

---

# Matching Bullets

**Eric Hare, Heike Hofmann**

Department of Statistics and Statistical Laboratory, Iowa State University

---

**F**orensic examiners try to determine whether two guns were fired from the same barrel by analyzing them under a microscope in controlled conditions. Can we make the task simpler?

**Keywords:** Data visualization, Statistical graphics, Statistical computing.

## Contents

<b>1</b>	<b>Introduction</b>
<b>2</b>	<b>Methodology</b>
2.1	Cylindrical Fit . . . . .
2.2	Loess Fit . . . . .
<b>3</b>	<b>Web Application</b>
<b>4</b>	<b>Approach to automatic matching</b>
4.1	Horizontal Alignment . . . . .
4.2	Varying crosscuts . . . . .
4.3	Varying smoothing factor . . . . .
4.4	Automated matching algorithm . . . . .
<b>5</b>	<b>Evaluation</b>
<b>6</b>	<b>Conclusion</b>

## 1 Introduction

Bullet matching is a process in forensic science done to determine whether two bullets were fired from the same gun barrel. This process has broad applicability in terms of the conviction of criminals in the United States criminal justice system. Ballistics identification has long been considered an accepted and reliable procedure, but in the past ten years has undergone more significant scrutiny. In 2005, in *United*

*States vs. Green*, the court ruled that the expert could not confirm that the bullet casings came from a specific weapon with certainty, but could merely describe other casings which are similar. Further court cases in the late 2000s expressed caution about the use of ballistics identification evidence [7].

The new scrutiny of the procedure culminated with the 2009 National Academy of Sciences report [16]. The report concluded that minimal research had been done exploring the reliability of these ballistic identification methods. In fact, the report says the forensic evidence “including, for example, bite marks and firearm and toolmark identification is introduced in criminal trials without any meaningful scientific validation, determination of error rates, or reliability testing to explain the limits of the discipline.”

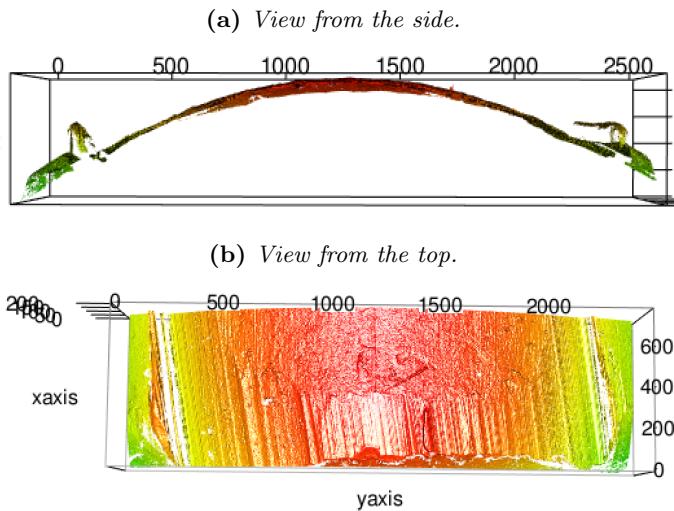
1      2      3      4      5      6      7      8      9  
Rifling in the barrel creates striation marks on the bullet during the firing process (and grooves). These marks are claimed to be unique to the barrel, as described in a 1992 AFTE article [1]. “The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the *unique surface contours* of two toolmarks are in sufficient agreement” The article goes on to state that “Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows.”

10     11     12     13     14  
The primary goal in this type of forensic analysis is to identify markings in an attempt to match two bullets. Since gun barrels should produce similar markings on bullets fired consecutively, the careful analysis of these surface images can help us determine whether two bullets are a match. However, this process has traditionally required forensic experts to perform the analysis, and required physical examination of the bullets. We’ve produced a framework which, in software, allows for the analysis of the surface topologies, and the transcription of the markings into a 2D plotting framework which we believe makes identification of matches easier for the lay person.

## 2 Methodology

Throughout this paper, we will work with images from the James Hamby Consecutively Rifled Ruger Barrel Study [8]. Ten consecutively rifled Ruger P-85 pistol barrels were obtained from the manufacturer and fired to produce known test bullets and unknown bullets for comparison.

3D topographical images of each bullet were taken using a NanoFocus lens at 20x magnification and made publicly available on the NIST Ballistics Database Project<sup>1</sup> in a format called x3p (XML 3-D Surface Profile). x3p is a container format conforming to the ISO5436-2 standard<sup>2</sup>, implemented to provide a simple and standard conforming way to exchange 2D and 3D profile data. x3p was adopted by the OpenFMC (Open Forensic Metrology Consortium<sup>3</sup>), a group of academic, industry, and government Firearm Forensics researchers whose aim is to establish best practices for researchers using metrology in the forensic sciences.

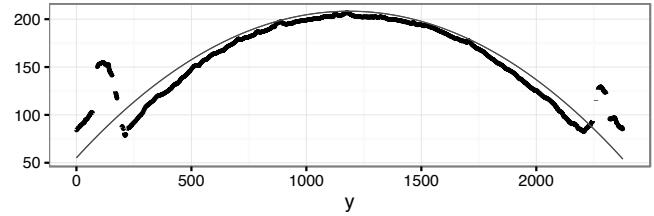


**Figure 1:** An example of a scan of a bullet land from one groove to the next.

Each fired bullet is provided in form of a set of six x3p files, where each file is a surface scan between adjacent grooves on the bullet, called a “land”. An example of plotting one of these lands is given in Figures 1a and 1b, which show side and top profiles of the land respectively. The tilt of the lines to the left in Figure 1b is not an artifact, but a direct and expected consequence of the spin induced by the rifling during the firing process. Depending on whether a barrel is rifled clockwise or counter-clockwise, the striations have a left or right tilt. The direction of the rifling is a class characteristic.

An initial naive approach leads us to fixing a particular value of a coordinate, and then producing a plot of the

heights across values of the other coordinate. Figure 2 gives a plot resulting from such a procedure. It can be seen that the global structure of the land dominates the appearance of the plot. The grooves can be clearly identified on the left and right side, and the curvature of the surface is the most visible feature in the middle.



**Figure 2:** Side profile of the surface measurements (in  $\mu\text{m}$ ) of a bullet land across a fixed  $x$  value. Note that the global features dominate any deviations, corresponding to the individual characteristics of striation marks.

It is clear that our next approach must be to model the overall structure in order to focus on the deviations, given that the deviations should correspond to bullet markings.

### 2.1 Cylindrical Fit

One reasonable approach is to fit a circle to the curve and compute the residuals from this fit.

For this, assume that  $n$  data points are given in form of data tuples  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  that are (approximately) located on a circle. We want to estimate the location of the center and radius of the best fitting circle using a least squares approach.

We minimize the following expression:

$$D = \sum_{i=1}^n (r^2 - (x_i - a)^2 - (y_i - b)^2)^2 \quad (1)$$

We do this by differentiating  $D$  with respect to  $r, a$ , and  $b$ : let us assume that  $x_i$  and  $y_i$  are centralized (i.e.  $\sum x_i = \sum y_i = 0$ ). Note, if they are not, make a note of the current means, subtract them now and add them to  $(\hat{a}, \hat{b})$  at the end.

The derivate of  $D$  with respect to  $r$  is:

$$\begin{aligned} \frac{d}{dr} D &= 2 \sum_i (r^2 - (x_i - a)^2 - (y_i - b)^2) 2r = \\ &= 4r \left( nr^2 - \sum_i (x_i - a)^2 - \sum_i (y_i - b)^2 \right) \end{aligned}$$

For a minimum then holds:

$$\frac{d}{dr} D = 0 \iff r \neq 0 \quad nr^2 = \sum_i (x_i - a)^2 + \sum_i (y_i - b)^2 \quad (2)$$

<sup>1</sup><http://www.nist.gov/forensics/ballisticsdb/hamby-consecutively-rifled-barrels.cfm>

<sup>2</sup><http://sourceforge.net/p/open-gps/mwiki/X3p/>

<sup>3</sup><http://www.openfmc.org/>

The derivative of  $D$  with respect to  $a$  is:

$$\begin{aligned}\frac{d}{da}D &= 2 \sum_i (r^2 - (x_i - a)^2 - (y_i - b)^2) 2(x_i - a) = \\ &= -4 \left[ a \cdot nr^2 + \sum_i (x_i - a)^3 + \sum_i (x_i - a)(y_i - b)^2 \right]\end{aligned}$$

Using (2) for  $nr^2$  in the equation above we get:

$$\begin{aligned}\frac{d}{da}D &= -4 \left[ \sum_i a(x_i - a)^2 + \sum_i a(y_i - b)^2 + \right. \\ &\quad \left. \sum_i (x_i - a)^3 + \sum_i (x_i - a)(y_i - b)^2 \right] = \\ &= -4 \left[ \sum_i (x_i - a)^2(a + x_i - a) + \right. \\ &\quad \left. \sum_i (x_i - a + a)(y_i - b)^2 \right] = \\ &= -4 \left[ \sum_i (x_i - a)^2 x_i + \sum_i x_i (y_i - b)^2 \right] \quad \sum_i x_i = 0 \\ &= -4 \left[ \sum_i x_i^3 + \sum_i x_i y_i^2 - 2as_{xx} - 2bs_{xy} \right],\end{aligned}$$

where  $s_{xx} = \sum_i x_i^2$ ,  $s_{xy} = \sum_i x_i y_i$  and  $s_{yy} = \sum_i y_i^2$ .

