

# A Robust Approach to Automatically Locating Grooves in 3D Bullet Land Scans

Kiegan Rice \*

Department of Statistics, Iowa State University  
and

Ulrike Genschel

Department of Statistics and the Center for Statistics and Applications in Forensic Ev  
and

Heike Hofmann

Department of Statistics and the Center for Statistics and Applications in Forensic Ev

August 26, 2019

## Abstract

Land engraved areas (LEAs) provide evidence to address the same source-different source problem in forensic firearms examination. Advances in technology have led to new research on applying image-analysis algorithms to the automated, quantitative analysis of bullet evidence. One prominent example is an algorithm developed by Hare et. al (1) based on 3D imaging data of LEAs. Currently accepted best practice for collecting 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 comparisons. 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 developed a method for automated removal of GEA data based on robust locally weighted regression. This automated method was tested on high resolution 3D scans of LEAs from two bullet test sets. This separation method outperforms current methods at separating LEA and GEA data.

*Keywords:* land engraved areas (LEAs), groove engraved areas (GEAs), 3D scans, bullet identification, automatic matching

---

\*We would like to thank the efforts of the Center for Statistics and Applications in Forensic Evidence (CSAFE) and the Roy J. Carver High Resolution Microscopy lab in scanning the Hamby set 44 and providing the scans to us. This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement No. 70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.

# 1 Reviewer Comments

Below are the reviewer comments with proposed responses. Proposed changes in accordance with reviewer comments within the text are colored teal.

Separate out Table 1 to be for Hamby/Houston data separately. Also, update conclusions section to be correct with updated data.

## 1.1 Reviewer 1

1. “Currently accepted best practice...”

- We have rephrased in accordance with the reviewer’s comment.

2. “2D crosscuts” should be “1D profiles”.

- Profiles are recorded in two dimensions - location and height. Traditionally in imaging, 3D topographies such as these might be considered 2.5-D, which would make profiles 1.5-D. In order to avoid discussion of dimensionality here, we choose to remove dimension entirely and refer to them as simply “crosscuts” and “profiles”.

3. Duplicate paragraph.

- Noted and removed.

4. Describe rollapply, whether it is publicly available, etc.

- An additional paragraph describing the rollapply method and where it is available has been added to the background section.

5. Figure 8 (boxplots).

- We have rewritten to clarify what we are plotting. The description of Figure 8 within the text has been updated to clarify this point, as well as the figure caption.

6. Figure 9/Table 1 are duplicated.

- They are in fact based on the same numbers. However, the figure gives an overview of trend while the table provides the specific quantitative information. For this reason, we believe they both provide the audience with useful insight. A sentence clarifying that they are sourced from the same information has been added.

7. Defend choice of cutoffs for results.

- The binning could be done differently. What we want to see is essentially a rough categorization of different levels of deviation from the “ground truth” shoulder location. To clarify this, we have renamed the categories as “small deviation”, “medium deviation”, and “large deviation”.

8. X3P should be all caps

- The ISO standard, referenced in the “Data Source” section, defines x3p as “x3p”, with lowercase “x” and “p”.

9. Caliber, ammo brand, bullet material

- We have updated the data source section to include this information.

10. Why was 0.645 chosen as the resolution?

- This is the resolution of our instrument at 20x magnification, so we are bound by 0.645. For reference, Hamby sets scanned at NIST and posted publicly on the NIST Ballistics and Toolmark Research Database (NBTRD) are gathered at a resolution of 1.5625 microns/pixel. Our machine’s resolution is higher than that.

11. Is rollapply only other method? Is it current-best?

- Rollapply is the only other automated approach available in the literature for methods which perform LEA-to-LEA comparisons.

12. Fig. 5 and 6 should be same data, or clarify if they are not.

- We have added information about which LEA is shown in Figures 5, 6, and 7 for clarity.

- Figure 5 has also been updated to be the same LEA data as Figures 5 and 6 for clarity.

### 13. Why is right shoulder harder to find?

- We added additional details to support our hypothesis on this point.

## 1.2 Reviewer 2

### 1. Title is confusing

- We have updated the title for clarification.

### 2. “Visual feature comparison” vs “pattern matching”

- We use “visual feature comparison” over “pattern matching” because we want to emphasize the process of comparison, not the result. However, for clarification, we have re-worded to simply “visual comparison”.

### 3. “Striated tool marks”

- Reworded to use the phrase striated tool marks instead of striations.

### 4. Citation for “same source-different source problem”.

- Added a citation.

### 5. Spell out LEA. Use “toolmarks produced on fired bullets”.

- Spelled out LEA on its first appearance.
- Added a statement to clarify that we are referring to toolmarks on fired bullets.

### 6. “3rd paragraph...”

- We are assuming you would like us to rephrase. We have done so.

### 7. “Impressed” should be “engraved”.

- You are correct. We have rephrased.

8. “Automated Computer Vision Techniques”

- We were referring to automated computer vision techniques in general. However, to clarify for this specific application, we have rephrased.

9. “Data source...”

- Yes, there are some papers that assume operators will manually trim. This is not what the Hare et. al. paper does. Rollapply is an automated approach to this in that paper.

10. “Tank rash”

- Have rephrased.

11.

12. Misread of process in Fig. 7

- Note that our process, described step-by-step on page 8, is simply to identify shoulder locations. All LEA data inside those shoulder locations are preserved and remain as part of the data included in further analysis.
- Lines to reference where identified shoulder locations would be have been added to Figs. 6d and 7d, and clarifications have been added to the figure captions.

13. Bullets with smooth GEA

- This particular method is meant for sharp-edged rifling. Smooth GEAs would not cause the same structural issues as sharp-edged rifling. In downstream analysis, a LOESS fit to the bullet structure can capture the more gradual changes present with a smooth GEA and still produce a useful signature.

14. Shoulders not aligned along longitude could cause a problem

- We have noted this limitation in our conclusions section.

15. “MUST rely on operator to manually crop”

- There is not really a comment or suggested change that we can address.

16. “Comparison of manually cropped vs. automated using CCF”

- There are too many factors to account for to properly evaluate this within the scope of this paper. We understand that downstream analysis is the “gold standard” but the CCF alone is not good enough to evaluate performance here. Calculating a CCF for two profiles will almost always result in a very high value due to the underlying curvature present in LEA profiles. There are multiple methods in the literature to extract a LEA signature, and they do not result in the same CCF values. To demonstrate this property, see the images below. For this reason, we believe discussing the downstream analysis would be out of scope for this paper.

