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 or different sources. 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. This algorithm relies on removal of the overall curve of the bullet to obtain what [Hare et al., 2017] refer to as a "signature". Cross-correlation functions, consecutive matching straie, consecutive non-matching striae and other features are then calculated based on pairwise comparisons of these signatures. The currently established best practice for collection of 3D images of bullet LEAs requires capturing portions of the neighboring groove engraved areas (GEAs). Dealing with LEA and GEA data separately is crucial to aim for high accuracy and precision in subsequent feature calculations. However, 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. The following work describes a better solution to this pre-processing problem based on robust statistical models.

TODO:

1 Background

Forensic firearms examiners analyze bullets through a process of visual feature comparison. The main objective of visual feature comparison is to determine whether two patterns were generated by the same object. This problem is known in forensic science as the *same source-different source problem*, i.e. the problem of determining whether two patterns originated from the same source or different sources. In forensic firearms analysis, the same source-different source problem is focused on whether two bullets were propelled through the same gun barrel.

Striation marks, which are engraved on the surface of a bullet through contact with the rifling of a gun barrel, provide evidence to address the same source-different source problem. Alternating sections of the bullet that make the closest contact with the barrel are called land engraved areas (LEAs). 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.

Evidence is examined using a comparison microscope: two bullets in question are placed under the microscope and firearms examiners make a decision according to the AFTE Theory of Identification [AFTE Glossary, 1998].

In the last several decades, advances in technology have led to an increase in research focused on developing image-analysis algorithms to complete automated, quantitative analyses of bullet evidence. The main technological development that has created a pathway for image-analysis techniques is the introduction of high resolution 3D scanning technology to the field of forensic science [e.g. De Kinder et al., 1998, De Kinder and Bonifanti, 1999, Bachrach, 2002]. These 3D data have since been used in the development of several methods of varying complexity for automated comparison of land engraved areas [e.g. Ma et al., 2004, Chu et al., 2010, 2013, Hare et al., 2017].

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.

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. Features such as cross-correlation function, consecutive matching striae [see Biasotti, 1959], and consecutive non-matching striae are calculated based on pairwise comparisons of these signatures. These calculated features are the information used in a random forest model to assess the similarity of the corresponding pair of signatures.

Extraction of the aforementioned features depends on the step which translates a profile to a signature: the removal of the global strucutre. This step is complicated by the 3D data collection process.

While there is not yet an established standard operating procedure, the currently established best practice for collection of 3D images of bullet LEAs requires that scanning across the object must begin and end in the neighboring groove engraved areas (GEAs) as shown in Figure 1. However, this practice introduces a challenging data structure. The presence of extraneous GEA data dictates the most significant step in data pre-processing: correctly identifying between data from LEAs and GEAs.

Dealing with these two areas separately is crucial to aim for high accuracy and precision in subsequent processing steps. Removal of data from groove engraved areas is necessary to properly calculate features used in automated comparisons. In order to distinguish between these areas, we aim to identify "shoulder locations", the locations at which the LEA ends and the GEAs begin.

Distinguishing between land and groove engraved areas is a problem at which human vision excels, but it is quite challenging for automatic procedures due to the nature of the data collected: the bullet curvature presents the main structure in the data, but the abrupt change between land and groove engraved areas introduces a competing structure. The atypical structure overwhelms standard statistical modeling techniques which cannot distinguish between the structures. An early solution based on data smoothing (described in [Hare et al., 2017]) can result in misidentification of deep striae as shoulder locations. The following work describes a better solution to this pre-processing problem based on robust statistical methods.

