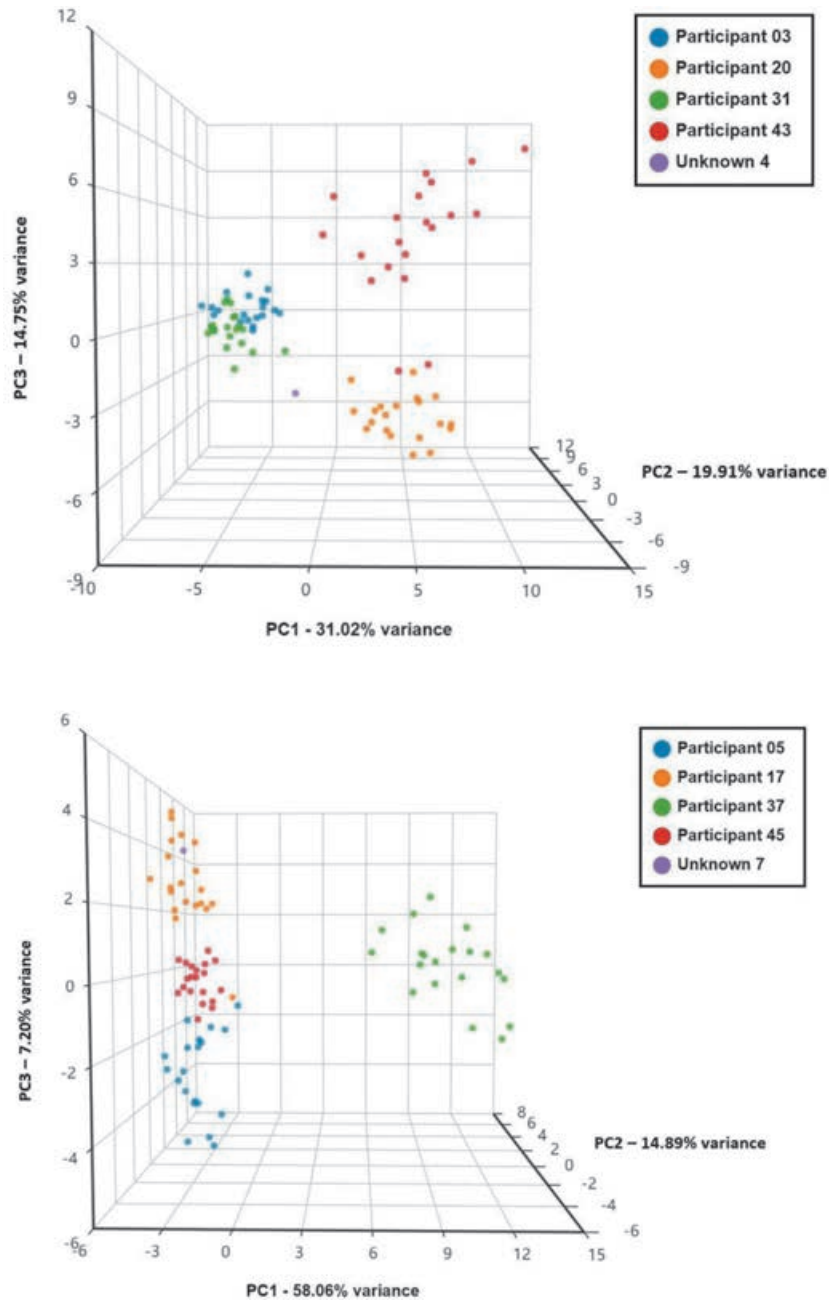


Fig. 4. PCA plots showing true negative unknowns with participant sample sets yielding low Mahalanobis distance values. Top: true negative showing no overlap with participant data (three principal components accounted for 65.68% of variance); Below: true negative showing overlap with one set of participant data (unknown 7 overlaps with participant 17; three principal components accounted for 80.15% of variance).

data (Table 6). In the second examination, the practitioner correctly identified the set from which the supplied unknown originated. This conclusion was also supported by the Mahalanobis distance data (Table 7).

4. Discussion

The results of this study show the strengths and limitations of using a multivariate statistical model to assess authorship of writing in a questioned document. Given the natural variation in the handwriting of any person, handwriting comparison is different than other pattern disciplines that rely on comparing fixed points, such as in a fingerprint or footwear impressions. It is therefore important to determine the level of variation

expected in the writing of an individual. Given that handwriting is dependent on neuromuscular processes that may change over time due to factors such as health and age, this is a difficult challenge, particularly when exemplar and evidentiary writing are not contemporaneous (Hilton, 1982). Although the basis of individuality in handwriting is well established (Osborn, 1929), it is difficult to imagine, given the variation in individual handwriting, that determination of authorship can be made to near 100 percent confidence

Table 5

Mean Mahalanobis distance data (four significant figures) of traced and simulated handwriting samples. Set 20 was the original handwriting sample being traced/simulated for females and set 17 was the original handwriting sample being traced/simulated for males. Participant numbers represent participants from the original dataset that performed the traced and simulated writing

Traced					
Female			Male		
	Set 20	Participant exemplar set		Set 17	Participant exemplar set
Participant 4	2.943	216.2	Participant 22	11.62	54.97
Participant 30	6.122	77.42	Participant 39	9.605	71.89
Simulated					
Female			Male		
	Set 20	Participant exemplar set		Set 17	Participant exemplar set
Participant 4	10.45	78.38	Participant 22	43.65	11.07
Participant 30	37.17	19.51	Participant 39	32.20	8.239

Table 6

Mahalanobis distance data (four significant figures) indicating two handwriting samples* came from a common source. The bold numbers indicate samples that are most similar to that particular set. Therefore, samples C and D, which are from a common source, best classify with participant eight's set of handwriting samples

	Set 4	Set 5	Set 8	Set 9	Set 10
A	0.09718	37.08	115.1	269.9	102.1
B	17.10	0.2467	35.81	153.9	41.45
C	89.22	28.72	0.1424*	61.63	2.928
D	79.79	24.01	0.05431*	67.52	3.952
E	345.6	79.69	81.02	0.03631	89.77
F	133.1	48.29	6.811	41.18	0.2873

Table 7

The unknown has the smallest Mahalanobis distance to exemplar 3, indicating that it is the most similar to exemplar 3 (true source)

