# A Robust Approach to Automatic Groove Identification in 3D Bullet Land Scans

Kiegan Rice, M.Sc., Heike Hofmann, Ph.D., Ulrike Genschel, Ph.D.

**Abstract**

Forensic firearms examiners analyze bullets through a process of visual feature comparison to determine whether two bullets originate from the same source. Striation marks found on land engraved areas (LEAs) provide evidence to address this same source-different source problem. Advances in technology have led to an increase in research focused on applying image-analysis algorithms to the automated, quantitative analysis of bullet evidence. One prominent example is an algorithm developed by [Hare et al., 2017] based on 3D imaging data of LEAs. The currently accepted best practice for collection of 3D images of bullet LEAs requires capturing portions of the neighboring groove engraved areas (GEAs). Analyzing LEA and GEA data separately is imperative to achieve high accuracy and precision in subsequent feature calculations. However, existing standard statistical modeling techniques fall short when applied to the atypical structure of 3D bullet data, often failing to adequately separate LEA and GEA data. We propose a method based on robust locally weighted regression and show that this method outperforms current methods at separating LEA and GEA data.

TODO: - Redo abstract
- Update scores to match area of misidentification formula

## 1 Background

Forensic firearms examiners analyze bullets through a process of visual feature comparison to determine whether two bullets originate from the same source. Two bullets in question are placed under a comparison microscope and firearms examiners evaluate similarities and differences between the bullets' striation marks according to the AFTE Theory of Identification [AFTE Glossary, 1998] guidelines resulting in a decision about whether both bullets were fired through the same gun barrel. In forensic science, this problem is known as the same source-different source problem and focuses on establishing quantitative evidence whether two bullets were fired through the same gun barrel.

Recent advances in technology, particularly wider access to high resolution 3D microscopy tools, have led to an increase in research focused on image-analysis algorithms for automated, quantitative analyses of bullet evidence. The introduction of this scanning technology to the field of forensic science allows for capture of high resolution 3D images of bullet LEAs, depicted in Figure 1 [see De Kinder et al., 1998, De Kinder and Bonifanti, 1999, Bachrach, 2002]. The resulting 3D images have since been used in the development of several methods for automated comparison of land engraved areas [e.g. Ma et al., 2004, Chu et al., 2010, 2013, Hare et al., 2017].

In this paper, we will focus only on barrels with traditional sharp-edged lands and grooves (i.e., no polygonal rifling). Sections of the bullet that make the closest contact with the barrel are called land engraved areas (LEAs). Those alternate with groove engraved areas (GEAs). Microimperfections in the barrel introduce striae on the bullet during the firing process. The resulting striation marks provide evidence to address the same source-different source problem. A guiding principle in forensic firearms analysis is that two bullets fired through the same barrel will bear more similar striation marks on their LEAs than two bullets fired from different barrels. Hare et al. [2017] proposes a matching algorithm based on 3D imaging data of LEAs. Horizontal slices of the 3D images, called profiles, provide a detailed representation of striae impressed on the surface at a horizontal cross-section of each LEA. A current limitation of this algorithm is that it can not deal with a mix of striae from both LEA and GEAs. For the human visual system, separating the two areas is straightforward. However, the same cannot be said for automated computer vision techniques.

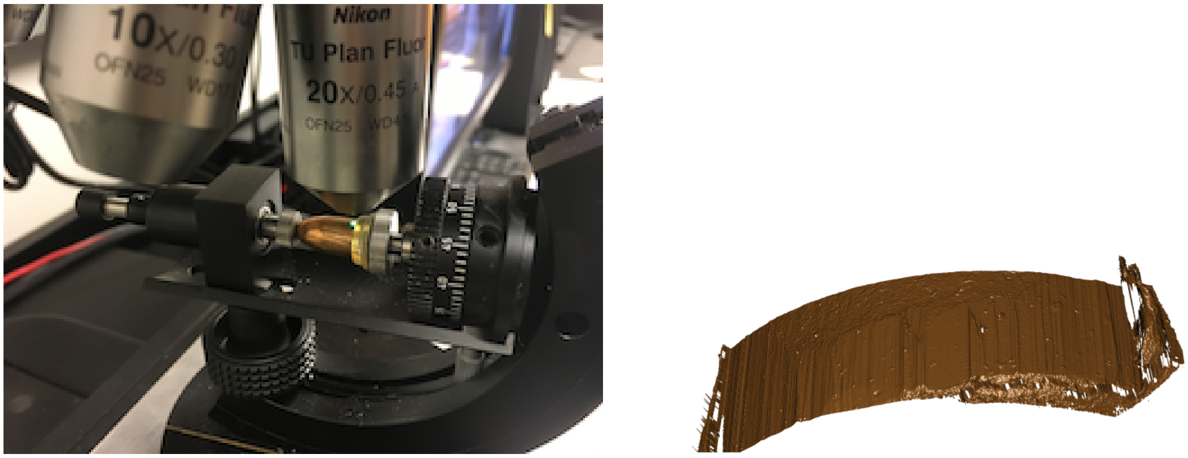



Figure 1: (Left) Close-up view of a bullet staged in a confocal light microscope. The green light marks the focal view of the capture area. (Right) Computer-rendered image of the scanned land engraved area with prominent striation marks.

A correct identification of LEAs is vital to achieve high accuracy and precision in the subsequent downstream analysis. The purpose of this paper is to discuss different automated methods for identifying so-called shoulder locations, the locations at which the land engraved areas end and the groove engraved areas begin.

The structure of the paper is as follows: ...

## 2 Data Source

All currently published automated methods rely on high resolution 3D scans of bullet land engraved areas. Currently accepted best practice for collecting 3D images of bullet LEAs requires that bullets are staged such that striae appear vertically in the scan. Scanning across the LEA must begin and end in the neighboring groove engraved areas as shown in Figure 2. Parts of the breakoff are captured as a visual reference for orientation.

