Statistical Approaches to Automated Groove Engraved Area Identification in 3D Bullet Land Scans

Kiegan Rice *
Department of Statistics, Iowa State University
and
Nathaniel Garton
Department of Statistics, Iowa State University
and
Ulrike Genschel
Department of Statistics and CSAFE, Iowa State University
and
Heike Hofmann
Department of Statistics and CSAFE, Iowa State University

April 16, 2019

Abstract

Keywords: 3 to 6 keywords, that do not appear in the title

*The authors gratefully acknowledge ...

1 Background

In forensic firearms analysis, visual feature comparison is used to analyze bullets to address the same source-difference source problem. Striation marks act as features which provide evidence to help determine whether two bullets were propelled through the same gun barrel. Land engraved areas (LEAs), alternating sections of the bullet that make the closest contact with the gun barrel, are the primary source of striation marks.

A cornerstone of 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. The AFTE Theory of Identification (1) is used to make decisions when comparing two bullets under a comparison microscope: Add additional details from AFTE Theory of Identification here.

Technological advances coupled with concerns about the objectivity of visual comparison have produced several image-analysis algorithms which aim to complete automated, quantitative analyses of bullet evidence.

Many of these efforts have capitalized on the introduction of high resolution 3D scanning technology to forensic science (see 2; 3; 4). 3D scans of land engraved areas, such as the one seen in Figure 1, have been used to develop several methods for the automated comparison of bullet lands (e.g. 5; 6; 7; 8).

One such method, proposed by Hare et al. (8), is a random forest algorithm based on features calculated from 2D horizontal slices of the 3D images. These slices, called profiles, have a data structure which is dominated by the global structure of the bullet land: they are curved. Feature comparison in firearms analysis is concerned with the similarity of straitions impressed by the barrel; whether two bullets are both curved does not answer the forensic inquiry. As such, a crucial step is to remove the global structure and distill the information down to the deviations from the curve. The remaining pattern of peaks and valleys - called a signature - is a much more useful representation of the striations. The process of translating a 3D scan into a 2D signature is demonstrated in Figure 2.

While removal of a curve from data is typically a straightforward statistical problem, 3D scans of LEAs contain a unique data structure which obfuscates this task. Currently established best practice for collection of 3D images of bullet LEAs dictates that scanning

begin and end slightly past the edges of the LEA, in the neighboring groove engraved areas (GEAs). This introduces a secondary structure that needs to be removed before algorithms can be reliably applied to data.

Correctly separating LEA and GEA data is one of the most challenging and important aspects of data pre-processing. Without removing GEA data, algorithms will be using extraneous data which can misrepresent the character of the striations present on a given LEA. This type of misrepresentation is demonstrated in Figure 3. In order to distinguish between LEA and GEA data, we aim to identify "shoulder locations", the locations at which the LEA ends and the GEAs begin.

The method for shoulder location identification described in Hare et al. (8), based on data smoothing and local minima, fails to reliably separate the GEA data from LEA data. The following work first describes a robust methodology for fitting the global curvature of 2D profiles. Once global structure is removed, two different approaches to identify shouler locations are presented. Performance of both proposed shoulder location identification methods are then compared with current approaches.

2 Data Source

The data used consist of high resolution 3D scans of bullet LEAs from three separate test sets: Hamby Set 44 (9), Phoenix PD set, and Houston-test set.

Hamby set 44 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. 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 land engraved areas. 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 Phoenix PD set consists of 33 bullets fired from 8 barrels more information on the type of barrels?. There are three known bullets for each of the eight barrels, as well as 9 additional questioned bullets. There are 6 LEAs for each of the fired bullets, producing a

total of 198 individual land engraved areas.

The Houston-test set consists of 45 bullets fired from at least 10 barrels more information on the type of barrels?. There are three known bullets for each of the ten barrels, as well as 24 questioned bullets that are fired from a combination of the 10 known barrels and additional, out-of-set barrels. There are 6 LEAs for each of the fired bullets, producing a total of 414 individual land engraved areas.

The 3D scans of each test set 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 (10). 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.

A crosscut was extracted from each scan by identifying an optimal crosscut using `x3p_crosscut_optimize` in the `bulletxtrctr` package in `R` (11). Then, an averaged crosscut was calculated by averaging across ten consecutive crosscuts which fall directly to either side of the optimal crosscut identified. Shoulder location methods are applied to these averaged crosscut, for a total of 820 crosscuts across the three test sets.

3 Methodology

3.1 Global Structure Removal

The first stage of our GEA removal process focuses on removing the bullet curvature from each crosscut. GEAs on the edge of each LEA represent a change in the data structure, so removal of the primary structure will result in two distinct structures remaining in the residuals. These two structures can then be separated using features of the remaining residuals.

Although bullets are circular, the pressure they are subjected to in the process of being fired out of a gun can often result in LEAs that are not perfectly circular. In addition, the angle at which the LEAs are scanned by operators can mean crosscuts are tilted or mis-

shapen. A quadratic model may then be unreasonable to remove the structure accurately across a whole crosscut. Thus, non-parametric locally weighted regression (LOESS) is a natural choice for removing the structure.

To mitigate boundary effects when applying LOESS, we employ a robust version (12). Robust LOESS fits models by iteratively downweighting high residual values to reduce the influence of outliers which deviate the most from the main data structure. In the context of LEA crosscuts, robust LOESS will iteratively downweight data points with high residuals, typically those from the GEA which deviate most severely from the overall bullet curvature.

Traditional robust LOESS can be implemented using the `locfit.robust` function in the `locfit` package in `R` (13). An example of the difference in both predicted values and residual structure between `loess` and `locfit.robust` can be seen in Figure 4.

When applied to LEAs from Hamby set 44, the traditional robust LOESS procedure appears to work well. However, the method fails to mitigate effects when striations are deeper or GEA structures are more pronounced, such as those in the Houston-test set, seen in Figure 5.

Thus, we instead apply an adapted version of the traditional robust LOESS procedure which focuses on iteratively reducing the influence of only positive residuals. The procedure is as follows:

1. Fit a LOESS model with $span = 1$ to an entire LEA crosscut to predict *height* using values of x. Assign weights of 1 to each data point for this fitting procedure.

2. Obtain predicted values of *height* from the model fit in step 1.

3. Calculate residual values using the predicted *height* values.

4. Calculate bisquare weights for each residual value using the following formula:

$$\max(1 - (residual/(6 * MAR))^2, 0)^2,$$

   where MAR is the median absolute residual for the crosscut.

5. Assign weights to each data point according to its residual value. If the residual value is positive, assign the bisquare downweight. If the residual is zero or negative, leave the weight at 1.

6. Repeat steps 1-5 with updated weights at each iteration for k iterations, with 20 iterations as the default.

7. After k iterations of updating the weight vector, fit a LOESS model with $span = 1$ and obtained predicted and residual values for $height$.

This procedure is attractive in the LEA crosscut context because the main influence of GEA data is to pull predictions up towards the higher GEA height values. Downweighting high residuals for only those that fall above the fitted line will iteratively bring the fitted line down towards the LEA structure. While this is a small procedural difference, the results are dramatic. The same LEA pictured in Figure 5 can be seen again with updated predictions in Figure 6.

The utility of this adaptation can be seen in Figure 7, which shows the mean shift in predicted values from the robust LOESS to the adapted robust LOESS for our three different test sets of interest. While the difference is almost imperceptible for Hamby set 44, the Phoenix PD and Houston-test sets demonstrate visible downward shifts near both the left and right boundaries.

Once the adapted robust LOESS procedure has been applied to remove the bullet curvature of the LEA, the resulting residuals should have relatively different distributions. Points in the LEA will fall more closely to the fitted line, while points in the GEA will typically have high, positive residuals.

The subsequent prediction methods for shoulder location are based on the residuals calculated from the fit to the global structure of each land. One method uses penalized two-class classification techniques to classify each data point into "LEA" or "GEA", while the second uses Bayesian changepoint analysis to predict the data points at which the shoulders begin on either side.

3.2 Two-Class Classification

Shoulder location can be predicted by classifying data points as one of two classes: "LEA" or "GEA". A shoulder location prediction can then be found by gathering the range of values classified as "LEA" points.

Classification into "LEA" or "GEA" is first approached by a process of feature engineering based on adapted robust LOESS residuals. While the residuals themselves should have different height patterns, residuals alone are not enough to classify data with high accuracy.

As already demonstrated by the differences in Figure 4 and Figure 5, certain ammunition or barrel types may result in more pronounced striae on an LEA. Thus, standardization of features is imperative for transferability of fitted model parameters. The LEA scan context requires some non-traditional standardization practices.

For example, consider the distribution of residual values resulting from adapted robust LOESS. There is reason to believe that the distribution will be quite skewed, which means a standard deviation will not be a good proxy for the spread of the distribution. Thus, rather than the standard deviation, we consider instead the standard deviation of residual values from the middle 50% of x values present on the scan. This alternative acts as a proxy for the depth of striae on each LEA, with higher standard deviations found for lands with deeper striae. Standardizing residual values by this proxy puts all residuals on a similar scale.

For variables which instead deal with differences in the x direction, such as depth from the center of a scan, values will be mapped to a $(0, 1)$ range, with the maximum x location on the scan acting as the divisor.

The full list of features developed are as follows:

`rlo_resid_std`: Robust LOESS residual value, standardized by dividing by standard deviation of residual values from middle 50% of **x** values.

`(rlo_resid_std)`$^{\mathbf{2}}$: Squared term of `rlo_resid_std`.

`side`: Whether data point is to left or right of median x value.

`depth_std`: Distance of data point from median x value, standardized by dividing by maximum x value (a proxy for the range of x).

`side:depth_std`: Interaction between `side` and `depth_std` variables.

`xint1_std`: Predicted x intercept of robust LOESS on left side of land, standardized by dividing by maximum x value (a proxy for the range of x).

`xint2_std`: Predicted x intercept of robust LOESS on right side of land, standardized by dividing by maximum x value (a proxy for the range of x).

`range_50_std`: Range of residual values within a 50-point window around data point, standardized by dividing by standard deviation of residual values from middle 50% of x values.

`numNA_50`: Number of missing values within a 50-point window around data point.

`ind_2mad`: Indicator of whether `rlo_resid` is greater than `2*MAD(rlo_resid)`.

`numpos_50`: Number of positive residual values within a 50-point window around data point.