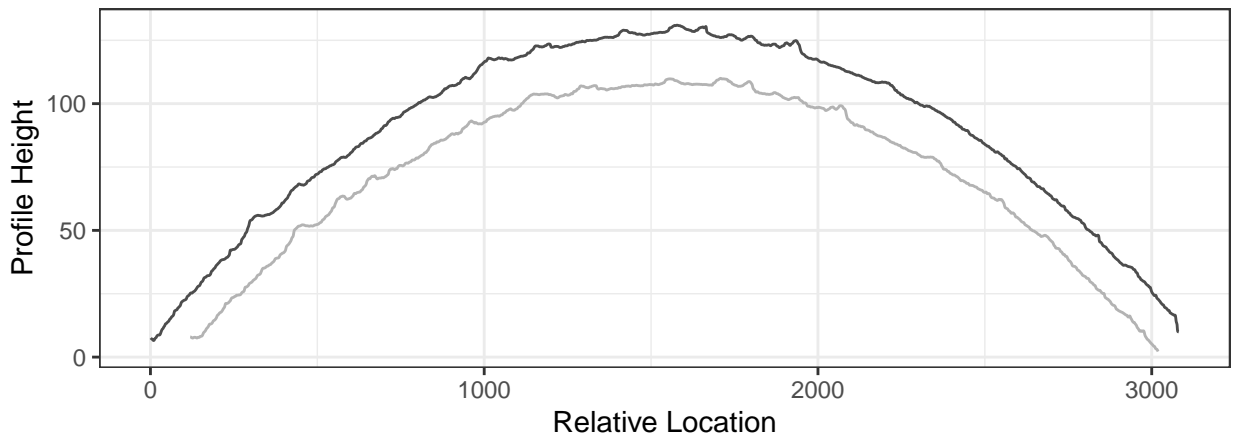
17. “Diversity in scan sets”

- We have added an additional test set, Houston-test, to our results section. This has also been detailed in the data source section.
- The figures have been updated to include the Houston-test set, but the table has not. Should the table for Hamby and Houston be kept separate?

## Known Match

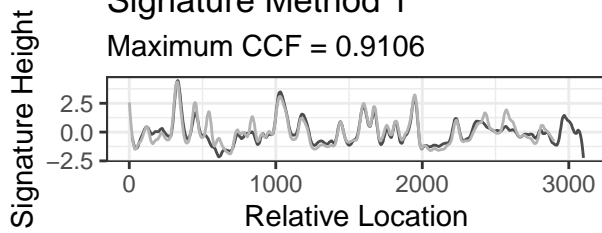
### Profile

Maximum CCF = 0.9986



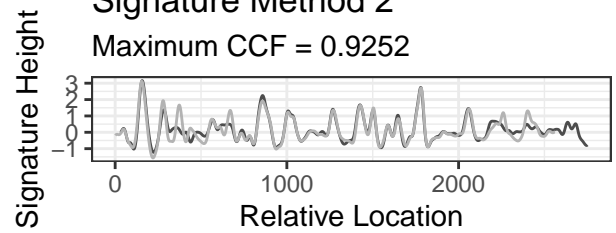
### Signature Method 1

Maximum CCF = 0.9106



### Signature Method 2

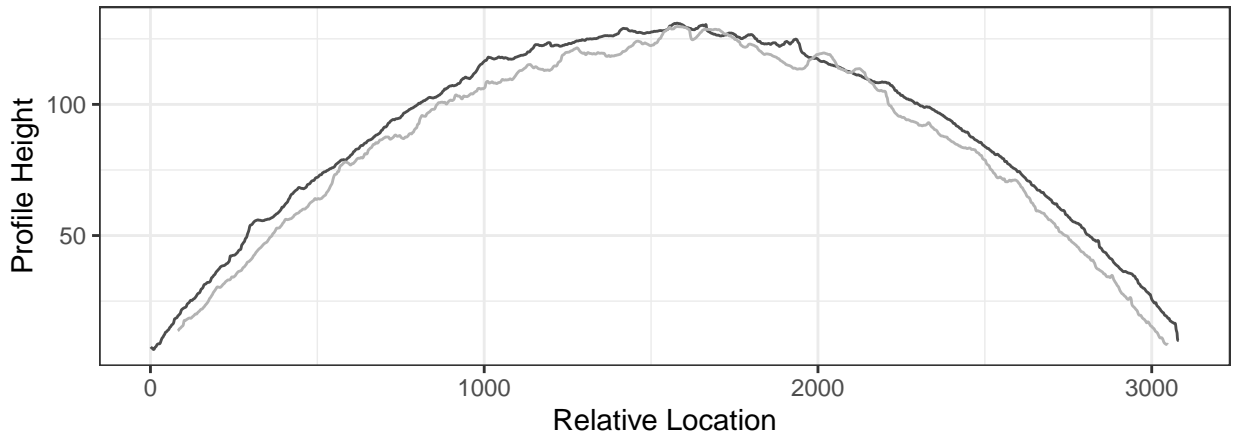
Maximum CCF = 0.9252



## Known Non-Match

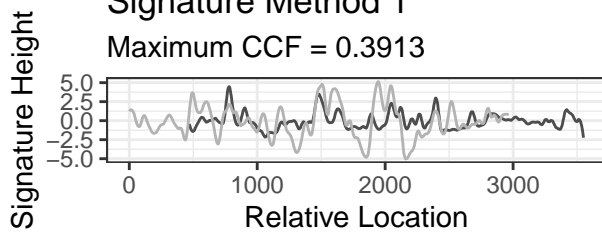
### Profile

Maximum CCF = 0.9964



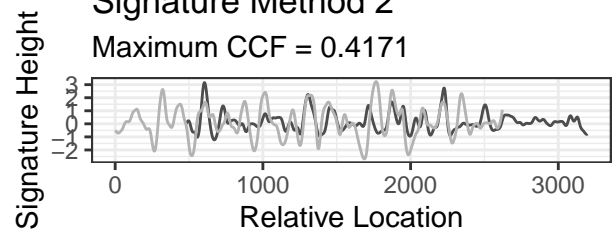
### Signature Method 1

Maximum CCF = 0.3913



### Signature Method 2

Maximum CCF = 0.4171



## 2 Background

Forensic firearms examiners analyze bullets through a process of [visual 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 [striated toolmarks produced on fired bullets from rifled barrels](#). [The assessment of these patterns follows](#) the AFTE Theory of Identification (2) 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* (see 3) and it 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 on image-analysis algorithms for automated, quantitative analyses of bullet evidence. The introduction of such scanning technology to the field of forensic science allows for capture of high resolution 3D images of bullet land engraved areas (LEAs), depicted in Figure 1 (see 4; 5; 6). The resulting 3D images have since been used in the development of several novel methods for automated comparison of land engraved areas (e.g. 7; 8; 9; 1).

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 [contact with high points inside the barrel](#) are called land engraved areas (LEAs) and alternate with low points called [groove engraved areas \(GEAs\)](#). Micro imperfections 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. (1) proposes a matching algorithm based on 3D imaging data of LEAs. Horizontal slices of the 3D images, called crosscuts, provide a detailed representation of striae engraved on the surface at a horizontal cross-section of each LEA. A current limitation of this algorithm is that it cannot 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](#)



toolmark comparison techniques.

