JOURNAL OF FORENSIC SCIENCES

*J Forensic Sci*, March 2018, Vol. 63, No. 2
doi: 10.1111/1556-4029.13529
Available online at: onlinelibrary.wiley.com

**PAPER**

**CRIMINALISTICS**

*Eric M. Warren,*[1] *Ph.D.; and H. David Sheets,*[2] *Ph.D.*

# Statistical Approaches to Type Determination of the Ejector Marks on Cartridge Cases

**ABSTRACT:** While type determination on bullets has been performed for over a century, type determination on cartridge cases is often overlooked. Presented here is an example of type determination of ejector marks on cartridge cases from Glock and Smith & Wesson Sigma series pistols using Naïve Bayes and Random Forest classification methods. The shapes of ejector marks were captured from images of test-fired cartridge cases and subjected to multivariate analysis. Naïve Bayes and Random Forest methods were used to assign the ejector shapes to the correct class of firearm with success rates as high as 98%. This method is easily implemented with equipment already available in crime laboratories and can serve as an investigative lead in the form of a list of firearms that could have fired the evidence. Paired with the FBI's General Rifling Characteristics (GRC) database, this could be an invaluable resource for firearm evidence at crime scenes.

**KEYWORDS:** forensic science, Glock, Smith & Wesson Sigma, ejector marks, morphometrics, Naïve Bayes classification

Determining the type or manufacturer of a firearm based on the class characteristics present on recovered bullets has been documented as early as 1862 (1). However, the use of class characteristics on cartridge cases was not investigated until the 1920s (2,3). While the use of class characteristics on bullets (General Rifling Characteristics) to determine a list of firearms that could have fired them has seen widespread use in crime laboratories, the same cannot be said for the use of class characteristics on cartridge cases. The value of recognizing that certain types of firearms produce specific class characteristics was revisited numerous times over the years but has never enjoyed extensive use (4–6). These types of class characteristics include the shape of the firing pin aperture, the type of breechface machining, locations of the extractor and ejector, and the size and shape of the extractor and ejector. In a case with a bullet where no firearm is submitted, reporting a General Rifling Characteristics (GRC) list can assist the investigator in searching for the correct weapon. This "type determination" can also be applied to cartridge cases to provide the investigator with a list of firearms that are known to leave similar marks to those observed by the firearm examiner.

A good example of "type determination" is distinguishing between the first-generation Smith & Wesson Sigma series firearms ("Sigmas") and Glock firearms. This is important for firearm examiners because these two firearms both leave similar characteristic firing pin and firing pin aperture shapes on the cartridge cases they fire. This makes recognizing the class characteristics on cartridge cases from these two firearms easy for firearm examiners but makes distinguishing between the two of them difficult. In fact, firearm examiners have struggled with distinguishing between cartridge cases fired in Glocks and cartridge cases fired in Sigmas for several years (7,8). It is not uncommon to receive cartridge case evidence with no firearm and, after determining they are fired in the same unknown firearm, report something to the effect of "class characteristics present on these cartridge cases are common to firearms manufactured by Glock and some Smith & Wesson Sigma series pistols" without being able to distinguish any further. This is due to the fact that the firing pin aperture on the breechface of both of these firearms is machined using very similar methods, which lead to a very similar appearance (personal communication with William Carmichael, Glock). In fact, Smith & Wesson was sued by Glock over the similarities and forced to make changes due to an out-of-court settlement (9). The breechface and firing pin impression are the most common areas examined because of the prevalence of individual characteristics which allow the examiner to form a conclusion as to the source of the cartridge cases. However, by examining other marks, often those that receive less attention due to the dearth of individual characteristics, it is possible to make a conclusion about a type or group of firearms that could have fired the cartridge cases in question.

Type determination is possible because of repetitive and reproducible class characteristics. These occur once a manufacturer decides on a design to mass-produce a particular firearm. Once the tooling and machining operations are determined, it is very time-consuming, costly, and inefficient for the manufacturer to make major changes. However, type determination on cartridge cases has not enjoyed widespread use because it has thus far depended on examiners with keen eyes for these types of similarities and good memories to recognize and remember the types of marks left by each type of firearm they have examined. Additional challenges to type determination of cartridge cases include the use and abuse of a firearm as well as the use of different

[1]Firearm and Toolmark Identification Unit, Tennessee Bureau of Investigation, Nashville, TN 37216.
[2]Department of Physics, Canisius College, 2001 Main Street, Buffalo, NY 14208.

Received 27 Oct. 2016; and in revised form 10 Feb. 2017; accepted 3 April 2017.

© 2017 American Academy of Forensic Sciences

types of ammunition, which can affect the quality and location of these marks. To solve the former problem, card files (6) and electronic files (5) were developed so that the examiner could diagram and code the types of marks they observed and then search the files at a later date. Proposed here is a statistical approach focusing on the examination of ejector mark class characteristics.

## Materials and Methods

The firearms examined for this study came from the Tennessee Bureau of Investigation Firearms Reference Collection (FRC) as well as the National Integrated Ballistic Information Network (NIBIN) database operated by the Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF). These firearms consisted of Generation 1 through Generation 4 Glock pistols as well as Smith & Wesson Sigma series pistols manufactured before 1997 (Sigma pistols manufactured after this date no longer had breechface cutouts similar to Glock pistols). This resulted in a data set consisting of test-fired cartridge cases from 100 pre-Generation 4 Glock pistols, 32 Generation 4 Glock pistols, 46 Smith & Wesson Sigma pistols, and 22 Glock model 42/43 pistols. These pistols varied in caliber from 380 Auto to 45 Auto, depending on the models that were manufactured.

### *Shape Measurement for Classification*

The approach employed here to quantify the shape of the ejector marks and assign the measured shapes of the ejector marks to different manufacturers and firearm models is based on methods commonly used in biological research. While human biometry focuses largely on identifying specific individuals, morphometrics is often used in biological research to describe or detect differences between specimens from different species or populations. While many methods in human biometry attempt to identify an individual, we focus here on methods meant to assign an individual to a group, effectively working with what are called class characteristics within the forensic community.