`ind_edges`: Indicator of whether data point is to the left of `xint1` or to the right of `xint2`. Values between `xint1` and `xint2` receive a value of 0, while values on the outside of the two values receive a value of 1.

Examples of the distributions of some of these features can be seen in Figure 8.

The developed features can be used in the fit of a logistic LASSO model, which is a form of penalized regression. LASSO parameter values for p covariates are found by identifying:

$$\hat{\beta}_\lambda^q = \overset{\arg\min}{\beta \in \mathbb{R}^p} \left\{ (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|^q \right\}$$

which adds a penalty to the traditional ordinary least squares minimization problem, and uses a tuning parameter λ.

Cross-validated LASSO models were fit using the `cv.glmnet` function in the `glmnet` package (14). For each model, parameter values from the model with λ_{1se} were used. λ_{1se}, a standard when using LASSO, is the tuning parameter which results in the simplest model that still has cross-validation error within one standard deviation of the best model.

Two separate models were fit: "LASSO Basic", which uses each of the features listed above, and "LASSO Full", which uses each of the features along with pairwise interactions

for each of them. These parameter values were trained on the Hamby set 44 data. The resulting parameter values were used to calculate predicted values between 0 and 1; the closer to 1, the higher probability of membership in the "GEA" class.

Traditional two-class classification techniques call for finding an **equal error rate cutoff** to classify the predicted values for each data point into each of the two classes; i.e., values above a certain cutoff are classified as part of the "GEA" class, and values below the cutoff are classified as part of the "LEA" class. However, since scans are primarily of the land engraved area, there is significant class imbalance in our response categories. This unbalanced response means that any criterion used to determine a reasonable cutoff value for classification needs to be adjusted to account for differences in the size of the response classes. The larger class size of "LEA" means an equal error rate would allow for more false negatives (i.e., more "GEA" data points being classified as "LEA") - exactly what we are trying to avoid.

Thus, instead of looking for the cutoff that gives raw equal error rate (where sensitivity and specificity are equal), we found an equal error rate based on the overall number of data points. This allows for an equal **number** of errors in each category, rather than equal **percentage** of errors. This tactic more fairly penalizes points that should be classified "GEA" but are predicted to be "LEA".

The final process for shoulder location identification is as follows:

1. Use `robust_loess_fit` to remove bullet curvature.
2. Calculate features based on x locations and residual height values from `robust_loess_fit`.
3. Use fit parameter values from either (a) lassobasic or (b) lassofull models to calculate probabilities of membership in GEA class.
4. Apply cutoff of (a) .33 or (b) .34 to classify higher probabilities as GEA data points, and lower probabilities as LEA data points.
5. Gather the x range of points classified as LEA data points. These are the final shoulder location preditions.

This process results in two methods for shoulder location identification in the `R` package `bulletxtrctr`: "lassobasic" and "lassofull".

3.3 Bayesian Changepoint Analysis

The idea behind the changepoint approach is that within either the left GEA, right GEA, or the LEA, the global structure is consistent and can either be described by a line with zero slope, a line with positive slope for the right GEA, or a line with negative slope for the left GEA. Finding the points where the GEAs and LEA meet is treated as a problem of model selection. That is, the best fitting statistical model, in terms of the magnitude of the likelihood, should be the one which assumes that the points at which the global structure changes align with where the GEAs and LEA meet. This approach was proposed in the more general context of Bayesian changepoint detection in Stephens (15). Thus, the points of global structural change are what we will call changepoints. Thus, our model will be defined in a piecewise fashion. In practice there are also complex additional patterns which may exist for a number of reasons, but this large scale structural assumption remains generally reasonable. The complex smaller scale patterns can be thought of as the dependence in the data after accounting for the global structure. Because of the nature of the model which we consider, it becomes necessary for computational reasons to perform a couple of additional data preprocessing steps. Specifically, we will scale the residuals from the robust LOESS procedure, and we will impute missing values. In the next section, we describe the model that we will use to identify changepoints, after which we will describe the estimation procedure which we use. Details of the additional data preprocessing steps can be found in the appendix.

3.3.1 Bayesian Model Formulation

Before introducing the model, we introduce some notation. First, let $\{Y(x_i) : i = 1, 2, ..., n\}$ denote the set of random variables representing the residuals from the robust LOESS procedure at the values x_i. For simplicity, also assume that $x_1 < x_2 < ... < x_n$. Also, let c_l be the value of the left changepoint and c_r be the value of the right changepoint. Here, the left changepoint is where the left GEA meets the LEA, and the right changepoint is where

the right GEA meets the LEA. Also, denote the median centered x values as $x_i' = x_i - \tilde{x}$ where $\tilde{x}$ is the median x value. As mentioned in the previous paragraph, the complex small scale patterns, such as the striae, will be modeled through a covariance structure on the data that will be allowed to differ between each GEA and between the GEAs and LEA. We will construct the covariance matrices from the exponential covariance function $K(x, x'; \sigma, \ell) = \sigma^2 e^{-\frac{|x-x'|}{\ell}} = cov(Y(x), Y(x'))$. The differences in covariance matrices for the GEAs and LEA will be reflected in the parameters σ and ℓ. The data model that we consider is then,

$$(Y(x_1), Y(x_2), ..., Y(x_{k_1}))^\top \sim N(\beta_{01}\mathbb{1} + \beta_{11}x_{1:k_1}, \Sigma_1(\sigma_1, \ell_1)) \quad (1)$$

$$(Y(x_{k_1+1}), Y(x_{k_1+2}), ..., Y(x_{k_2}))^\top \sim N(0, \Sigma_2(\sigma_2, \ell_2)) \quad (2)$$

$$(Y(x_{k_2+1}), Y(x_{k_2+2}), ..., Y(x_n))^\top \sim N(\beta_{02}\mathbb{1} + \beta_{12}x_{k_2+1:n}, \Sigma_3(\sigma_3, \ell_3)), \quad (3)$$

where $x_{k_1} < c_l \leq x_{k_1+1}$ and $x_{k_2} < c_r \leq x_{k_2+1}$ Here, $x_{1:k}$ denotes the column vector $(x_1, x_2, ..., x_k)^\top$, and $\mathbb{1}$ denotes the vector of ones. Indpendence is assumed between each of these three distributions for simplicity. The parameters that need to be estimated include the four mean parameters in the GEAs, the six covariance parameters (two for each of the three areas), and the two changepoint parameters, c_l and c_r.

The above model encapsulates the essence of the approach. However, there are a few difficulties. The first difficulty is that there are not always two GEAs in a particular land. There may be one GEA, or the land may only consist of the LEA. Thus, the above model is actually conditional on there being two GEAs in the data. We also define models for when there is one GEA on the left, one GEA on the right, or no GEAs. The models are defined in an essentially identical way. Conditional on there being only one GEA, the left GEA model is defined as,

$$(Y(x_1), Y(x_2), ..., Y(x_k))^\top \sim N(\beta_0\mathbb{1} + \beta_1 x_{1:k}, \Sigma_1(\sigma_1, \ell_1)) \quad (4)$$

$$(Y(x_{k+1}), Y(x_{k+2}), ..., Y(x_n))^\top \sim N(0, \Sigma_2(\sigma_2, \ell_2)), \quad (5)$$

and the right GEA model is defined as,

$$(Y(x_1), Y(x_2), ..., Y(x_k))^\top \sim N(0, \Sigma_1(\sigma_1, \ell_1)) \tag{6}$$

$$(Y(x_{k+1}), Y(x_{k+2}), ..., Y(x_n))^\top \sim N(\beta_0 \mathbb{1} + \beta_1 x_{k+1:n} \Sigma_2(\sigma_2, \ell_2)). \tag{7}$$

Finally, conditional on there being no GEAs in the data, the model is simply

$$(Y(x_1), Y(x_2), ..., Y(x_n))^\top \sim N(0, \Sigma(\sigma, \ell)). \tag{8}$$

We see that estimating the changepoint locations also involves selecting the most appropriate model. In order to avoid confusion, we have slightly abused notation and, for example, $\Sigma_1(\sigma_1, \ell_1)$ as it is estimated in the two changepoint model is *not* the same as $\Sigma_1(\sigma_1, \ell_1)$ from either of the one changepoint models, and $\Sigma_1(\sigma_1, \ell_1)$ is also *not* the same between the two one changepoint models. As another example, β_0 is *not* the same between each of the one changepoint models. So, to be clear, duplication of notation in *different* models is not meant to imply that those parameters are shared between models.

Ultimately, these above four models are each individually fitted, and each model above is given a prior. From there, we do model selection in the formal Bayesian way, selecting number and location of changepoints by maximizing the estimated posterior distribution.

In order to complete a Bayesian model specification, we need priors on each of the parameters in each model as well as each model itself. We will assume independence between each parameter a priori. For each length scale ℓ, we will assume $\ell \sim \text{Gamma}(3, 5)$. For each standard deviation, we will assume $\sigma \sim \text{Half-Normal}^+(0, 1)$, where $\text{Half-Normal}^+(\cdot, \cdot)$ is notation for the normal distribution restricted to the positive real numbers. For intercept parameters, $\beta_{01}, \beta_{02}, \beta_0 \sim N(0, 10)$. For the slope parameters, the preceding trend deviates slightly. For any slope that corresponds to the *left* GEA, β_1 or β_{01}, we will assume that the slope can not be positive. That is, $\beta_1, \beta_{01} \sim \text{Half-Normal}^-(0, 10)$, where $\text{Half-Normal}^-(\cdot, \cdot)$ is notation for the normal distribution restricted to the negative real numbers. Contrastingly, for any slope that corresponds to the *right* GEA, β_1 or β_{02}, we will assume that the slope can not be negative. That is, $\beta_1, \beta_{01} \sim \text{Half-Normal}^+(0, 10)$. For the changepoint locations, we assume a uniform prior $\pi(c_l, c_r) \propto I(a < c_l < c_r - \gamma < b - \gamma)$. Here, a and b are some values close to the edges of the data. How close those values are to the edges is a

parameter that is set manually. Further, we include another hyperparameter, γ, which can be set so that the changepoints are not allowed to be too close to each other. This is also a parameter that is set manually. Lastly, we assume a uniform prior over all four models.