2 Data Source

The data used in this paper consist of high resolution 3D scans of bullet land engraved areas (LEAs). The scanned bullets come from Hamby Set 44 [Hamby et al., 2009]. Each Hamby set consists of 35 bullets fired from 10 consecutively rifled Ruger P85 barrels. There are two known bullets for each of the ten barrels, as well as 15 additional questioned bullets in each set. 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 Unknowns, Bullet L, Land 5 – were removed from consideration due to "tank rash". Tank rash results from a bullet striking the bottom of 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 with a Sensofar Confocal light microscope at 20x magnification resulting in a resolution of 0.645 microns per pixel. These LEAs were scanned at Iowa State University's High Resolution Microscopy Facility, and the scans are stored in 3D format as x3p files, conforming to the ISO5436-2 standard [ISO 5436-2:2012, en]. A visualization of the data gathered for a single LEA is seen in Figure 1. 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 motivation for removal of GEA data lies in the algorithm proposed by Hare et al. [2017]; as such, the focus will be on horizontal slices (profiles) of the 3D scans. These slices capture the striation pattern horizontally across the surface, as seen in **??**.

The final data consists of 2D profiles gathered from 3D imaging. The height values in the profiles were averaged over several crosscuts spaced out along the 3D image. This ensures predicted locations will be relatively applicable across the depth of the LEA scan.

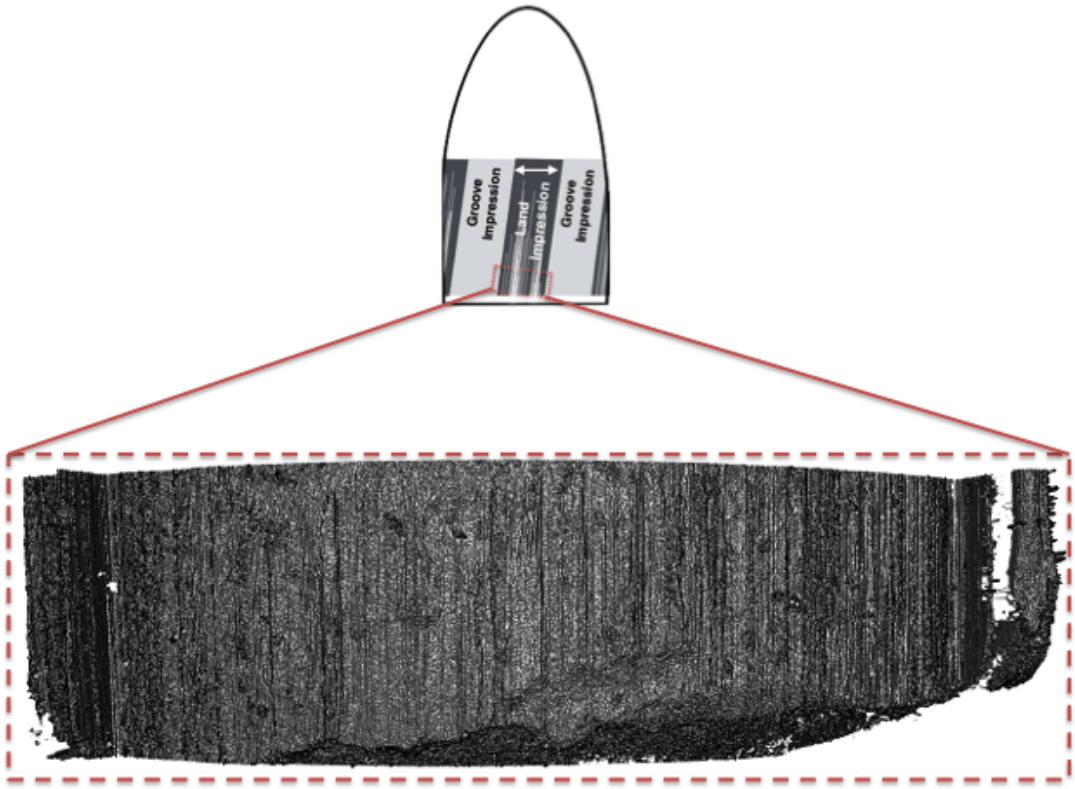

Figure 1: Visualization of 3D data collected through high resolution scanning of a land engraved area. Striations on the surface of the object can be seen by viewing this data from "above", as presented here.