	Exemplar 1	Exemplar 2	Exemplar 3	Exemplar 4
Unknown	80.79	146.6	0.4021*	79.17

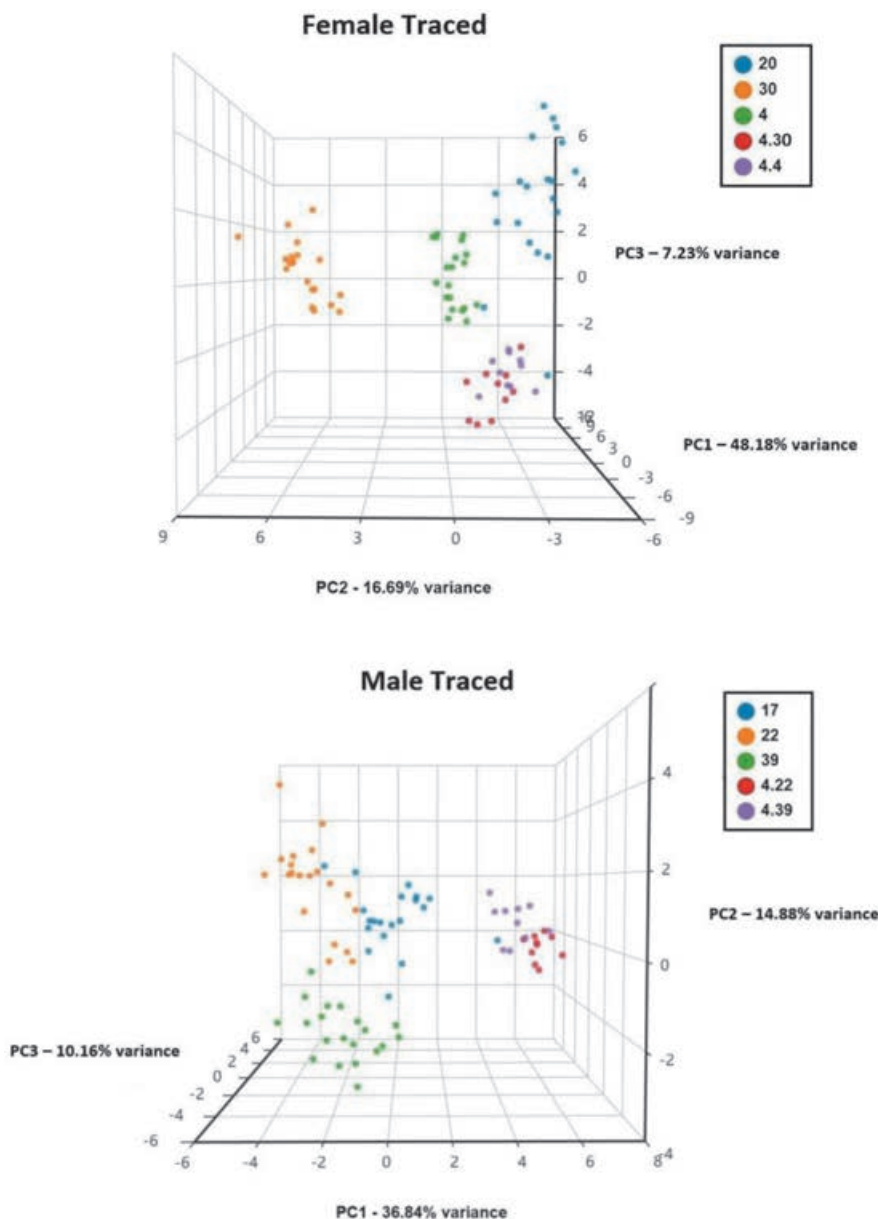


Fig. 5a. PCA plots of traced writing. Top: Female, exemplars = 4, 20, 30; 4.4 = participant 4 tracing participant 20; 4.30 = participant 30 tracing participant 20 (three principal components accounted for 72.10% of variance). Bottom: Male, exemplar = 17, 22, 39; 4.22 = participant 22 tracing participant 17, 4.39 = participant 39 tracing participant 17 (three principal components accounted for 61.88% of variance).

as previous work has claimed (Srihari, 2002). It was never the intention of this work to develop a method that was dispositive, but rather to provide an objective examination that can be used in conjunction with traditional methodology. Perhaps increased discrimination could have been achieved by using a maximal number of principal components in linear discriminant analysis but given the potential for unaccountable parameters, any increased discrimination may prove to be simply anecdotal. It was the intention of the study

to maximize variance in sets of handwriting in the first three principal components and to display them graphically. In addition, to ensure that overfitting of data did not occur, better and more commonly used discrimination methods such as linear discriminant analysis was not used after application of PCA. Furthermore, PCA has been shown to be useful in detecting outlier data in several different realms (Brown, Goddard, Brown, 2009; Saha, Roy, Mukherjee, Sarkar, 2016).