Likewise, we get for the derivative of  $D$  with respect to  $b$ :

$$\frac{d}{db}D = -4 \left[ \sum_i y_i^3 + \sum_i x_i^2 y_i - 2as_{xy} - 2bs_{yy} \right]$$

For the minimum we therefore get a system of two linear equations in  $a$  and  $b$ :

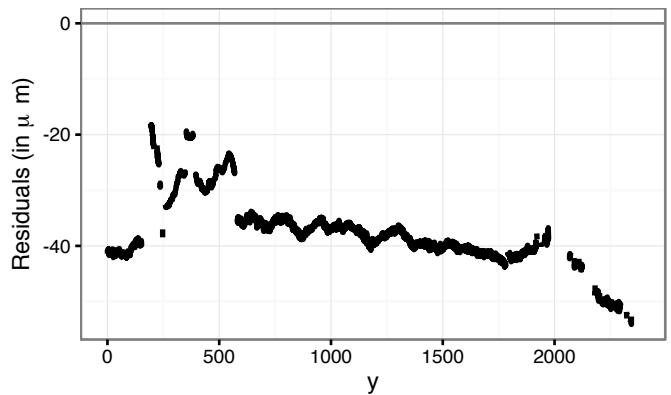
$$\begin{aligned}2s_{xx}a + 2s_{xy}b &= c_1 \quad \text{with } c_1 = \sum_i x_i^3 + x_i y_i^2 \\ 2s_{xy}a + 2s_{yy}b &= c_2 \quad \text{with } c_2 = \sum_i x_i^2 y_i + y_i^3.\end{aligned}$$

This resolves to

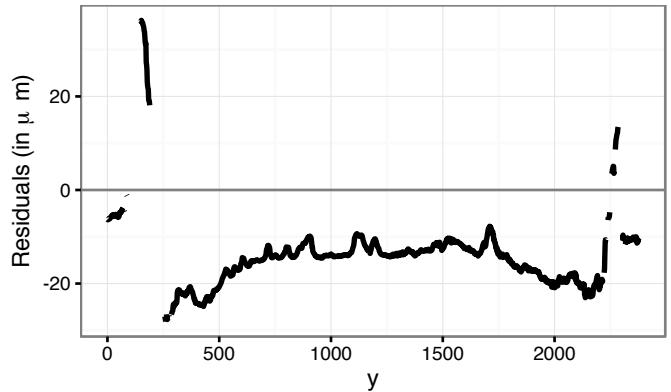
$$\begin{aligned}\hat{a} &= \frac{c_1 s_{yy} - c_2 s_{xy}}{2s_{xx}s_{yy} - 2s_{xy}^2}, \\ \hat{b} &= \frac{c_2 s_{xx} - c_1 s_{xy}}{2s_{xx}s_{yy} - 2s_{xy}^2}, \text{ and} \\ \hat{r}^2 &= \frac{1}{n} s_{xx} + \frac{1}{n} s_{yy} + \hat{a}^2 + \hat{b}^2.\end{aligned}$$

The scatterplot in Figure 3 shows the residuals of such a fit. In this instance, the radius is estimated as  $\hat{r} = 4666.49\mu\text{m} = 4.67\text{mm}$  and the land covers about 29.5 degrees. Both of these estimates are consistent with a 9 mm bullet fired in a Ruger P-85. The residuals are dominated, as expected, by the grooves, which show up as large positive residuals. For

(a) Residual structure of a surface crosscut at  $x = 1.5625$  (bottom of the bullet).



(b) Residual structure of a surface crosscut at  $x = 100.00$



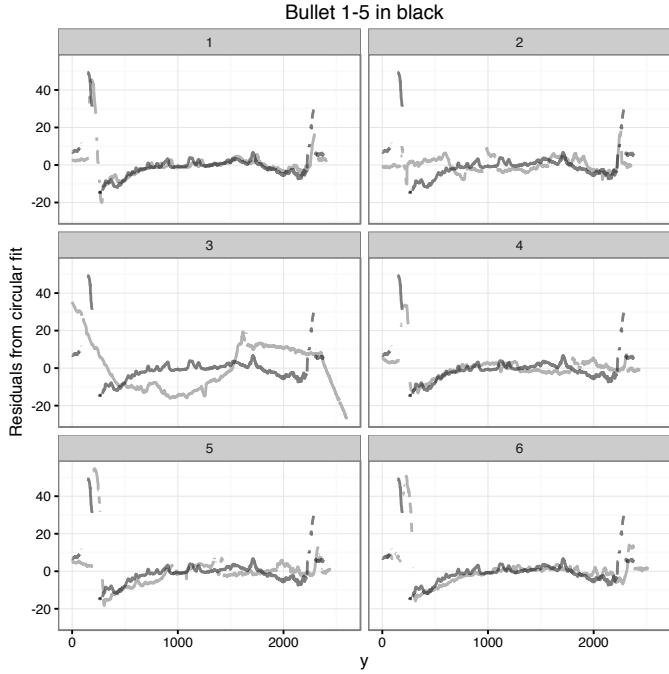
**Figure 3:** Residual structure of circular fits at two different cross-section. Both residual plots show systematic structures, indicating that a circular fit is not entirely appropriate.

a surface cross cut at  $x = 100$  there is a residual circular structure that does not show up for all cross sections.

A single cylinder as a fit is unlikely to be a particularly good fit, because there seem to be quite massive deformations in the vertical direction. Even when we fit a circle to each cross section of the bullet, as in Figure 4, we can not address all of these issues. While the wider circumference at the base of the bullet can be resolved by individual circular fits, the systematic residual structure in Figure 3b stays the same.

## 2.2 Loess Fit

In an approach to improve upon the previous algorithm, we first have to eliminate the grooves from the images under the assumption that they do not contain relevant information for determining a match. Fortunately, the location and appearance of the grooves in these 2d plots is quite consistent. Surface measurements reach local maxima around the peak of the groove at either end of the range of  $y$ , we can then follow the descent of the surface measurements inwards to



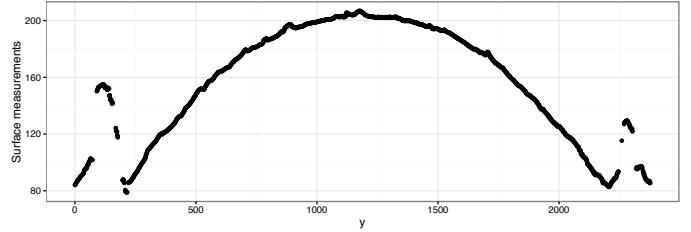
**Figure 4:** Circle fit for each cross section of bullet 2, with bullet 1-5 overlaid. Bullet 1-5 matches best with the bullet 2-1.

the valley of the groove. This marks the point at which we should trim the image. The procedure can be described as follows:

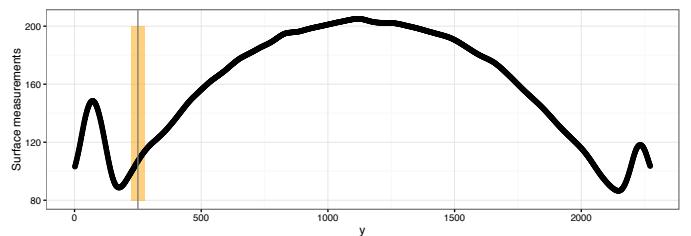
1. Fix an  $x$  value (Figure 5a, with  $x$  set to  $243.75\mu\text{m}$ ).
2. For each  $y$  value, smooth out any deviations occurring near the minima by twice applying a rolling average with a pre-set **smoothing factor**. (Figure 5b, smoothing factor of 35, corresponding to  $55\mu\text{m}$ ).
3. Determine the location of the peak of the left groove by finding the first double-smoothed value  $y_i$  that is the maximum within its smoothing window (e.g. such that  $y_i > y_{i-1}$  and  $y_i > y_{i+1}$ ,  $i = 1, \dots, 17$ ). We call the location of this peak  $p_{left}$  (see Figure 5c).
4. Similarly, determine the location of the valley of the left groove by finding the first double-smoothed  $y_j$  that is the minimum within its smoothing window. Call the location of this valley  $v_{left}$ .
5. Reverse the order of the  $y$  values and repeat the previous two steps to find the peak and valley of the right groove, ( $p_{right}, v_{right}$ ).
6. Trim the surface measurements to values within the two grooves (i.e. remove all records with  $y_i < v_{left}$  and  $y_i > v_{right}$ ). (Figure 5c).

The “smoothing factor” described in the algorithm represents the window size to use for a rolling average. Higher values of the smoothing factor therefore lead to more smoothed

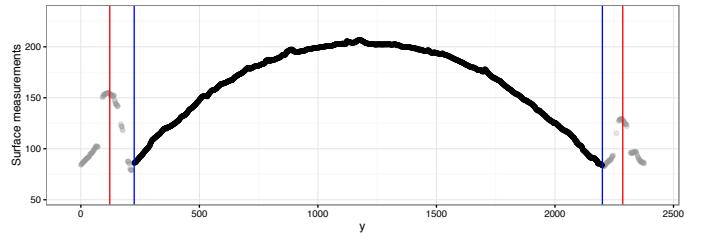
(a) Step 1 the Loess fit algorithm. For a fixed crosscut ( $x$  held constant at  $243.75\mu\text{m}$ ) of surface measurements for bullet 1-5 are plotted across different values of  $y$ .



(b) Step 2 of the Loess fit algorithm. The height values are smoothed twice with a smoothing factor of 35. The orange rectangle shows an example of the smoothing window. Groove and peak will be detected, if they are not within the same window.