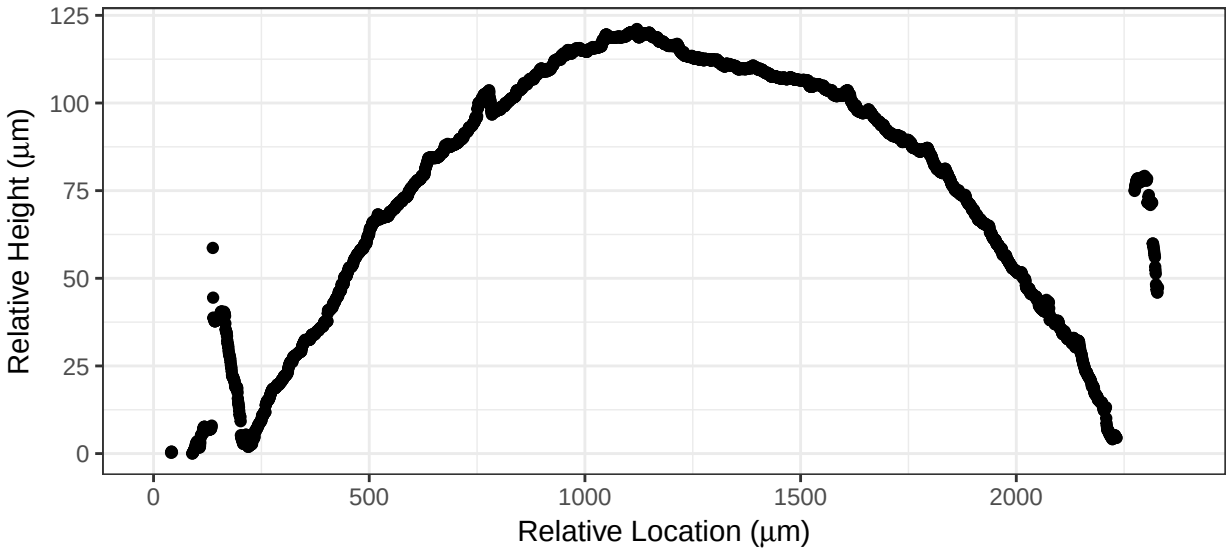

Figure 2: Single profile of 3D bullet land data. The main data structure, located in the center, is comprised of the land engraved area. The groove engraved areas occur to the left and right sides of the profile.

3 Methodology

The nature of the data structure is such that much of the variability found within the data is due to the global structure of the physical object; that is, the curve of a bullet. Since the goal is to identify where the global data structure changes (where the shoulder location is), methods need to be able to separate out the LEA structure from the GEA structure. This can most effectively be accomplished by fitting a line to the curve of the bullet and analyzing the pattern of deviations from that curve. That is, fitting a statistical model to the curve of the bullet and examining the residual values.

This 2-step process can be achieved with a statistical method which treats the secondary structure of the GEA as outlying data and fits the curve of the LEA alone. We examine and compare two methods for fitting the LEA structure. While the two approaches differ in methodology, they are both rooted in the ability to mitigate undue influence caused by outlying data.

Once an appropriate fit is obtained, the fitted curve is used to obtain residual values and determine a good cutoff which separates the two competing structures within the residuals.

3.1 Robust Linear Models

Due to the curved nature of the bullet, a quadratic linear model is a good candidate model. Linear models are based on minimizing the vertical squared distance between each data point and a fitted line. This means that if there are data points in unusual places, a linear model will fit a line that is pulled towards outlying data in order to minimize the overall sum of squared distances. In this particular data environment, the fitted curve can be easily influenced by GEA data, as is seen in **??**.

The robust approach under the linear model framework focuses on minimizing the least absolute deviations. This method of minimization is less influenced by possibly large outlying values present in the GEA data. Due to the overwhelming majority of data being from the LEA structure, minimizing the least absolute deviations will favor fitting the LEA structure and allowing GEA data to have large residual values. This is preferable to the traditional linear model, which corrects for the presence of GEA data by compromising between the two competing structures, and does not fit either structure appropriately. A striking example of the difference in results from these two model frameworks is seen in **??**.