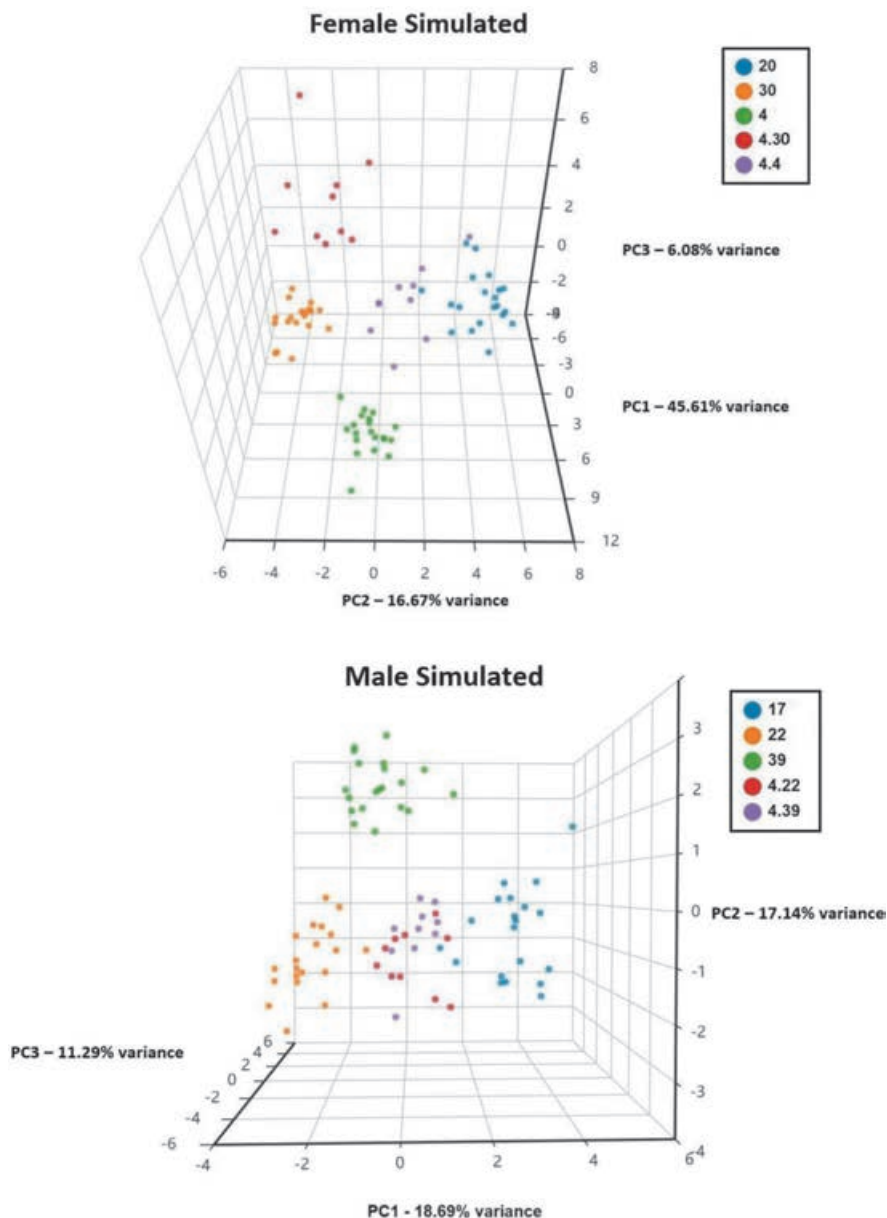


Fig. 5b. PCA plots of simulated writing. Top: Female, exemplars = 4, 20, 30; 4.4 = participant 4 simulating participant 20; 4.30 = participant 30 simulating participant 20 (three principal components accounted for 68.36% of variance). Bottom: Male, exemplars = 17, 22, 39; 4.22 = participant 22 simulating participant 17, 4.39 = participant 39 simulating participant 17 (three principal components accounted for 47.12% of variance).

In Part 1 of this study, clustering of data points in 3-dimensional PCA plots showed that the writing of 38 of the 50 participants could be differentiated. The percent variation of the first three principal components was 83.69% which is significantly higher than similar work previously reported (Al Haddad et al., 2002). Given that the content of the writing was exactly the same from sample to sample, the variables used in this study may be more correlated than in the previous study. The wide range of Mahalanobis distance

values as reported in Table 2 between sets of participant handwriting shows the difference in the variation of writing between participants. The fact that writing was obtained in several sittings over a span of several weeks likely accounted for at least some of the variation. Some participants were simply better at reproducing their handwriting than others. An average Mahalanobis distance value (2.893) for all fifty sets was calculated and served as a reference for other parts of the study.

Part 2 of the study examined how well the study could identify true positives. A single check serving as an unknown was collected from ten participants that had participated in Part 1 and compared with the dataset from the ten participants. PCA showed that each unknown except one clustered within their expected group (Fig. 3). However, Mahalanobis distance data only correctly classified 7 of the 10 unknowns with their correct group. The seven include the sample that did not cluster with their correct group (unknown 6) on the PCA plot (which explains the very high Mahalanobis distance value of 260.6, which would likely lead to the conclusion that the unknown does not belong to any of the ten groups if applied in an actual case situation). Of the three that did not classify with their correct group, only one yielded a Mahalanobis distance value below the mean value of the sets in Part 1 (unknown 22, Mahalanobis distance 2.837). Given the combined examination of both the PCA and Mahalanobis distance results, 7 of the 10 unknowns (not including unknown 6 but including unknown 22) provided results consistent with their correct group. Although a sample size of ten unknowns is too small to firmly establish a true positive rate of 70%, it does show that true positives may not always be identified using just this method. Furthermore, classifications using Mahalanobis distance data will ultimately depend on the variation of the data in the sets used in the model. Mahalanobis distance data calculated with sets of data with similar values is likely to lower the true positive rate.

The chance for a false positive using the methodology in the study is very low. In Part 3, Mahalanobis distances were calculated to compare the fifty datasets from Part 1 to each of the seven checks from individuals not in the original dataset. Of these 350 comparisons, only 23 could be wrongly grouped into one of the fifty sets since they yielded a Mahalanobis distance below the average value calculated in Part 1 (Table 4). Examining these 23 comparisons with PCA, only 4 were found to cluster with one of the original fifty sets of data (unknown 1 with set 25, unknown 7 with set 17, and unknown 5 with sets 19 and 25). Within the context of this study, a false positive rate of 1.1% (4/350) was achieved when combining the visual analysis of PCA with Mahalanobis distance.

In Part 4, correlations between disguised writing (both traced and simulated) and exemplar writing (either that of the forger or the participant whose writing was attempted to be imitated) were not observed although forged and non-forged writing could be distinguished. This observation is consistent with earlier published work which used both PCA and descriptive

statistics to determine forged writing (Al Haddad et al., 2002).

Part 5 perhaps shows the possible benefit of incorporating the model developed in this study into common forensic handwriting examination practice. Although it was only one example, the disparate conclusions reached between using Mahalanobis distance and the subjective erroneous determination of a forensic handwriting examiner in examining checks for common origin (Table 6), shows that an additional objective component in these types of comparisons is likely warranted.

5. Conclusions

The need for objective measures in forensic science has been strongly advocated in recent years (National Research Council, 2009; Kafadar, 2019). Objective criteria can help increase the consistency of conclusions across experts, reduce the potential for contextual and confirmational bias, and provide guideposts for proper method validation. This study examined the potential use of two common multivariate statistical techniques, PCA and Mahalanobis distance, in providing objective measures for the determination of authorship of handwriting. PCA and Mahalanobis distance are well established methods and may be more user friendly than machine learning techniques previously described (Hecker, 1993; Franke et al., 2003; Srihari et al., 2007; Marquis et al., 2005). The study shows that application of mathematical approaches could help to deliver new variables (samples descriptions) which could be used for solving comparison/classification problem(s) by application of suitable techniques such as linear discriminant analysis or SIMCA[®]. However, since the success of any statistical model for handwriting may vary depending on such factors as character selection, any mathematical approach is not immune to errors. The true positive rate of only 70% determined in Part 2 of the study shows this. Only some quantitative features were used in this study, creating a major limitation as forensic handwriting examiners will use quantitative and qualitative features (e.g. line properties or character structure) to reach conclusions. Visual examination of participant six's handwriting samples indicated that the knowns and "unknown" have common features that would have allowed an examiner to make a common source conclusion between the two even when they were clearly separated on the PCA plot (Fig. 3). This proves that subjective analysis in handwriting comparison based on standard accepted practices is still very much needed. However,

errors with the use of standard practices, such as the example shown in Part 5 of this study, can happen and have happened (Kam et al., 1997; Sita et al., 2002). The incorporation of mathematical models into standard practice can serve as a means of verification and may be the best approach to the problem.