Within the biological sciences, the general class of methods known as morphometrics has a long history dating back to Galton (10) and Pearson (11). The more basic methods use simple measurements of lengths, widths, and angles measured along an organism or artifact, while more complex methods capture sets of landmark measurements or outlines on the organism (12). Agricultural uses include the detection and identification of invasive insect pests (13) and the quantification of the shape of fruits and vegetables (14). The same methods have been used to classify human-produced artifacts, such as projectile points (15,16). Forensic uses of this technology have included the estimation of probable ancestry and sex from cranial measurements (17,18) and the detection of evidence of fetal alcohol exposure (19).

The process of using the shape of a structure, such as the mark left by the ejector on a cartridge case, is a four-step process. The shape must first be quantified so that the information about the shape is coded as a set of measurements independent of random factors such as orientation of the image, the scale of the image, and the location of the mark relative to the boundaries of the image. Next, the reliability of the measurement process must be verified: The measurement error (or variation within the measurement) must be determined and shown to be smaller than the variation between specimens (variation within the group) and between different groups of specimens (variation between the groups). If these conditions are met, then it is possible to construct a mathematical method of classifying specimens, such as a linear discriminant function as used by Wright (18), Canonical Variates Analysis as used by Yee et al. (13), or a Naïve Bayes classification method (20–22), as was found to be most effective in this case. Once the classification method has been created, its effectiveness should be tested using a cross-validation method, in which a portion of the data is held back as a test set.

To capture the shape of the ejector marks appearing on the test-fired cartridge cases, a set of statistical tools collectively known as geometric morphometrics was used (23–25). In this approach, the outline of the periphery of the ejector mark was traced from a photograph of the cartridge case, starting at the point closest to the center of the cartridge case (the primer) and proceeding around the perimeter in a counter-clockwise direction. This traced outline was then mathematically reduced to 100 evenly spaced points, hereafter referred to as landmarks, around the perimeter of the outline. The choice to use 100 points is arbitrary, although previous work (26) has indicated that the classification of shapes is not highly dependent on the number of points used, and MacCleod (27) has suggested one approach to estimating a minimum reasonable of points along an outline. Using more points on an outline does require more specimens to estimate the variance–covariance matrix in the data. With 100 landmark points, it is very difficult to obtain a good estimate of the variance–covariance matrix. Each of the landmark points is a Cartesian coordinate pair, so there are 200 total variables in the study, and the variance–covariance would require 1000 specimens if the rule of thumb that at least five specimens per estimated parameters is applied.

Two different approaches are used to deal with the difficulties posed by the high number of landmark points. The simplest is to simply reduce the number of landmarks to by removing every other landmark, to produce alternate forms of the data set with 50, 25, and 13 landmarks. The analyses used will then be carried out each of the reduced data sets to examine the impact of the number of points used on the analysis. The other tactic used is to rely heavily on nonparametric statistical approaches which do not require an estimation of the variance–covariance structure of the landmarks or require that assumptions be made about the nature of the probability distribution functions that describe the variation in landmark points.

It is common to assume multivariate normality of landmarks in geometric morphometric studies (28), particularly as the multivariate normal model is thought to be generally robust to minor violations of the assumption of the form of the distribution. In practice, it is often difficult, if not impossible, to effectively test to see whether the data points are in fact multivariate normal without having very large numbers of specimens available. It is thus hard to determine whether the violations of the assumptions are minor or not. Within the set of methods collectively referred to as geometric morphometrics, the approach often used is to carry out interference on a distance measure (28–34) called the Procrustes Distance, a univariate distance measure between specimens, rather than on the individual landmarks themselves. The nonparametric statistical tests applied to the data (such as the permutation MANOVA discussed later) are based on the distribution of the interspecimen distances (the Procrustes distances) not the landmark coordinates themselves (also referred to as Procrustes coordinates, or Procrustes residuals). Minor changes in the number of landmark coordinates used to describe a curve do not seem to have a large impact on the relative distances between specimens, and thus on the inferences based on those distances.

The use of nonparametric resampling methods (35–38) used in this study means that there is no explicit assumption made of the form of the underlying probability distribution function (pdf) of either the landmark coordinates, or the Procrustes distances derived from those coordinates. Instead of making assumptions about the nature of these distributions and using probability models on these assumed distributions, the nonparametric resampling methods used preserve the underlying distributions of individual coordinates and simply permute group membership to ascertain significance. Permutation methods date back to Fisher (39), but only became practical with the advent of inexpensive personal computers.

As discussed in more detail later, the study will rely primarily on nonparametric, distance-based inference, and on classification methods Naïve Bayes (40) and Random Forest (41) classifiers, which are generally viewed as highly robust to violations of the assumption of multivariate normality. The performance of the classifiers will be assessed using numerical cross-validation methods, which are again nonparametric in form.

The tracing and reduction to evenly spaced landmarks were performed using a specialized software program designed for collecting geometric morphometric data, the tpsDig program written and distributed by Rohlf (42).

After a collection of specimens was completed, each specimen was represented by 100 landmarks and recorded as Cartesian coordinate pairs, all obtained at the same magnification level. These specimens were then subjected to a registration protocol known as a Procrustes Superimposition (25), which requires that all specimens have a common size, orientation, and average location. While the mathematics may appear complex, this is an embodiment of what common sense tells us about the shape of an object. The shape does not change when the object is moved from one location to another, when it is rotated, or if a zoomed in photograph of it is created, altering its scale. The mathematics used is a least squares process that minimizes the summed squared distances between corresponding points on each specimen (see discussion in 25). The average location is the centroid or average of all coordinates of the specimen, and the size is measured as centroid size, the root mean square distance of all landmarks from the centroid. Specimens are then rotated iteratively to minimize the distance of each landmark from the corresponding landmark on each other specimen. The resulting set of measurements is said to be in Procrustes Superimposition. This process produces both a set of multivariate coordinates describing the shape and a univariate measure of differences between two shapes. When two shapes are in Procrustes Superimposition, the square root of the summed squared distances between the corresponding landmarks of two specimens is called the Procrustes Distance. This Procrustes Distance is thus a univariate measure, or distance metric, of the extent of difference between two specimens. A Procrustes Distance of zero indicates no differences in shape between sets of landmarks; increasing distances indicate progressively more different shapes.