Once a model has been fit for each crosscut, residual values are calculated. The model, having fit the structure of the LEA, results in small residuals scattered around zero in the LEA zone, and larger, mostly positive residuals in the GEA zones. Thus, the magnitudes of the residuals themselves can serve as one indicator of whether a data point is part of the LEA structure or the GEA structure. A cutoff value for the magnitude to separate the residuals from the two zones can be employed.

The median absolute deviation (MAD), is a robust metric for the spread of points, similar to the standard deviation. Let m be the median:

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

Since we are dealing with unbalanced residual values, the MAD is preferable to the standard deviation to quantify the spread. Removing unusually high residual values is effective at removing the GEA data once the global structure has been fit. Residual values that are considered unusually high, or outlying, are values that are larger than 4*MAD. Any residual value larger than 4 times the median absolute deviation value can be seen as a "large" residual.

Shoulder location predictions are calculated for each crosscut in the following manner:

1. Fit a robust linear model of order 2 (i.e., quadratic) to the averaged crosscut.
2. Calculate a residual value for each data point on the crosscut.
3. Calculate the median absolute residual (MAD) for the crosscut.

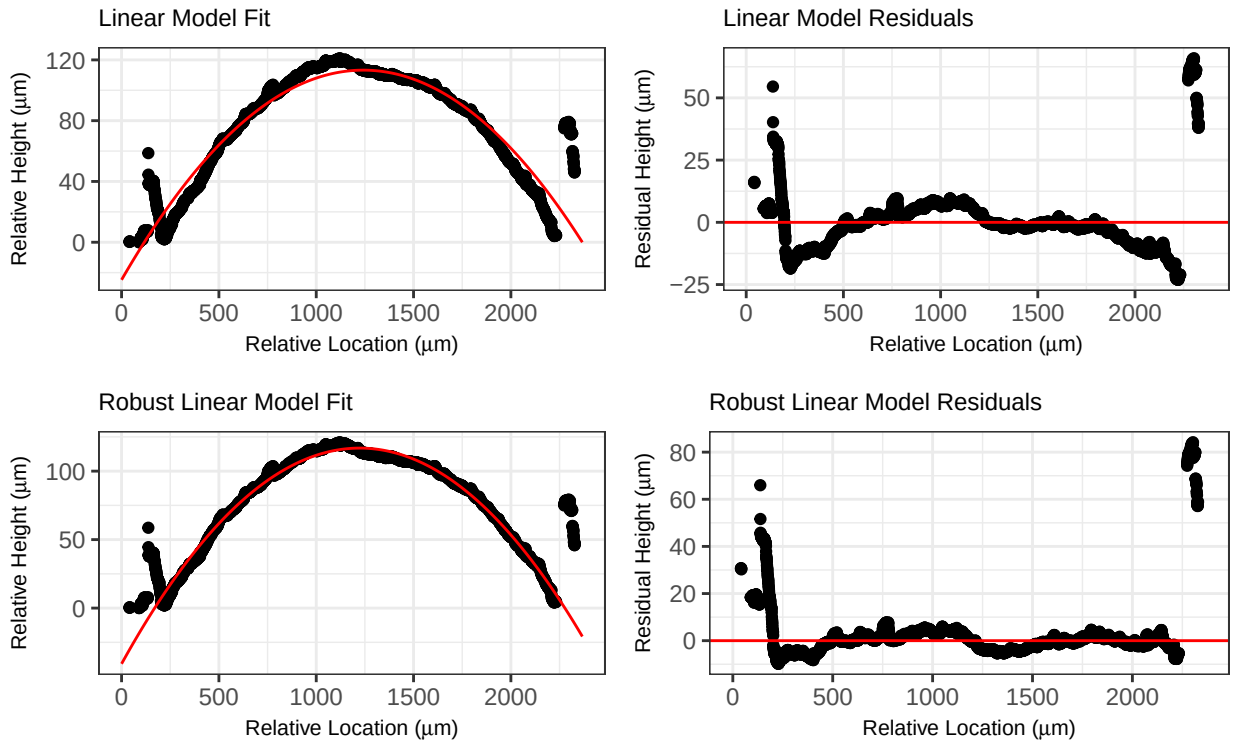

Figure 3: Example of a quadratic linear model fit and resulting residuals (top) compared to a robust quadratic linear model fit and residuals (bottom) for a single crosscut. The robust model is able to more effectively capture the curved structure of the LEA without being influenced by the GEA.