Scans are exported from the microscope as x3p files, conforming to the ISO5436-2 standard [ISO 5436-2:2012, en]. Each scan is recorded as a matrix of $(x, y)$ locations with a measured relative height value $z$ recorded for each $(x, y)$ location on the LEA.

The algorithm proposed by Hare et al. [2017] uses so-called crosscuts, height measurements along $x$ for a fixed $y$. Removal of the overall curve of the bullet – the global structure captured in the 3D scanning process – transforms these profiles into to what Hare et al. [2017] refer to as signatures (see Figure 3). The assessment of similarity between two LEAs is then based on a set of extracted features such as cross-correlation function, number of consecutively matching striae [see Biasotti, 1959] and maximum number of consecutively non-matching striae. Successful extraction of this set of features depends on how well we can remove the global bullet structure to translate from a crosscut to the corresponding signature.

In this paper, we are introducing and comparing two methods for identifying shoulder locations. In order to assess the performance of these methods, we are applying the methods on 3D scans of LEAs from Hamby set 44 [Hamby et al., 2009]. Each Hamby set consists of 35 bullets fired from 10 consecutively rifled Ruger P85 barrels.

Each fired bullet in Hamby Set 44 has 6 LEAs; every LEA was scanned for each of the 35 bullets, producing data for 210 individual lands. Two lands – Barrel 9, Bullet 2, Land 3 and Questioned Bullet L, Land 5 – were removed from consideration due to "tank rash". Tank rash results from a bullet striking the bottom of

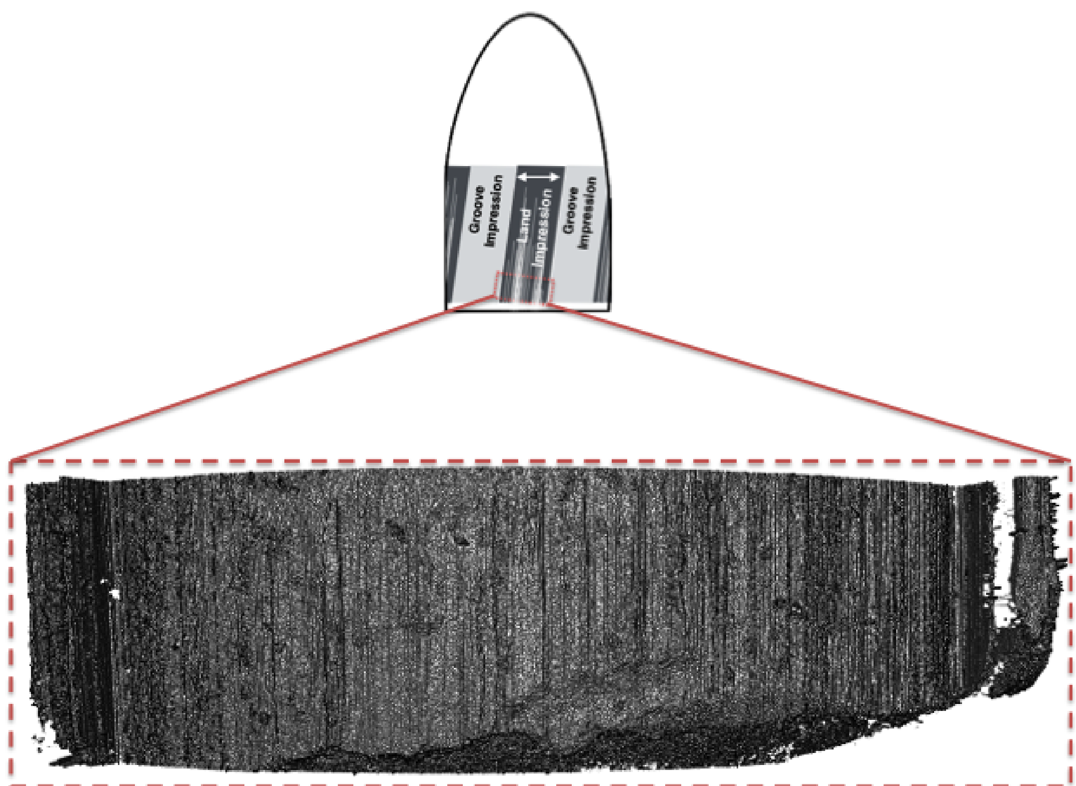
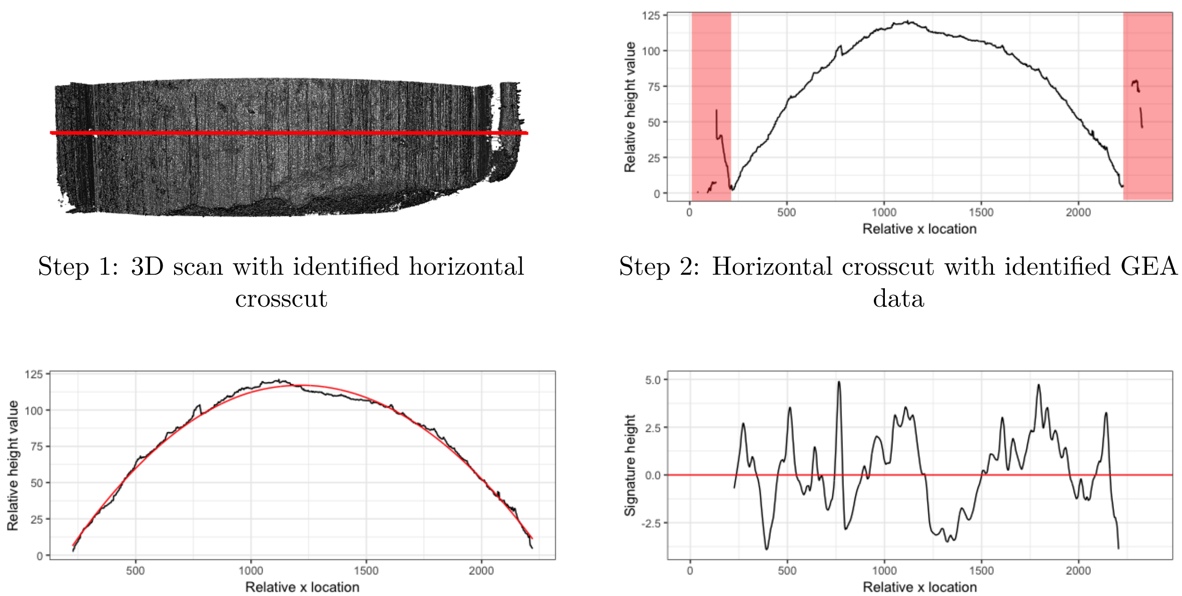