3.3.2 Bayesian Model Estimation

As was noted in Stephens (15), for any model including a changepoint, the likelihood is not a smooth function of the changepoint location. This is because, holding all other parameters fixed, shifting the changepoint value will result in zero change to the likelihood until it crosses the nearest point to the right or left, at which point the likelihood makes a jump. This makes maximum likelihood estimation in the standard way infeasible, but Bayesian estimation can be done in a fairly straightforward way via Markov chain Monte Carlo (MCMC). The basic idea is that, for each model, we can construct a two step Gibbs sampler. In step 1 we sample from the posterior distribution of the mean and covariance parameters given the changepoint locations, and in step 2 we sample from the changepoint locations given the mean and covariance parameters. Because of the non-conjugacy in our model, we perform both sampling steps using a random walk Metropolis-Hastings (RWMH) step with Gaussian proposals. For details on Gibbs sampling and the Metropolis-Hastings algorithm see Gelman et al. (16). It is also worth mentioning that the zero changepoint model does not require Gibbs sampling at all, and we perform estimation there using a RWMH algorithm.

We now provide the two basic steps of the Gibbs sampler for the two changepoint case. The algorithms to sample from the other three models are omitted, and are nearly identical except for the a smaller number of parameters need to be sampled. Denote collection of mean and covariance parameters for the left GEA as θ_1, the LEA as θ_2, and the right GEA as θ_3. Then, at iteration t after warmup

1. given changepoint locations $(c_l^{(t-1)}, c_r^{(t-1)})$, sample $(\theta_1^{(t)}, \theta_2^{(t)}, \theta_3^{(t)})$ using independent RWMH steps for each θ_i
2. given $(\theta_1^{(t)}, \theta_2^{(t)}, \theta_3^{(t)})$, sample $(c_l^{(t)}, c_r^{(t)})$ using a single RWMH step.

After running the MCMC for each model, parameter estimates and the most likely model are jointly chosen according to the largest joint posterior value. That is, we arrive

at estimates $(\hat{\theta}, \hat{M}) = \underset{(\theta,M)}{\text{argmaxlog}}(p(\theta, M|Y))$, where M is the random variable associated with the choice of model, θ is the associated parameter vector for the appropriate model, and Y is all of the available data. Additional MCMC details can be found in the appendix.

The idea behind the changepoint approach is that within either the left GEA, right GEA, or the LEA, the global structure is consistent and can either be described by a line with zero slope, a line with positive slope for the right GEA, or a line with negative slope for the left GEA. Finding the points where the GEAs and LEA meet is treated as a problem of model selection. That is, the best fitting statistical model, in terms of the magnitude of the likelihood, should be the one which assumes that the points at which the global structure changes align with where the GEAs and LEA meet. This approach was proposed in the more general context of Bayesian changepoint detection in Stephens (15). Thus, the points of global structural change are what we will call changepoints. Thus, our model will be defined in a piecewise fashion. In practice there are also complex additional patterns which may exist for a number of reasons, but this large scale structural assumption remains generally reasonable. The complex smaller scale patterns can be thought of as the dependence in the data after accounting for the global structure. Because of the nature of the model which we consider, it becomes necessary for computational reasons to perform a couple of additional data preprocessing steps. Specifically, we will scale the residuals from the robust LOESS procedure, and we will impute missing values. In the next section, we describe the model that we will use to identify changepoints, after which we will describe the estimation procedure which we use. Details of the additional data preprocessing steps can be found in the appendix.

3.3.3 Bayesian Model Formulation

Before introducing the model, we introduce some notation. First, let $\{Y(x_i) : i = 1, 2, ..., n\}$ denote the set of random variables representing the residuals from the robust LOESS procedure at the values x_i. For simplicity, also assume that $x_1 < x_2 < ... < x_n$. Also, let c_l be the value of the left changepoint and c_r be the value of the right changepoint. Here, the left changepoint is where the left GEA meets the LEA, and the right changepoint is where the right GEA meets the LEA. Also, denote the median centered x values as $x_i' = x_i - \tilde{x}$

where $\tilde{x}$ is the median x value. As mentioned in the previous paragraph, the complex small scale patterns, such as the striae, will be modeled through a covariance structure on the data that will be allowed to differ between each GEA and between the GEAs and LEA. We will construct the covariance matrices from the exponential covariance function $K(x, x'; \sigma, \ell) = \sigma^2 e^{-\frac{|x-x'|}{\ell}} = cov(Y(x), Y(x'))$. The differences in covariance matrices for the GEAs and LEA will be reflected in the parameters σ and ℓ. The data model that we consider is then,

$$(Y(x_1), Y(x_2), ..., Y(x_{k_1}))^\top \sim N(\beta_{01}\mathbb{1} + \beta_{11}x_{1:k_1}, \Sigma_1(\sigma_1, \ell_1)) \tag{9}$$

$$(Y(x_{k_1+1}), Y(x_{k_1+2}), ..., Y(x_{k_2}))^\top \sim N(0, \Sigma_2(\sigma_2, \ell_2)) \tag{10}$$

$$(Y(x_{k_2+1}), Y(x_{k_2+2}), ..., Y(x_n))^\top \sim N(\beta_{02}\mathbb{1} + \beta_{12}x_{k_2+1:n}, \Sigma_3(\sigma_3, \ell_3)), \tag{11}$$

where $x_{k_1} < c_l \leq x_{k_1+1}$ and $x_{k_2} < c_r \leq x_{k_2+1}$ Here, $x_{1:k}$ denotes the column vector $(x_1, x_2, ..., x_k)^\top$, and $\mathbb{1}$ denotes the vector of ones. Indpendence is assumed between each of these three distributions for simplicity. The parameters that need to be estimated include the four mean parameters in the GEAs, the six covariance parameters (two for each of the three areas), and the two changepoint parameters, c_l and c_r.

The above model encapsulates the essence of the approach. However, there are a few difficulties. The first difficulty is that there are not always two GEAs in a particular land. There may be one GEA, or the land may only consist of the LEA. Thus, the above model is actually conditional on there being two GEAs in the data. We also define models for when there is one GEA on the left, one GEA on the right, or no GEAs. The models are defined in an essentially identical way. Conditional on there being only one GEA, the left GEA model is defined as,

$$(Y(x_1), Y(x_2), ..., Y(x_k))^\top \sim N(\beta_0\mathbb{1} + \beta_1 x_{1:k}, \Sigma_1(\sigma_1, \ell_1)) \tag{12}$$

$$(Y(x_{k+1}), Y(x_{k+2}), ..., Y(x_n))^\top \sim N(0, \Sigma_2(\sigma_2, \ell_2)), \tag{13}$$

and the right GEA model is defined as,

$$(Y(x_1), Y(x_2), ..., Y(x_k))^\top \sim N(0, \Sigma_1(\sigma_1, \ell_1)) \tag{14}$$

$$(Y(x_{k+1}), Y(x_{k+2}), ..., Y(x_n))^\top \sim N(\beta_0 \mathbb{1} + \beta_1 x_{k+1:n} \Sigma_2(\sigma_2, \ell_2)). \tag{15}$$

Finally, conditional on there being no GEAs in the data, the model is simply

$$(Y(x_1), Y(x_2), ..., Y(x_n))^\top \sim N(0, \Sigma(\sigma, \ell)). \tag{16}$$

We see that estimating the changepoint locations also involves selecting the most appropriate model. In order to avoid confusion, we have slightly abused notation and, for example, $\Sigma_1(\sigma_1, \ell_1)$ as it is estimated in the two changepoint model is *not* the same as $\Sigma_1(\sigma_1, \ell_1)$ from either of the one changepoint models, and $\Sigma_1(\sigma_1, \ell_1)$ is also *not* the same between the two one changepoint models. As another example, β_0 is *not* the same between each of the one changepoint models. So, to be clear, duplication of notation in *different* models is not meant to imply that those parameters are shared between models.

Ultimately, these above four models are each individually fitted, and each model above is given a prior. From there, we do model selection in the formal Bayesian way, selecting number and location of changepoints by maximizing the estimated posterior distribution.

In order to complete a Bayesian model specification, we need priors on each of the parameters in each model as well as each model itself. We will assume independence between each parameter a priori. For each length scale ℓ, we will assume $\ell \sim \text{Gamma}(3, 5)$. For each standard deviation, we will assume $\sigma \sim \text{Half-Normal}^+(0, 1)$, where $\text{Half-Normal}^+(\cdot, \cdot)$ is notation for the normal distribution restricted to the positive real numbers. For intercept parameters, $\beta_{01}, \beta_{02}, \beta_0 \sim N(0, 10)$. For the slope parameters, the preceding trend deviates slightly. For any slope that corresponds to the *left* GEA, β_1 or β_{01}, we will assume that the slope can not be positive. That is, $\beta_1, \beta_{01} \sim \text{Half-Normal}^-(0, 10)$, where $\text{Half-Normal}^-(\cdot, \cdot)$ is notation for the normal distribution restricted to the negative real numbers. Contrastingly, for any slope that corresponds to the *right* GEA, β_1 or β_{02}, we will assume that the slope can not be negative. That is, $\beta_1, \beta_{01} \sim \text{Half-Normal}^+(0, 10)$. For the changepoint locations, we assume a uniform prior $\pi(c_l, c_r) \propto I(a < c_l < c_r - \gamma < b - \gamma)$. Here, a and b are some values close to the edges of the data. How close those values are to the edges is a

parameter that is set manually. Further, we include another hyperparameter, γ, which can be set so that the changepoints are not allowed to be too close to each other. This is also a parameter that is set manually. Lastly, we assume a uniform prior over all four models.

3.3.4 Bayesian Model Estimation

As was noted in Stephens (15), for any model including a changepoint, the likelihood is not a smooth function of the changepoint location. This is because, holding all other parameters fixed, shifting the changepoint value will result in zero change to the likelihood until it crosses the nearest point to the right or left, at which point the likelihood makes a jump. This makes maximum likelihood estimation in the standard way infeasible, but Bayesian estimation can be done in a fairly straightforward way via Markov chain Monte Carlo (MCMC). The basic idea is that, for each model, we can construct a two step Gibbs sampler. In step 1 we sample from the posterior distribution of the mean and covariance parameters given the changepoint locations, and in step 2 we sample from the changepoint locations given the mean and covariance parameters. Because of the non-conjugacy in our model, we perform both sampling steps using a random walk Metropolis-Hastings (RWMH) step with Gaussian proposals. For details on Gibbs sampling and the Metropolis-Hastings algorithm see Gelman et al. (16). It is also worth mentioning that the zero changepoint model does not require Gibbs sampling at all, and we perform estimation there using a RWMH algorithm.