4. Remove all data points on the crosscut whose absolute residual value is greater than 4*MAD.
5. Identify the range of the remaining X values - these are the predicted shoulder locations for that crosscut.

3.2 Robust LOESS

Locally weighted regression, known as LOESS, is an approach that is not restricted by the need for perfect quadratic curvature. This is advantageous when working with bullets, as it is unrealistic to expect a flawless circular shape to remain after the bullet has been subjected to the forces of a gun barrel and striae have been impressed upon it.

LOESS fits many models to small subsets of the data and combines them into one non-parametric fit of the data, rather than focusing on the overall structure of the data. This allows for greater flexibility. However, it also means that LOESS models are affected by GEA structures in a much more unpredictable manner. A model fitted on a subset of data that mainly falls in the GEA structure will look vastly different than another model fit with data from the LEA. This results in a combined prediction that misrepresents much of the data near one or both shoulder locations (see **??**).

Just as the nature of LOESS models differs from traditional linear models, the robust approach must differ too. Robust LOESS utilizes an iterative process focused on re-weighting [see Cleveland, 1979]. First, an initial LOESS fit is created. This is followed by a step which gives smaller weights to data points with high residual values, and a subsequent LOESS fit with new weights applied. The down-weighting of values with high residual values slowly reduces the influence of a secondary structure within the data; here, the GEA data. This iterative process results in a non-parametric fit to the LEA structure that treats GEA data as less important, which is desirable in this context.