References

1. Aeria, G., Claes, P., Vandermeulen, D., Clement, J. G. (2010). Targeting specific facial variation for different identification tasks. *Forensic Science International*, 201(1–3), 118–124.
2. Al Haddad, A., White, P. C., Cowell, A. M. (2011). The use of principal component analysis to provide objective methods for the examination of Arabic signatures. *Journal of the American Society of Questioned Document Examiners*, 14(1), 3–18.
3. Brown, R. J. C., Goddard, S. L., Brown, A. S. (2010). Using principal component analysis to detect outliers in ambient air monitoring studies. *International Journal of Environmental Analytical Chemistry*, 90(10), 761–772.
4. Caprihan, A., Pearlson, G. D., Calhoun, V. D. (2008). Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements. *Neuroimage*, 42(2), 675–682.
5. Franke, K., Schomaker, L., Veenhuis, C., Taubenheim, C., Guyon, I., Vuurpijl, L., van Erp, M., Zwarts, G. (2003). Wanda: A generic framework applied in forensic handwriting analysis and writer identification. (In) A. Abraham, M. Koppen, K. Franke (Eds.), *Design and application of hybrid intelligent systems. Proceedings 3rd International Conference on Hybrid Intelligent Systems (HIS03)* (pp. 927–938). Amsterdam: IOS Press.
6. Hecker, M. (1993). *Forensic information system for handwriting (fish)*. Technical Document from the Kriminaltechnisches Institut, Bundeskriminalamt.
7. Hilton, O. (1982). *Scientific examination of questioned documents*. New York: Elsevier.
8. Kafadar, K. (2019). The need for objective measures in forensic evidence. *Significance*, 16(2), 16–20.
9. Kam, M., Fielding, G., Conn, R. (1997). Writer identification by professional document examiners. *Journal of Forensic Sciences*, 42(5), 778–786.
10. Marquis, R., Bozza, S., Schmittbuhl, M., Taroni, F. (2011). Handwriting evidence evaluation based on the shape of characters: application of multivariate likelihood ratios. *Journal of Forensic Sciences*, 56(s1), S238–S242.
11. Marquis, R., Schmittbuhl, M., Mazzella, W. D., Taroni, F. (2005). Quantification of the shape of handwritten characters: a step to objective discrimination between writers based on the study of the capital character O. *Forensic Science International*, 150(1), 23–32.
12. National Research Council. (2009). *Strengthening forensic science in the United States: a path forward*. Washington DC: The National Academies Press.
13. Osborn, A. S. (1929). *Questioned documents*, second edition. Albany NY: Boyd Printing.
14. Petraco, N. D. K., Gambino, C., Kubic, T. A., Olivio, D., Petraco, N. (2010). Statistical discrimination of footwear: a method for the comparison of accidentals on shoe outsoles inspired by facial recognition techniques. *Journal of Forensic Sciences*, 55(1), 34–41.
15. Saha, P., Roy, N., Mukherjee, D., Sarkar, A. K. (2016). Application of principal component analysis for outlier detection in heterogenous traffic data. *Procedia Computer Science*, 83, 107–114.
16. Sita, J., Found, B., Rogers, D. (2002). Forensic handwriting examiners' expertise for signature comparison. *Journal of Forensic Sciences*, 47(5), 1117–24.
17. Sonnex, E., Almond, M. J., Bond, J. W. (2016). Enhancement of latent fingerprints on fabric using the cyanoacrylate fuming method followed by infrared spectral mapping. *Journal of Forensic Sciences*, 61(4), 1101–1106.
18. Srihari, S. N., Cha, S.-H., Arora, H., Lee, S. (2002). Individuality of handwriting. *Journal of Forensic Sciences*, 47(4), 856–872.
19. Srihari, S., Srinivasan, H., Desai, K. (2007). Questioned document examination using CEDAR-FOX. *Journal of Forensic Document Examination*, 14, 1–20.
20. Sulner, A. (2018) Critical issues affecting the reliability and admissibility of handwriting identification opinion evidence – how they have been addressed (or not) since the 2009 NAS report, and how they should be addressed going forward: a forensic document examiner tells all. *Seton Hall Law Review*, 48, 631–717.
21. Taroni, F., Marquis, R., Schmittbuhl, M., Biedermann, A., Thiery, A., Bozza, S. (2012). The use of likelihood ratio for evaluation and investigative purposes in comparative forensic handwriting examination. *Forensic Science International*, 214(1–3), 189–94.
22. Taroni, F., Marquis, R., Schmittbuhl, M., Biedermann, A., Thiery, A., Bozza, S. (2014). Bayes factor for investigative assessment of selected handwriting features. *Forensic Science International*, 242, 266–273.

Corresponding author

Prof. Lawrence Quarino
Forensic Science Program,
Cedar Crest College, Allentown, PA, USA
e-mail: laquarin@cedarcrest.edu

POTENCJALNE ZASTOSOWANIE ANALIZY GŁÓWNYCH SKŁADOWYCH I ODLEGŁOŚCI MAHALANOBISA W PORÓWNYWANIU PISMA RĘCZNEGO

1. Wstęp

W epoce postdaubertowskiej odnotowano wiele przypadków zakwestionowania przez sądy USA wiarygodności analizy porównawczej pisma. Mimo iż większość orzeczeń uznawała je za wiarygodne, istnieją przypadki zgłaszania przez sądy wątpliwości odnośnie do wiarygodności dowodów (*USA vs. Prime*, 220 F. Supp. 2d 1203, 2002) i w kilku przypadkach dowody w postaci analizy porównawczej pisma zostały zupełnie odrzucone (*USA vs. Fujii*, 152 F. Supp. 2d 939, 2000; *USA vs. Saelee*, 162 F. Supp. 2d 1097, 2001). Indywidualność pisma bywa kwestionowana, gdyż istnieje pogląd, iż porównywanie pisma jest zbyt narażone na błędy kontekstowe z uwagi na subiektywny charakter tradycyjnej analizy porównawczej pisma (Sulner, 2018). Co więcej, naukowcy ustalili, że porównywanie pisma ręcznego obarczone jest wysokim wskaźnikiem błędu (Kam, Fielding, Conn, 1997; Sita, Found, Rogers, 2002). Mimo iż Raport National Academy of the Sciences z 2009 roku, poświęcony doskonaleniu technik kryminalistycznych w Stanach Zjednoczonych, uznał wartość, jaką niesie porównywanie pisma, stwierdzając prawdopodobieństwo istnienia naukowej podstawy do porównywania pisma ręcznego, niewiele jest badań potwierdzających jej istnienie (National Research Council, 2009).

Ostatnie kilka dziesięcioleci przyniosło poszukiwania modeli mających określić stopień pewności lub prawdopodobieństwa w celu zapewnienia bardziej obiektywnej podstawy porównania pisma, choć pomysł ten nie jest nowy. Albert Osborn, pionier w dziedzinie kryminalistycznej ekspertyzy pisma ręcznego, zaproponował zastosowanie zasady prawdopodobieństwa Newcomba dla stworzenia statystycznych ram dla prób pisma, co do których stwierdzono, że pochodzą z tego samego źródła (Osborn, 1929).