Figure 2: Visualization of 3D data collected through high resolution scanning of a land engraved area. Striations on the surface of the object can be seen by viewing this data from "above", as presented here.




Step 1: 3D scan with identified horizontal crosscut

Step 2: Horizontal crosscut with identified GEA data

Step 3: Non-parametric curvature estimation

Step 4: Extracted LEA signature

Figure 3: The process of extracting a 2D signature from a 3D LEA scan described by Hare et al. [2017]. GEA removal between Step 2 and Step 3 is critical to ensure precise signature extraction.

a water recovery tank after exiting the barrel, thereby creating marks on the land that are not due to the contact with the barrel.

The 3D scans of Hamby Set 44 were captured at Iowa State University's Roy J. Carver High Resolution Microscopy Facility with a Sensofar Confocal light microscope at 20x magnification resulting in a resolution of 0.645 microns per pixel. Physically, each land is approximately 2 millimeters in width; as such, data structures for a single LEA can contain more than 3 million individual data points.

The data used to assess performance of the two methods consists of 2D crosscuts gathered from the 3D scans, as shown in Figure 4.

# 3 Methodology

The structure in the 2D crosscuts is dominated by the curvature of the physical object (the bullet). To assess the similarity of features from two land engraved areas, this curvature has to be removed.

Non-parametric methods suggested in the literature, such as a LOESS fit [Hare et al., 2017] or a Gaussian filter [Chu et al., 2010] are effective for removing the curvature to extract a signature. However, they are prone to boundary effects, which cause mischaracterizations of data patterns near the boundaries of the data domain. In the case of crosscuts, the boundaries are often dominated by values originating from the GEA structure. This structure exaggerates existing boundary effects because GEAs introduce a secondary structure different from the main curvature of the LEA, as shown in Figure 4 and Figure 5. Figure 4 shows how much a non-parametric LOESS fit is affected by including GEA data. Figure 5 illustrates the effect this same inclusion has on extracted signatures. If included, GEA data result in strong boundary effects in the signatures. Statistically, the GEA data introduce outliers into the LEA data. In the next sections, we introduce two methods for fitting the LEA structure. Both methods aim to describe the relationship between horizontal position and relative height on a crosscut. While the two approaches differ in methodology, they are both rooted in the ability to mitigate influence caused by outlying data.

In the following, we will describe the horizontal position on a crosscut of a scan as $x_i$ and measured relative

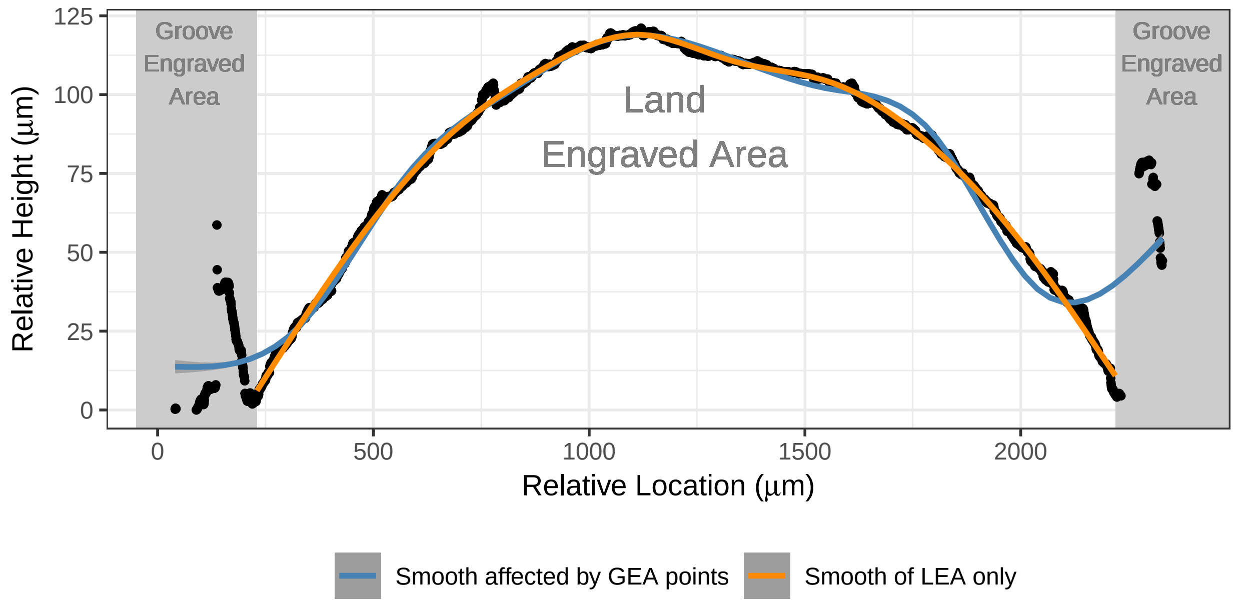



Figure 4: The black points show measured heights for a single crosscut of a 3D LEA scan. The main data structure, located in the center, is comprised of the land engraved area. The groove engraved areas are found on the left and right sides of the crosscut. The lines show fits of two non-parametric LOESS smooths, with and without GEA data. When GEA data is included, the smooth fails to estimate the main LEA structure near the boundaries.