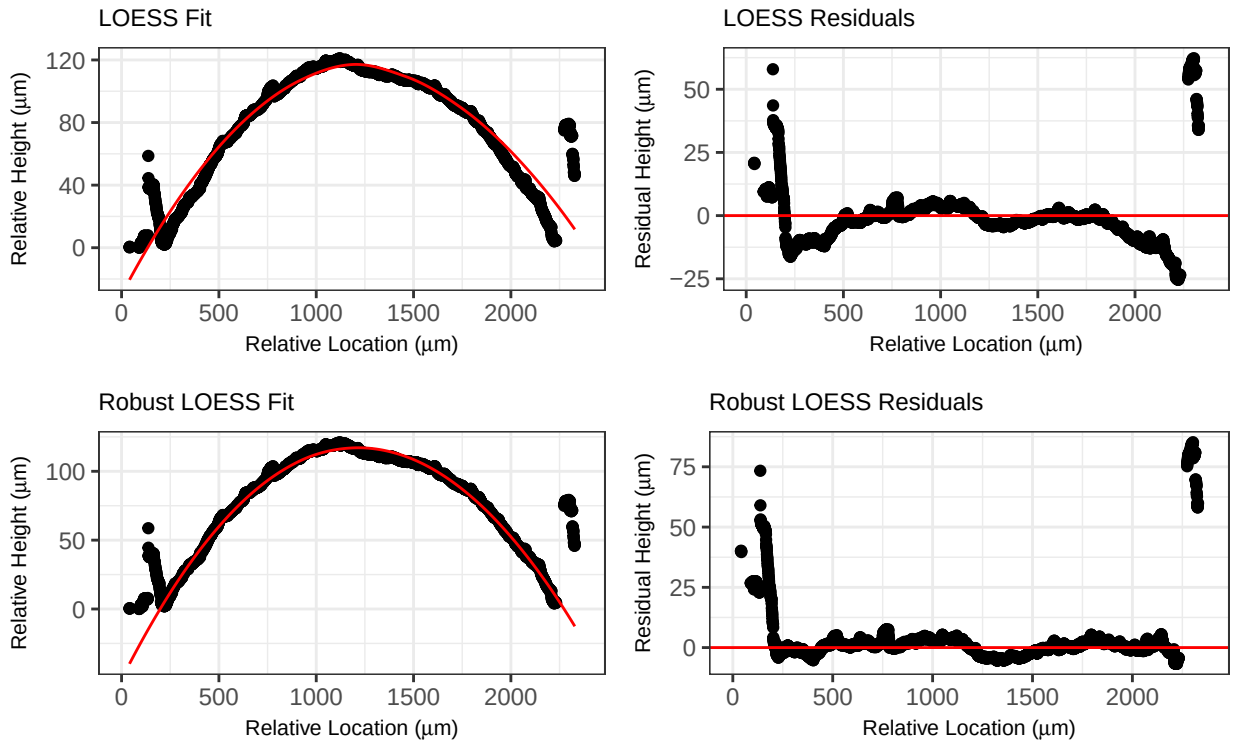

Figure 4: Example of a LOESS model fit and residuals (top) compared to a robust LOESS model fit and residuals (bottom) for a single crosscut. The robust model is again able to more effectively capture the curved structure of the LEA without being influenced by the GEA.

While robust LOESS methods are more flexible than robust linear models, a model that is accurately fit to the LEA structure will result in the same expected residual structure as with robust linear models: positive and negative residuals scattered around zero in the LEA zone, and positive, possibly large residuals in the GEA zones. A similar approach as with the robust linear model cutoff value to distinguish between "large" residual values and reasonable ones; however, because this model is more flexible and fits more closely to the specific data at hand, our cutoff will be lower. A cutoff that performs well on the Hamby set 44 is twice the median absolute deviation (2*MAD).

Shoulder location predictions are calculated for each crosscut in the following manner:

1. Fit a robust LOESS model with a span of 1 to the averaged crosscut. This is fit using the 'locfit.robust' function in the 'locfit' package in R.
2. Calculate a residual value for each data point on the crosscut.
3. Calculate the median absolute deviation (MAD) for the crosscut.
4. Remove all data points on the crosscut whose absolute residual value is greater than 2*MAD.
5. Identify the range of the remaining Y values - these are the predicted shoulder locations for that crosscut.

4 Results

In order to assess the accuracy of these predictions, we take a unique approach. To calculate a quantitative measure for the overall performance of predictions, we first manually identified "ground truth" shoulder locations.

The numerical comparison of predicted and manually identified locations can be tricky; distance metrics alone do not suffice to represent the true character of a prediction's accuracy. For example, a predicted shoulder location that falls 10 data points away from the manually identified shoulder location could be caused by noise in the data, missing data points, or simply the scale of the data. Note that a span of 10 data points represents only 6.45 microns in physical space. Alternatively, a distance of 10 points could indeed be part of the groove engraved area, and thus being incorrectly identified could potentially cause problems in subsequent analyses.

Thus, a more suitable measure is to investigate the residual values resulting from the robust LOESS model that fall in the range spanned by the predicted and manually identified shoulder location. This method penalizes shoulder location predictions that are too far out to the side and leave GEA data in the main structure.

Because the robust LOESS most reliably approximates the curved shape of the bullet land due to its flexibility, we want to use residual values resulting from that model to assess final performance of all methods. Residual values from the GEA will not necessarily be uniformly large, but are expected to be positive as their structure and the modeling technique dictate that they would fall above the fitted line from robust LOESS.

Given the assumption of positive residuals in the GEA, even a 10-point difference can quickly add to a large residual sum. However, a 10-point difference within the land engraved area will be balanced out by the presence of both positive and negative residual values and remain closer to zero.

For this reason, gathering the sum of residuals between the predicted location and the manually identified location is appropriate. This residual sum is referred to as an "inaccuracy score" for which higher values indicate a higher level of inaccuracy. An inaccuracy score was calculated separately for the left hand side and right hand side predictions for each crosscut in the data set.