In working with landmarks along outlines, a semilandmark procedure can be used to mathematically compensate for the arbitrary placement of discrete landmarks along an outline. Mathematically, a line is an infinite set of points, so the reduction in this infinite set to some limited number of points has some arbitrary aspects to it. Semilandmark alignment methods have been developed which either minimize the implied differences in specimens (43) or assume that the differences in outlines arose from mathematically smooth changes (44). In this study, these methods caused unacceptably high levels of distortion due to wide variation in shape between the two groups of extractor outlines. Consequently, semilandmark alignment was not used in this study.

### *Classification Methods*

Once the measurements of the specimens have been subjected to a Procrustes Superimposition, the resulting coordinates can be analyzed as a set of multivariate statistical measurements. In this study, these landmark coordinates and the measured centroid size are the inputs to the Naïve Bayes and Random forest classification methods used to classify ejector marks, unlike the MANOVA results which are based on the interspecimen Procrustes Distance. The Procrustes Distance between specimens is helpful in discussing the nature of the measurement, in that it may be used to quantify the variance within a group of measurements. By repeatedly measuring a particular specimen, the variation or error in the measurement process can be determined as the squared Procrustes distances of the repeated measurements about the mean shape. It is not expected that two or more tracings will be absolutely identical; the human operator will choose slightly different pixels in the image as the tracing proceeds. The use of the summed squared Procrustes distances as a measure of variance allows this operator variance to be quantified and tracked (45). Similarly, the variation within ejector marks within specimens from a single model of firearm and the variation between firearm models from different manufacturers can be determined.

To ascertain whether there are statistically significant differences in the mean shape of ejector marks on cartridge cases from different models of firearms, a nonparametric permutation-based multivariate analysis of variance (MANOVA) procedure is used, using statistics based on a distance metric between specimens, rather than on raw landmark measurements. The MANOVA process determines the variation between the two (or more) groups of ejector marks and compares the variation between the groups to the "unexplained" variation within the groups. The test statistic used is an F-ratio of the variance between groups divided by the variance within groups. In this study, the Adonis function within the Vegan package (46) in R (47) was used to compute the MANOVA using the permutation methods developed by Anderson (29–32,34). In this permutation approach to MANOVA, interspecimen distances (Procrustes distances in this case) are used to compute the sums of squares used in the MANOVA, rather than relying on the landmark coordinates themselves. As this is a nonparametric permutation method, it is not necessary to assume any particular form of the distribution of either the landmark coordinates or the Procrustes distances, as discussed earlier. The permutation is performed at the specimen level, so that the underlying variance–covariance structure of the landmarks is preserved. As a further check on the impact of arbitrary choice of how many landmarks to use in the analysis, the MANOVA will be repeated for the reduced data sets of 50,25, and 13 landmarks as well as the original 100 landmark data set. To have a successful classification, there must be a meaningful difference in the mean ejector mark, but this is not necessarily an adequate condition if there is a large variation within each group of firearms.

To determine whether the shapes of the ejectors could be used to assign traced ejector marks to the correct firearm model, the Naïve Bayes and Random Forest classification methods implemented in the caret package (48) in R were used. The entire set of all Cartesian landmark coordinates and the centroid size measure for each specimen was used as the input to both classifiers.

Bayesian methods are known to be highly effective in assigning data into categories, most notably as a spam filter for e-mail. The Naïve Bayes method simply computes the probability that each individual variable (landmark in our case) belongs to a given group using an independent normal model of variation for each variable (landmark). The term “Naïve” in the description of this method arises from this assumption of independent normal distributions, which is seldom actually satisfied. However, like many methods based on least squares or normal models (bell curves), the Naïve Bayes method works well in practice despite known violations of independent normality in the data. The Naïve Bayes method is also known to work well when there are a lot of data available, even when the individual measurements are highly correlated with one another (49), as is the case with the data presented here. The Random Forest method used is an extension of the regression tree (50) method. Both are nonparametric data classification approaches based on machine learning methods that seek to find an optimal set of rules for classifying specimens. Regression trees are relatively simple approaches, but have a strong tendency to overfit the data; the Random Forest method uses a consensus approach based on a set of many regression trees fitted to different subsets of the data, thus reducing overfitting and achieving more robust classification schemes.

There are a wide range of other possible classification methods available which were not examined in detail; an initial attempt with Canonical Variates Analysis was not as effective as the Naïve Bayes method, and comments from initial reviewers prompted examination of the Random Forest approach. In the initial examination of the performance of the classifiers, they were applied to the reduced data sets (100, 50, 25, and 13 landmarks).

The effectiveness of the Naïve Bayes and Random Forest methods was assessed using a cross-validation method (51) rather than using any parametric statistical approaches to assessing the effectiveness of the classifier. In the cross-validation process, ten percent of the specimens chosen at random were set aside as a “test set,” and the remaining ninety percent of the specimens formed the “training set.” This process is intended to mimic or simulate how the test would perform on additional data. As long as the new data are similar in its overall characteristics to the “test set,” the process should work as well on new data as it did on the test data. The Naïve Bayes or Random Forest method was applied to the training set to build the classification method, which was used to assign the specimens in the test set to a firearm model, and the results of this classification were stored. A new training set was then created at random from all of the specimens, and the process was repeated multiple times. As the randomization is at the specimen level, the covariation structure present in the data is retained throughout the process, regardless of the nature of the covariation structure.

After many test sets have been classified, the rate of correct assignment was then calculated and reported in the form of a “confusion matrix.” In a confusion matrix, the true source of the specimens is shown along columns of the table and the firearm models assigned by the numerical classification method are arrayed along rows of the table. The counts in each cell of the table indicate how many ejector marks from a given source are assigned to a particular category by the classification method. A perfect assignment method would have entries only along the diagonal, while off-diagonal elements indicate incorrect assignments and may be informative about the nature of errors made in assigning ejector marks to specific firearm models. We carried out the classification at several levels of detail, attempting initially to classify only by manufacturer, but also extending this to classification of specific models of firearm.

## Results

The mean shape of ejector marks on 9 mm and 40S&W/10 mm cartridge cases fired from Glock and Sigma firearms differed slightly between caliber and within their respective manufacturer (Fig. 1). However, the differences between cartridge cases fired from firearms made by different manufactures were very noticeable (Fig. 2). Compared to Generation 1-3 Glocks, the Gen 4 Glock ejectors appear to have been straightened out, while the tip has been twisted and widened. So while the location on the cartridge case has changed slightly (from closer to the primer toward the edge of the cartridge case rim), the shape of the ejector mark remains similar. Unsurprisingly, using this statistical approach, ejector marks from the Gen 4 Glocks were similar to those from earlier generations, but the ejector marks from Glock models 42 and 43 looked quite different from ejector marks produced by the other Glock models or Sigma firearms.