Poszukując sposobów na zwiększenie wiarygodności porównywania pisma, począwszy od lat 90. wypracowano kilka metod uczenia maszynowego i metod informatycznych, które umożliwiają rozróżnianie źródeł pisma. Wśród nich można wymienić FISH (Hecker, 1993), Wanda (Franke, 2003), CEDAR-FOX (Srihari, Srinivasan, Desai, 2007) oraz analizę Fouriera (Marquis, Schmittbuhl, Mazzella, Taroni, 2005). Ponadto modele Bayesa i obliczenia współczynników prawdopodobieństwa pozwoliły wykazać nie tylko zdolność do rozróżniania pisma ręcznego w badaniach porównawczych (Marquis, Bozza, Schmittbuhl, Taroni, 2011), lecz także

w celach dochodzeniowych (takich jak wnioskowanie o płci i ręczności; Taroni i in., 2012, 2014).

Prostszym podejściem, które wykorzystano w badaniu innych dowodów opartych na wzorcach, jest połączenie analizy głównych składowych (PCA) i odległości Mahalanobisa. PCA jest techniką redukcji danych niepodlegającą kontroli, często stosowaną w celu lepszego zrozumienia dużych nieskorelowanych. Jest ona często stosowana przed wykorzystaniem bardziej różnicujących algorytmów klasyfikacji nieliniowej lub analizy składowych niezależnych, ale może również służyć do oddzielenia różnych zbiorów danych (Caprihan, Pearlson, Calhoun, 2008). Odległość Mahalanobisa jest wielowymiarową miarą odległości pomiędzy punktem danych a średnią (centroidem) wielowymiarowych danych wygenerowanych z rozkładu referencyjnego. Technika ta porównuje pojedynczy punkt danych ze znaną średnią, w rezultacie jej stosowania otrzymujemy odległość opisującą relację pomiędzy próbami. W czasie gdy pojawiły się obiecujące prace dotyczące zastosowania PCA (Al Haddad, White, Cowell, 2011) i odległości Mahalanobisa (Marquis i in., 2005) oddzielnie w porównywaniu pisma ręcznego, inne dyscypliny oparte na wzorcach wykorzystywały jedną lub obydwie metody, uzyskując zadowalające wyniki. Na przykład przeprowadzono analizę przypadkowych parametrów obuwia przy użyciu metody PCA, która wykazała grupowanie punktów danych dla poszczególnych prób (Petraço, Gambino, Kubic, Olivio, Petraço, 2010). Ta sama metoda pozwoliła uzyskać bardziej wyraźne obrazy odcisków palca przez odseparowanie punktów danych dotyczących podłoża (Sonnex, Almond, Bond, 2016). Otrzymując dzięki PCA odległości Mahalanobisa, zastosowano ją także do porównania nieznanego skanu twarzy ze znanym skanem twarzy oraz identyfikacji (Aeria, Claes, Vandermeulen, Clement, 2010).

Celem niniejszego opracowania jest stwierdzenie, czy PCA i odległość Mahalanobisa można z powodzeniem zastosować w kryminalistycznej ekspertyzie pisma ręcznego, przyjmując jako cel wzbogacenie tradycyjnej metody porównawczej pisma o prosty, ale obiektywny element. Cechy ilościowe, takie jak wysokość i szerokość liter, zostaną uwzględnione w badaniu i wprowadzone do wielowymiarowego modelu analityczno-statystycznego.

2. Metody

Badaniem objęto pięćdziesięciu uczestników w wieku 18–60 lat, zarówno kobiety, jak i mężczyźni, którzy zostali wybrani niezależnie od dominacji prawej lub lewej ręki. Osoby te nie uskarżały się na żadne choroby, które mogłyby mieć wpływ na ich pismo. Uzyskano zgodę Institutional Review Board na przeprowadzenie procedur niezbędnych do uzyskania od nich prób pisma. Poinformowano wszystkich uczestników badania o szczegółach procedury zbierania prób oraz o tym, że mogą się wycofać z udziału w badaniu na każdym jego etapie. Otrzymali oni dwadzieścia przykładowych prób kontrolnych pochodzących z banku na ogólnodostępnej stronie internetowej (creativetemplate.net). Używając nazwisk z zakwestionowanych wyborów prezydenckich w Stanach Zjednoczonych w XIX wieku, przez kilka tygodni na każdej próbie dokonywano wpisów (niezgodnych z prawdą historyczną), stosując pismo pochyłe (Ryc. 1). Aby zapewnić naturalny przebieg i zmniejszyć sztywność, cechy podpisu zostały przećwiczone przez uczestników przed dokonaniem wpisów na próbach kontrolnych. Aby zminimalizować ryzyko rozmazania, używano zwykłych długopisów, w których tusz szybko schnie. Informacja o pseudonimie została umieszczona na odwrocie każdej próby, aby chronić tożsamość uczestników i pomóc w porównaniach.

Po zebraniu prób kontrolnych od uczestników papierowe wersje zostały zdigitalizowane przy użyciu skanera (Ricoh MP4054, Tokyo, Japan) i wprowadzone do programu Adobe® Photoshop® (Adobe® Photoshop® 2015, San Jose, USA) w celu dokonania pomiarów. Narzędzia pomiarowe Adobe zastosowano, by zmierzyć cechy charakterystyczne wskazane w tabeli 1. Przed zakończeniem pomiarów na próbach proces pomiarowy został zweryfikowany przy użyciu dodatkowej próby kontrolnej pobranej od uczestnika, aby upewnić się, że pomiary cyfrowe były dokładne i spójne w przypadku każdej próby. Dodatkowa próba została zmierzona ręcznie, następnie cyfrowo, oraz porównana. Dane z pomiarów zostały wykorzystane do przeprowadzenia analizy głównych składowych (BioVinci 2.0.0, San Diego, CA, USA) i obliczenia odległości Mahalanobisa (RStudio 1.1.38, Boston, MA, USA) w stosownych przypadkach.