(c) Steps 3 – 6 of the Loess fitting algorithm. After smoothing the  $y$  values left and right minima are detected (marked by vertical lines, red indicating peaks and blue indicating valleys). Values outside these boundaries are removed (shown in grey)

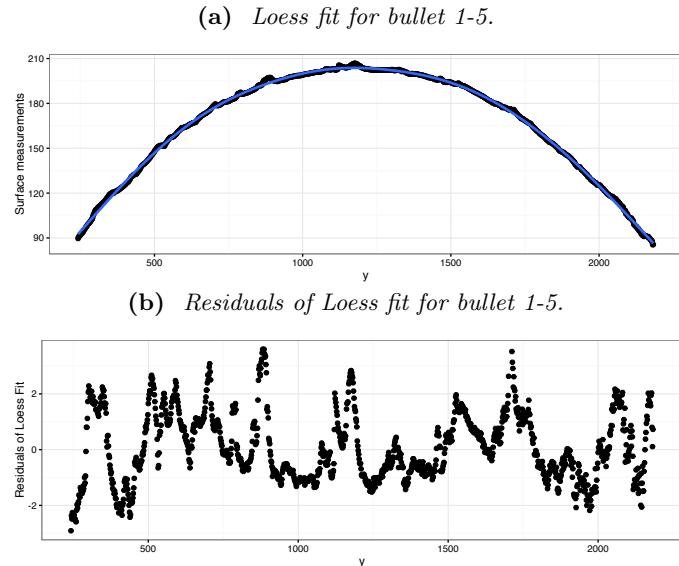


**Figure 5:** Overview of all six steps of the smoothing algorithm to identify and remove grooves from the bullet images.

results. In the groove detection portion of this algorithm, we wish to remove most of the deviation but still maintain enough signal in the groove to detect it. Empirically, a value of 35 for the smoothing factor seems to work quite well (later in the manuscript we investigate varying the smoothing factor). It is important to note that the smoothing pass is done twice. That is, the smoothed data is once again smoothed by computing a new rolling average with the same smoothing factor. This bears some similarities to the ideas of John Tukey in his book Exploratory Data Analysis, where he describes a smoothing process called “twicing” in which a second pass is made on the residuals computed from the first pass and then added back to the result [24]. This has the effect of introducing a little bit more variance into the smoothed data. We instead performed a second smoothing pass on the smoothed data, which has the effect of weighting observations near the center highest,

while the weights linearly drop off as you reach the end of the smoothing window or smoothing factor.

Next, we fit a Loess regression to the data. Figure 6a provides a look at the Loess fit, in blue, overlaid against the processed image of part 1 of bullet 1. The fit seems to do a reasonable job of capturing the structure of the image. Figure 6b shows the residuals from this fit. We can reproduce a similar figure to the one from the circular fits by overlaying the residuals from each loess fit for bullet 2 onto an image of the loess residuals from bullet 1-5. This is shown in Figure 7. Bullet 2-1 is still the best match for bullet 1-5, but now the details of the match are more pronounced.

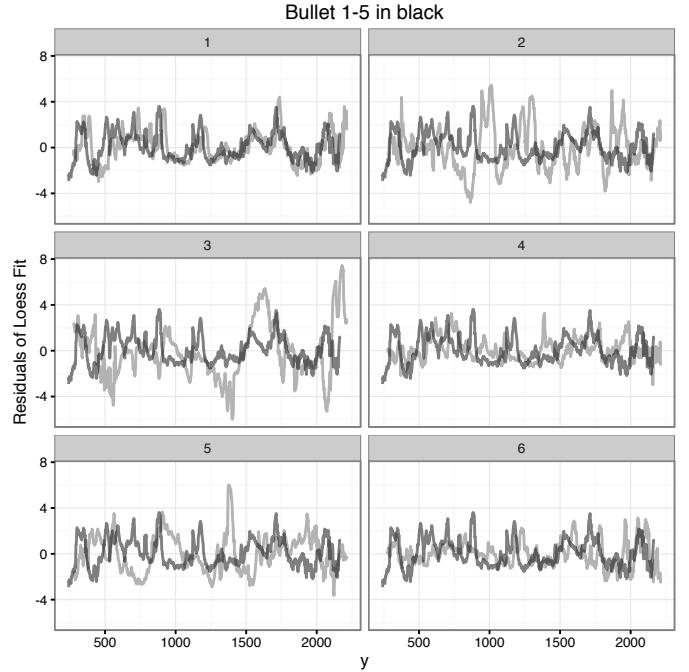


**Figure 6:** Fit and residuals of a loess fit to bullet 1-5 (Barrel 1).

### 3 Web Application

We have implemented a web interface to automate many of the steps outlined in the previous section. The web interface, developed using Shiny [5], allows a user to upload two x3p files representing scans of a bullet between two grooves. A 3D viewer, using Plotly [21], allows the user to rotate and zoom to examine surface features of the images. A screenshot of the application is available in Figure 8. The application is available at the url <http://erichare.shinyapps.io/x3prproto>.

The mouse is the primary control input for interacting with the surface plots. One can mouse over the plots to view the coordinate values for each point. Clicking and dragging anywhere on the plot will allow rotation. Scrolling up and down allows the user to zoom out and in respectively. There are also a set of controls available at the top right of the page, which will allow the selection of some other rotation methods, taking of screenshots, and resetting the camera to the default view.



**Figure 7:** Residuals of a loess fit to bullets 1 and 2 from barrel 1. Bullet 1-5 is shown in black.

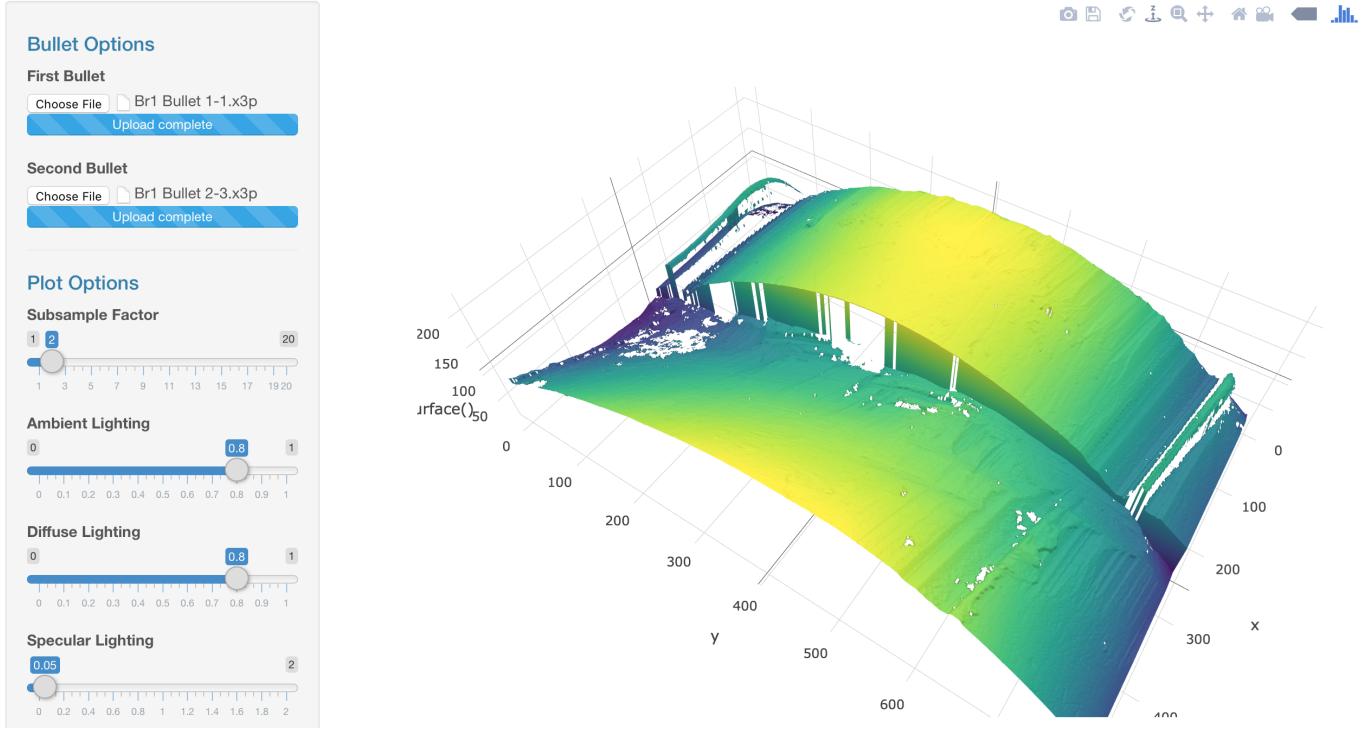
Several lighting options are available to tweak in order to highlight markings in on the bullets more strongly:

1. **Ambient Lighting** - Lighting which is used to simulate the light naturally present in the environment.
2. **Diffuse Lighting** - Setting this value higher will cause the surface to re-emit a higher proportion of light.
3. **Specular Lighting** - The lighting which appears on the surface when illuminated.
4. **Roughness Lighting** - The amount of light reflected by the roughness of the surface.
5. **Fresnel Lighting** - Higher values create a wider, softer lighting across the surface.

The application includes a button which allows the automatic creation of a residual plot from the Loess fits of two bullets. An x coordinate value can be selected as a cross cut, and a cross-section of the two bullet at the x value will be taken. The previously described Loess regression procedure will be applied to the two cross sections. A residual plot will be produced beneath the surface plots allowing for an assessment of the similarity of the markings. It is intended that users will perform this procedure at different x values in order to see whether the markings indicate a match at different points along the bullet.

Enhancements to the web application are still in progress. In particular, much of the automatic matching functionality described in the next section will be added in the future.

# Bullet Rotation Prototype



**Figure 8:** Screenshot of the web interface for analyzing bullet images.

## 4 Approach to automatic matching

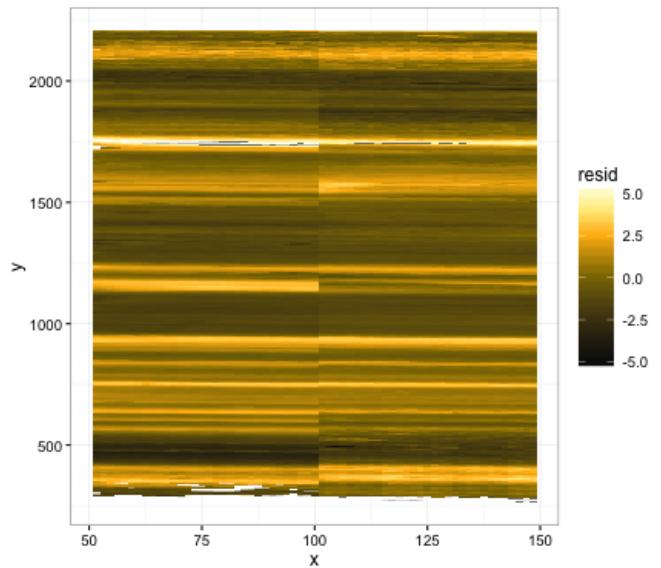
Applying the loess fit to a range of different surface cross cuts (see Figures 9 and 10, with cross cuts from 50 to 150) shows the striation marks across two bullets. Cross cuts of bullet 1 are shown on the left (cross cuts with values below 100) and cross cuts of bullet 2 are shown on the right (cross cuts with values above 100). A lot of the striation marks align, allowing for an easy visual assessment of the shared similarities between the two bullets.