A correct identification of LEAs is vital to achieve high accuracy and precision in the subsequent downstream analysis and our goal is to present and 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 best currently available method to identify these shoulder locations is the “roll-apply” method proposed by @Hare1. This method is publicly available through the `bulletxtrctr` package for the open source statistical computing language R. The authors propose first applying a rolling average to each profile to smooth out bumps in data, followed by identifying the local minima closest to the edges of each smoothed profile.

The structure of the paper is as follows: a description of the data storage format and data set, an explanation of methodology to predict shoulder locations, and finally an illustration of improved prediction accuracy results.

### 3 Data Source

All currently published automated methods rely on high resolution 3D scans of bullet land engraved areas. Our approach to the collection of 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 (see Figure 1).

Scans are exported from the microscope as x3p files, conforming to the ISO5436-2 standard (10). Each scan is recorded as a matrix of prespecified  $(x, y)$  locations with a measured relative height value  $z$  for each  $(x, y)$  location on the LEA.

The algorithm proposed by Hare et al. (1) 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 crosscuts into to what Hare et al. (1) 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 11) 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 the following, we are introducing and comparing two methods for identifying shoulder locations. In order to assess the performance of these methods, we are applying both methods on 3D scans of LEAs from Hamby set 44 (12) and the Houston-test set. Each Hamby set consists of 35 Winchester 9mm copper bullets fired from 10 consecutively rifled Ruger P-85 9mm Luger barrels. The Houston-test set consists of XXX bullets fired from XXX type of barrels...

Each fired bullet in Hamby Set 44 and Houston-test set has 6 LEAs; every LEA was scanned for each of the 35 (104) bullets, producing data for 210 624 individual LEAs. Two LEAs from Hamby Set 44 – Barrel 9, Bullet 2, Land 3 and Questioned Bullet L, Land 5 – were removed from consideration because they were deemed unsuitable for comparison. These two LEAs contained significant abrasians created by contact with the bottom of a water recovery tank after exiting the barrel. These abrasians are thus marks present on the LEAs that are not due to contact with the barrel itself.

All 3D scans of Hamby Set 44 and Houston-test were captured at Iowa State University’s Roy J. Carver High Resolution Microscopy Facility with a Sensofar Confocal light microscope at 20 times magnification resulting in a resolution of 0.645 microns per pixel (1 micron =  $1\mu\text{m}$  = 0.001mm). 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.

Data used to assess performance of the two methods consist of crosscuts gathered from the 3D scans, as shown in Figure 4.

## 4 Methodology

The structure in the 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 (1) or a Gaussian filter (8) 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 the inclusion of the same has on extracted signatures. If included, GEA data result in strong boundary effects in the signatures. Statistically, 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 potential 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 height as  $z_i$ , where  $i = 1, \dots, n$ , the number of data points along a crosscut.

## 4.1 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  $\sigma^2$ . Parameter values of  $\beta_0, \beta_1, \beta_2$  are estimated by finding the values which minimize:

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

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:

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

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

Estimation based on taking squared deviations 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 deviations reduces the degree of compromise in favor of fitting the majority structure, here LEA. 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 data, similar to the standard deviation. It is preferable to the standard deviation to quantify the spread in situations with large outlying observations, such as the residuals in the groove engraved area.

Let  $m$  denote the median function. Then the MAD is defined as

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

Any residual value larger than  $4 \times \text{MAD}$  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.
3. Calculate the residual median absolute deviation (MAD) for the crosscut.
4. Remove all data points on the crosscut with a residual magnitude greater than  $4 \times \text{MAD}$ .
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.

## 4.2 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

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

where  $w_k(x_i)$  is a weight assigned to each data point  $x_k$  based on its 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. A LOESS model can also be described as a non-parametric weighted average of many parametric models fit to subsets of the data.

Although this approach allows for greater flexibility, LOESS models are affected more by GEA structures than robust linear models are. 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 exists a robust approach to LOESS to adjust for these boundary effects.

The robust approach to LOESS uses an iterative re-weighting process to reduce the influence of outlying data points (see 13). First, an initial LOESS is fit. This step 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

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

else, weights are set to 0. These new weights are re-applied in (1) and updated predictions are obtained. 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.

A robust LOESS model which accurately fits the LEA structure should result in a similar residual structure as that expected for the robust linear model: small residuals scattered near zero for  $x_i$  locations in the land engraved area, and positive, possibly large residuals for  $x_i$  locations in the groove engraved areas. The increased flexibility of LOESS models lead to a closer fit to the curvature and thus the robust LOESS more reliably results in the aforementioned residual pattern. We can thus use a lower cutoff for separation of the residuals via magnitude. 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 to the crosscut (14).
2. Calculate a residual value  $e_i$  for each data point on the crosscut.
3. Calculate the residual median absolute deviation (MAD) for the crosscut.
4. Remove all data points on the crosscut with a residual magnitude greater than  $2 \times MAD$ .
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.

## 5 Results

To quantitatively assess the predictive ability of the two alternative models, we first identified “ground truth” of shoulder locations by visual inspection for each of the 208 crosscuts in Hamby set 44 [and each of the 414 crosscuts in Houston-test](#).

As a performance measure, we define 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,  $e_{ij}$  corresponds to 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 [and Houston-test](#), this distance is equal to  $0.645\mu m$  for all locations.

Similarly, the 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 misidentification are thus in terms of microns and represent the area of loss we incur from incorrectly identifying a shoulder location.

Quantifying the results as an area is preferable to a distance metric as it captures not only the width of the profile area that is misidentified, but also the relative heights of the data. Larger misidentified areas 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 suggest 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 of the 208 +414 profiles in the data set, where predictions were based on the Robust Linear Model and Robust LOESS methods, as well as the Rollapply method suggested in Hare et al. (1).

To assess the performance of all three methods, we consider the distribution of the areas of misidentification across all 208 +414 lands of Hamby set 44 and Houston-test. Figure 8 presents each distribution as a boxplot. The box encompasses the middle 50% of values, spanning the 25th to 75th percentile. The line inside each box represents the median value, and individual data points are unusually large values we call outliers. A distribution starting at or close to zero with minimal spread is ideal as 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. Some of the outliers are so extreme, statistics appear very close to one another or on top of each other.

Because the raw distributions are difficult to visually compare due to extreme outliers, we categorized areas of misidentification as small, medium, and large deviations. Scores under 100 are considered small deviations, scores between 100 and 1000 are medium, and scores above 1000 are considered large deviations (see Figure 9). Cases with large deviations are the most likely to cause poor or flawed results in subsequent analyses. Note that Figure 9(a) and Table 1, as well as Figure 9(b) and Table 2 are based on the same numbers but provide two different presentations of the results.