To determine the measurement error and the variation within cartridge case firings, a repeated measurement study was carried out in which five cartridges were fired from one Glock and from one Sigma firearm. Each ejector mark was then traced five times, for a total of 50 specimens in a balanced design. The permutation MANOVA procedure was carried out (Table 1) and clearly shows statistically significant differences between the mean ejector shape of these two firearms, as well as statistically significant differences among the five cartridge cases fired from each firearm, regardless of the number of landmarks (100, 50, 25, or 13) used to characterize the ejector mark. The differences between the firearms accounted for 92.6% of the total variance, variance within the cartridge cases from the same firearm was 4.7% of total error, and measurement error accounted for the remaining 2.7% of the total variation in the 100 landmark set. In the reduced versions of the data set, the differences between firearms were roughly 84% of the variance, the variance within firearms was typically 13.5%, and random noise was 2.5%. In this case, the measurement error is clearly far less than the differences between the two firearms, and less than the variation in the ejector mark from one firing to the next of the same firearm. The form of the data with the highest number of landmarks did appear to detect more variance between firearms, but also seemed to have higher random noise.

The initial attempt at Naïve Bayes and Random Forest classification used the 100 Glock firearms of various calibers as a single group and the 47 Sigma firearms of various calibers as a second group, so the classification was to the manufacturer level, with the Gen 4 and model 42/43 cartridge cases not included as separate categories. Table 2 shows the results of 100 cross-validation trials of the methods, in which 15 specimens (roughly 10%) were held out as the test set. The on-diagonal elements in this table show correct assignments of cartridges using the Naïve Bayes method; off-diagonals are incorrect assignments. Both classification methods were applied to the full data set of 100 landmarks, and to the reduced data sets of 50, 25, and 13 landmarks. The rate of correct classification did not depend strongly on the number of landmarks used or on the method (Naïve Bayes or Random Forest). Thus, when implementing this method, it is recommended to use a smaller number of landmarks when tracing features. While not shown, a parametric Canonical Variates Analysis (CVA) was less effective as a

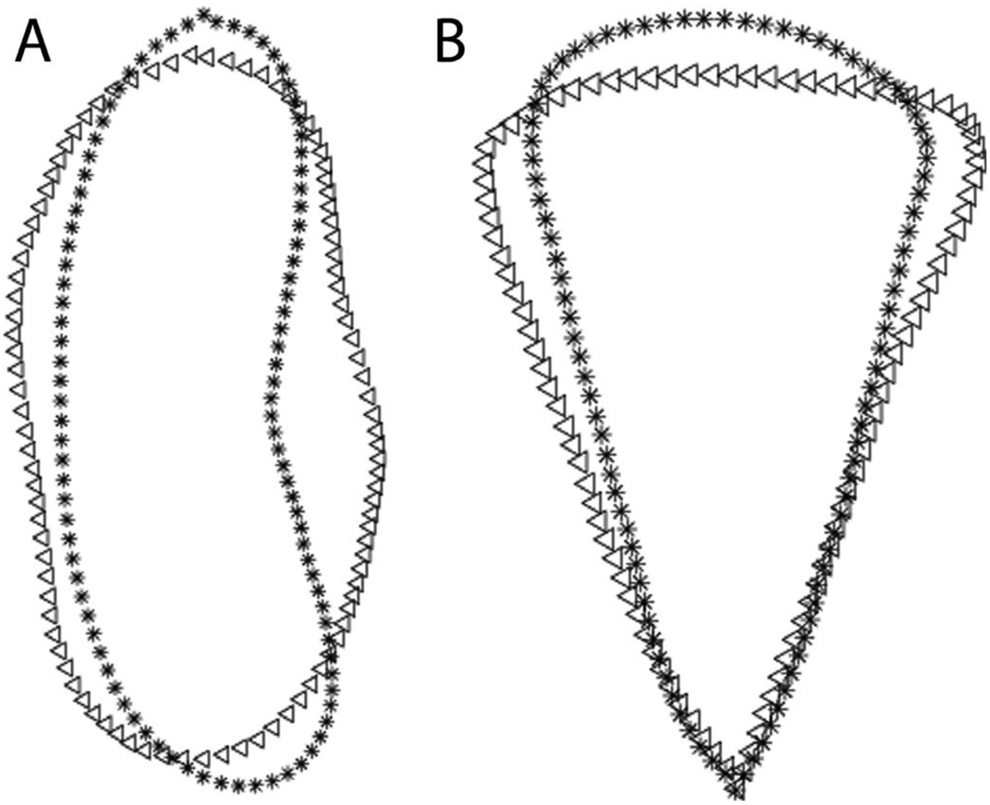


FIG. 1—*(A) Mean shape of ejector marks from 9 mm Glocks (triangles) and 40S&W/10 mm Glocks (asterisks) (B) Mean Shape of ejector marks from 9 mm Sigmas (triangles) and 40S&W Sigmas (asterisks).*

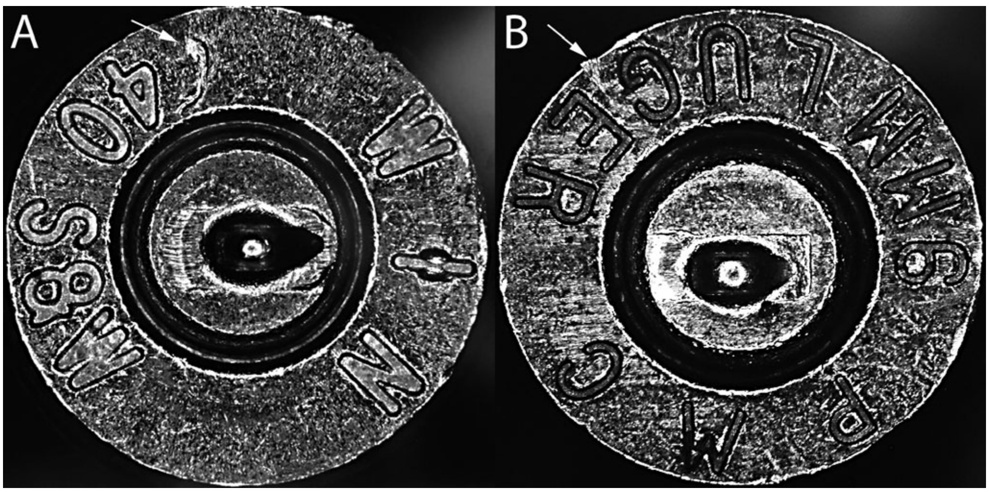


FIG. 2—*Photographs of a typical Glock ejector mark (A) and a typical Sigma ejector mark (B). Red arrow indicating the ejector marks.*

classifier, perhaps due to the difficulties of accurately estimating the variance–covariance structure in the data, which is necessary in the CVA.