Of interest are the distributions of these inaccuracy scores across all 208 lands used in the study (see Figure 5). 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 inaccuracy scores 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 inaccuracy scores 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 6). Unsatisfactory cases are the most likely to cause mistakes in subsequent analyses.

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 could 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, crosscuts can have tilted shapes relative to the x-axis which a quadratic linear model would fail to address. In these situations, LOESS excels.

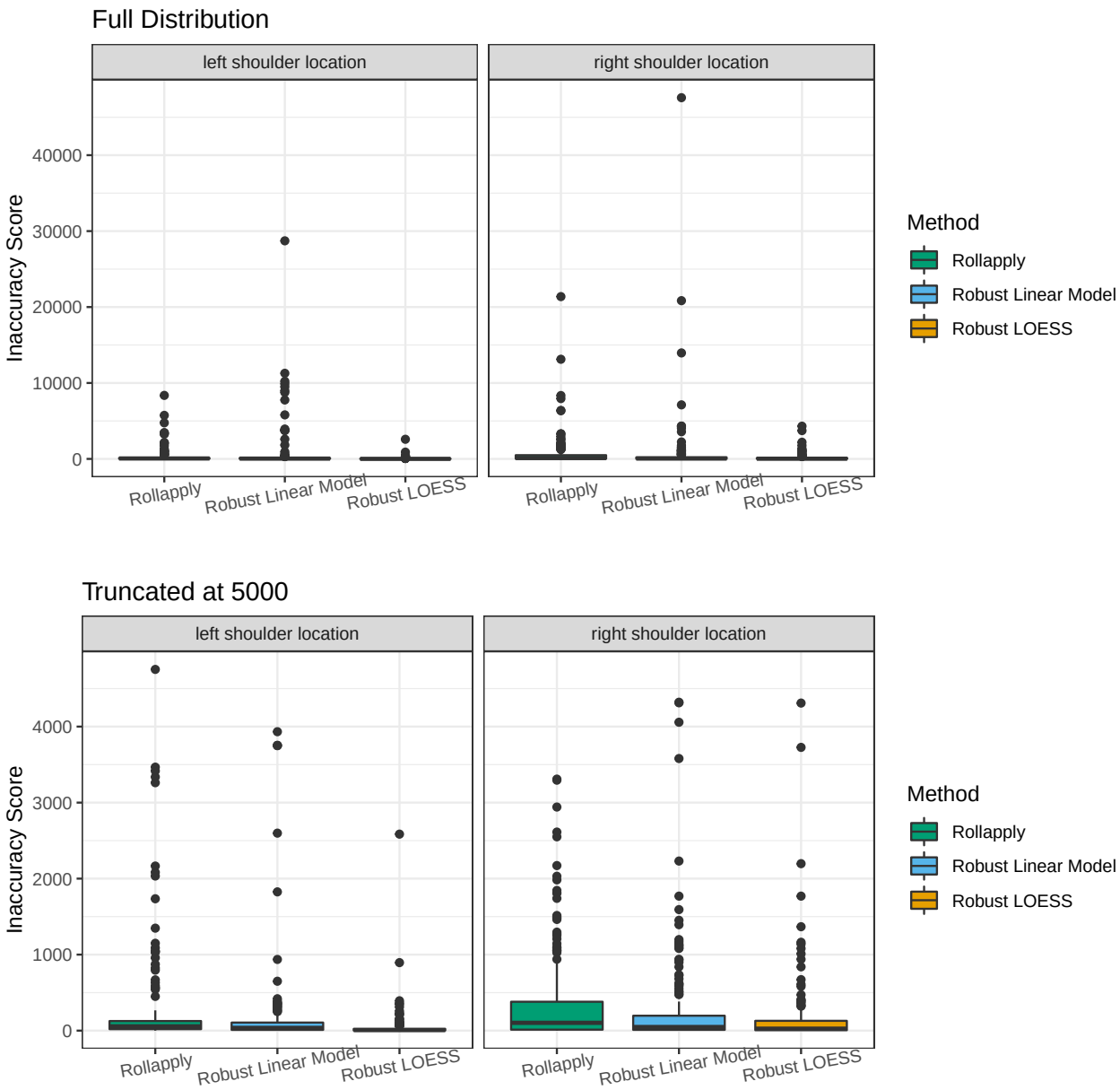

Figure 5: Distribution of inaccuracy scores for data smoothing method, robust linear model method, and robust LOESS method, separated by left and right shoulder locations. A tight distribution with few high values indicates good performance across the LEAs in the data set.