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. This may be due to the left shoulder corresponding to the leading edge in the twist of a bullet as it is propelled through a gun barrel. 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.



## 6 Conclusions

The robust LOESS approach clearly outperforms both the robust linear model and the rollapply method for the right shoulder location. Both the robust linear model and robust LOESS approaches outperform currently implemented solutions based on data smoothers for both the left and right shoulder locations on both Hamby set 44 and the Houston test set. While the robust linear model approach outperforms rollapply for the right shoulder location on Hamby set 44, it performs the worst for both sides of the Houston test data and the left shoulder location of Hamby set 44. 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 proposed cutoff values as multiples of MAD work well on Hamby set 44 and Houston test, 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. LEAs with shoulders which are not aligned horizontally due to tilt may also require slight alterations to the method. 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

1. Hare E, Hofmann H, Carriquiry A. Automatic matching of bullet land impressions. Ann Appl Stat 2017 12;11:2332–2356.

2. AFTE Glossary. Theory of Identification as it Relates to Toolmarks. AFTE Journal 1998;30(1):86–88.
3. Ommen DM, Saunders CP. Building a unified statistical framework for the forensic identification of source problems. Law Probab Risk 2018 05;17(2):179–197. <https://doi.org/10.1093/lpr/mgy008>.
4. De Kinder J, Prevot P, Pirlot M, Nys B. Surface topology of bullet striations: an innovating technique. AFTE Journal 1998;30(2):294–299.
5. De Kinder J, Bonifanti M. Automated comparison of bullet striations based on 3D topography. Forensic Sci Int 1999;101:85–93.
6. Bachrach B. Development of a 3D-based Automated Firearms Evidence Comparison System. J Forensic Sci 2002;47(6):1253–1264.
7. Ma L, Song J, Whitenton E, Zheng A, Vorburger T, Zhou J. NIST bullet signature measurement system for RM (Reference Material) 8240 standard bullets. J Forensic Sci 2004;49(4):649–59.
8. Chu W, Song T, Vorburger J, Yen J, Ballou S, Bacharach B. Pilot study of automated bullet signature identification based on topography measurements and correlations. J Forensic Sci 2010;55(2):341–47.
9. Chu W, Thompson RM, Song J, Vorburger TV. Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria. Forensic Sci Int 2013;231(1-3):137–41.
10. Geometrical product specifications (GPS) – Surface texture: Profile method; Measurement standards – Part 2: Software measurement standards. Geneva, CH: International Organization for Standardization; 2012.
11. Biasotti AA. A Statistical Study of the Individual Characteristics of Fired Bullets. J Forensic Sci 1959;4(1):34–50.

12. Hamby JE, Brundage DJ, Thorpe JW. The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries. *AFTE Journal* 2009;41(2):99–110.
13. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. *JASA* 1979;74(368):829–836.
14. Loader C. locfit: Local Regression, Likelihood and Density Estimation.; 2013. R package version 1.5-9.1. <https://CRAN.R-project.org/package=locfit>.

# List of Figures

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. Breakoff is seen visually on the bottom right hand side of the scan.	22
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. . . . .	23
3	The process of extracting a signature from a 3D LEA scan described by (1). GEA removal between Steps 2 and 3 is critical to ensure precise signature extraction. . . . .	24
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. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2. . . . .	25
5	An example of the impact failure to remove GEA data can have on an extracted signature. Even though there are only very few points in the GEA structure, the extracted signatures are dominated by boundary effects. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2. . . . .	26
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 horizontal line in (d) is drawn at 4 x MAD. Values above the dashed line are considered outliers. The vertical lines in (d) are drawn where the left and right shoulder locations would be identified. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2. . . . .	27
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. The vertical lines in (d) are drawn where the left and right shoulder locations would be identified. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2. . . . .	28
8	Distribution for all 622 crosscuts, presented here as a boxplot, of areas of misidentification for rollapply (data smoothing) method, robust linear model method, and robust LOESS method, separated by left and right shoulder locations. A dense distribution with few high values indicates good performance across the LEAs in the data set. . . . .	29

9	Distribution of areas of misidentification for rollapply (data smoothing) method, robust linear model method, and robust LOESS method, separated by left and right shoulder locations. Areas of misidentification are placed into three categories: less than 100 microns (small deviations), 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. Results are split between Hamby set 44 and the Houston-test set. . . . .	30
---	---	----

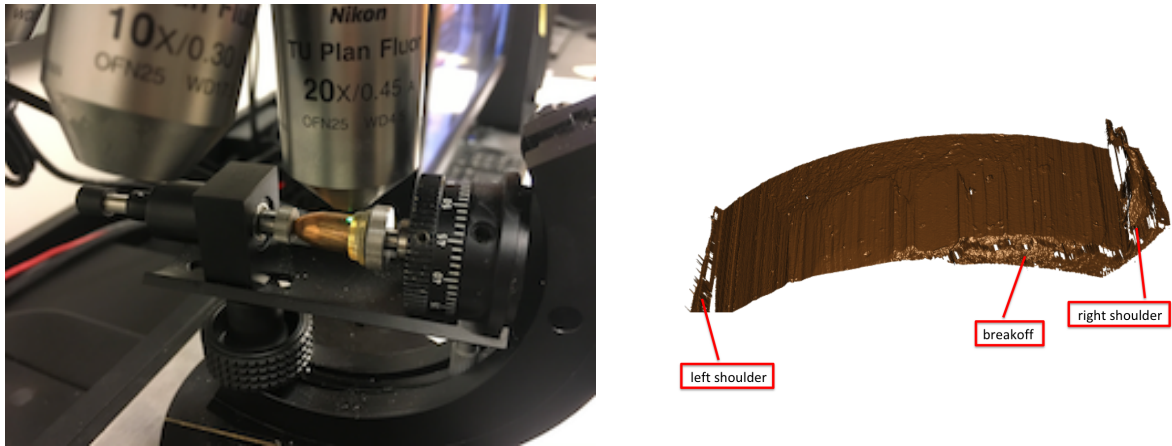


Figure 1: (Left) Close-up view of a bullet staged in a confocal light microscope. The green light marks the focal view of the capture area. (Right) Computer-rendered image of the scanned land engraved area with prominent striation marks. Breakoff is seen visually on the bottom right hand side of the scan.