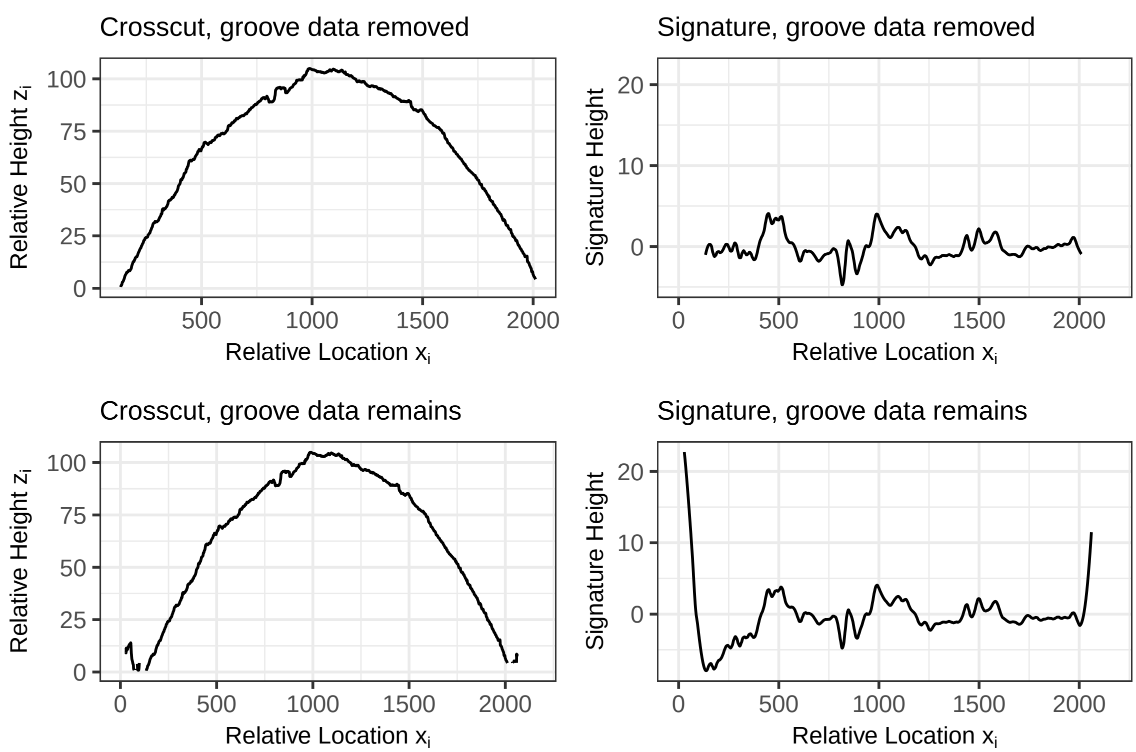



Figure 5: An example of the impact failure to remove GEA data can have on an extracted 2D signature. Even though there are only very few points in the GEA structure, the extracted signatures are dominated by boundary effects.

height as $z_i$, where $i = 1, \ldots, n$, the number of data points along a crosscut.

### 3.1 Robust Linear Models

A natural candidate for a curved structure is a quadratic linear model of the form

$$z_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i,$$

where all error terms $\epsilon_i$ are considered independent and normally distributed with mean 0 and variance $\sigma^2$. The parameters $\beta_0, \beta_1, \beta_2$ are estimated by finding the values which minimize:

$$\arg\min_{\beta} \sum_{i=1}^{n} \left( z_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2) \right)^2 ,$$

the vertical squared distance between each measured height value and the fitted curve. Figure 6 shows that the presence of outliers near the boundaries pulls the resulting curve upwards towards groove engraved area data.

Alternatively, the curve can be fit by minimizing the absolute deviations in place of squared deviations:

$$\arg\min_{\beta} \sum_{i=1}^{n} \left| z_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2) \right| .$$

This method of minimization, known as a robust linear model, is less influenced by large outliers present in the GEA data.

When the estimation is based on squared distances, the method seeks to balance the fit between the LEA structure and the GEA structure, resulting in a fit which compromises between both structures without adequately fitting either. The use of absolute distances reduces the degree of compromise in favor of fitting the majority structure. Figure 6(a) and (c) show the fitted curves from each model framework. Figure 6(b) and (d) display the differences between predicted height and observed height at each location $x_i$, known as the observed model residuals $e_i$. We will utilize the observed residuals to separate the two structures. The robust model in Figure 6 fits the LEA structure more closely and better captures the curvature, allowing for a more accurate separation between GEA and LEA structures.

Because the robust approach results in residual values scattered near zero in the land engraved area and larger, mostly positive residuals in the groove engraved area, we will use high residual magnitude as an indicator of GEA membership (see Figure 6). High residual magnitude is determined using the median absolute deviation (MAD) of all residuals from a crosscut.

The MAD is a robust metric for the spread of points, similar to the standard deviation. It is preferable to the standard deviation to quantify the spread in situations with large outlying residuals, such as this.

Let $m$ denote the median function:

$$MAD(\mathbf{e}) = m(|e_i - m(\mathbf{e})|) \quad \forall\ e_i \in \mathbf{e}.$$

Any residual value larger than 4×MAD is considered an outlier and likely a member of the GEA structure.

Shoulder location predictions are then calculated for each crosscut in the following manner (**Linear Shoulder Location Prediction**):

1. Fit a robust linear model of order 2 (i.e., quadratic) to the crosscut.
2. Calculate a residual value $e_i$ for each data point on the crosscut.