We now provide the two basic steps of the Gibbs sampler for the two changepoint case. The algorithms to sample from the other three models are omitted, and are nearly identical except for the a smaller number of parameters need to be sampled. Denote collection of mean and covariance parameters for the left GEA as θ_1, the LEA as θ_2, and the right GEA as θ_3. Then, at iteration t after warmup

1. given changepoint locations $(c_l^{(t-1)}, c_r^{(t-1)})$, sample $(\theta_1^{(t)}, \theta_2^{(t)}, \theta_3^{(t)})$ using independent RWMH steps for each θ_i

2. given $(\theta_1^{(t)}, \theta_2^{(t)}, \theta_3^{(t)})$, sample $(c_l^{(t)}, c_r^{(t)})$ using a single RWMH step.

After running the MCMC for each model, parameter estimates and the most likely

model are jointly chosen according to the largest joint posterior value. That is, we arrive at estimates $(\hat{\theta}, \hat{M}) = \underset{(\theta,M)}{\text{argmaxlog}}(p(\theta, M|Y))$, where M is the random variable associated with the choice of model, θ is the associated parameter vector for the appropriate model, and Y is all of the available data. Additional MCMC details can be found in the appendix.

4 Results

To assess the degree of improvement in automated shoulder location identification, we want to quantify the impact each groove prediction method has on the automated bullet matching algorithm's accuracy. Five different shoulder location identification methods will be inserted into the automated bullet matching algorithm process:

(1) rollapply, the method proposed by (8),

(2) LASSO basic,

(3) LASSO full,

(4) Bayesian changepoint, and

(5) manual identifications, the "gold standard" for identification.

Shoulder location predictions are found using each of these methods, and used to remove GEA data from each crosscut. This is followed by extracting a signature, and using extracted signatures for each land engraved area to calculate pairwise similarity scores for all LEA signatures in each test set. Pairwise similarity scores, calculated using the random forest algorithm in the `bulletxtrctr` packages, are based on pairwise features such as consecutively matching striae and cross-correlation function for two aligned signatures.

Random forest scores should be high for land-to-land comparisons between signatures that originate from the same land, and lower for land-to-land comparisons between signatures from different lands.

We will investigate these scores for each individual test set: Hamby set 44, Phoenix PD set, and Houston-test set. All pairwise comparisons within a test set are completed. We can look at both visual representations of the random forest score distributions as

well as investigate the random forest method's accuracy in determining whether two lands originate from the same source.

There should be visual separation between "same source" and "different source", as is seen for Manual ID predictions for all three test sets, seen in Figure 9, Figure 10, and Figure 11. LASSO basic, LASSO full, and Bayesian changepoint all show improvement in separation for all three test sets; however, there is still much room for improvement when compared to Manual ID distributions in the Phoenix PD and Houston-test sets.

Quantitatively, there is also significant improvement in AUC values for all three test sets, as seen in Table 1, Table 2, and Table 3. Of interest is the classification accuracy with a fixed false positive rate. Since false positives - in this case identifying two land engraved areas as same source when they are different source - are the worst possible mistake, we set a cutoff rate for random forest scores based on a controlled false positive rate of .01.

Accuracy is overall high regardless of shoulder location identification method; however, there was a significant reduction in number of false negatives for the LASSO basic, LASSO full, and Bayesian changepoint methods for both Hamby set 44 and Phoenix PD sets. Bayesian changepoint improves upon rollapply in the Houston-test set, but lags behind the LASSO basic and LASSO full methods in AUC and reduction of false negatives.

5 Conclusions

All three proposed approaches show significant improvement both visually and quantitatively over the "rollapply" method proposed by (8). However, both LASSO methods show greater improvement than Bayesian changepoint on the Houston-test set. While manual identification of shoulder locations is still the most accurate method, and considered the "gold standard", the reduction in time to get predictions when using LASSO methods is advantageous and allows for less human involvement in the overall automated matching process.

While improvement is apparent on all three test sets using the LASSO and Bayesian changepoint methods, future work on the LASSO methods should include re-training LASSO models on a wider variety of LEA types rather than just the Hamby set 44 to avoid over-fitting to a specific type of LEA.

Nate, could you add some conclusions about BCP stuff here? A sentence or two?

6 References

7 Appendix

7.1 MCMC Details

As a practical note, it turns out that the posterior distribution is almost always multimodal, and it can happen that the sampler gets stuck in a suboptimal mode for a large number of iterations. It is also the case that the suboptimal modes need not even be close to the groove locations. It has, however, been our experience that the optimal mode corresponds well to the actual groove locations, which are often somewhat close to the edges of the data. With this in mind, starting values and the RWMH proposal variances play a very important role in the success of the sampling algorithm. Fortunately, it seems to be the case that by setting the initial changepoint values close to the edges of the data and making the proposal variance small (around 100 seems to work well) allows the sampler to wander inwards, and even with a modest number of iterations (say 5000), typically pass through the largest mode corresponding to the groove locations. This is not always the case, and it is possible that increasing the number of iterations produces better results.

In our implementation of this algorithm, the sampling functions were originally written with the intention of tuning the proposal variances to potentially accelerate convergence, and thus several warmup iterations are required for this purpose. This turns out to be a bad idea in this context for two reasons. The first reason is that the warmup iterations allow the sampler to wander past the global modes and get stuck in suboptimal modes far from the groove locations, from which the sampler may or may not find its way back to the optimal modes in just a few thousand iterations. Secondly, if the sampler does wander past the optimal modes, which are usually on the edges of the data, the tuned proposal variance can be quite large. The large proposal variance might not be a huge problem if it weren't for the fact that the width of the modes are almost always quite small. This means that it can take a very, very long time for the sampler to move from a suboptimal mode to

the global mode. In order to mitigate this problem, we are currently setting the number of warmup iterations to be relatively small (somewhere in 100 to 500 seems to work well). In future, our implementation of the algorithm will not require any warmup iterations.

Initially, the Metropolis proposal variance for each θ_i is diagonal with diagonal elements all equal to 1/2. The proposal variance for (c_l, c_r) is initially set to be diagonal with elements equal to 10^2. Note that because of the currently necessary warmup iterations, the variances after warmup for each θ_i becomes $\frac{2.4^2}{d}\hat{Var}(\theta_i^{(1:w)}) + \text{diag}(0.1)$, where d is the dimension of θ_i (which is not constant between GEAs and LEA), and $\hat{Var}(\theta_i^{(1:w)})$ is the estimated variance covariance matrix from the w warmup iterations. Note that the addition of a diagonal matrix with entries 0.1 is to avoid the case when most or all warmup iterations have the same value. Similarly, the proposal variance for (c_l, c_r) after warmup becomes $\frac{2.4^2}{2}\hat{Var}((c_l, c_r)^{(1:w)}) + \text{diag}(1)$.

7.2 Data Preprocessing for MCMC

Before running the MCMC to do the changepoint detection, we first perform two data preprocessing steps. The first step is to scale the residuals from the robust loess procedure by the standard deviation calculated from the entire set of residuals. The reason for this is simply to make priors for standard deviation and slope parameters easier to specify. For example, ensuring that the residuals are scaled to have standard deviation one means that the standard deviation parameters in our model should also be close to one. This scaling also ensures that slopes values are not very large.

The second preprocessing step is a bit more involved. In order to enable the algorithm to run reasonably fast, we need to take advantage of the sparse precision matrix structure that is induced by the exponential covariance function. Indeed, this was the reason for choosing this covariance function in the first place. Unfortunately, it is challenging to do this unless the observations are evenly spaced in the domain. In our case, this would be true if there were no missing values. In order to remedy this problem, we impute the missing data, but only in the case that there exist nonmissing observations outside of the missing values. In the case that the missing values exist on the edges of the data, we simply do not consider those domain values in the model.

We perform the imputation by treating the observations as coming from an unknown function, and infer the missing values from the known function values. In order to do this, we model the data with a Gaussian process and the squared exponential covariance function. That is, we suppose that

$$Y(x) \sim \mathcal{GP}(0, K(x, x'; \sigma^2, \ell)),$$

where now $K(x, x'; \sigma^2, \ell) = \sigma^2 e^{-(x-x')^2/(2\ell^2)}$ is the squared exponential covariance function. We emphasize for clarity that this is a different covariance function than we use in the changepoint model. The main reason for this is that in imputing values, it seems desireable to allow dependencies beyond immediately neighboring points to influence predictions as the function that we are trying to predict generally has a smooth global structure. For all of our experiments, we set $\sigma = 0.8$ and $\ell = 15$. These values were chosen from doing maximum likelihood estimation for a representative bullet.

When we impute the missing values, we compute the conditional mean of the missing values. To be clear, denote the distribution of the observed and missing data as

$$(Y, Y^*)^\top \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yy^*} \\ \Sigma_{y^*y} & \Sigma_{y^*y^*} \end{bmatrix}\right).$$

Here, Y is observed data and Y^* is the missing data, and the covariance matrix above is constructed from the squared exponential covariance function. We then use normal distribution theory to calculate the imputed values

$$E(Y^*|Y = y) = \Sigma_{y^*y}\Sigma_{yy}^{-1}y$$

.

References

1. AFTE Glossary. Theory of Identification as it Relates to Toolmarks. AFTE Journal 1998;30(1):86–88.

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

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

4. Bachrach B. Development of a 3D-based Automated Firearms Evidence Comparison System. Journal of Forensic Sciences 2002;47(6):1253–1264.

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

6. 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. Journal of Forensic Sciences 2010;55(2):341–47.

7. Chu W, Thompson RM, Song J, Vorburger TV. Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria. Forensic Science International 2013;231(1-3):137–41.

8. Hare E, Hofmann H, Carriquiry A. Automatic matching of bullet land impressions. The Annals of Applied Statistics 2017 12;11:2332–2356.

9. 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.

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

11. Hofmann H, Vanderplas S, Krishnan G. bulletxtrctr: Analyze bullet striations using nonparametric methods (Computer Program]; 2019. https://heike.github.io/bulletxtrctr/.

12. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 1979;74(368):829–836.

13. Loader C. locfit: Local Regression, Likelihood and Density Estimation (Computer Program]; 2013. https://cran.r-project.org/web/packages/locfit/locfit.pdf.

14. Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian J. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models (Computer Program]; 2018. https://cran.r-project.org/web/packages/glmnet/glmnet.pdf.

15. Stephens DA. Bayesian Retrospective Multiple-Changepoint Identification. Journal of the Royal Statistical Society Series C (Applied Statistics) 1994;43(1):159–178. https://www.jstor.org/stable/2986119.

16. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Chapman and Hall; 2013.

List of Figures

1 An example of a high-resolution LEA scan captured on a Sensofar Confocal Light Microscope. 27

2 The process of extracting a 2D signature from a high-resolution 3D scan of a land engraved area (LEA) on a bullet. Extraction of a signature is dependent on both the choice of horizontal crosscut and accuracy of GEA data removal procedure. 28

3 An example of the impact failure to remove GEA data can have on an extracted 2D signature. Important data features are obfuscated in the signature by remaining GEA structure. 29

4 An example of the difference between traditional LOESS fit and robust LOESS fit to an LEA crosscut from Hamby set 44. While the GEA structure present is relatively minor, even the small improvement in model fit with robust LOESS results in significantly different residual structures. . . 30

5 An example of the difference between traditional LOESS fit and robust LOESS fit to an LEA crosscut from the Houston-test set. While the smaller GEA structure on the right side is accounted for using the robust procedure, the larger GEA structure on the left is still problematic. 31

6 An example of the difference between robust LOESS fit and adapted robust LOESS fit to an LEA crosscut from the Houston-test set. Iteratively down-weighting only positive residuals results in a significantly different fit which accounts for GEA structures on both the left and right. 32

7 Mean shift in predictions when applying the adapted robust LOESS procedure in place of the traditional robust LOESS procedure. For Hamby set 44, predictions are, on average, very similar. The Phoenix PD and Houston-test sets have more significant downwards shifts in predictions near the left and right boundaries. 33

8 Example distributions of features used in two-class classification. While depth shows the most clear separation between GEA and LEA data, it alone will not suffice to classify data correctly. While other distributions are relatively tight for LEA, there is still significant overlap with the wider GEA distributions. 34

9 Random forest score distributions for same source and different source land-to-land comparisons for Hamby set 44. Distributions should ideally separate between same source and different source pairs. LASSO Basic, LASSO Full and Bayesian changepoint all demonstrate significant improvement over Rollapply. 35

10 Random forest score distributions for same source and different source land-to-land comparisons for the Phoenix PD set. LASSO Basic, LASSO Full, and Bayesian changepoint all demonstrate significant improvement over Rollapply, but are still not as well separated as the Manual ID distributions. . 36

11 Random forest score distributions for same source and different source land-to-land comparisons for the Houston-test set. LASSO Basic and LASSO Full both demonstrate improvement over Rollapply, but are still not as well separated as the Manual ID distributions. Bayesian changepoint demonstrates minor improvement, but does not improve as much as the LASSO methods. 37

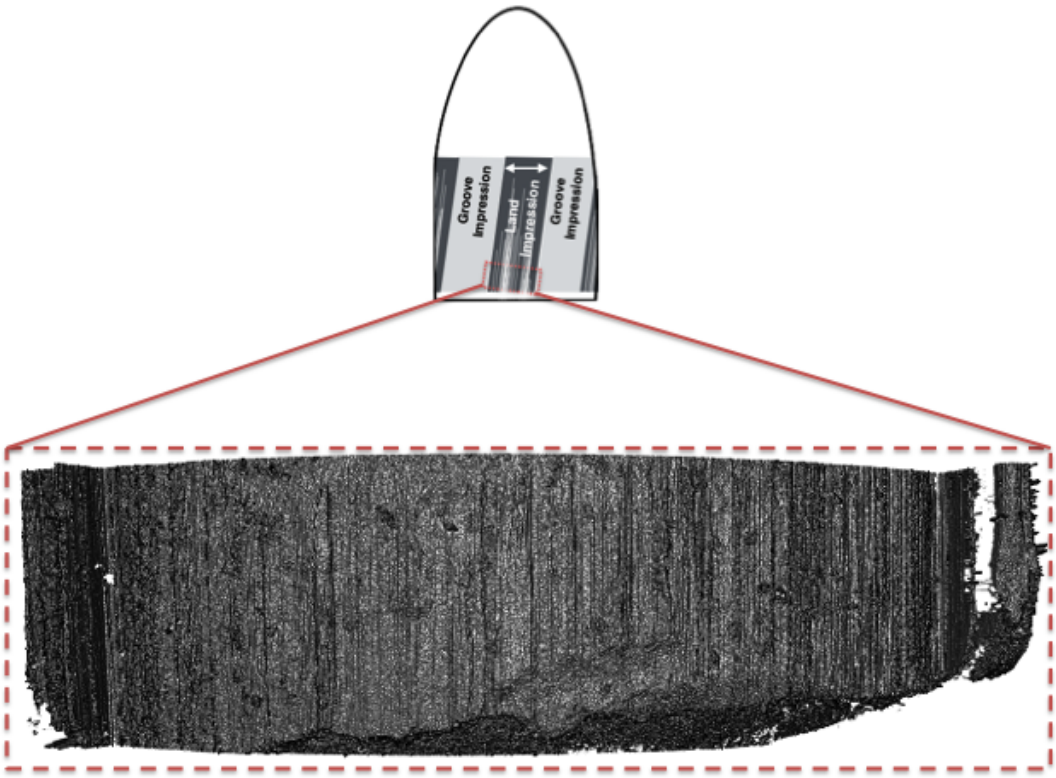
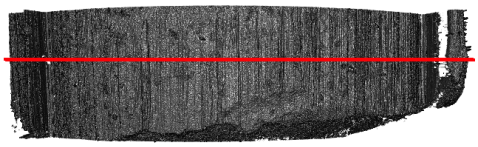
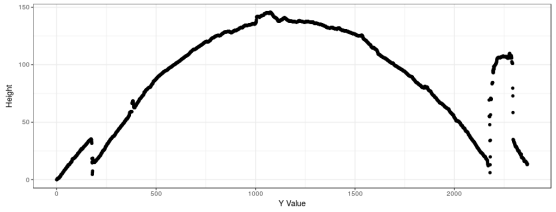


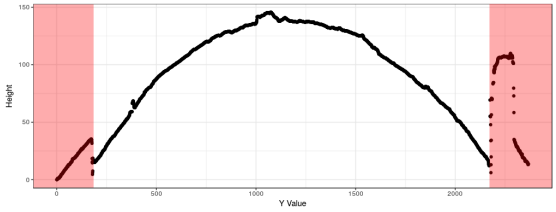
Figure 1: An example of a high-resolution LEA scan captured on a Sensofar Confocal Light Microscope.



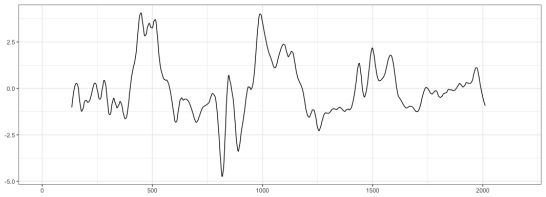
Step 0: High-resolution 3D scan



Step 1: Horizontal crosscut



Step 2: GEA data removal



Step 3: Extracted signature

Figure 2: The process of extracting a 2D signature from a high-resolution 3D scan of a land engraved area (LEA) on a bullet. Extraction of a signature is dependent on both the choice of horizontal crosscut and accuracy of GEA data removal procedure.

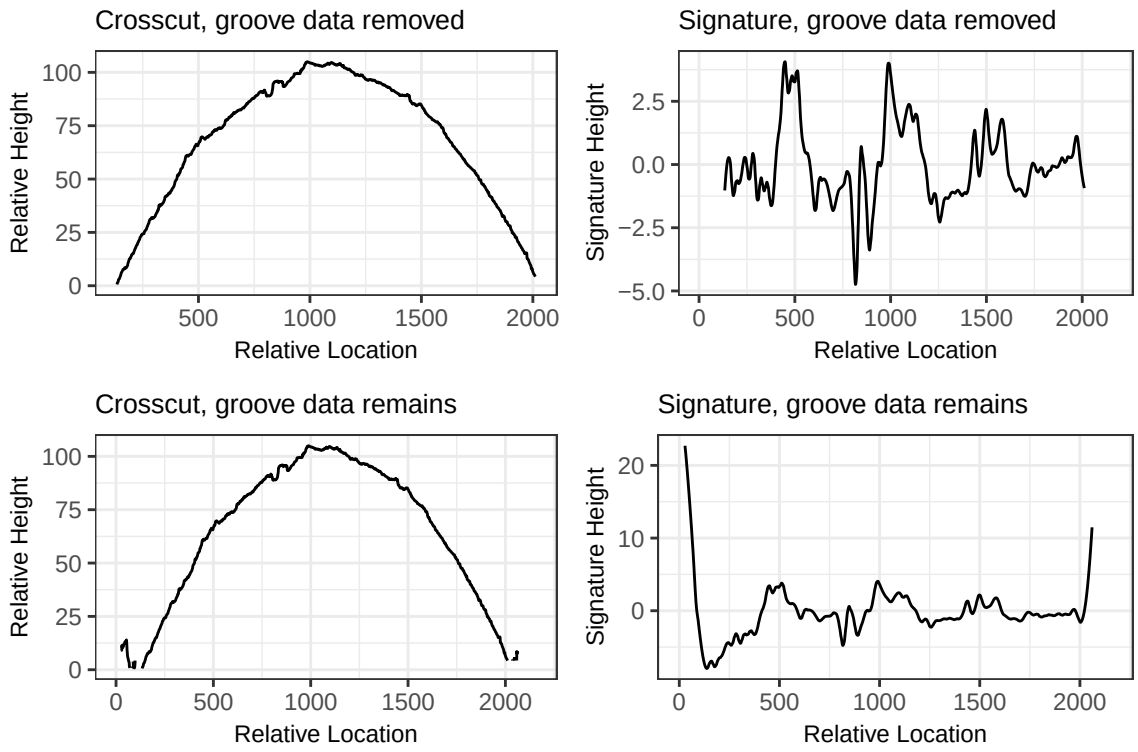


Figure 3: An example of the impact failure to remove GEA data can have on an extracted 2D signature. Important data features are obfuscated in the signature by remaining GEA structure.