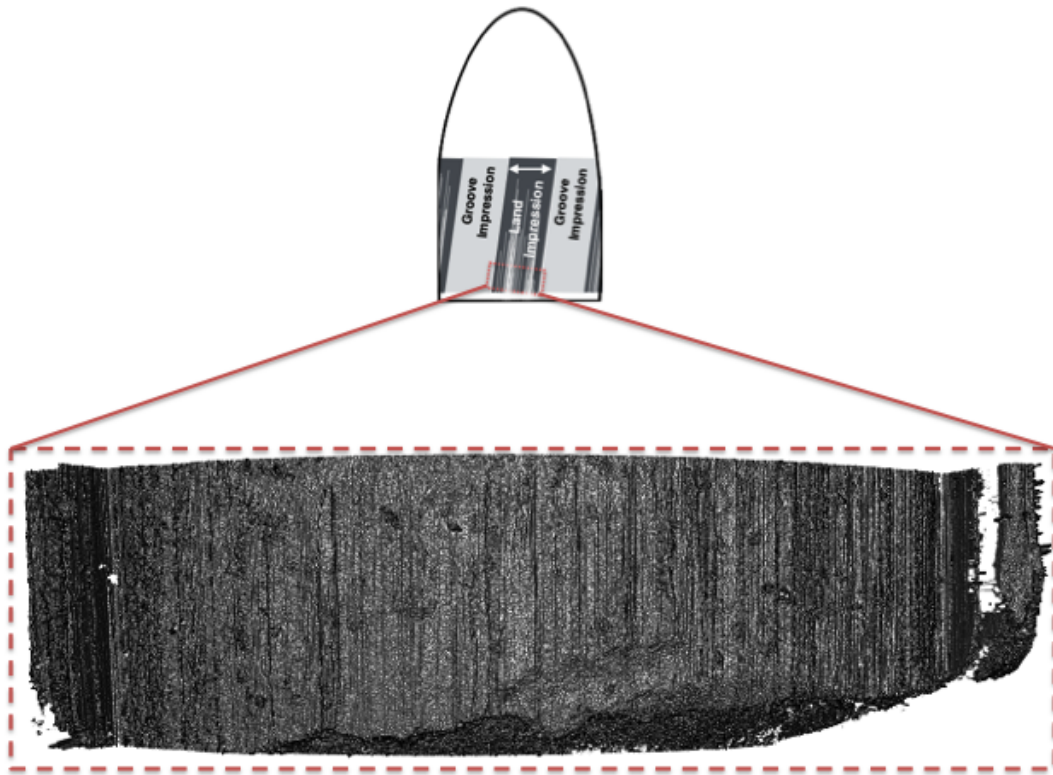
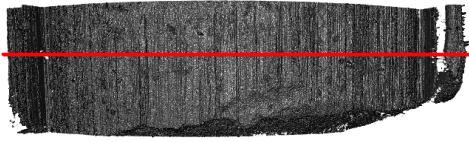
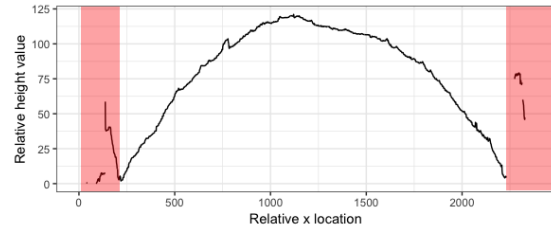


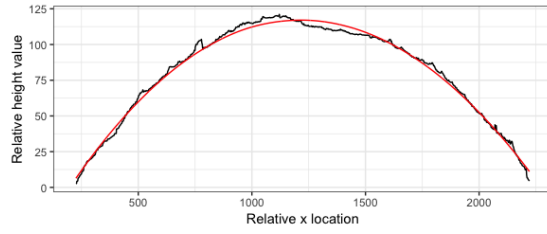
Figure 2: Visualization of 3D data collected through high resolution scanning of a land engraved area. Striations on the surface of the object can be seen by viewing this data from "above", as presented here.



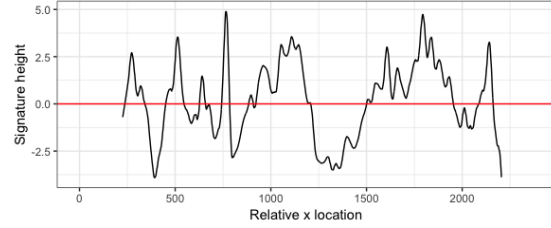
Step 1: 3D scan with identified horizontal crosscut



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 signature from a 3D LEA scan described by (1). GEA removal between Steps 2 and 3 is critical to ensure precise signature extraction.



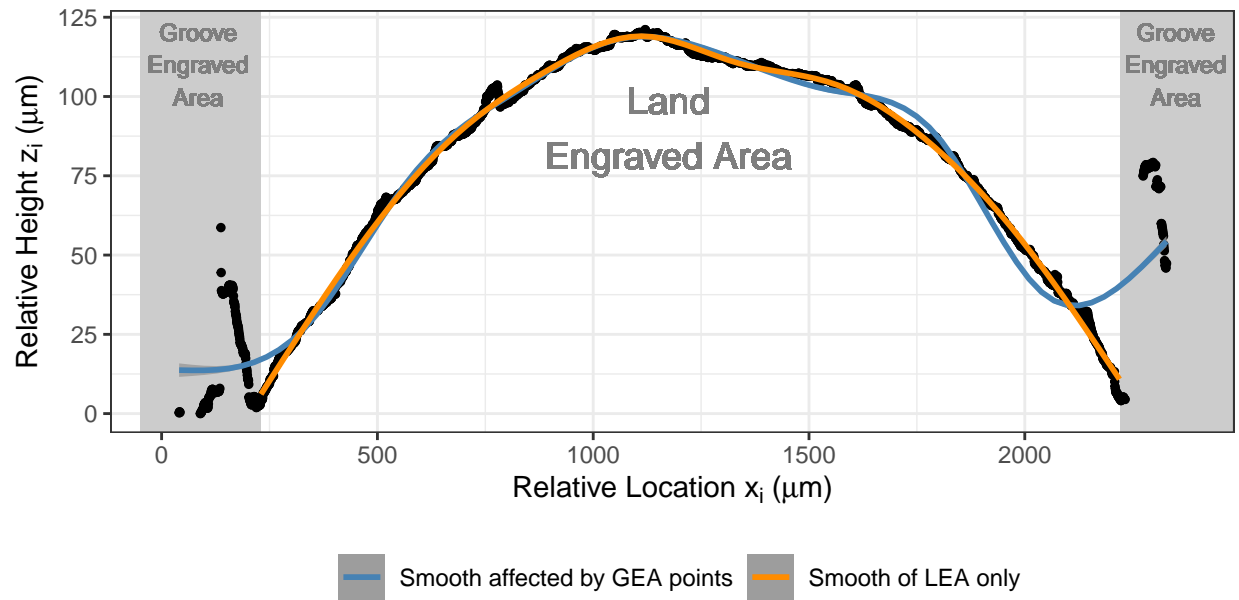


Figure 4: The black points show measured heights for a single crosscut of a 3D LEA scan. The main data structure, located in the center, is comprised of the land engraved area. The groove engraved areas are found on the left and right sides of the crosscut. The lines show fits of two non-parametric LOESS smooths, with and without GEA data. When GEA data is included, the smooth fails to estimate the main LEA structure near the boundaries. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2.