### 4.1 Horizontal Alignment

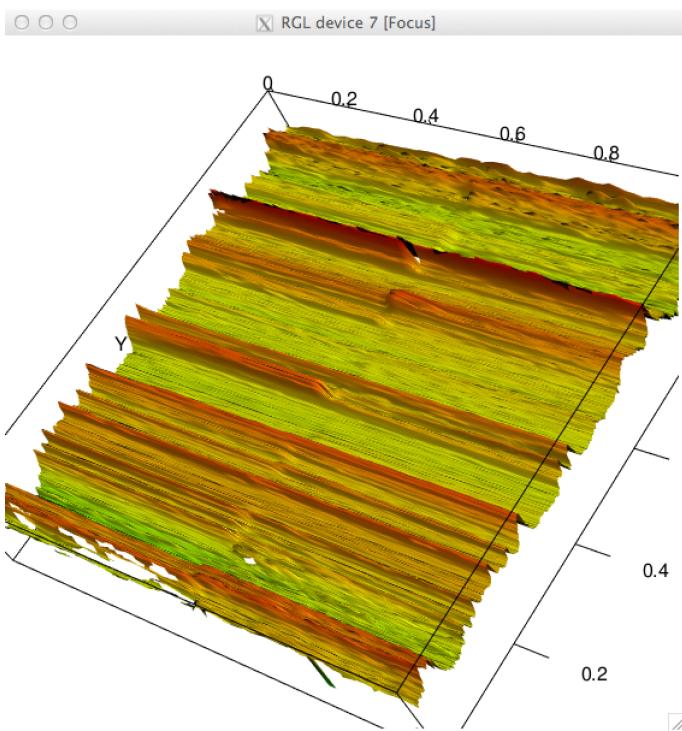
Using (for now) a single crosscut across the surface, we extract the values and remove the class characteristics using a loess fit as before. This enables us to focus on the individual characteristics. Striation marks show up in this representations as peaks and valleys. Figure 11 shows lines of the surface crosscut at  $x = 100$ . A horizontal shift of one of the lines (result shown in (b)) emphasizes the strong similarities. This horizontal shift is based on the cross-correlation between the two surface crosscuts. Let  $f(t)$  and  $g(t)$  define the values of the crosscut at  $t$ , then the cross-correlation between  $f$  and  $g$  at lag  $k$  is defined as

$$(f * g)(k) = \sum_t f(t+k)g(t),$$

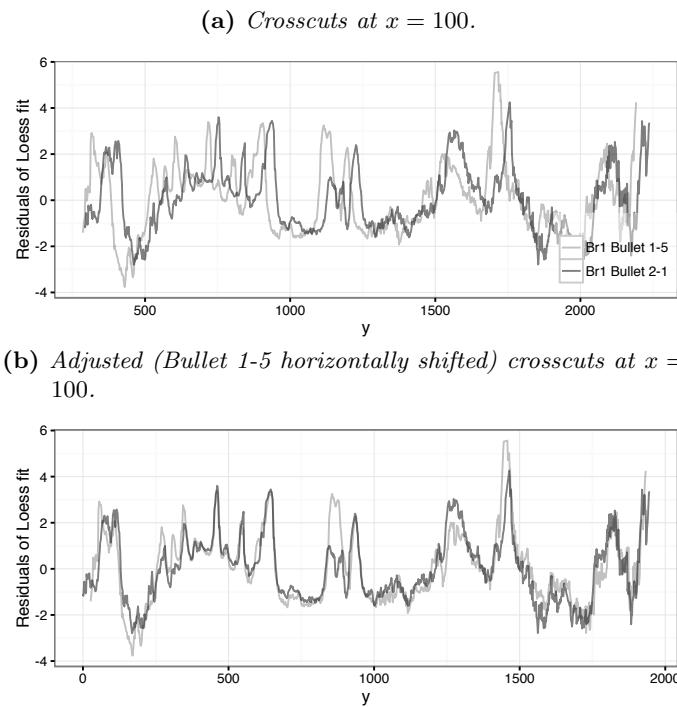
with suitable defined limits of the summation.



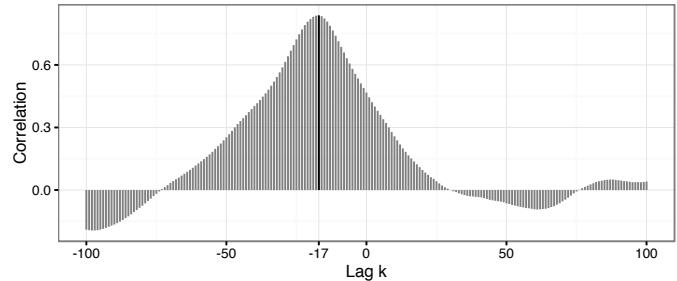
**Figure 9:** Manually adjusted side-by-side comparison of land five of bullet 1 and land 1 of bullet 2.



**Figure 10:** Different view of the manually adjusted side-by-side comparison of land five of bullet 1 and land 1 of bullet 2.



**Figure 11:** Lines showing surface crosscuts at  $x = 100$ . A horizontal shift of the values of Bullet 1-5 to the right shows the similarity of the striation marks.

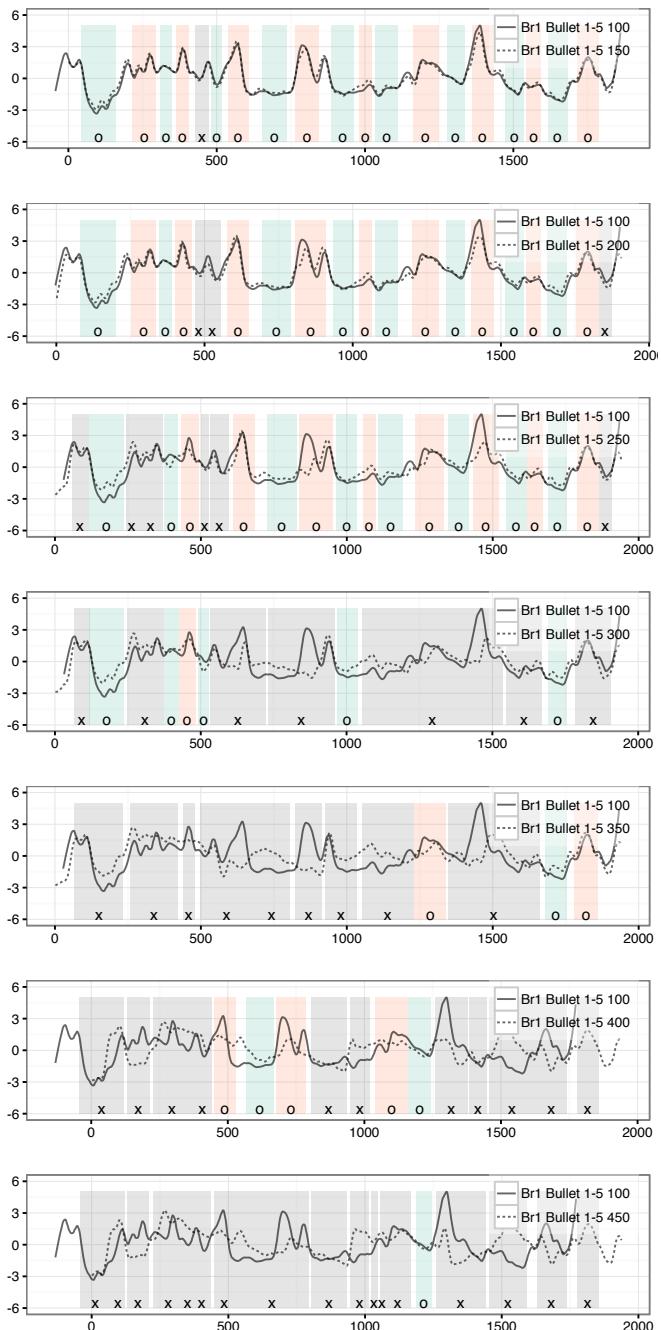


**Figure 12:** Cross-correlation between the two crosscuts shown in Figure 11a at lags between -100 and 100. At a lag of -10 the correlation peaks, indicating the largest amount of agreement between the crosscuts. Figure 11b shows the lag-shifted crosscuts.