The same classification models were then used to assign the Glock Gen 4 models and the Glock model 42 and 43 specimens, without using the Gen 4, model 42, or model 43 as part of the training set. All 32 of the Glock Gen 4 models were classified as being Glocks, and all 22 of the model 42/43 ejector marks were also classified as Glocks, when using the 100 landmark data set.

When four categories of firearm were included in the assignment procedure (Glock Gen 1-3, Glock Gen 4, Glock model 42/43, and Sigma models) using the 100 landmark data set, the Naïve Bayes assignment method did not perform as well at making correct assignments to the four categories (the on-diagonal entries). It was 60.75% effective to the individual category (Table 3), and 87.5% correct as to the manufacturer. It appears in this calculation that it was quite difficult to tell the different Glock models from one another; in particular, the Glock Gen 4 models appeared to be roughly evenly distributed

TABLE 1—*Permutation MANOVA results from repeated-measurements study of variation.*

| | df | Sum of Squares | Mean Square | $F$ | $R^2$ | Pr (>$F$) |
|---|---|---|---|---|---|---|
| 100 Landmark Set | | | | | | |
| Firearm Type | 1 | 3.208 | 3.208 | 1376.74 | 0.92599 | 0.001 |
| Firings | 8 | 0.1632 | 0.0204 | 8.75 | 0.04711 | 0.001 |
| Residuals | 40 | 0.0932 | 0.0023 | | 0.0269 | |
| Total | 49 | 3.4644 | | | 1 | |
| 50 Landmark Set | | | | | | |
| Firearm Type | 1 | 2.58222 | 2.58222 | 1411.73 | 0.84176 | 0.001 |
| Firings | 8 | 0.41228 | 0.05153 | 28.17 | 0.13439 | 0.001 |
| Residuals | 40 | 0.07316 | 0.00183 | | 0.02385 | |
| Total | 49 | 3.06766 | | | 1 | |
| 25 Landmark Set | | | | | | |
| Firearm Type | 1 | 2.58308 | 2.58308 | 1405.81 | 0.84162 | 0.001 |
| Firings | 8 | 0.41261 | 0.05158 | 28.07 | 0.13444 | 0.001 |
| Residuals | 40 | 0.0735 | 0.00184 | | 0.02395 | |
| Total | 49 | 3.06919 | | | 1 | |
| 13 Landmark Set | | | | | | |
| Firearm Type | 1 | 2.58823 | 2.58823 | 1381.6 | 0.83871 | 0.001 |
| Firings | 8 | 0.4228 | 0.05285 | 28.21 | 0.13701 | 0.001 |
| Residuals | 40 | 0.07493 | 0.00187 | | 0.02428 | |
| Total | 49 | 3.08596 | | | 1 | |

Little difference is seen in the permutation MANOVA results carried out on the versions of the data set reduced from 100 to 50, 25, or 13 landmarks. For the 100 landmark version of the data set, the difference between the cartridge cases of different firearms explained 92.6% of the variance, differences between cartridge cases from the same firearm was 4.7% of variance, and the measurement error was 2.7% of total variance. In the smaller data sets, differences between the firearms were roughly 84% of variance, between firings was 13%, and measurement error was under 2.5%.

TABLE 2—*Naïve Bayes and regression forest confusion matrices for Glock vs Sigma cartridge cases.*

| 100 Landmark Set | Naïve Bayes Glock | Sigma | | Regression Forest Glock | Sigma |
|---|---|---|---|---|---|
| Glock | 1035 | 1 | Glock | 1004 | 8 |
| Sigma | 6 | 458 | Sigma | 10 | 478 |
| 50 Landmark Set | Glock | Sigma | | Glock | Sigma |
| Glock | 1003 | 14 | Glock | 1001 | 8 |
| Sigma | 4 | 479 | Sigma | 5 | 486 |
| 25 Landmark Set | Glock | Sigma | | Glock | Sigma |
| Glock | 1035 | 4 | Glock | 1001 | 8 |
| Sigma | 14 | 447 | Sigma | 5 | 486 |
| 13 Landmark Set | Glock | Sigma | | Glock | Sigma |
| Glock | 1014 | 6 | Glock | 1011 | 3 |
| Sigma | 20 | 460 | Sigma | 10 | 476 |

Confusion matrix showing the results of 100 cross-validation trials of the Naïve Bayes and the Random Forest assignment methods of ejector marks to the manufacturer (a two-category classification), using all four landmark configurations (100, 50, 25, and 13). In each trial, 15 specimens formed the test set, and the remaining 132 were used as the training set. The table indicates that in six cases of the 1500 trials of the Naïve Bayes method on the 100 landmark set, for example, cartridge cases from a Sigma model were incorrectly classified as being from a Glock, while one Glock cartridge case was classified as being from a Sigma. The Naïve Bayes method was 99.9% correct overall in this trial at 100 landmarks.

over the Glock models. The Random Forest method applied to the same data set was 86% effective at the model level and 97% effective at the manufacturer level. The presence of additional categories did apparently make it more difficult to tell the Glock and Sigma firearms apart, particularly with the Naïve Bayes approach. However, these difficulties with the Glock model 42/43 firearms would be easily overcome by examination of the breechface characteristics due to differences in the firing pin aperture shape.