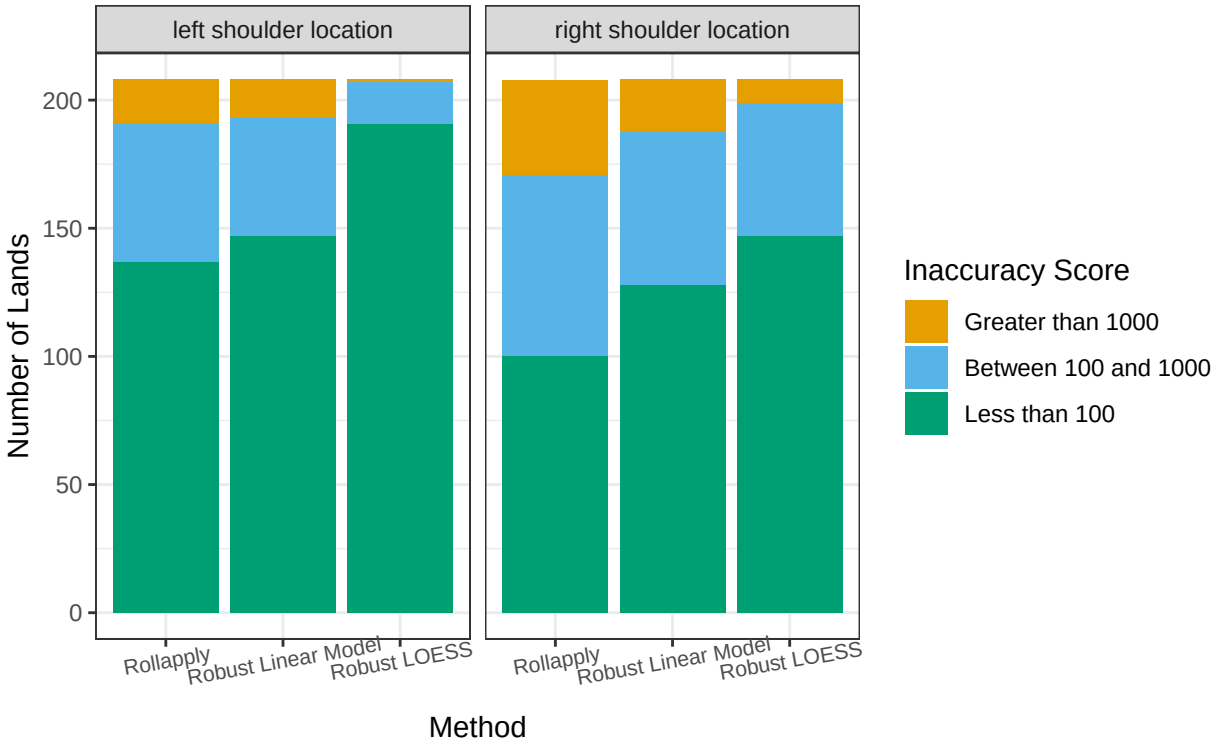

Figure 6: Distribution of inaccuracy scores for data smoothing method, robust linear model method, and robust LOESS method, separated by left and righth shoulder locations. Inaccuracy scores are placed into three categories: less than 100 microns (satisfactory), between 100 and 1000 microns, and greater than 1000 microns. A larger proportion of inaccuracy scores under 100 microns indicates good performance across LEAs in the data set.

While the cutoff values presented work well on Hamby set 44, additional cross-validation will need to be implemented on a variety of bullet 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 a robust LOESS data pre-processing strategy on overall automated image-analysis methods will need to be addressed. 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.

6 References

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.