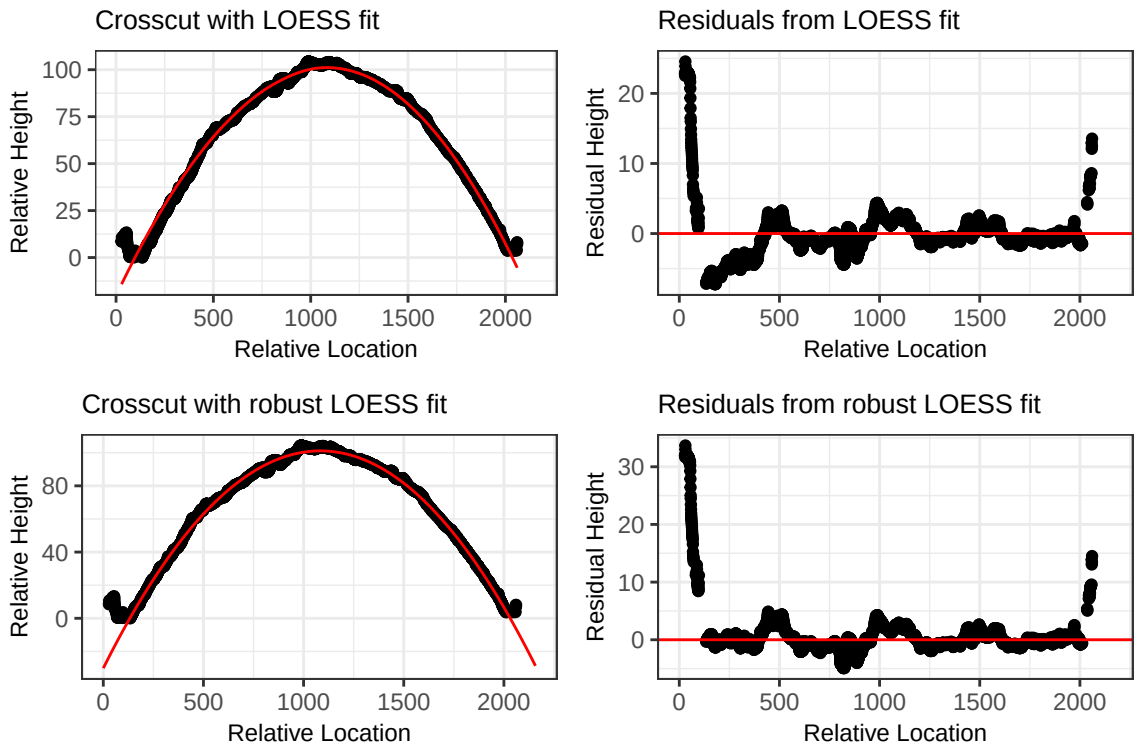


Figure 4: An example of the difference between traditional LOESS fit and robust LOESS fit to an LEA crosscut from Hamby set 44. While the GEA structure present is relatively minor, even the small improvement in model fit with robust LOESS results in significantly different residual structures.

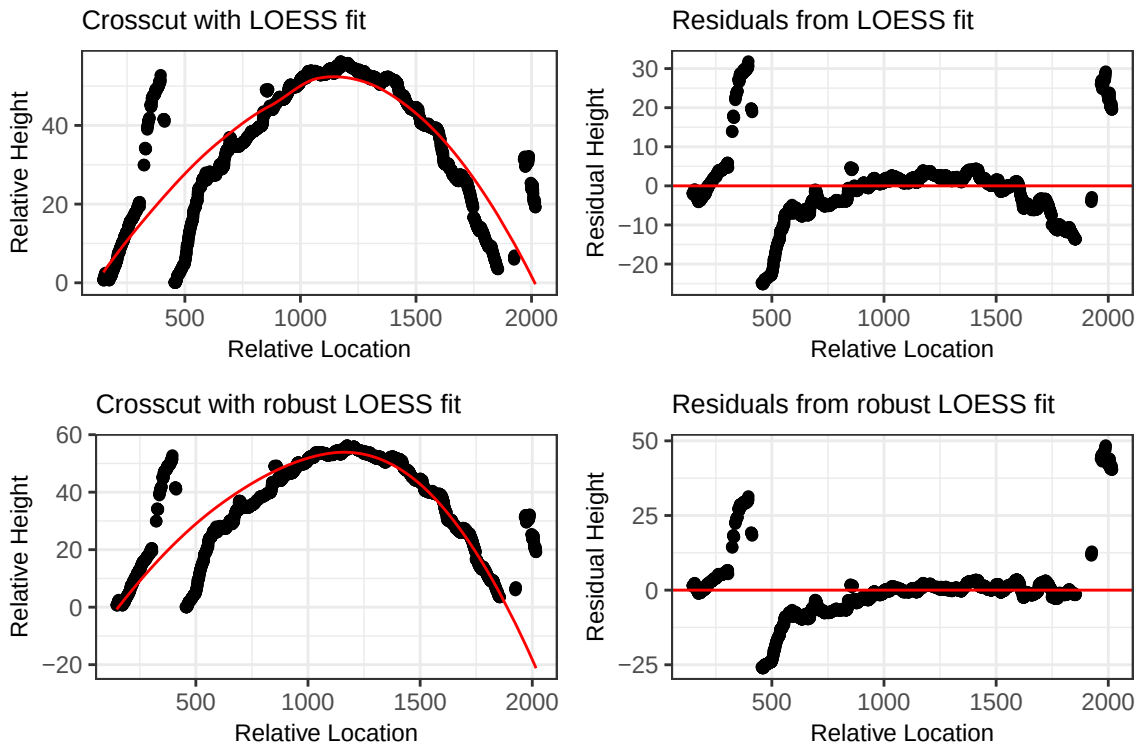


Figure 5: An example of the difference between traditional LOESS fit and robust LOESS fit to an LEA crosscut from the Houston-test set. While the smaller GEA structure on the right side is accounted for using the robust procedure, the larger GEA structure on the left is still problematic.

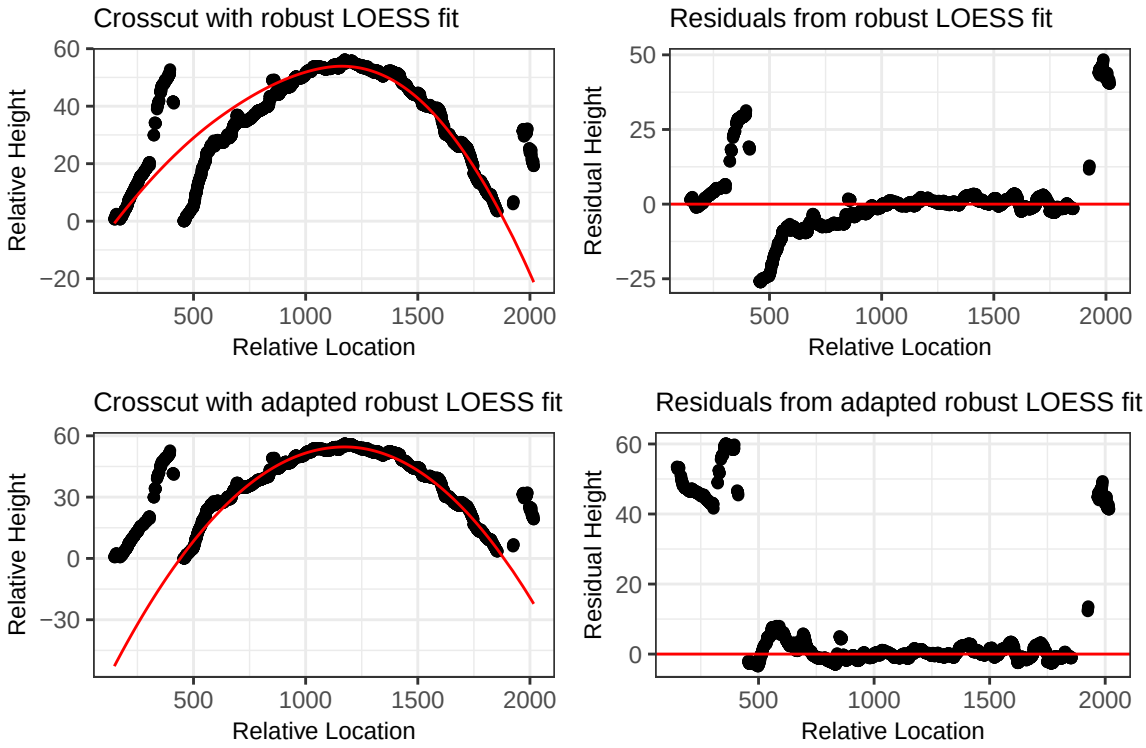


Figure 6: An example of the difference between robust LOESS fit and adapted robust LOESS fit to an LEA crosscut from the Houston-test set. Iteratively downweighting only positive residuals results in a significantly different fit which accounts for GEA structures on both the left and right.