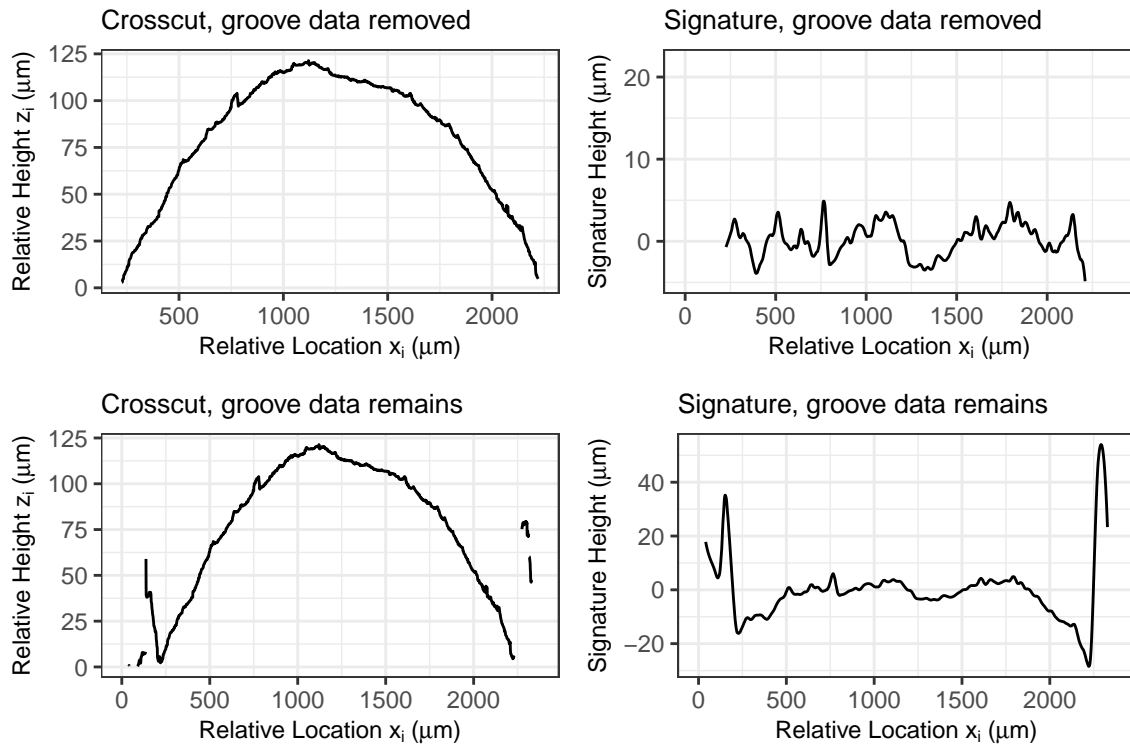


Figure 5: An example of the impact failure to remove GEA data can have on an extracted signature. Even though there are only very few points in the GEA structure, the extracted signatures are dominated by boundary effects. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2.

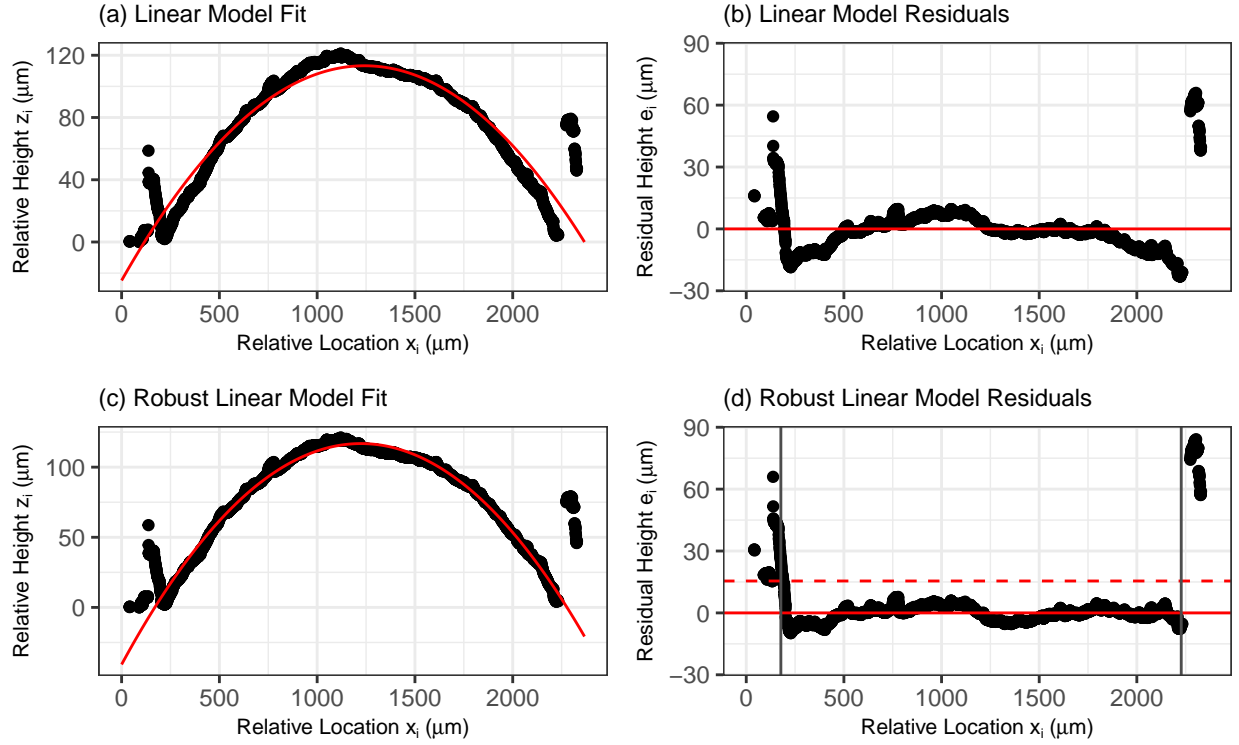


Figure 6: Example of a quadratic linear model fit and resulting residuals (a, b) compared to a robust quadratic linear model fit and residuals (c, d) for a single profile. The robust model is able to more effectively capture the curved structure of the LEA without being influenced by the GEA. The dashed horizontal line in (d) is drawn at  $4 \times \text{MAD}$ . Values above the dashed line are considered outliers. The vertical lines in (d) are drawn where the left and right shoulder locations would be identified. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2.