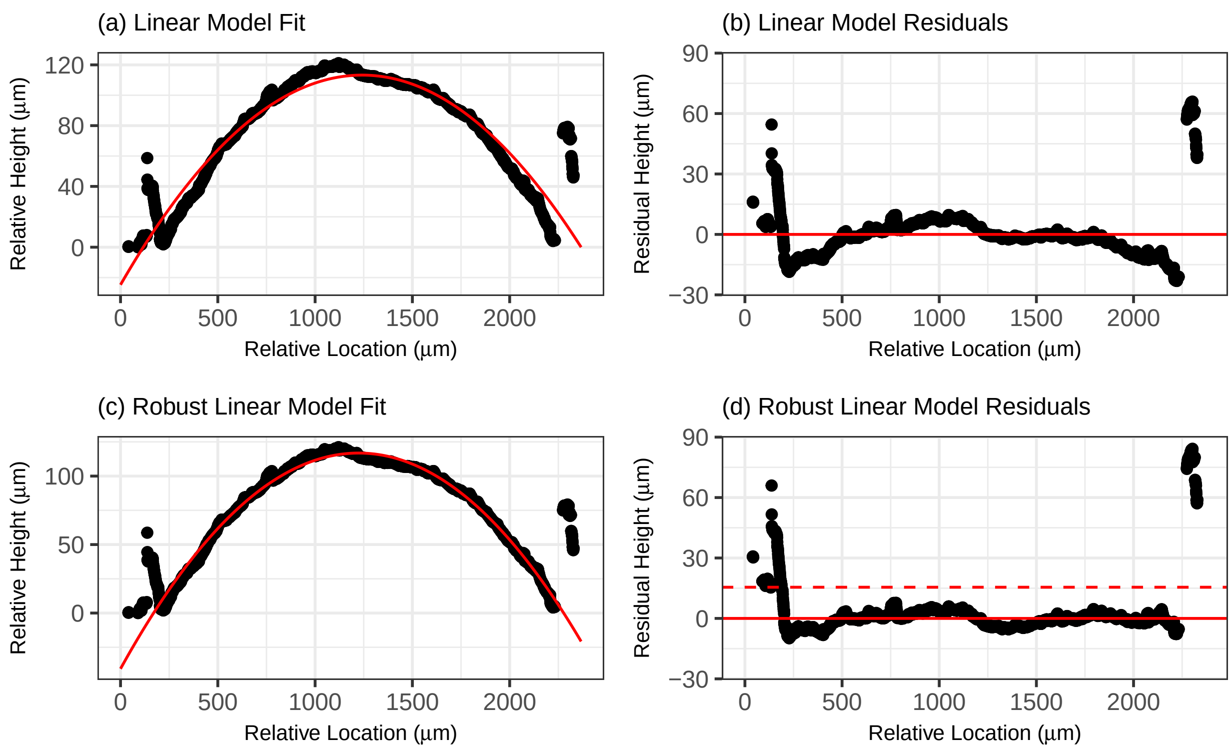



Figure 6: Example of a quadratic linear model fit and resulting residuals (a, b) compared to a robust quadratic linear model fit and residuals (c, d) for a single profile. The robust model is able to more effectively capture the curved structure of the LEA without being influenced by the GEA. The dashed line in (d) is drawn at 4 x MAD. Values above the dashed line are considered outliers.

3. Calculate the median absolute deviation (MAD) for the crosscut.
4. Remove all data points on the crosscut with a residual magnitude greater than 4×MAD.
5. Identify the minimum $x_L$ and maximum $x_R$ of the remaining $x_i$ values. Then, $x_L$ and $x_R$ are the predicted left and right shoulder locations for that crosscut.

## 3.2 Robust LOESS

One of the drawbacks of the linear approach is its rigidity in the shape of the curve. Locally weighted regression, known as LOESS, is a more flexible approach. This is advantageous when working with bullets, as it is unrealistic to expect a circular shape to remain after the bullet has been subjected to the forces of a gun barrel.

LOESS models estimate a predicted value $\hat{z}_i$ for each height $z_i$ corresponding to location $x_i$ by estimating values $\beta_0, \beta_1$ which minimize:

$$\arg\min_{\beta} \sum_{k=1}^{n} w_k(x_i)\left(z_k - (\beta_0 + \beta_1 x_k)\right)^2,$$

where $w_k(x_i)$ is a weight assigned to each data point $x_k$ based on proximity to $x_i$. Weights $w_k$ decrease as the distance to $x_i$ increases, so that data points closest to $x_i$ influence the prediction $\hat{z}_i$ most. This can also be described as a non-parametric weighted average of many parametric models fit to subsets of the data.

This approach allows for greater flexibility. However, it also means that LOESS models are affected by GEA structures more. Data points near and in the GEA structure are most influenced by other GEA data rather than the overall global structure. This results in a set of predictions which misrepresents much of the data near the boundaries (see Figure 7).

Similar to the linear approach, there is a robust approach to LOESS to deal with these boundary effects.

The robust approach to LOESS uses an iterative re-weighting process to reduce the influence of outlying data points [see Cleveland, 1979]. First, an initial LOESS is fit. This is followed by a redistribution of the weights $w_k(x_i)$ based on residual values, $e_i = (z_i - \hat{z}_i)$. New weights are calculated as

$$\left(1 - \left(\frac{e_k}{6 \times MAD}\right)^2\right)^2 \times w_k(x_i) \qquad \text{if } \left|\frac{e_k}{6 \times MAD}\right| < 1.$$

Otherwise, weights are set to 0. These new weights are applied and updated predictions are calculated. This reduces the influence of data points with large residual values $e_k$ in subsequent iterations. In the context of LEA crosscuts, this re-weighting reduces the influence of GEA data.

While robust LOESS methods are more flexible than robust linear models, a model that is accurately fit to the LEA structure results in the same expected residual structure as with robust linear models: positive and negative residuals scattered around zero in the land engraved areas, and positive, possibly large residuals in the groove engraved areas. We are using a similar approach as with the linear shoulder location predictions, i.e. we use a cutoff value to identify large magnitude residual values. The LOESS approach fits more closely to the crosscut and allows us to choose a lower cutoff for separation. A cutoff that performs well on the Hamby set 44 is twice the median absolute deviation ($2 \times MAD$).