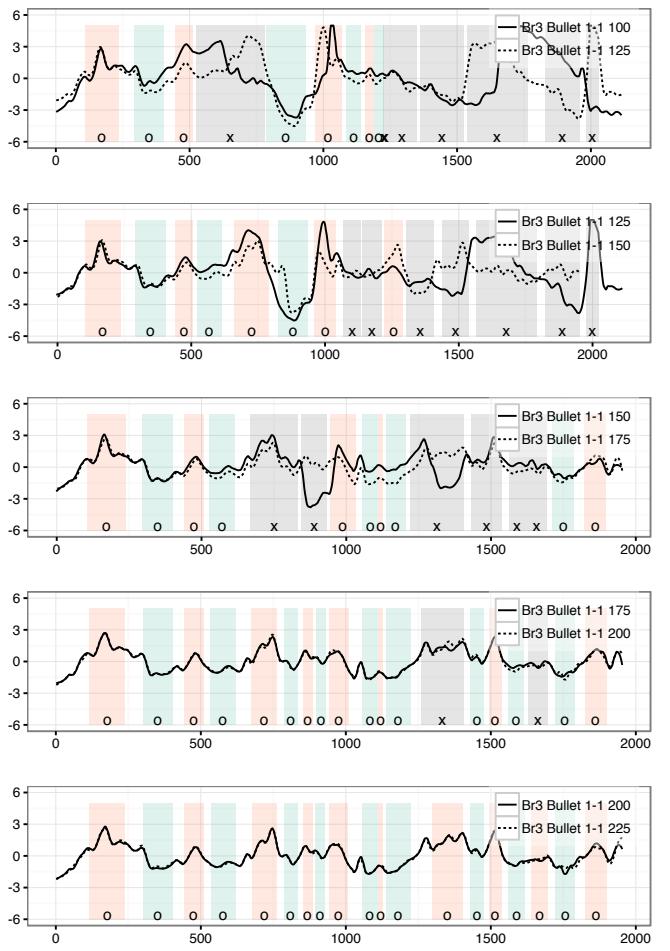
## 4.2 Varying crosscuts

The crosscut at which a comparison is done matters – as crosscuts are further apart, more pronounced differences between the crosscuts show up. This poses both a caveat to matching attempts as well as an opportunity: on the one hand we have to be aware of which crosscut level was used in a matching, but on the other hand we can use multiple crosscuts of the same bullet for an initial assessment of its quality. Both of these situations are assessed and discussed in more detail below: Figure 13 shows a sequence of crosscuts for bullet land 1-5 (barrel 1) that are taken  $50\mu\text{m}$  apart, between  $150\mu\text{m}$  and  $450\mu\text{m}$ . These are compared to a crosscut at  $100\mu\text{m}$ . Initially this comparison constitutes an almost perfect match (consolidated by a CMS of fourteen) between the two crosscuts. However, the match quickly deteriorates with increasing distance to non-matches at a crosscut of  $x = 300\mu\text{m}$  or higher. Only if the crosscut is within  $150\mu\text{m}$  do we get a good visual match even when we know that the same bullet surface is being used. Given that we have to expect some variation in nominally the same crosscut levels due to manual alignments in microscopes, we should take crosscut levels into account in the automatic matching routine by evaluating matches at several cross sections.

An opportunity that arises from comparing multiple crosscuts of a bullet's surface measurements is the following: by comparing cross sections that are not too far apart – maybe 25 or  $50\mu\text{m}$  for example – we get an indication whether the crosscut is in a rapidly changing section of the surface, indicative of a break-off, or in a stable section, where we can expect to find matches to other surfaces reasonably well. In the approach here, we keep increasing the crosscut level until we find a section with a stable pattern. This process is shown in Figure 14 at the example of land 1-1 from barrel 3. ‘Stability’ is here defined as two aligned crosscuts chosen  $12.5\mu\text{m}$  apart having a cross-correlation of at least 0.9.



**Figure 13:** Overview of the variations in the surface along different crosscuts. Crosscut at  $x = 100$  is compared to crosscuts every  $50\mu\text{m}$ . With every step away from the original crosscut site, the number of differences between the crosscuts increases, and the number of maximum CMS decreases from initially 14 to 3 or fewer at  $x = 300\mu\text{m}$  and higher.



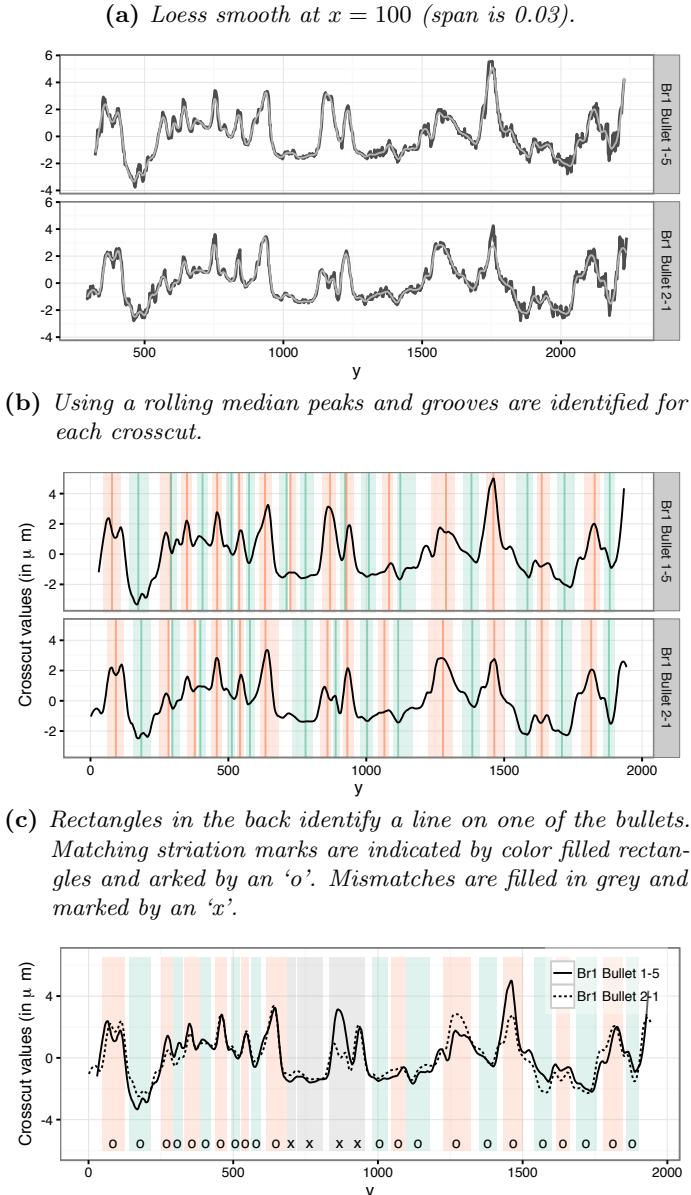
**Figure 14:** Varying crosscuts for barrel 3, bullet 1. Initially, the match between crosscuts  $25\mu\text{m}$  apart is affected strongly by break off at the bottom of the bullet. At a crosscut level of  $175\mu\text{m}$  the bullet pattern stabilizes. For this land, matches should not be attempted at lower crosscuts.

### 4.3 Varying smoothing factor

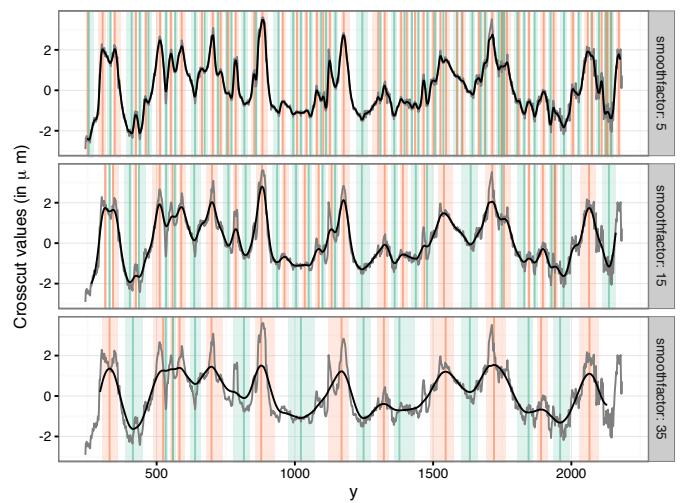
As mentioned earlier, the algorithm for peak/valley detection depends on the selection of a smoothing window, called the smoothing factor. In particular, for detecting and removing the grooves prior to fitting a Loess regression we selected a smoothing factor of 35, while for detecting the peaks/valleys of the loess residuals a smoothing factor of 15 seems more appropriate. We can investigate the performance of peak/valley detection across different smoothing factors. Figure 16 displays the resulting peaks and valleys across smoothing factors of 5, 15, and 35, respectively. The dark line corresponds to the smoothed values, while the grey line in the back shows the unsmoothed values of the surface measurements. The choice of smoothing factor is a classical decision of a bias/variance trade-off. It is immediately clear that a small smoothing factor like 5 is a poor choice. It results in a significant amount of noise in the data such that

even just a point or two can skew the rolling average enough for a peak or valley to be detected. Given that the striation patterns are typically much larger, we are in effect muddying the waters by performing such minimal smoothing.

A larger smoothing factor on the other hand, like 35, seems to be a more plausible option. Most of the peaks/valleys present which are detected by a smoothing factor of 15 are also detected at 35. However, some notable issues arise. Notice that the valley on the right hand side of the image is smoothed out, and thus not detected. On the left hand side, a double peak is detected - that might be a questionable decision - but there are several peaks in the middle, that are being smoothed out, for example the peak at around  $y = 750$  or the triple-peak pattern around  $y = 1,100$ . That is, in many cases, large windows are smoothing out some of the structure that we wish to see. Furthermore, it can be seen that the peaks/valleys are often shifted relative to their position in the original loess residuals, or in the smoothed data with smaller smoothing factors.



**Figure 15:** Matching striation marks: smooth (a), identify peaks and valley (b), and match peaks and valleys between crosscuts (c).



**Figure 16:** Peak/valley detection at smoothing factors of 5, 15, and 35, respectively. Note that a smoothing factor of 5 yields enough noise that many very minimal overlapping peaks and valleys are detected, while a smoothing factor of 35 might over-smooth and cause the peaks/valleys to either disappear or be horizontally shifted from their position in the original loess residuals.

#### 4.4 Automated matching algorithm

Figure 15 gives an overview of the automated matching routine: We first identify a stable region for each bullet and pick the lowest crosscut level in this region, because typically, individual characteristics are best express at the lower end of the bullet.

All of the other steps are done on pairs of bullet lands:

- Smooth the two crosscuts by a loess with a very small span (see Figure 15a).