Badanie podzielono na pięć części. Część pierwsza polegała na analizie różnorodności pisma wszystkich pięćdziesięciu uczestników na trójwymiarowych wykresach analizy głównych składowych. Dodatkowo dla próby otrzymanej od każdego uczestnika wygenerowano odległość Mahalanobisa. Odległość Mahalanobisa obliczono, usuwając losowo jedną kontrolną z zestawu dwudziestu prób każdego uczestnika za pomocą generatora liczb losowych, a następnie porównując odległość euklidesową z pojedynczej próby z centroidem pozostałych dziewiętnastu prób w zestawie. Część 2

miała w założeniu wykazać, czy metodologia zastosowana w badaniu sprawdziła się w poszukiwaniu wyników prawdziwie dodatnich. W części 2 PCA i odległość Mahalanobisa zastosowano do porównania jednej dodatkowej próby z dziesięciu pochodzących od uczestników badania z danymi zebranymi od tych samych dziesięciu uczestników w części 1. Dziesięć zebranych prób kontrolnych potraktowano jako niewiadome, a wykonujący pomiary był całkiem nieświadomy ich pochodzenia. Część 3 wykazała, czy metodologia zastosowana w badaniu była podatna na fałszywe wyniki dodatnie. Próby pisma ręcznego uzyskano od siedmiu osób poza oryginalnym zbiorem danych. Wykorzystując PCA i odległość Mahalanobisa, jedną dodatkową próbę kontrolną uzyskaną od każdej z tych siedmiu osób porównano ze zbiorem danych zebranych w części 1. Część 4 miała pomóc w ustaleniu wpływu pisma maskowanego na metodologię zaproponowaną w tym badaniu poprzez porównanie kopiowanego i naśladowanego pisma, które otrzymano od czterech z pięćdziesięciu uczestników w zbiorze danych utworzonym w części 1. Wykorzystując światło przechodzące, wyprodukowano próby kopiowane, które powstały poprzez nałożenie pustej próby kontrolnej na próbę z zestawu danych. Zebrano cztery zestawy kopiowane (10 prób kontrolnych w każdym zestawie), dwa od kobiet i dwa od mężczyzn. Próby pisma naśladowanego uzyskano, gdy uczestnik próbował naśladować pismo ręczne innej osoby. Uzyskano cztery zestawy naśladowane (10 prób w każdym zestawie), dwa od kobiet i dwa od mężczyzn. Na potrzebę obydwu badań stworzono wykresy PCA zawierające dane pochodzące z prób naśladowanych i prób kopiowanych wraz z zestawami danych odpowiednich prób wzorcowych z części 1. Średnią odległość Mahalanobisa w zestawach wzorcowych określono dla prób kontrolnych naśladowanych i kopiowanych uzyskanych od każdego uczestnika. Oddzielenie uczestników od uczestniczek w tym badaniu było wynikiem wcześniejszych prób, które wykazały, że na podstawie pisma ręcznego w oparciu o statystykę bayesowską dla celów dochodzeniowych można wnioskować o płci (Taroni i in., 2012, 2014). Autorzy niniejszego opracowania nie formułują twierdzenia dotyczącego różnic w piśmie ręcznym wynikających z płci. Część 5 to porównanie danych dotyczących odległości Mahalanobisa z wynikami uzyskanymi przez pracującego w laboratorium w sektorze publicznym grafologa z ponad trzydziestoletnim doświadczeniem. Eksperta poproszono o sformułowanie wniosków dotyczących dwóch scenariuszy. Najpierw otrzymał on sześć prób kontrolnych wykorzystanych w zbiorze danych uzyskanych w części 1; miał stwierdzić, czy którakolwiek z nich miała wspólne źródło. Kolejnym zadaniem grafologa było porównanie próby niewiadomej z czterema próbami uzyskanymi od różnych uczestników tego badania.

3. Wyniki

PCA przeprowadzono w odniesieniu do każdego zestawu z 20 prób kontrolnych pochodzących od wszystkich 50 uczestników w części 1 badania. Wykorzystując tę metodę, zdołano rozróżnić wizualnie pismo 38 z 50 uczestników. Ryc. 2 pokazuje trójwymiarowe wykresy PCA dla pierwszych trzech składowych głównych, ukazane w dwóch różnych aspektach dla uczestników 31–40 (dla przejrzystości zaprezentowano punkty danych pochodzące od reprezentatywnej liczby uczestników). Choć występują przypadki pokrywania się punktów danych pochodzących od różnych uczestników, obserwuje się klastry takich punktów pochodzących od każdego z nich. Różnicę dotyczącą zmienności w obrębie pisma jednego autora dla każdego uczestnika obserwowano na podstawie odległości Mahalanobisa obliczonej dla każdego zestawu (Tabela 2). Odległość Mahalanobisa wahała się od najniższej wartości zmienności wynoszącej 0,0383 (zestaw 4) do najwyższej wartości zmienności 19,95 (zestaw 8) przy średniej wartości 2,893.

W części 2 badania uzyskano pojedynczą próbę kontrolną od dziesięciu uczestników (uczestnicy 1, 4, 6, 7, 9, 10, 12, 14, 19, i 22), którzy brali udział w części 1. Te 10 uzyskanych prób kontrolnych potraktowano jak niewiadome i porównano z zestawem 20 prób kontrolnych zebranych w części 1 od tych samych uczestników. Wykorzystując PCA, dziewięć z dziesięciu prób niewiadomych wizualnie pogrupowano do zestawów właściwych uczestników (Ryc. 3). Dla zwiększenia przejrzystości wykresy PCA podzielono na dwa, z których każdy demonstrował wyniki uzyskane z pięciu niewiadomych. Obliczenie odległości Mahalanobisa dało jednak wynik w postaci siedmiu poprawnie sklasyfikowanych na dziesięć niewiadomych (odsetek wyników prawdziwie dodatnich wyniósł 70%). Niewiadome 1, 6, 7, 10, 12, 14 oraz 19 wykazały najmniejszą odległość Mahalanobisa w kontekście właściwie rozpoznanego zestawu. Spośród tych sklasyfikowanych niewłaściwie dwa uzyskały drugą najmniejszą odległość Mahalanobisa (niewiadome 9 i 22) na tle właściwego zestawu, a jedna (niewiadoma 4) wykazała trzecią najmniejszą wartość odległości Mahalanobis w porównaniu z właściwym zestawem (Tabela 3).

Analizując odsetek wyników fałszywie dodatnich (część 3), odległości Mahalanobisa obliczano dla każdej z siedmiu prób kontrolnych pochodzących od osób nie w ramach oryginalnego zestawu wszystkich pięćdziesięciu zestawów z części 1. Z 350 porównań 327 (93,43%) uzyskało odległość Mahalanobisa powyżej średniej wartości 2,893 wygenerowanej na podstawie danych zawartych w tabeli 2. W oparciu jedynie o odległości Mahalanobisa (Tabela 4) sześć spośród prób kontrolnych mogło być fałszywie przypisane do zakresu od dwóch do pięciu zestawów testowych (łącznie 23) z części 1. Jednakże metoda PCA poszczególnych niewiadomych

w kontekście załączonych zestawów wykazała większe zróżnicowanie. Spośród sześciu niewiadomych, które mogły być fałszywie przypisane, trzy (niewiadome 3, 4 i 6) nie zostały przypisane do żadnego zestawu. Dla pozostałych trzech dwie próby wykazały nakładanie się z jedną grupą (niewiadoma 1 z zestawem 25; niewiadoma 7 z zestawem 17), a jedna wykazała nakładanie się z dwiema grupami (niewiadoma 5 z zestawami 25 i 19). Reprezentatywne wykresy PCA demonstrujące obydwa typy przykładów (niewiadoma 4 i 7) pokazano na rycinie 4.