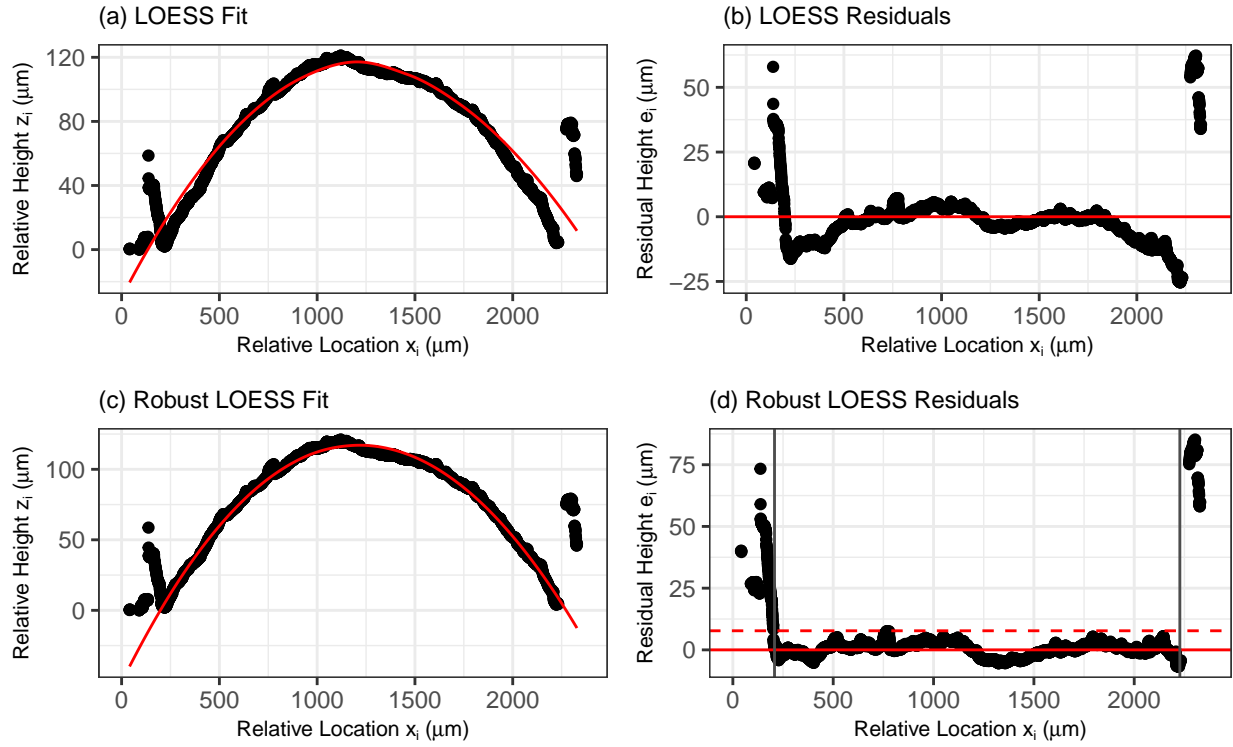


Figure 7: Example of a LOESS model fit and residuals (a, b) compared to a robust LOESS model fit and residuals (c, d) for a single profile. The robust model is again able to more effectively capture the curved structure of the LEA without being influenced by the GEA. The dashed line in (d) represents a cutoff of  $2 \times \text{MAD}$ . Values above the dashed line are considered outliers. The vertical lines in (d) are drawn where the left and right shoulder locations would be identified. The LEA pictured here is Hamby 44, Barrel 10, Bullet 2, Land 2.

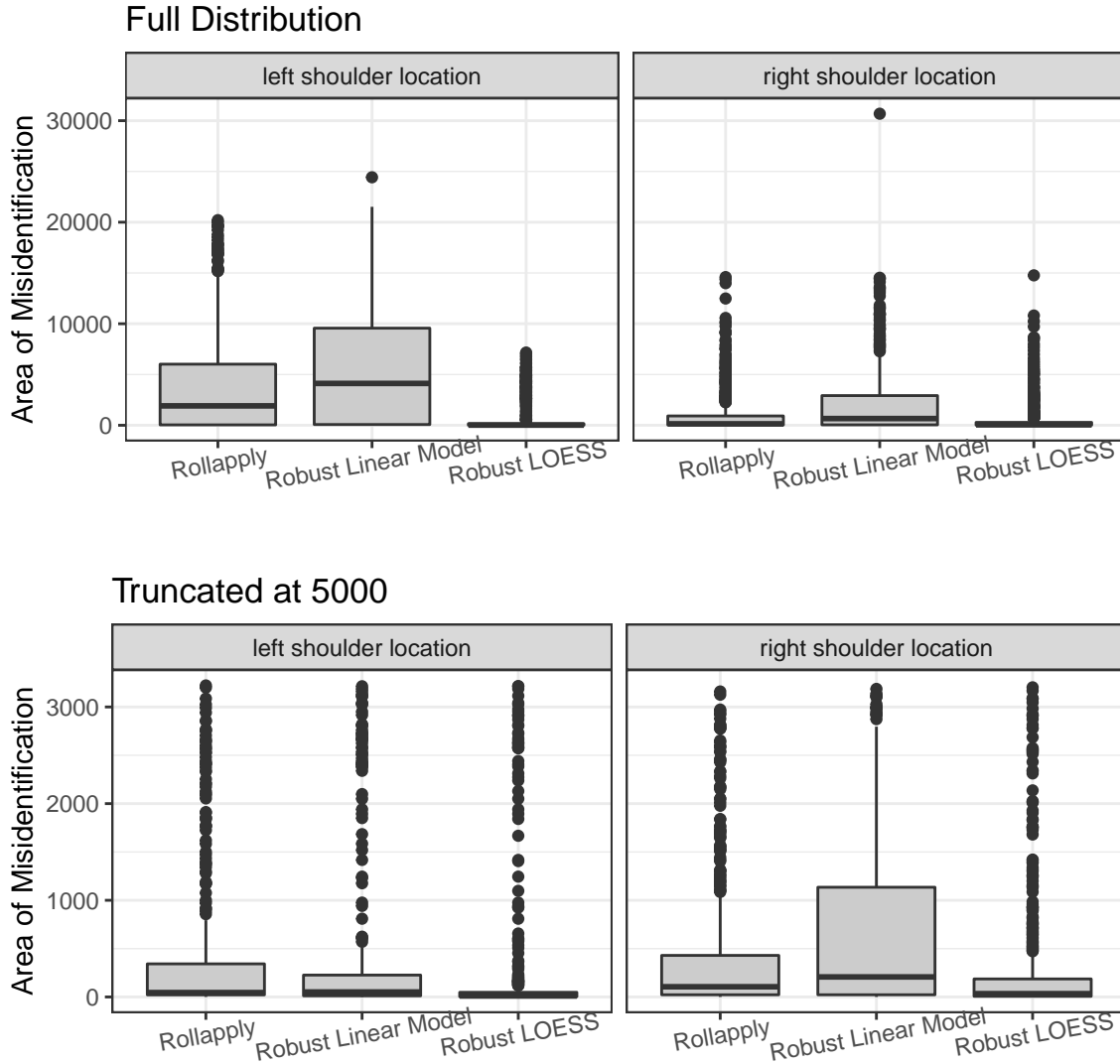


Figure 8: Distribution for all 622 crosscuts, presented here as a boxplot, of areas of misidentification for rollapply (data smoothing) method, robust linear model method, and robust LOESS method, separated by left and right shoulder locations. A dense distribution with few high values indicates good performance across the LEAs in the data set.