2. Use cross-correlation to find the best alignment along the crosscuts: Shift one of the crosscuts by the lag indicated by the cross-correlation function (see Figure 12 for the cross-correlation function and Figure 11b for the resulting shift).
3. Using a rolling average, identify peaks and grooves for each of the crosscuts. We define a region around the extrema on each side as one third of the distance to the next extrema (see Figure 15b).
4. Match grooves and peaks between crosscuts: based on the regions around the extrema as defined above, we identify joint regions between two crosscuts as those areas, in which at least two of the individual regions touch. A joint region is defined as the smallest interval that covers all of the touching regions. A joint region is called a match between the crosscuts, if all of the crosscut-specific regions are of the same type of extrema, i.e. either all peaks or all valleys. In Figure 15 all matches are colored corresponding to their type of extrema. Non-matching regions are left grey.
5. Identify the maximal number of CMS (consecutive matching striae) between the two crosscuts by counting consecutive matching lines. In the example of Figure 15, the number of consecutive matching striations (CMS) is fifteen, a high number indicative of a match between the bullets. For lead bullets, such as this, Biasotti [2] considered four or more consecutive lines to be sufficient evidence of a match, and which are part of current practice in assessing bullet matches [17, 18, ?].

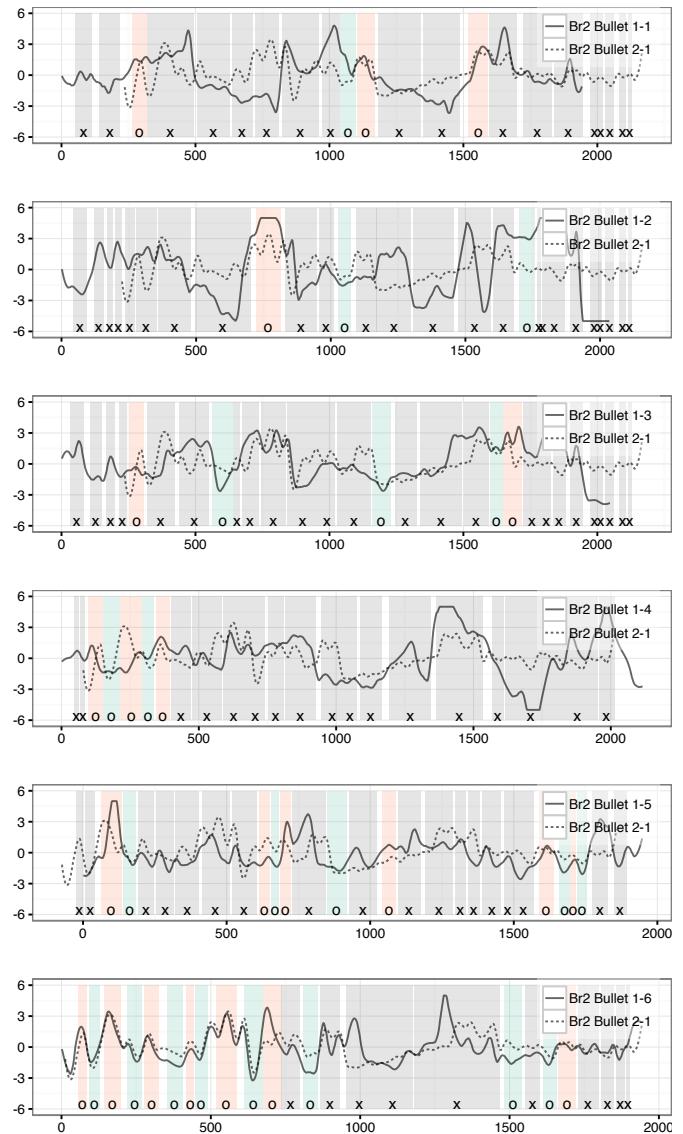
In order to get a better understanding of how this matching works in known matches and non-matches, we next investigate the algorithm's performance in a test scenario.

**XXX what would be good is to get the line identifications and project them back onto the surface.**

## 5 Evaluation

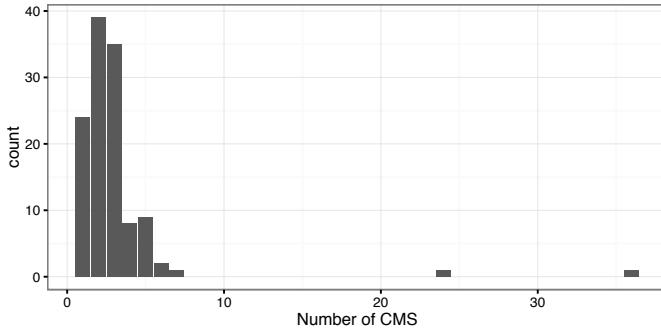
Using the James Hamby study, we run all unknown bullets against all known bullets for matches. The algorithm is using land-to-land matches, i.e. we are running  $15 \times 6 = 90$  lands from unknown bullets against  $2 \times 10 \times 6 = 120$  lands from known bullets, yielding a total of 10,800 comparisons with 180 known land-to-land matches. When things go perfectly, they look like the results in Figure 18: Figure 18a shows the number of maximum consecutive matching striae between land D-5 and all 120 known lands. Two lands show a high CMS. These correspond to the known matches, shown in Figures 18b and c. Unfortunately, not all results are as clear cut. It is not reasonable to assume that we can match all lands, but we can try to maximize the number of matches to get an overview of what we can reasonably expect from an automated match.

Figure 19 shows the strong connection between the maximal number of consecutive striae and matches in the Hamby

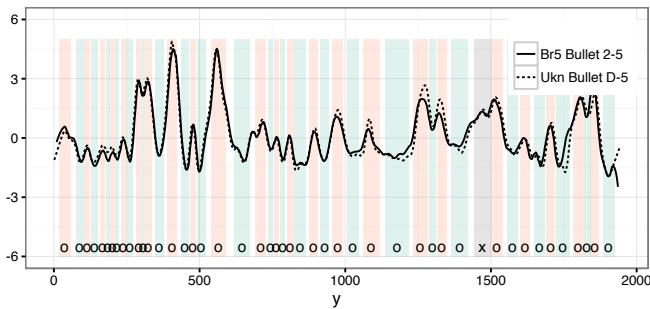


**Figure 17:** All lands from the first bullet shot from barrel 2 are compared to the first land of the second bullet shot from barrel 2. The largest number of CMS is eleven, observed with land six of the first bullet.

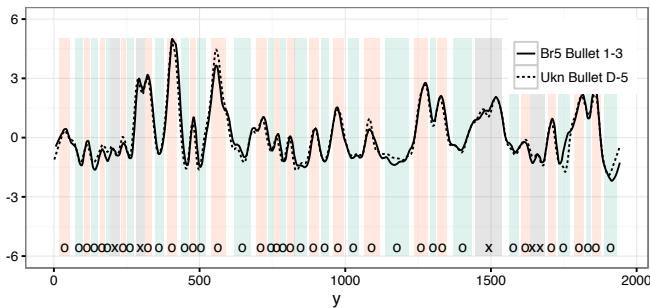
- (a) Maximal number of CMS between unknown Bullet D land 5 and all of the 120 other (known) lands. For two lands the number of maximum CMS is high.



- (b) Overlaid lines of D-5 and the land with the top matching CMS.

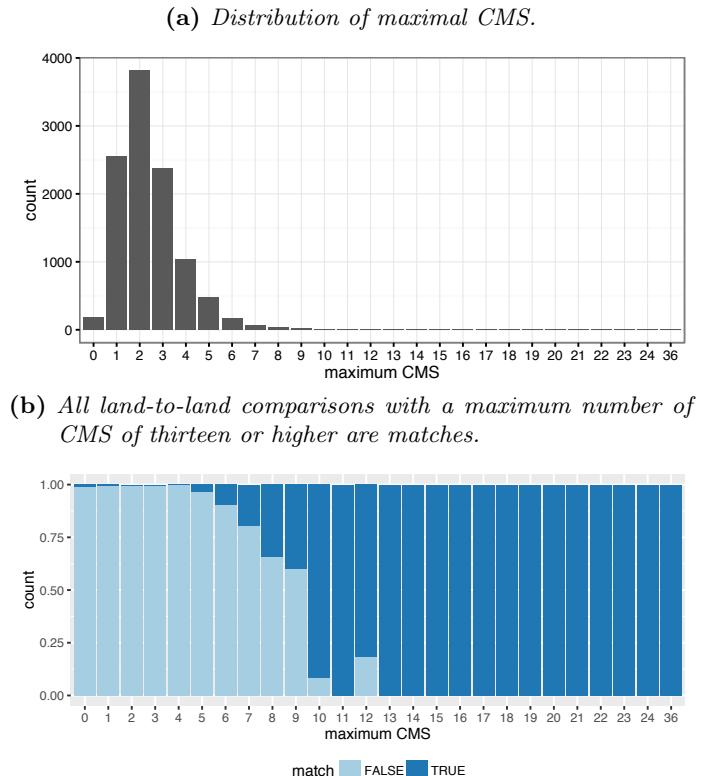


- (c) Overlaid lines of D-5 and the land with the second top matching CMS.



**Figure 18:** Showcase scenario when matching with CMS works perfectly. Unfortunately the matches are not always that convincing.

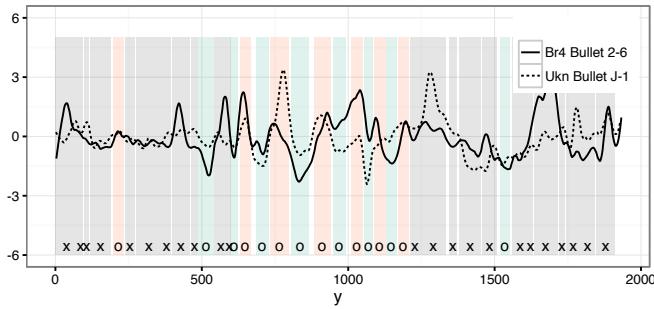
study. All 36 pairs of lands with at least sixteen CMS in common are matches. There are two things that should be noted at this point: the automated algorithm finds a relatively high number of CMS even for non-matches. On average, there are 2.38 maximal CMS between known non-matches (with a standard deviation of 1.3. Known matches share on average 8.31 maximal CMS, with a standard deviation of 5.9. Secondly, a large number of maximal CMS by itself is not indicative of a match, as was previously pointed out by Miller [15]. Figure 20 shows a known mismatch, between two lands that share twelve consecutively matched striae. Visually we can easily tell these two lands do not match well.