Okazało się, że analiza składowych głównych i obliczanie odległości Mahalanobisa miały ograniczoną przydatność podczas badania pisma kopiowanego i naśladowanego. PCA wykazała, że wszystkie cztery zestawy kopiowane (dwa źródła męskie i dwa żeńskie) nie zostały zgrupowane ani do wzorcowego zestawu autora, ani do zestawu, który próby kopiowane miały powielić (Ryc. 5a), we wszystkich czterech znaczonych zestawach odległość Mahalanobisa pozwoliła jednak uzyskać niższą wartość, ale we wszystkich czterech kopiowanych zestawach odległość Mahalanobisa dawała niższą wartość dla zestawu kopiowanego niż dla zestawu wzorcowego autora pisma (Tabela 5), choć wartości te były znacznie powyżej średniej wartości Mahalanobisa (2,893) wyznaczonej dla zestawu oryginalnego w części 1. Podobne wartości Mahalanobisa zaobserwowano w przypadku prób pisma naśladowanego; wyjątkiem była próba uzyskana od jednej kobiety, której pismo uzyskało niższą wartość w stosunku do zestawu naśladowanego. Na wykresach PCA zaobserwowano grupowanie naśladowanego pisma dwóch mężczyzn z zestawem naśladowanym, ale nie z ich własnym pismem wzorcowym (Ryc. 5b). Dwie próby naśladowanego pisma kobiecego nie wykazały grupowania ani z zestawem, który miał być zduplikowany, ani z własnym zestawem wzorców (Ryc. 5b).

Ostatecznie wyniki sformułowane przez grafologa w części 5 były częściowo rozbieżne w stosunku do wyników uzyskanych przy zastosowaniu metody opisanej w niniejszym opracowaniu. W pierwszym badaniu stwierdził on błędnie, że żadna z sześciu badanych prób kontrolnych nie pochodziła ze wspólnego źródła. W rzeczywistości dwie z prób kontrolnych pochodziły ze wspólnego źródła, co potwierdziły dane uzyskane na podstawie odległości Mahalanobisa (Tabela 6). W ramach drugiego badania ekspert poprawnie zidentyfikował zestaw, z którego pochodziły dostarczone niewiadome. Również dane dotyczące odległości Mahalanobisa potwierdziły powyższe wyniki (Tabela 7).

4. Dyskusja

Wyniki niniejszego badania pokazują zarówno mocne strony, jak i ograniczenia wynikające z zastosowania wielowymiarowego modelu statystycznego dla określenia autorstwa pisma w zakwestionowanym dokumencie. Gdy się weźmie pod uwagę naturalne zróżnicowanie w obrębie pisma ręcznego każdej osoby, można stwierdzić, że porównanie pisma różni się od innych dyscyplin opartych na wzorcach, które opierają się na stałych niezmiennych punktach, takich jak odcisk palca czy obuwia. Niezwykle zatem istotne jest określenie poziomu zróżnicowania oczekiwanego w piśmie poszczególnych osób. Mając na względzie to, że pismo uzależnione jest od procesów nerwowo-mięśniowych, które z upływem czasu podlegają zmianom z powodu czynników takich jak stan zdrowia czy wiek, należy uznać, że jest to szczególnie trudne wyzwanie, zwłaszcza gdy materiał wzorcowy i dowodowy nie są jednoczesne (Hilton, 1982). Pomimo iż podstawa indywidualnej natury pisma ręcznego jest już dość ugruntowana (Osborn, 1929), trudno sobie wyobrazić, biorąc pod uwagę zmienność indywidualnego pisma ręcznego, że ustalenia autorstwa można dokonać z niemal stuprocentową pewnością, jak postulowano we wcześniejszych pracach (Srihari, 2002). Celem niniejszego opracowania nigdy nie było opracowanie metody o charakterze dyspozytywnym, a raczej dostarczenie obiektywnej metody badawczej, którą można stosować w połączeniu z tradycyjną metodologią. Być może zwiększoną skuteczność rozróżniania można byłoby osiągnąć poprzez zastosowanie maksymalnej liczby głównych składowych w liniowej analizie dyskryminacyjnej, jednakże gdy się weźmie pod uwagę istnienie niemożliwych do skorygowania parametrów, to ta zwiększona skuteczność dyskryminacji może nie mieć poparcia w faktach. Założeniem badania było maksymalne zwiększenie wariantowości w zestawach pisma ręcznego w trzech pierwszych składowych głównych i przedstawienie ich w formie graficznej. Dodatkowo – aby nie nastąpiło przepasowanie danych, po zastosowaniu PCA nie posłużono się lepszymi i częściej używanymi metodami dyskryminacji, takimi jak liniowa analiza dyskryminacyjna. Ponadto wykazano, że metoda PCA jest pomocna w wykrywaniu danych odstających w kilku różnych dziedzinach (Brown, Goddard, Brown, 2009; Saha, Roy, Mukherjee, Sarkar, 2016).

W części 1 niniejszego badania grupowanie punktów danych w trójwymiarowych wykresach PCA wykazało, że pismo 38 spośród 50 uczestników można różnicować. Procentowa zmienność pierwszych trzech składowych głównych wyniosła 83,69%, co stanowi wartość znacznie wyższą niż w podobnych wcześniej opisywanych pracach (Al Haddad i in., 2002). Jako że treść pisma była dokładnie taka sama we wszystkich próbach, zmienne zastosowane w tym badaniu mogą być w większym

stopniu skorelowane niż w poprzednim. Szeroki zakres wartości odległości Mahalanobisa między zestawami pisma uczestników wskazany w tabeli 2 świadczy o różnicy w stopniu zróżnicowania pisma pomiędzy poszczególnymi uczestnikami. Fakt, że pismo uzyskiwano podczas kilku sesji na przestrzeni kilku tygodni, z dużym prawdopodobieństwem przyczynił się do przynajmniej częściowej wariancji. Niektórzy uczestnicy byli zwyczajnie lepsi w odtwarzaniu pisma ręcznego niż inni. Obliczono średnią wartość odległości Mahalanobisa (2,893) dla wszystkich pięćdziesięciu zestawów, posłużyła ona jako punkt odniesienia w pozostałych częściach badania.