Shoulder location predictions are calculated for each crosscut in the following manner (**LOESS shoulder location prediction**):

1. Fit a robust LOESS model with a span of 1 to the crosscut. This can be fit using the ‘locfit.robust‘ function in the ‘locfit‘ package in R.
2. Calculate a residual value $e_i$ for each data point on the crosscut.

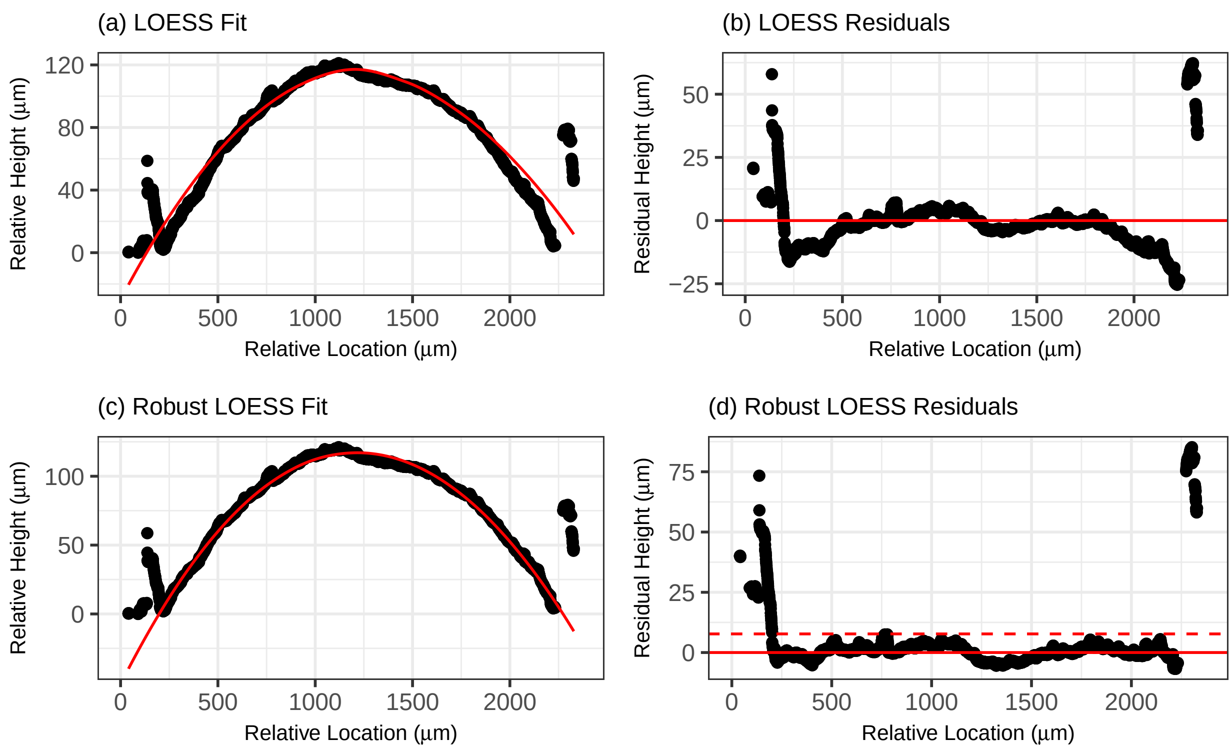



Figure 7: Example of a LOESS model fit and residuals (a, b) compared to a robust LOESS model fit and residuals (c, d) for a single profile. The robust model is again able to more effectively capture the curved structure of the LEA without being influenced by the GEA. The dashed line in (d) represents a cutoff of 2 x MAD. Values above the dashed line are considered outliers.

3. Calculate the median absolute deviation (MAD) for the crosscut.
4. Remove all data points on the crosscut with a residual magnitude greater than 2×MAD.
5. Identify the minimum $x_{L'}$ and maximum $x_{R'}$ of the remaining $x_i$ values. Then, $x_{L'}$ and $x_{R'}$ are the predicted left and right shoulder locations for that crosscut.

# 4 Results

In order to assess the accuracy of the two alternative models based on a quantitative measure for their overall performance of predictions, we first identified "ground truth" of shoulder locations by visual inspection for each of the 208 crosscuts in Hamby set 44.

In order to assess performance of each method, we define a measure for the error as the "area of misidentification", i.e. the area of each crosscut which is identified incorrectly by the method. This metric is calculated for the left shoulder location as:

$$\hat{A}_{jL} = \sum_{e_{ij} \in \tilde{X}_{jL}} |e_{ij}| \times (x_{i+1} - x_i)\,,$$

where $\tilde{X}_{jL}$ is the set of points in crosscut $j$ that fall between the predicted and actual left shoulder location, $e_{ij}$ is the residual value at location $x_i$ from the robust LOESS fit to crosscut $j$, and $(x_{i+1} - x_i)$ is the distance between two subsequent locations in the crosscut. For the scans of Hamby set 44, this distance is equal to $0.645\mu m$ for all locations.

The analogous area is calculated for the right shoulder location as:

$$\hat{A}_{jR} = \sum_{e_{ij} \in \tilde{X}_{jR}} |e_{ij}| \times (x_{i+1} - x_i)\,,$$

where $\tilde{X}_{jR}$ is the set of points in crosscut $j$ that fall between the predicted and actual right shoulder location.

Both the left and right areas of misindentification are thus in terms of microns and represent the area of loss we incur from incorrectly identifying a shoulder location.

The quantification of results as an area is preferable to a distance metric as it captures not only the width of profile area that is misidentified, but also the relative heights of the data. Larger areas of misidentification indicate larger portions of the GEA remain included in a profile, and thus signal an area which is more likely to have influence on an extracted signature. Smaller areas of misidentification indicate minimal loss is incurred, and these areas will have minimal effect on an extracted signature.