TABLE 3—*Naïve Bayes and random forest confusion matrix of Glock Gen 1-3, Glock Gen 4, Glock 42/43, and Sigma cartridge cases.*

| | Glock | Glock Gen 4 | Glock 42/43 | Sigma |
|---|---|---|---|---|
| Naïve Bayes Method, 100 Landmark Set | | | | |
| Glock | **685** | 160 | 17 | 67 |
| Glock Gen 4 | 83 | **41** | 30 | 25 |
| Glock 42/43 | 129 | 116 | **165** | 46 |
| Sigma | 78 | 16 | 16 | **324** |
| Regression Forest, 100 Landmark Set | | | | |
| Glock | **977** | 144 | 0 | 9 |
| Glock Gen 4 | 85 | **199** | 0 | 8 |
| Glock 42/43 | 0 | 0 | **198** | 21 |
| Sigma | 10 | 0 | 16 | **433** |

Confusion matrix showing the results of 100 cross-validation trials of the Naïve Bayes and Regression Forest assignment methods of ejector marks (100 landmark set) to the manufacturer and model (four categories). In each trial, 20 specimens formed the test set, and the remaining 180 were used as the training set. The Naïve Bayes method was 60.8% correct to model in this trial and 87.5% correct to manufacturer. The Regression Forest method was 86% correct to model and 97% correct to the manufacturer.

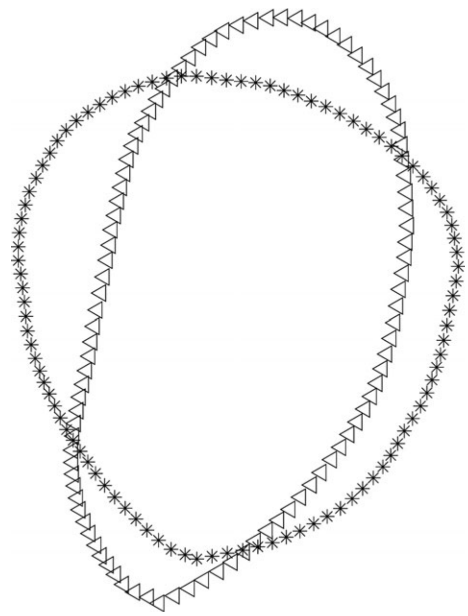

FIG. 3—*Mean shape of ejector marks from Glock Gen 4 (triangles) and Glock models 42 and 43 (asterisks).*

## Discussion

Glock and Sigma firearms leave differently shaped ejector marks on the cartridge cases. The Glock ejector marks typically have a rounded outline, while the Sigma ejector marks are wedge-shaped with sharp angles. The shapes of these marks may be measured with a degree of precision such that the measurement error is less than the variation in shape from one cartridge case to the next fired from the same firearm, and much less than the difference between Glock and Sigma firearms.

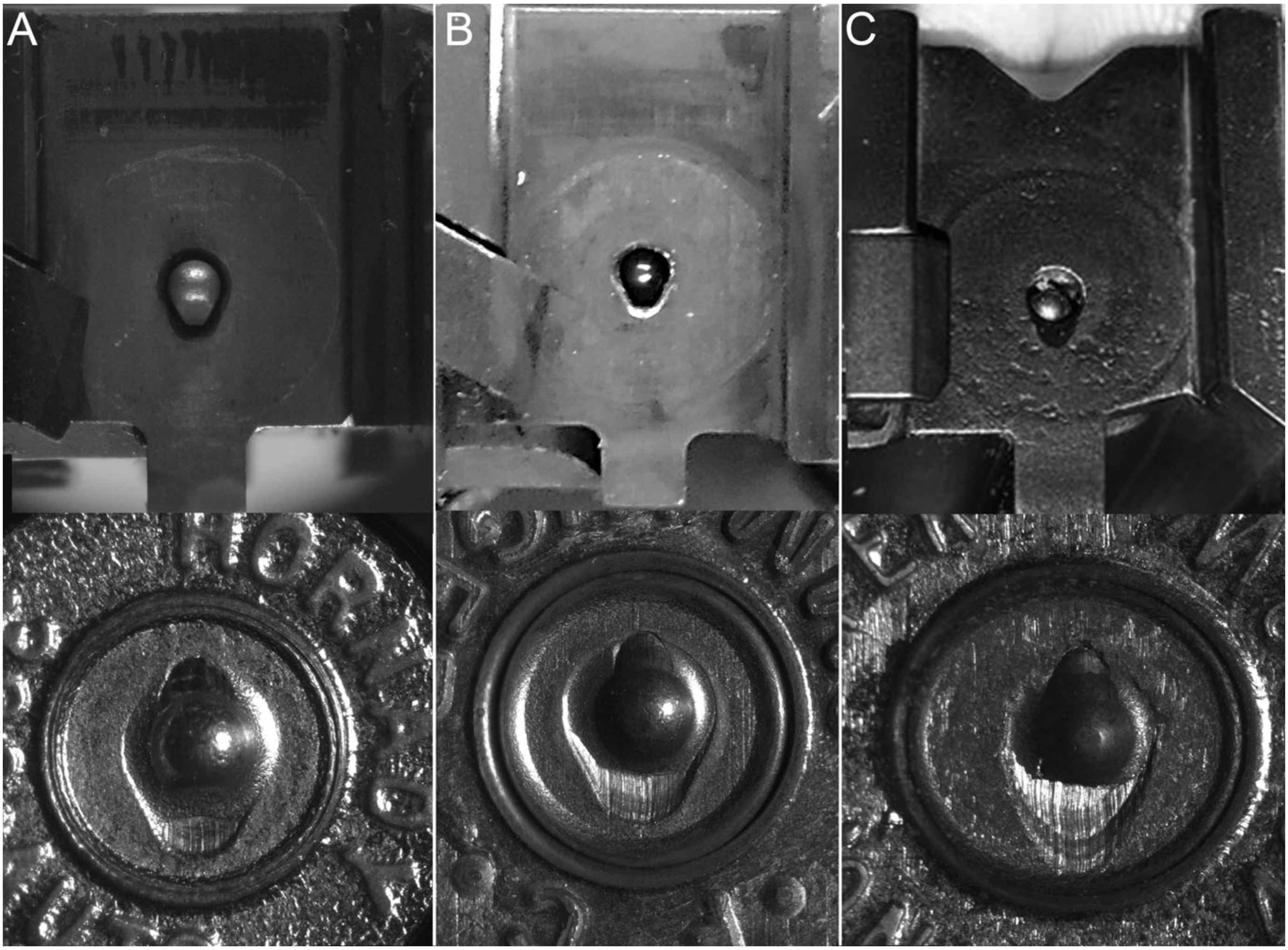


FIG. 4—*Photographs comparing the breechface of the firearm (top) and the firing pin aperture marks on cartridge cases (bottom) between Glock model 42 (A), Glock model 43 (B), and Smith & Wesson model M&P (C) test fired cartridge cases.*