**Figure 19:** Overview of the maximal number of consecutive maximal striae compared to matches.

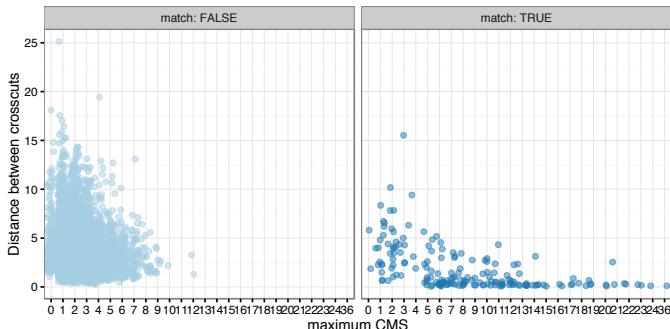
For smaller numbers of CMS, the percentage of false positives quickly increases. However, if we take other features of the image into account, we can increase the number of correct matches considerably: a high amount of cross-correlation between the two crosscuts (see Figure 21b) is indicative for a match; e.g. in the Hamby study only known matches have a cross-correlation of 0.75 or higher. There are 97 land-to-land comparisons with a cross-correlation of above 0.75.

Similarly, the Euclidean distance between two crosscuts across the range of loess residuals is strongly correlated to known matches. Out of the 48 pairs of land-to-land crosscuts with a distance of less than .25, 47 correspond to known matches (the single known mismatch consists of

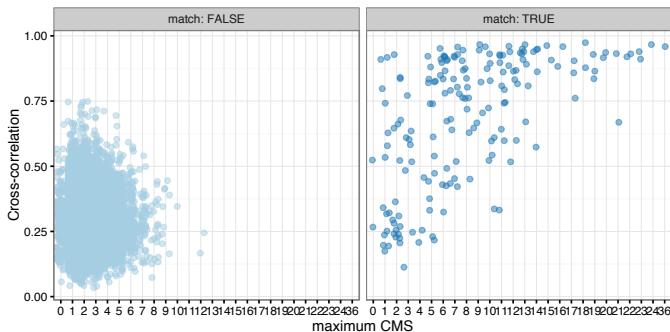


**Figure 20:** Known mismatch with a relatively large number of maximal consecutive matching striae (twelve) in the middle. Visual inspection easily reveals this as mismatch.

(a) Relationship between maximal CMS, distances between cross-cuts and matches.



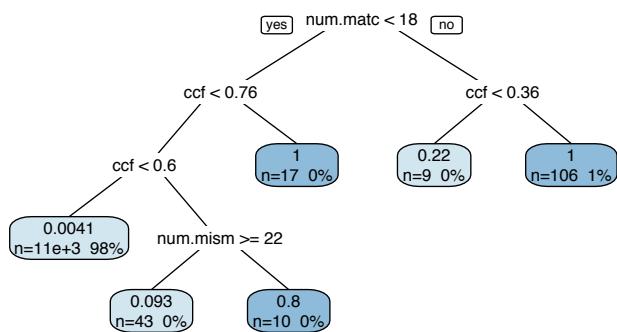
(b) Relationship between maximal CMS, cross-correlation between crosscuts and matches.



**Figure 21:** A small distance between crosscuts or a large cross-correlation between crosscuts are indicative of matches between lands.

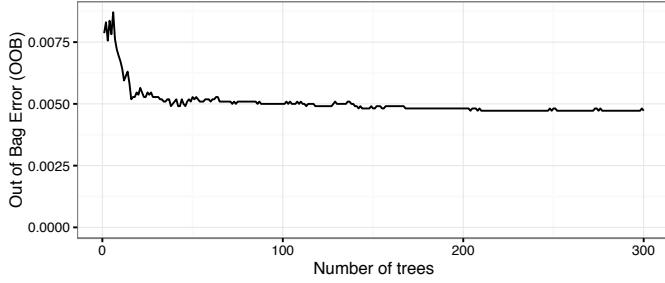
a pair of crosscuts that both are very flat with peaks and valleys of less than  $1.5\mu m$ ). The average distance between known matches in the Hamby study is  $1.86\mu m$  with a standard deviation of  $2.31\mu m$ ; while known mismatches have an average distance of  $3.27\mu m$  with a standard deviation of  $1.93\mu m$ . Both of these additional variables show large, if not significant, differences between matches and non-matches. Table 1 shows an overview in the averages and corresponding standard deviations of a set of feature variables derived from the 3d topological surface measurements. There are quite large differences in these averages, indicating that we can successfully employ machine learning methods to distinguish matches from non-matches.

Using recursive partitioning, we fit a decision tree [4, 22, 14] to predict matches between lands based on features derived from the image files. The resulting tree is shown in Figure 22. A total of 123 lands is being matched correctly. Interestingly, the number of consecutive matching striae does not feature in this evaluation. Instead of CMS, cross-correlation (ccf) between the crosscuts is very important in the matching process by the decision tree. Aside from cross correlation, the number of matches (num.matches) and mismatches (mis.matches) is involved in the decision. Within matches and non-matches, the values of CMS and cross-correlation seem to be almost independent (see Figure 21b), but both are highly predictive in distinguishing matches from non-matches. Out of the two, cross-correlation has higher predictive power. This does not contradict earlier findings emphasizing the value of CMS on visual assessments of bullet matches: in those papers, assessments were based on purely visual inspection of either actual bullets or 2d microscopic images of bullets. Neither one of these methods allows for an assessment of cross-correlations. This is one of the benefits of switching to a digitalized version of the images that preserves the 3d surface structure.



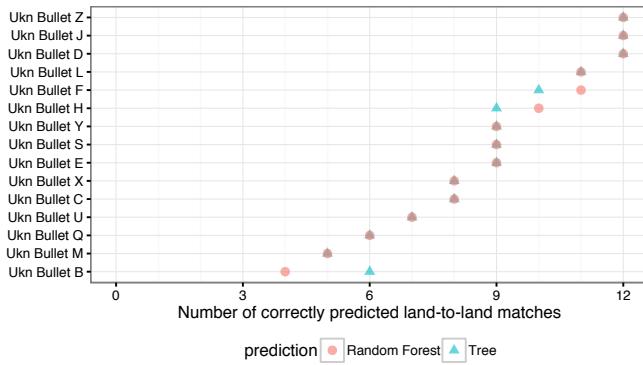
**Figure 22:** Decision tree of matching bullets based on recursive partitioning. The rectangular nodes are the leaves, giving a short summary consisting of the number of observations in the leaf (bottom left), the corresponding percentage of the total (bottom right). The number at the top shows the fraction of these observations that are a match. A 1 or a 0 therefore indicate a homogeneous (or perfect) node.

Another benefit is that we can not only apply a single decision tree, but several hundred of them to combine in a random forest [3, 13]. In a random forest multiple trees are fitted. For each of these trees, only two thirds of the observations are used for fitting, while predictions for the remaining third are used to evaluate the tree's predictive power and accuracy, or its reverse, the error rate. Because errors are determined from the one third of held-back observations, this error rate is called the out-of bag (OOB) error. Figure 23 shows the cumulative out-of-bag error (OOB) rate for 300 trees. After about 25 trees, the error



**Figure 23:** Cumulative out-of-bag error rate of a random forest fit to predict land-to-land matches from image features.

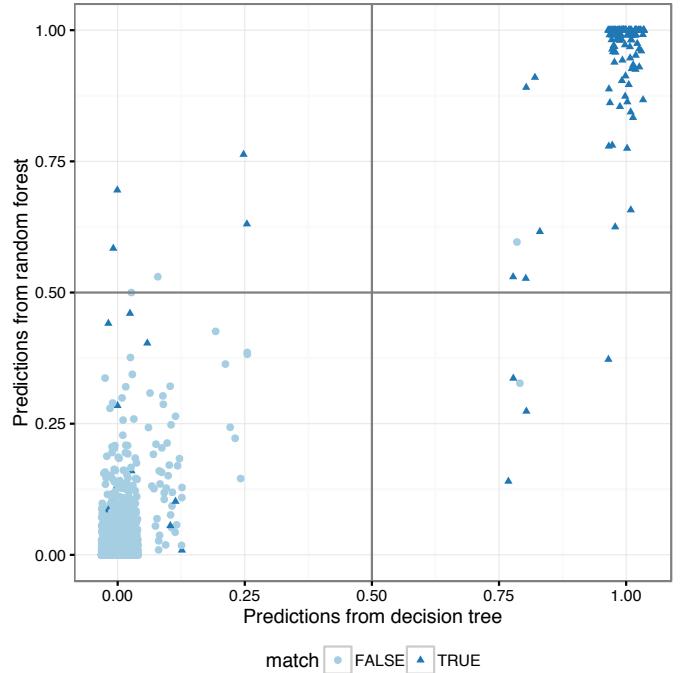
rate of land-to-land comparisons stabilizes at 0.0049. This is a weighted average between false positives of 0.00028 and false negatives of 0.2778. Note that these error rates are based on land-to-land comparisons and will be much lower for bullet-to-bullet matches. In the case of the Hamby study, the error rate is zero, if we require for a bullet match that at least two of a bullet's lands are matched (see Figure 24). However, this does not carry a lot of weight because of the



**Figure 24:** Overall predictions for each bullet based on the decision tree (maximum possible number is 12). Both results for the tree and the random forest are shown. For each bullet at least four lands are correctly matched. Unknown bullets B and M are the hardest for the algorithm to match.

few bullet comparisons in question: fifteen unknown bullets are successfully matched to two pairs of ten bullets. Matching bullets can only be tested realistically in a much bigger

test case. Another thing to note about the random forest's error rates is that they are based on probability cutoffs of 0.5, i.e. whenever the predicted probability of a match exceeds 0.5, a match is declared. Basing this decision on 0.5 is not necessary. In practice, examiners are allowed a third option of 'inconclusive'. On a probability spectrum of outcomes we can therefore introduce an interval of 'inconclusive' results in the middle of the spectrum. Figure 25 shows a comparison of the predicted probabilities of a match by the tree (x-axis) and the random forest (y-axis). Overall, there is a lot of agreement; out of 10,800 land-to-land comparisons, only 12 cases result in a different decision.