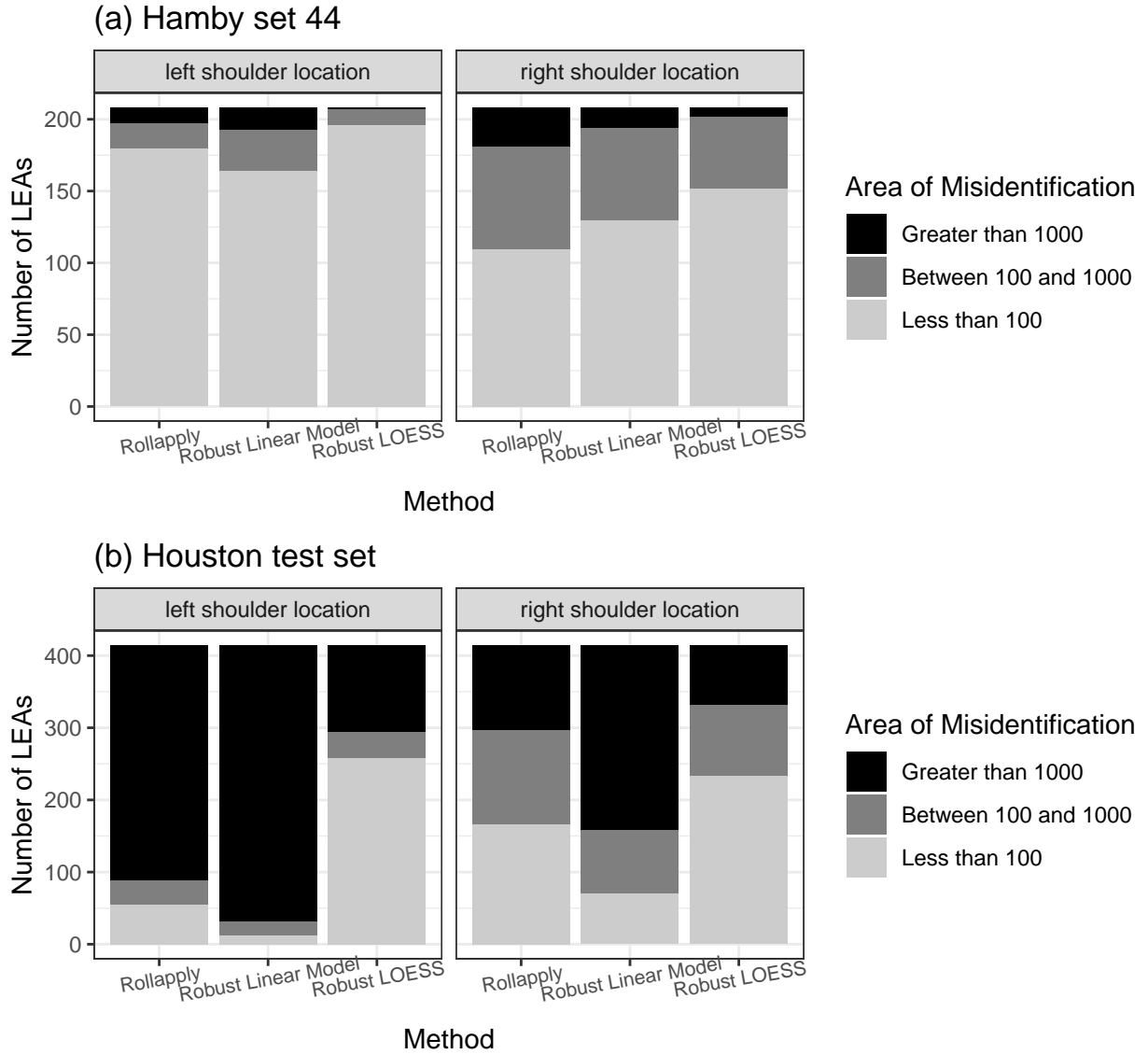


Figure 9: Distribution of areas of misidentification for rollapply (data smoothing) method, robust linear model method, and robust LOESS method, separated by left and right shoulder locations. Areas of misidentification are placed into three categories: less than 100 microns (small deviations), between 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. Results are split between Hamby set 44 and the Houston-test set.

## List of Tables

- 1 Table of results for areas of misidentification (in  $\mu m^2$ ) resulting from automated shoulder location identification methods applied to Hamby set 44. Results are presented as percentage of shoulder location predictions which fall into each category: small deviations (less than 100), medium deviations (between 100 and 1000), and large deviations (greater than 1000). . . . . 32
- 2 Table of results for areas of misidentification (in  $\mu m^2$ ) resulting from automated shoulder location identification methods applied to the Houston test set. Results are presented as percentage of shoulder location predictions which fall into each category: small deviations (less than 100), medium deviations (between 100 and 1000), and large deviations (greater than 1000). 33

<b>Method</b>	<b>Left Shoulder Location</b>			<b>Right Shoulder Location</b>		
	(0, 100)	(100, 1000)	(1000, $\infty$ )	(0, 100)	(100, 1000)	(1000, $\infty$ )
Rollapply	86.54	8.17	5.29	52.88	34.13	12.98
Robust Linear Model	78.85	13.94	7.21	62.50	30.77	6.73
Robust LOESS	94.23	5.29	0.48	73.08	24.04	2.88

Table 1: Table of results for areas of misidentification (in  $\mu m^2$ ) resulting from automated shoulder location identification methods applied to Hamby set 44. Results are presented as percentage of shoulder location predictions which fall into each category: small deviations (less than 100), medium deviations (between 100 and 1000), and large deviations (greater than 1000).



<b>Method</b>	<b>Left Shoulder Location</b>			<b>Right Shoulder Location</b>		
	(0, 100)	(100, 1000)	(1000, $\infty$ )	(0, 100)	(100, 1000)	(1000, $\infty$ )
Rollapply	13.29	8.21	78.5	40.10	31.64	28.26
Robust Linear Model	2.90	4.83	92.27	17.15	21.01	61.84
Robust LOESS	62.32	8.70	28.99	56.52	23.67	19.81

Table 2: Table of results for areas of misidentification (in  $\mu m^2$ ) resulting from automated shoulder location identification methods applied to the Houston test set. Results are presented as percentage of shoulder location predictions which fall into each category: small deviations (less than 100), medium deviations (between 100 and 1000), and large deviations (greater than 1000).