Furthermore, it is possible to use these measured shapes in mathematical classification processes (Naïve Bayes and Random Forests) to identify the class characteristics of the ejector shape with a success rate of up to 98%. Attempts to discriminate between Glock Gen 1-3, Glock Gen 4, Glock model 42/43, and Sigma firearms were less successful, although the Random Forest method appears to be more effective in this application than the Naïve Bayes. There appears to be little difference between Glock Gen 1-3 and Glock Gen 4, and only slightly more differences with the Glock models 42/43 (Fig. 3). The limited differences between the different Glock models and generations may not be reliably detectable, so the overall performance of the Naïve Bayes method was degraded by the attempt to subdivide the classification categories more finely than the data would actually support. However, in the areas where the assignments suffered, a visual examination will help solidify the examiner's conclusion. While the Sigma ejector is straight and leaves marks on the edge of the head of the cartridge, the Gen 1-3 Glock ejectors are angled in toward the primer of the cartridges and result in a mark near or on the primer itself. When the Gen 4 Glocks were introduced, the ejectors were changed such that the ejectors were straightened but with an axial twist. This results in the same curved shape as the Gen 1-3 Glocks, but with the placement occurring closer to the edge of the rim of the cartridge case. By comparing the ejector location on the cartridge, it may be possible to distinguish the newer ejector marks found in Gen 4 firearms from those that arise from Gen 1-3 Glocks (52).

Furthermore, when comparing the newest models 42 and 43, the differences are even more distinguished. Due to the design differences in the machining and final shape of the firing pin aperture, the class characteristics present on the primer are quite easy for an examiner to distinguish the Gen 1-4 Glocks from the models 42 and 43. It is interesting to note that, perhaps ironically, the Glock models 42 and 43 breechface class characteristics are very similar to the Smith & Wesson M&P series firearms (Fig. 4). It is also interesting to note that the Springfield line of XDs pistols is now being produced with similarly shaped firing pins and firing pin apertures. Therefore, this same approach could be extended and employed to distinguish between the newer Glock models and Smith & Weson M&P series firearms, and to distinguish between Glock and Springfield XDs series firearms based on differences in the ejector marks.

The type of statistical approach described here has been used to study handwriting samples as well as facial features (53,54), and it does not appear to have been applied to type determination using cartridge cases. Whenever two items of evidence are being compared to determine whether they share the same source, firearm examiners first examine the class characteristics present on these items. If the class characteristics do not match, then the two items are excluded as having the same source. However, if the class characteristics match, then the individual characteristics present on the items are examined to determine whether these unique marks share similarities. If they do, then the examiner can make a conclusion that the two items have the same source. Many articles have been published using statistical methods to try to automate this individualization of evidence to the source (reviewed in 55 and 56). However, this type of class determination approach to group cartridge cases based on the ejector shapes in order to determine firearms with similar class characteristics that could have fired them has not been done previously. A general location of the ejector mark has been used as a discriminating class characteristic (57), but this study presents an exciting additional tool for quantifying class characteristics and further demonstrates the importance of looking at ejector mark data when classifying cartridge cases. Doing this would allow the firearm examiner to provide the investigator or detective with a solid investigative lead of a list of possible firearms.

*Acknowledgments*

We would like to extend our gratitude and appreciation to the support specialists at Forensic Technology Inc. for their help and assistance in support of this research work.