**Figure 25:** Scatterplot of prediction results from the tree (horizontal) and the forest (vertical). Each point corresponds to a land-to-land comparison. Known matches are indicated in blue. Matches predicted by the decision tree are on the right half of the plot, matches predicted by the random forest are in the upper half.

Besides a probabilistic quantification of matches, we get from a random forest an assessment of the importance of each of the features derived from the bullets' 3d topological surface measurements. Table 1 shows an overview of the importance of each variable measured as the mean decrease in the gini index when the variable in question is included in a tree.

The rows in the table are ordered according to descending importance. The variable with the most predictive power by far is cross-correlation (ccf), followed by the number of matches (num.matches) and the sum of the height of joint peaks (sumpeaks). The maximal consecutively matching striae (cms) is found only in sixth position.

	Variable	Importance	Match	(sd)	Non-Match	(sd)
1	ccf	93.11	0.70	( 0.25)	0.30	( 0.10)
2	num.matches	74.58	19.10	(10.79)	5.40	( 2.87)
3	sumpeaks	43.09	23.20	(13.41)	6.60	( 3.93)
4	num.mismatches	42.19	17.60	( 8.91)	29.80	( 5.38)
5	distr.dist	27.45	1.90	( 2.31)	3.30	( 1.93)
6	cms	19.06	8.30	( 5.90)	2.40	( 1.30)
7	non_cms	15.66	7.40	( 6.20)	14.20	( 5.89)
8	distr.sd	15.04	2.80	( 0.80)	2.80	( 0.70)
9	lag	11.12	200.80	(49.88)	197.90	( 113.55)
10	x2	5.72	175.00	(71.59)	175.00	( 71.40)
11	x1	5.45	166.50	(81.07)	163.50	( 72.30)

**Table 1:** Table of features derived from bullet image ordered by importance in predicting matches. Importance is measured in terms of mean decrease in gini index when including the variable in a decision tree. Match and Non-Match show averages of each of the variables for matches and non-matches only. In parentheses behind the means, standard deviations are given

## 6 Conclusion

We have presented an algorithm which detects the most prominent but least relevant structure of a bullet from a ballistics identification perspective, removes these features, and produces residuals which allows for the easy identification of markings. We have generalized this algorithm to align the residuals from two bullets to automatically determine whether they are matches. We created a random forest model using many of the features to provide a probabilistic assessment of the quality of a match, along with the most relevant features. Finally, we have provided an easy-to-use open-source backend, and web frontend to many of these features.

The matching algorithm is sensitive to the parameter choices made. The distance between crosscuts (currently  $25\mu m$ ) to evaluate stability, as well as the cross-correlation factor (currently 0.9) we set as a minimum are affecting the final outcome. Another parameter is the amount of smoothing in identifying peaks and grooves (currently a window of  $23.4375\mu m$  is used, corresponding to a window of 7 values to the left and the right of an observation). While we tried to lay out in the paper the impact each of the parameter choices has on the matching, we are still far from an optimized scenario.

The Hamby study serves as our evaluation ‘database’. It consists of only 35 bullets – this is obviously not a particularly realistic scenario for automatic matching. But this is what at the moment is available to us. As the NIST database expands, this might get into more challenging territory.

The feasibility of a database of ballistic images was evaluated in a 2008 report [6] by the National Research Council. The evaluation was investigating the scalability of NIBIN (National Integrated Ballistic Information Network), which uses proprietary matching algorithms provided by IBIS. The bottomline of the report was that in spite of the many technical and practical hurdles, solutions to all but one problem could be found. The remaining problem is simply that statistically the quality of the matching algorithm could not

withstand a hugely increased number of records while still maintaining a reasonable workload for forensic examiners, who have to examine possible matches suggested by the system. The findings of the NRC report on ballistic imaging are based on two-dimensional greyscale images, which the committee argued were not reliable enough for distinguishing extremely fine marks. The report also found, that results from 2d images are improved upon by matches based on 3d images. This is consistent with the importance of features found here: out of the top six features in Table 1, only the number of matches and mismatches are available for a match based on 2d features.

By suggesting an automated algorithm that first identifies lands, then identifies peaks and valleys on this land, we reduce subjectivity and with it possible sources of bias. [T]he concept of counting striations is subjective and based on experience, but it does tend to be consistent within a single examiner’ [15].

For a fair assessment of the adequacy of an algorithm we need transparency. Our matching algorithm is open: the code is readily available in form of the R package x3rplus [9], and the code to produce this paper is available at <http://www.github.com/erichare/imaging-paper>. Obviously, this is only a first step, but in times of the Netflix prize [12, 23, 19] and the many Data Science challenges hosted by Kaggle<sup>4</sup>, we should not let the quality of the matching algorithm be the problem that stands in the way.

Only a much larger database and transparent communication of results will give us answers to whether automatic matching of ballistic evidence is really feasible. Whether there really is individualization of this evidence, as is often silently assumed. And in the case, that individualization is a reasonable assumption, whether there is a way to be able to distinguish between individual and subclass characteristics.

An approach that might help in determining this difference comes from Biometrics, is suggested by Jain et al [11, 10] and has been successfully applied in the Aadhar project (the effort by the Unique Identification Authority of India

<sup>4</sup>kaggle home:<https://www.kaggle.com/solutions/competitions>

(UIDAI) to provide a unique 12-digit identification number to approximately 1.2 billion residents of India based on ten fingerprints and two irises). This approach uses dictionary learning based on small square patches of 64x64 or 32x32 pixels. For matching bullets, a simpler, one-dimensional window approach might be sufficient to match striation marks because of the physical forces leading to the existence of these marks.

**Acknowledgment** Thanks to the men and women behind the software R [20], and the authors of the R packages knitr [26] and ggplot2 [25].

This work was funded in part by the Center for Statistics and Applications in Forensic Evidence (CSAFE).

## References

- [1] AFTE Criteria for Identification Committee. Theory of identification, range striae comparison reports and modified glossary definitions afte criteria for identification committee report. *AFTE Journal*, 24(2):336–340, 1992.
- [2] A. A. Biasotti. A statistical study of the individual characteristics of fired bullets. *Journal of Forensic Sciences*, 4(1):34–50, 1959.
- [3] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [4] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Inc., 1984.
- [5] W. Chang, J. Cheng, J. Allaire, Y. Xie, and J. McPherson. *shiny: Web Application Framework for R*, 2015. R package version 0.12.2.
- [6] Committee to Assess the Feasibility, Accuracy and Technical Capability of a National Ballistics Database; Committee on Law and Justice; Committee on National Statistics; National Materials Advisory Board; Division of Behavioral and Social Sciences and Education; Division on Engineering and Physical Sciences; National Research Council. *Ballistic Imaging*. The National Academies Press, Washington, DC, 2008.
- [7] P. C. Giannelli. Ballistics evidence under fire. *Criminal Justice*, 25(4):50–51, 2011.
- [8] J. E. Hamby, D. J. Brundage, and J. W. Thorpe. The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries. *AFTE Journal*, 41(2):99–110, 2009.
- [9] H. Hofmann and E. Hare. *x3prplus: Simple Models for x3d surface topology*, 2015. R package version 0.0.1.
- [10] A. K. Jain and K. Cao. Fingerprint image analysis: Role of orientation patch and ridge structure dictionaries. In I. L. Dryden and J. T. Kent, editors, *Geometry Driven Statistics*. John Wiley and Sons, New York, 2015.
- [11] A. K. Jain and A. Ross. Bridging the gap: From biometrics to forensics. *Philosophical Transactions of The Royal Society B*, 370(1674), 2015.
- [12] Y. Koren. The BellKor Solution to the Netflix Grand Prize, 2009.
- [13] A. Liaw and M. Wiener. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002.
- [14] S. Milborrow. *rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'*, 2015. R package version 1.5.3.
- [15] J. Miller. Criteria for identification of toolmarks. *AFTE Journal*, 30(1):15–61, 1998.
- [16] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, DC, 2009.
- [17] R. G. Nichols. Firearm and toolmark identification criteria: A review of the literature. *Journal of Forensic Sciences*, 42(3):466–474, 1997.
- [18] R. G. Nichols. Consecutive Matching Striations (CMS): Its Definition, Study and Application in the Discipline of Firearms and Tool Mark Identification. *AFTE Journal*, 35(3):298–306, 2003.
- [19] M. Piotte and M. Chabbert. The Pragmatic Theory Solution to the Netflix Grand Prize, in: Netflix prize documentation, 2009.
- [20] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.
- [21] C. Sievert, C. Parmer, T. Hocking, S. Chamberlain, K. Ram, M. Corvellec, and P. Despouy. *plotly: Create interactive web-based graphs via plotly's API*. R package version 2.0.2.
- [22] T. Therneau, B. Atkinson, and B. Ripley. *rpart: Recursive Partitioning and Regression Trees*, 2015. R package version 4.1-10.
- [23] A. Töscher, M. Jahrer, and R. M. Bell. The BigChaos Solution to the Netflix Grand Prize, 2009.
- [24] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.
- [25] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.
- [26] Y. Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. ISBN 978-1498716963.