An area of misidentification was calculated separately for the left hand side and right hand side predictions for each profile in the data set. This was calculated for the Robust Linear Model and Robust LOESS methods, as well as the Rollapply method suggested in Hare et al. [2017].

To assess the performance of the methods, we are investigating the distribution of the areas of misidentification across all 208 lands of Hamby set 44 (see Figure 8). A distribution that has a smaller spread and is close to zero is ideal; this suggests many of the predicted shoulder locations are very close to the manually identified locations, and predictions are removing many of the outlying GEA points. A distribution with a wider spread or many high, outlying areas of misidentification suggests a greater degree of uncertainty and inaccuracy for a particular method.

The raw distributions can be difficult to visually compare, so another way to inspect the results is to place areas of misidentification into categories: satisfactory, borderline, unsatisfactory. Scores under 100 are satisfactory, scores between 100 and 1000 are borderline, and scores above 1000 are unsatisfactory (see Figure 9). Unsatisfactory cases are the most likely to cause mistakes in subsequent analyses.

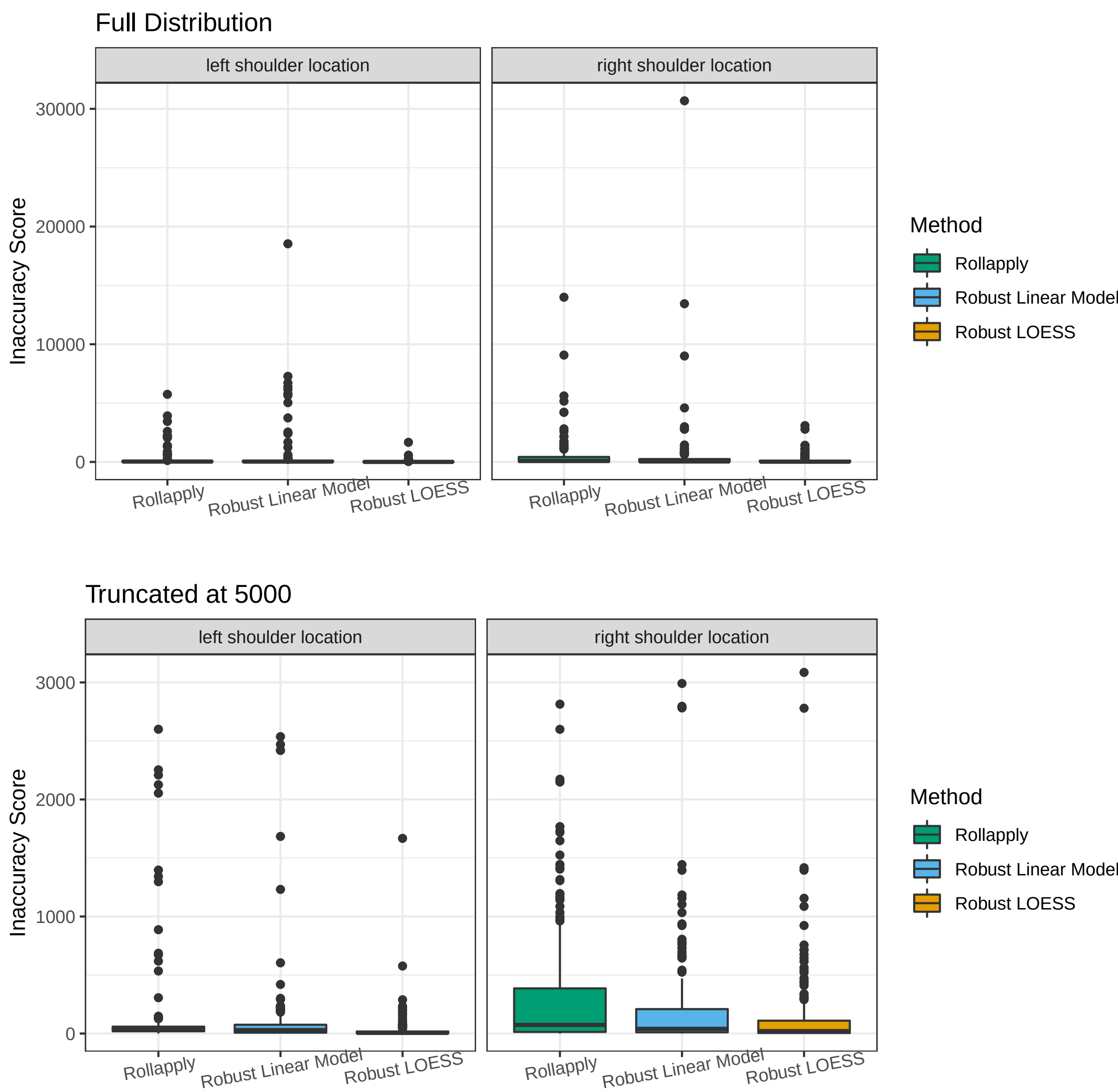



Figure 8: Distribution of areas of misidentification for data smoothing method, robust linear model method, and robust LOESS method, separated by left and right shoulder locations. A tight distribution with few high values indicates good performance across the LEAs in the data set.