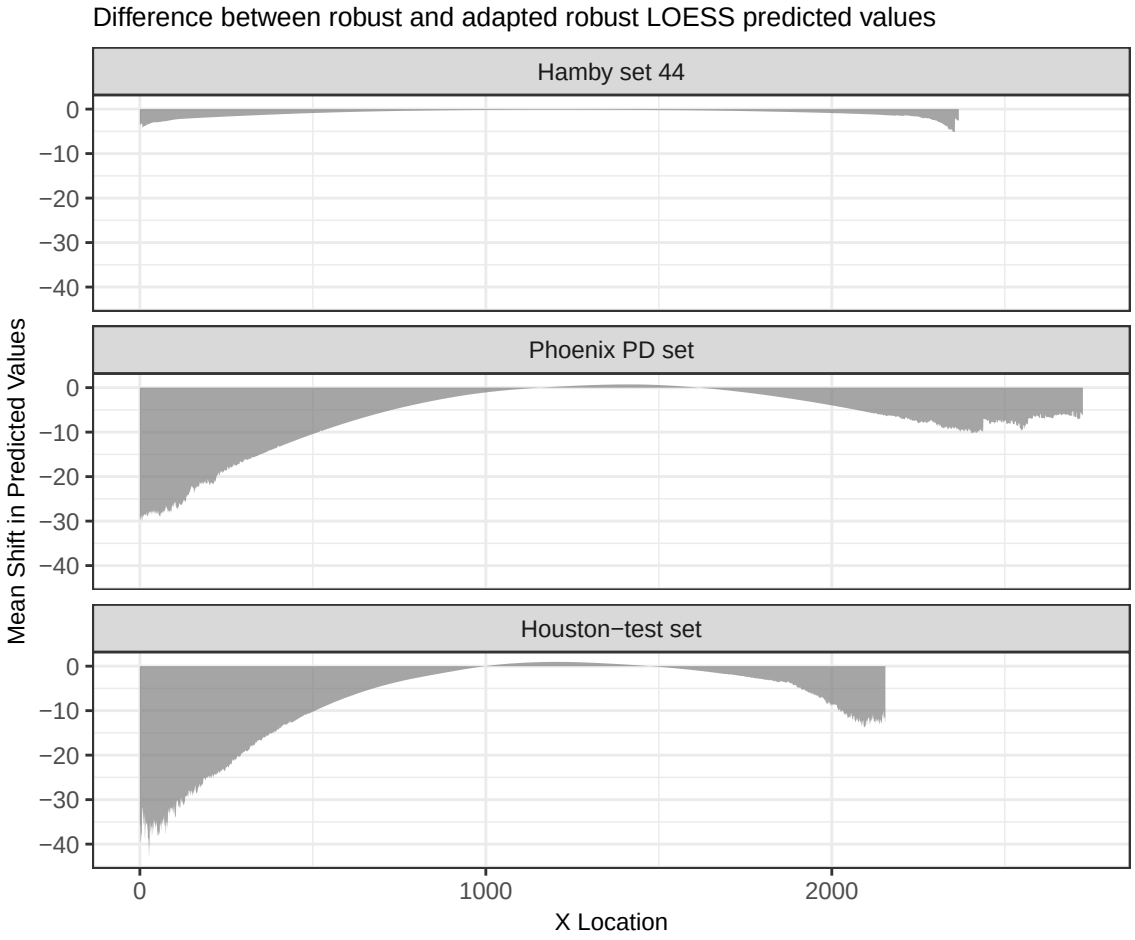


Figure 7: Mean shift in predictions when applying the adapted robust LOESS procedure in place of the traditional robust LOESS procedure. For Hamby set 44, predictions are, on average, very similar. The Phoenix PD and Houston-test sets have more significant downwards shifts in predictions near the left and right boundaries.

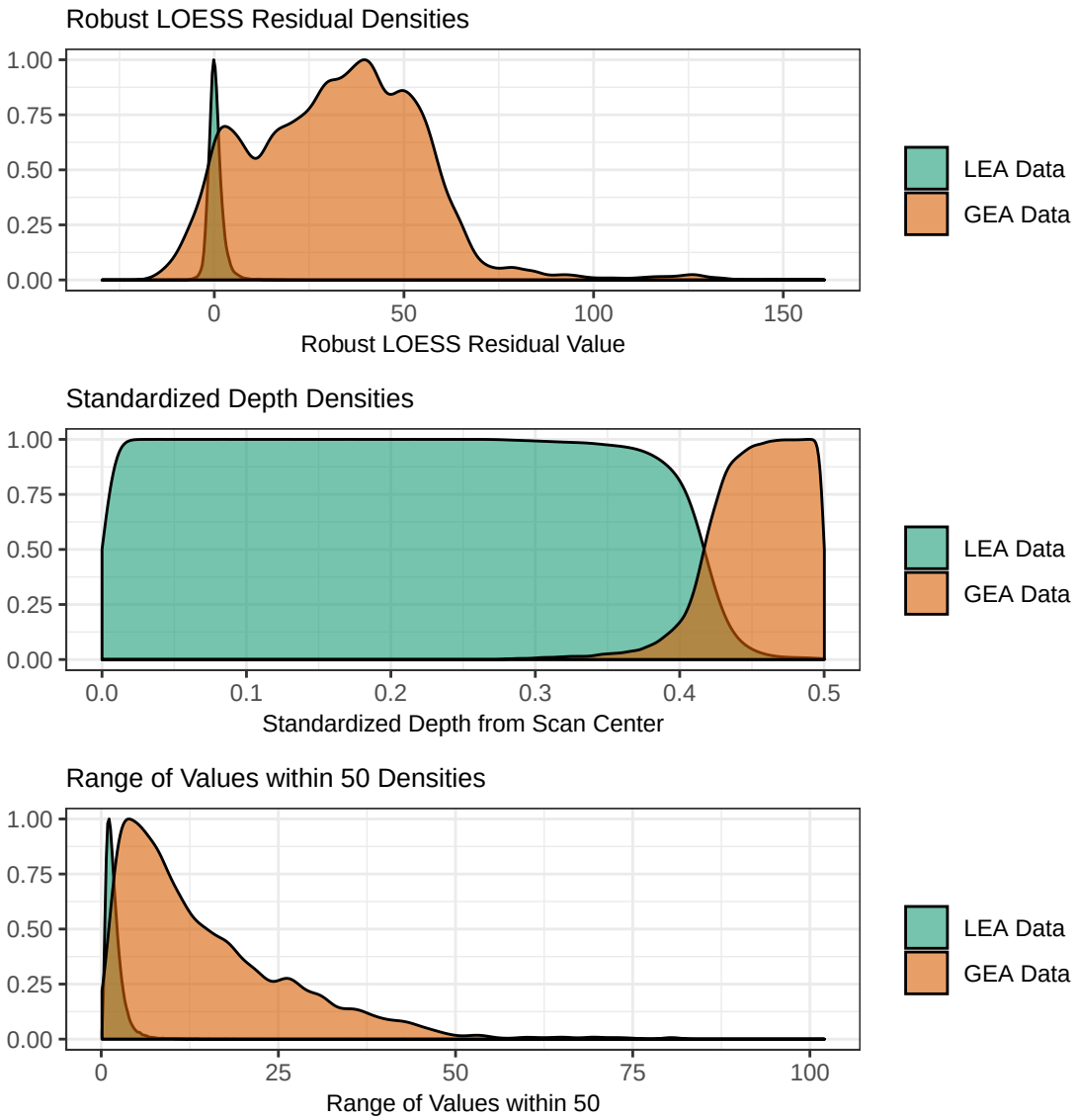


Figure 8: Example distributions of features used in two-class classification. While depth shows the most clear separation between GEA and LEA data, it alone will not suffice to classify data correctly. While other distributions are relatively tight for LEA, there is still significant overlap with the wider GEA distributions.

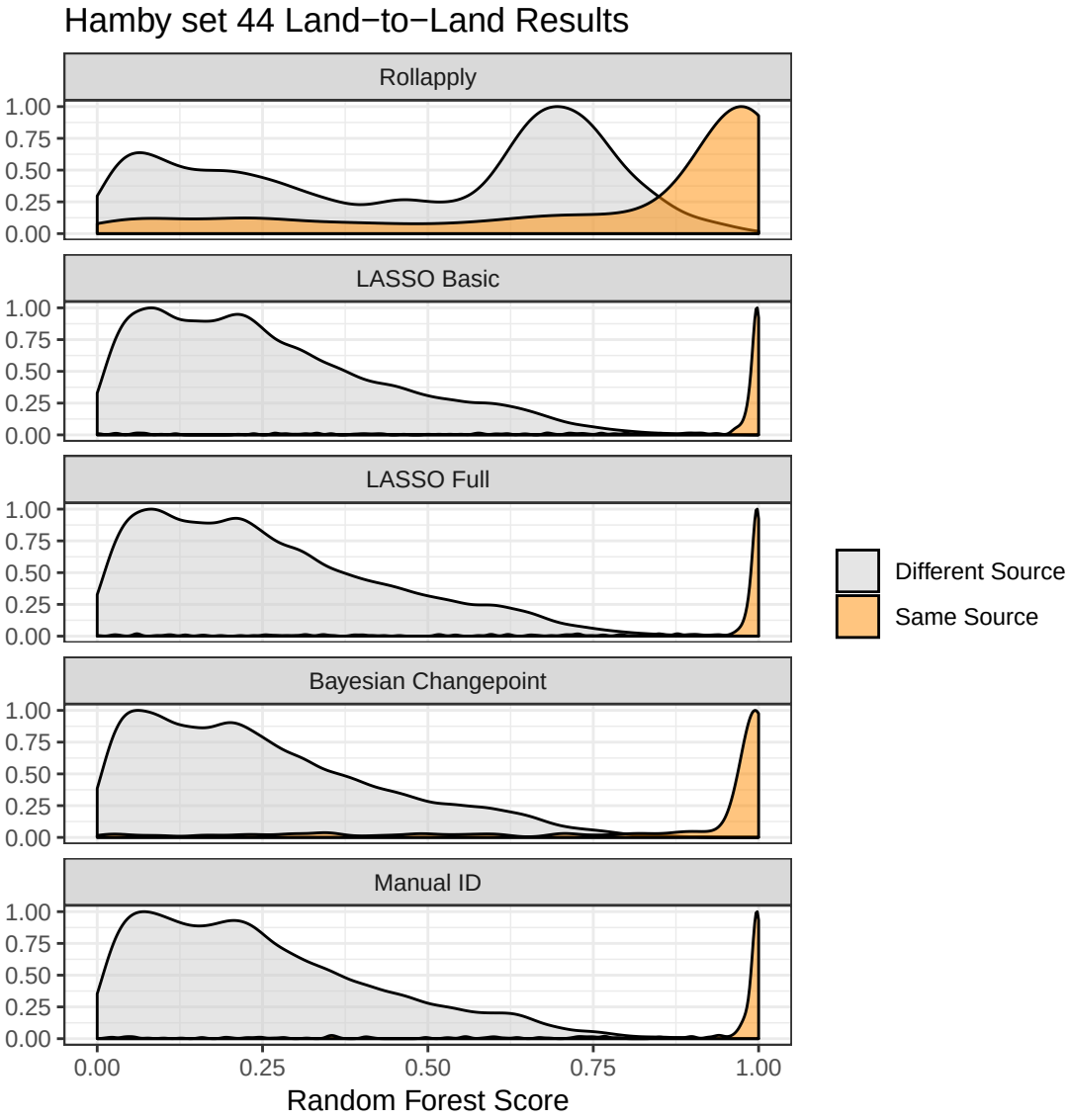


Figure 9: Random forest score distributions for same source and different source land-to-land comparisons for Hamby set 44. Distributions should ideally separate between same source and different source pairs. LASSO Basic, LASSO Full and Bayesian changepoint all demonstrate significant improvement over Rollapply.