## References

1. Hatcher JS, Jury FJ, Weller J. Firearms investigation, identification and evidence. Harrisburg, PA: Stackpole Co., 1957.
2. Mezger O. Über die entwicklung schießtechnischer untersuchungen im dienste der justiz. Dtsch Z Gesamte Gerichtl Med 1929;13(1):377–90.
3. Raestrup G. Die kriminaltechnische untersuchung von patronenhülsen und geschossen. Dtsch Z Gesamte Gerichtl Med 1926;7(1):242–58.
4. Sojat JG, Stauffer JC. A study of class characteristics of autoloading weapons. J Forensic Sci 1958;3(4):444–79.
5. Nennstiel R. A computer-supported method of firearm type determination. AFTE J 1986;18(4):4–32.
6. Kennington RH. The matrix, 9 mm Parabellum: an empirical study of type determination. Miami, FL: R.H. Kennington, 1992.
7. McCombs N. Fired cartridge case comparisons: 9 mm and .40 caliber Glock vs. Smith and Wesson Sigma series pistols. AFTE J 2004;36(2):150–4.
8. Nichols RG. Glock versus Smith & Wesson Model SW40F: comparison of markings on fired cartridge cases. AFTE J 1995;27(2):133–9.
9. Thurman R. Judge dismisses lawsuit against gunmaker, S&W and Glock settle suit. Shooting Industry, 1997; https://www.thefreelibrary.com/Judge+dismisses+lawsuit+against+gunmaker%2c+S%26W+and+Glock+settle+suit.-a019767928.
10. Galton F. Regression towards mediocrity in hereditary stature. J Anthropol Inst Great Britain and Ireland 1886;15:246–63.
11. Pearson K. Mathematical contributions to the theory of evolution. V. On the reconstruction of the stature of prehistoric races. Philos Trans R Soc Lond 1899;192:169–243.
12. Webster M, Sheets HD. A practical introduction to landmark-based geometric morphometrics. Quant Methods Paleobiology 2010;16:163–88.
13. Yee WL, Sheets HD, Chapman PS. Analysis of surstylus and aculeus shape and size using geometric morphometrics to discriminate *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae). Ann Entomol Soc Am 2011;104(2):105–14.
14. Costa C, Antonucci F, Pallottino F, Aguzzi J, Sun D, Menesatti P. Shape analysis of agricultural products: a review of recent research advances and potential application to computer vision. Food Bioproc Tech 2011;4 (5):673–92.
15. Thulman DK. Discriminating Paleoindian point types from Florida using landmark geometric morphometrics. J Archaeol Sci 2012;39(5):1599–607.
16. Cardillo M. Some applications of geometric morphometrics to archaeology. In: Elewa AMT, editor. Morphometrics for nonmorphometricians. Berlin/Heidelberg, Germany: Springer, 2010;325–41.
17. Ross AH, Slice DE, Williams SE. Geometric morphometric tools for the classification of human skulls. Washington, DC: BiblioGov, 2012.
18. Wright R. Guide to using the CRANID6 programs. CR6aIND: for linear and nearest neighbours discriminant analysis, 2010; http://www.box.net/shared/static/n9q0zg tr1y.EXE (accessed July 6, 2016).
19. Bookstein FL, Kowell AP. Bringing morphometrics into the fetal alcohol report: statistical language for the forensic neurologist or psychiatrist. J Psychiatry Law 2010;38(4):449–73.
20. Lewis DD. Naive (Bayes) at forty: the independence assumption in information retrieval. In: Nedellec C, Rouveirol C, editors. Proceedings of the 10th European Conference on Machine Learning (ECML '98); 1998 April 21–23; Chemnitz, Germany. Berlin/Heidelberg, Germany: Springer-Verlag, 1998;4–15.
21. Zhang H. The optimality of naive Bayes. In: Barr V, Markov Z, editors. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS); 2004 May 12–14; Miami Beach, FL. Menlo Park, CA: AAAI Press, 2004;562–7.
22. Hand DJ, Yu Y. Idiots Bayes - not so stupid after all? Int Stat Rev 2001;69:385–9.
23. Bookstein FL. Morphometric tools for landmark data: geometry and biology. Cambridge, U.K.: Cambridge University Press, 1997.
24. Dryden IL, Mardia KV. Statistical shape analysis. vol. 4. Chichester, U.K.: J. Wiley, 1998.
25. Zelditch ML, Swiderski DL, Sheets HD. Geometric morphometrics for biologists: a primer. Cambridge, MA: Academic Press, 2012.
26. Sheets HD, Covino KM, Panasiewicz JM, Morris SR. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. Front Zool 2006;3(1):15.
27. MacLeod N. Generalizing and extending the eigenshape method of shape space visualization and analysis. Paleobiology 1999;25(1): 107–38.
28. Goodall C. Procrustes methods in the statistical analysis of shape. J R Stat Soc Series B Stat Methodol 1991;53(2):285–339.
29. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol 2001;1:32–46.
30. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 2006;62(1):245–53.
31. Anderson MJ, Robinson J. Generalized discriminant analysis based on distances. Aust N Z J Stat 2003;3:301–18.
32. McArdle BH, Anderson MJ. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 2001;82(1):290–7.
33. Zapala MA, Schork NJ. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proc Natl Acad Sci U S A 2006;103(51):19430–5.
34. Anderson MJ. Permutation tests for univariate or multivariate analysis of variance and regression. Can J Fish Aquat Sci 2001;58(3):626–39.
35. Basso D, Pesarin F, Salmaso F, Solari A. Permutation test for stochastic ordering and ANOVA. New York, NY: Springer, 2009;1–35.
36. Manly BF. Randomization, bootstrap and Monte Carlo methods in biology. Boca Raton, FL: CRC Press, 2006.
37. Good P. Permutation tests: a practical guide to resampling methods for testing hypotheses. Berlin/Heidelberg, Germany: Springer, 2013.
38. Good P. A practitioner's guide to resampling for data analysis, data mining, and modeling. Boca Raton, FL: Chapman & Hall/CRC, 2011.
39. Fisher RA. The design of experiments. Edinburgh, Scotland: Oliver & Boyd, 1935.
40. Russell S, Norvig P. Artificial intelligence: a modern approach, 2nd edn. Upper Saddle River, NJ: Prentice Hall, 2009.
41. Ho TK. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 1988;20(8):832–44.
42. Rohlf FJ. tpsDig, version 2.10. Department of ecology and evolution. Stony Brook, NY: State University of New York-Stony Brook, 2006.
43. Sampson PD, Bookstein FL, Sheehan FH, Bolson EL. Eigenshape analysis of left ventricular outlines from contrast ventriculograms. In: Marcus LF, editor. Advances in morphometrics. Berlin/Heidelberg, Germany: Springer, 1996;211–33.
44. Bookstein FL. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med Image Anal 1997;1 (3):225–43.
45. Sheets HD, Bush PJ, Brzozowski C, Nawrocki LA, Ho P, Bush MA. Dental shape match rates in selected and orthodontically treated populations in New York state: a two-dimensional study. J Forensic Sci 2011;56(3):621–6.

46. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Package 'vegan,' 2016; https://cran.r-project.org/web/packages/vegan/vegan.pdf (accessed July 9, 2016).
47. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013; http://www.R-project.org/.
48. Kuhn M. Contributions from Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, et al. the R Core Team, 2016; https://cran.r-project.org/web/packages/caret/caret.pdf (accessed July 9, 2016).
49. Bickel PJ, Levina E. Some theory for Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. Bernoulli 2004;10(6):989–1010.
50. Rokach L, Maimon O. Data mining with decision trees: theory and applications. Hackensack, NJ: World Scientific Pub Co Inc., 2008.
51. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 1983;37(1):36–48.
52. Warren EM, Pitts SL. Using class characteristics to distinguish Glock from Smith & Wesson Sigma pistols. AFTE J 2017;49(2):70–78.
53. Bozza S, Taroni F, Marquis R, Schmittbuhl M. Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship. Appl Stat 2008;57(3):329–41.
54. Morecroft LC. A statistical approach to facial identification [PhD thesis]. Sheffield, England: University of Sheffield, 2009.
55. Gerules G, Bhatia SK, Jackson DE. A survey of image processing techniques and statistics for ballistic specimens in forensic science. Sci Justice 2013;53(2):236–50.
56. Vorburger TV, Song J, Petraco N. Topography measurements and applications in ballistics and tool mark identifications. Surf Topogr 2016;4(1):013002.
57. Smith CL. Multi-dimensional cluster analysis of class characteristics for ballistics specimen identification. Proceedings of the IEEE 35th International Carnahan Conference on Security Technology (Cat. No. 01CH37186); 2001 Oct 16-19; London, England. Piscataway, NJ: IEEE XPlore, 2001;115–21.

Additional information and reprint requests:
H. David Sheets
Department of Physics
Canisius College
2001 Main Street, Buffalo, NY 14208
E-mail: sheets@canisius.edu