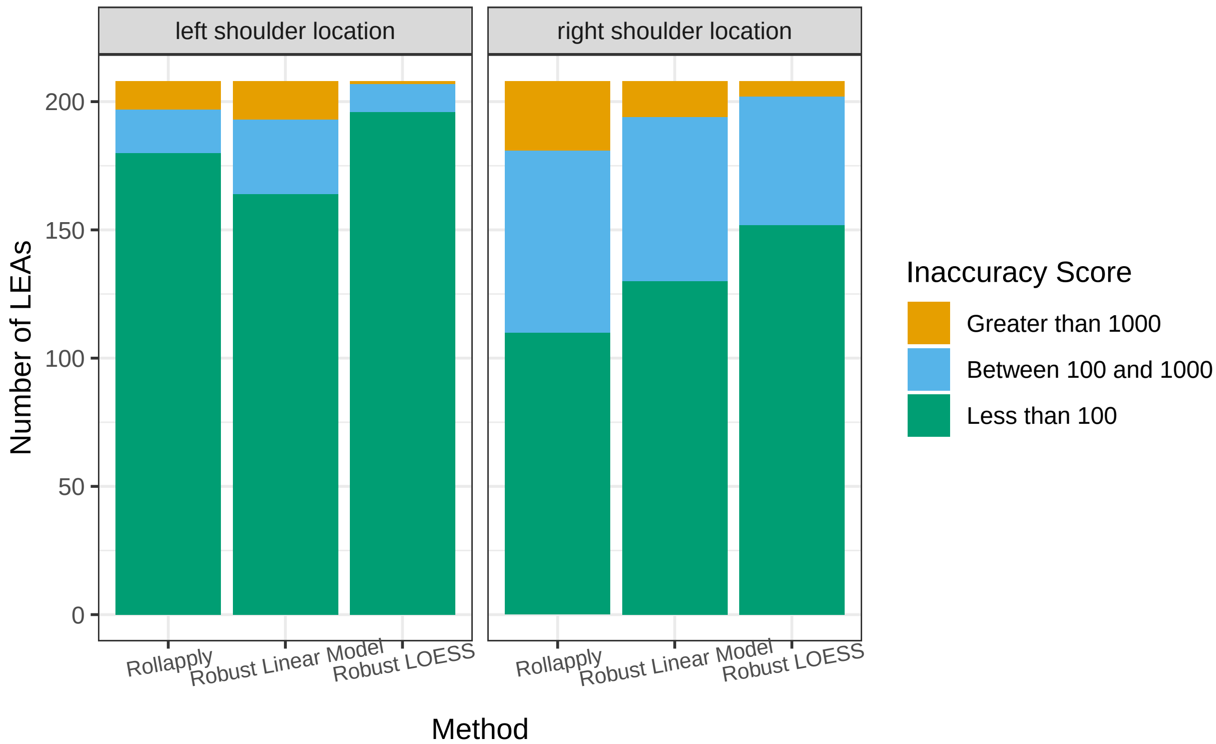



Figure 9: Distribution of areas of misidentification for data smoothing method, robust linear model method, and robust LOESS method, separated by left and righth shoulder locations. Areas of misidentification are placed into three categories: less than 100 microns (satisfactory), between 100 and 1000 microns, and greater than 1000 microns. A larger proportion of areas of misidentification under 100 microns indicates good performance across LEAs in the data set.

It is important to note that different results are expected for the left and right shoulder locations. Within Hamby set 44, almost all scans have a well-defined left groove. Left here is defined as visually left on the scan; this is the side the scan begins on, so a well-defined distinction between GEA and LEA is expected. Often, a less clear distinction is seen on the right side of the scan, with sometimes no apparent shoulder location visible. For this reason it is preferable to separate the left and right for visual inspection of results; a method may excel on one side but fall short on another.

# 5 Conclusions

Both the robust linear model and robust LOESS approaches outperform currently implemented solutions based on data smoothers. Of the two, the robust LOESS approach clearly outperforms the robust linear model. This hierarchy of performance is well within expectation given the strength of robust approaches in general as well as the flexibility of LOESS applied to this data type. Robust LOESS also readily handles variation introduced in the process of translating the physical bullet into a 3D object. If there is too much variability in how the bullet is placed relative to the plane of reference on the microscope, profiles can have tilted shapes relative to the x-axis which a quadratic linear model would fail to address. In these situations, LOESS excels.

While the cutoff values presented work well on Hamby set 44, additional validation will need to be executed on a wider variety of barrel types. Depth of striae, physical size of bullet due to caliber, and non-traditional rifling techniques may require alterations to this cutoff value. In addition, a study of the effect of implementing robust LOESS shoulder location identification on the downstream similarity assessment will need to be completed. Due to increased accuracy of predicted shoulder locations, the authors expect an increase in accuracy in bullet matching algorithms. However, this will need to be validated on a variety of data sets prior to implementation without human intervention in the automated process.

## References


AFTE Glossary. Theory of identification as it relates to toolmarks. AFTE Journal, 30(1):86–88, 1998.

Benjamin Bachrach. Development of a 3d-based automated firearms evidence comparison system. Journal of Forensic Sciences, 47(6):1253–1264, 2002.

Alfred A. Biasotti. A statistical study of the individual characteristics of fired bullets. Journal of Forensic Sciences, 4(1):34–50, 1959.

Wei Chu, T. Song, J. Vorburger, J. Yen, S. Ballou, and B. Bacharach. Pilot study of automated bullet signature identification based on topography measurements and correlations. Journal of Forensic Sciences, 55(2):341–47, 2010.

Wei Chu, Robert M Thompson, John Song, and Theodore V Vorburger. Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria. Forensic Science International, 231(1-3): 137–41, 2013.

William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74(368):829–836, 1979.

J. De Kinder and M. Bonifanti. Automated comparison of bullet striations based on 3d topography. Forensic Science International, 101:85–93, 1999.

J. De Kinder, P. Prevot, M. Pirlot, and B. Nys. Surface topology of bullet striations: an innovating technique. AFTE Journal, 30(2):294–299, 1998.

James E. Hamby, David J. Brundage, and James 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.

Eric Hare, Heike Hofmann, and Alicia Carriquiry. Automatic matching of bullet land impressions. The Annals of Applied Statistics, 11:2332–2356, 12 2017.

ISO 5436-2:2012(en). Geometrical product specifications (GPS) – Surface texture: Profile method; Measurement standards – Part 2: Software measurement standards. Standard, International Organization for Standardization, Geneva, CH, 2012.

L. Ma, J. Song, E. Whitenton, A. Zheng, T. Vorburger, and J. Zhou. NIST bullet signature measurement system for RM (reference material) 8240 standard bullets. Journal of Forensic Sciences, 49(4):649–59, 2004.