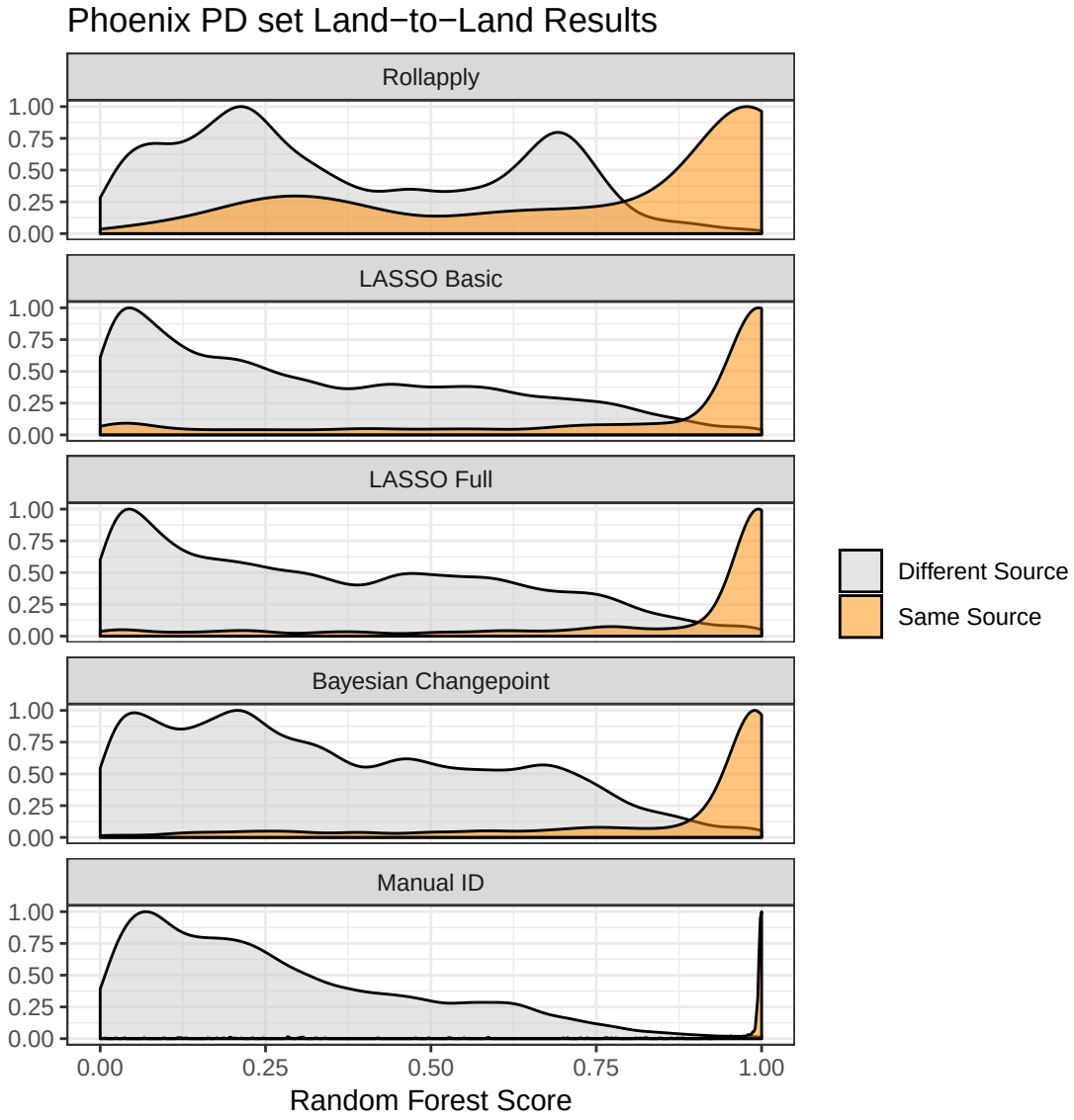


Figure 10: Random forest score distributions for same source and different source land-to-land comparisons for the Phoenix PD set. LASSO Basic, LASSO Full, and Bayesian changepoint all demonstrate significant improvement over Rollapply, but are still not as well separated as the Manual ID distributions.

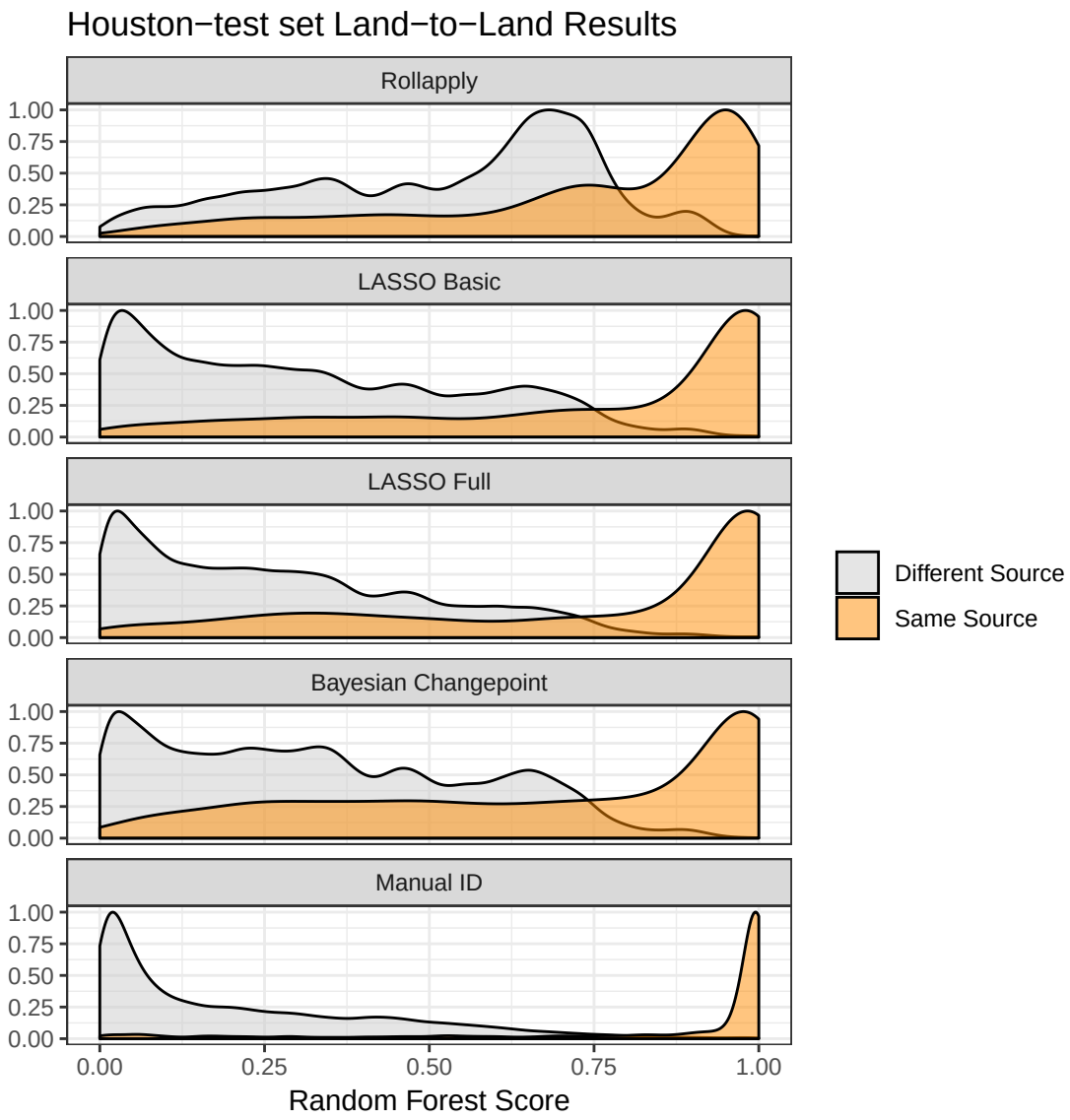


Figure 11: Random forest score distributions for same source and different source land-to-land comparisons for the Houston-test set. LASSO Basic and LASSO Full both demonstrate improvement over Rollapply, but are still not as well separated as the Manual ID distributions. Bayesian changepoint demonstrates minor improvement, but does not improve as much as the LASSO methods.

List of Tables

1 Land-to-land comparison results for Hamby set 44. 39
2 Land-to-land comparison results for the Phoenix PD set. 40
3 Land-to-land comparison results for the Houston-test set. 41

		Controlled FPR = .01					
Method	**AUC**	Cutoff	FN	TP	TN	Accuracy	**Time to Calculate**
Rollapply	0.8	0.92	332	420	42126	0.98	1 min.
LASSO basic	0.94	0.75	126	626	42098	0.99	6 min.
LASSO full	0.94	0.75	126	626	42106	0.99	6 min.
Bayesian Changepoint	0.93	0.74	152	600	42098	0.99	
Manual ID	0.94	0.74	124	628	42088	0.99	45 min.

Table 1: Land-to-land comparison results for Hamby set 44.

		Controlled FPR = .01					
Method	**AUC**	Cutoff	FN	TP	TN	Accuracy	**Time to Calculate**
Rollapply	0.818	0.9	364	378	38082	0.981	1 min.
LASSO basic	0.877	0.947	256	486	38088	0.9839	6 min.
LASSO full	0.893	0.953	238	504	38082	0.9842	6 min.
Bayesian Changepoint	0.903	0.937	256	486	38080	0.9837	
Manual ID	0.953	0.853	96	646	38084	0.9879	45 min.

Table 2: Land-to-land comparison results for the Phoenix PD set.

		Controlled FPR = .01					
Method	**AUC**	Cutoff	FN	TP	TN	Accuracy	**Time to Calculate**
Rollapply	0.761	0.91	1652	1016	167118	0.981	2 min.
LASSO basic	0.852	0.88	1274	1394	167136	0.9833	12 min.
LASSO full	0.858	0.823	1144	1524	167062	0.9836	12 min.
Bayesian Changepoint	0.795	0.863	1552	1116	167096	0.9814	
Manual ID	0.931	0.823	614	2054	167098	0.9869	75 min.

Table 3: Land-to-land comparison results for the Houston-test set.