Część 2 badania polegała na przetestowaniu jego skuteczności w rozpoznawaniu prawdziwie dodatnich wyników. Pojedynczą próbę kontrolną, służącą jako niewiadomą, pozyskano od dziesięciu uczestników, którzy brali udział w części 1, i porównano z zestawem danych tych dziesięciu uczestników. Analiza PCA wykazała, że, oprócz jednej, każda niewiadoma grupowała się w ramach oczekiwanej grupy (Ryc. 3), jednak dane dotyczące odległości Mahalanobisa w sposób poprawny sklasyfikowały jedynie 7 z 10 niewiadomych z ich właściwą grupą. W tych siedmiu zawiera się próba, która nie została sklasyfikowana z właściwą grupą (niewiadoma 6) na wykresie PCA (co tłumaczy niezwykle wysoką wartość odległości Mahalanobisa wynoszącą 260,6, a co doprowadziłyby najpewniej do wniosku, że niewiadoma nie należy do żadnej z dziesięciu grup, gdyby zastosowano ją w rzeczywistym przypadku). Spośród trzech, które nie zostały sklasyfikowane z ich właściwą grupą, tylko jedna uzyskała wartość odległości Mahalanobisa poniżej średniej wartości zestawów w części 1 (niewiadoma 22, odległość Mahalanobisa 2,837). Łączne rozpatrywanie wyników PCA i odległości Mahalanobisa pozwoliło na stwierdzenie, że 7 z 10 niewiadomych (nie licząc niewiadomej 6, ale wliczając niewiadomą 22) uzyskało wyniki zgodne z ich właściwą grupą. Chociaż próba dziesięciu niewiadomych jest zbyt mała, aby zdecydowanie ustalić wskaźnik prawdziwych wyników dodatnich na poziomie 70%, pokazuje ona, że prawdziwe wyniki dodatnie mogą nie zawsze być identyfikowane wyłącznie za pomocą tej metody. Co więcej, klasyfikacje wykorzystujące dane oparte na odległości Mahalanobisa będą ostatecznie zależały od zmienności danych w zbiorach użytych w modelu. Dane dotyczące odległości Mahalanobisa obliczone dla zestawów danych o podobnych wartościach prawdopodobnie obniżają współczynnik prawdziwych wyników dodatnich.

Szansa na uzyskanie wyniku fałszywie dodatniego przy zastosowaniu tej metodologii jest niewielka. W części 3 odległości Mahalanobisa obliczono, by porównać pięćdziesiąt zestawów danych z części 1 z każdą z siedmiu prób kontrolnych od osób spoza oryginalnego zestawu danych. W ramach tych 350 porównań tylko 23 mogły być błędnie zaklasyfikowane do jednego

z pięćdziesięciu zestawów, jako że uzyskały wartość odległości Mahalanobisa poniżej wartości średniej obliczonej w części 1 (Tabela 4). Badając te 23 porównania za pomocą analizy PCA, stwierdzono, że jedynie 4 z nich grupują się z jednym z oryginalnych pięćdziesięciu zestawów danych (niewiadoma 1 z zestawem 25, niewiadoma 7 z zestawem 17 i niewiadoma 5 z zestawami 19 i 25). W kontekście tego badania, łącząc analizę wizualną PCA z odległością Mahalanobisa, uzyskano odsetek wyników fałszywie dodatnich wynoszący 1,1% (4/350).

W części 4 nie zaobserwowano korelacji pomiędzy pismem maskowanym (zarówno kopiowanym, jak i naśladowanym) a pismem wzorcowym (czy to pochodzącym od osoby podrabiającej, czy uczestnika, którego pismo próbowano naśladować), chociaż można było odróżnić pismo podrabiane od tego, które takie nie było. Obserwacja ta jest spójna z pracą wcześniej publikowaną, w której zastosowano zarówno PCA, jak i statystykę opisową do rozpoznania pisma podrabianego (Al Haddad i in., 2002).

Część 5 być może ukazuje możliwe korzyści płynące z włączenia modelu opracowanego w tym badaniu do powszechnej praktyki kryminalistycznej ekspertyzy grafologicznej. Mimo iż był to tylko jeden przykład, odmienne wnioski uzyskane w wyniku zastosowania odległości Mahalanobisa i subiektywnej błędnej ekspertyzy grafologa badającego próby kontrolne pod kątem wspólnego pochodzenia (Tabela 6) pokazują, że dodatkowy obiektywny element w tego typu porównaniach jest prawdopodobnie uzasadniony.

5. Wnioski

W ostatnich latach silnie postuluje się potrzebę stosowania obiektywnych miar w kryminalistyce (National Research Council, 2009; Kafadar, 2019). Obiektywne kryteria mogą pomóc zwiększyć spójność wniosków wysuwanych przez ekspertów, zmniejszyć potencjalną liczbę błędów kontekstowych i konfirmacyjnych oraz zapewnić wytyczne dla właściwej weryfikacji metod. W niniejszej pracy zbadano możliwości wykorzystania dwóch popularnych technik analizy wielowymiarowej – PCA i odległości Mahalanobisa – w celu uzyskania obiektywnych miar do stwierdzenia autorstwa pisma ręcznego. Analiza PCA i odległość Mahalanobisa funkcjonują jako dobrze ugruntowane metody i mogą być bardziej przyjazne dla użytkownika niż wcześniej opisane techniki uczenia maszynowego (Hecker, 1993; Franke i in., 2003; Srihari i in., 2007; Marquis i in., 2005). Badanie pokazuje, że podejście matematyczne może pomóc w dostarczeniu nowych zmiennych (opisów prób), które mogą być wykorzystane do rozwiązania problemu (problemów) dotyczącego porównania/klasyfikacji poprzez zastosowanie odpowiednich technik, takich jak liniowa analiza

dyskryminacyjna lub SIMCA®. Zważywszy jednak, że powodzenie każdego modelu statystycznego dotyczącego pisma ręcznego może być różne w zależności od takich czynników jak np. wybór znaków, należy uznać, że żadne podejście matematyczne nie jest wolne od błędów. Pokazuje to odsetek wyników prawdziwie dodatnich, wynoszący jedynie 70%, uzyskany w części 2 tego badania. W niniejszym badaniu wykorzystano tylko niektóre cechy ilościowe, co stanowi poważne ograniczenie, ponieważ aby sformułować wnioski, eksperci badający pismo ręczne będą wykorzystywać cechy ilościowe i jakościowe (np. właściwości linii lub strukturę znaków). Wizualna analiza prób pisma ręcznego uczestnika numer 6 wykazała, że wiadome i niewiadome mają wspólne cechy, które pozwoliłyby badaczowi na sformułowanie wniosku o wspólnym pochodzeniu, nawet jeśli były one wyraźnie rozdzielone na wykresie PCA (Ryc. 3). Dowodzi to, że subiektywna analiza w ramach porównania pisma ręcznego w oparciu o standardowo przyjęte praktyki jest nadal bardzo potrzebna. Jednak błędy w stosowaniu standardowych praktyk, takie jak przykład przedstawiony w części 5 niniejszego opracowania, mogą się zdarzać i zdarzały się (Kam i in., 1997; Sita i in., 2002). Włączenie modeli matematycznych do standardowej praktyki może posłużyć jako metoda weryfikacji i może być najlepszym podejściem do tego problemu.