

# Threshold Bootstrap for Land Engraved Area Signatures

Susan VanderPlas \*

Department of Statistics, Iowa State University

June 4, 2021

## Abstract

abstract-text

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

---

\*This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #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 Introduction

A common question in forensics (whether statistical or examiner based) is the error rate of the assessment procedure. The term “error rate” includes the statistical concepts of false positives (declaring a match when the two pieces of evidence are from different sources) and false negatives (declaring that two pieces are from different sources when they are from the same source). Statistical methods for forensic comparisons take the following form:

1. Reduce the evidence to some set of numeric data used for comparison
2. Use the numeric data to calculate a match score
3. Compare the match score to reference distributions derived from known matches (KM) and known non matches (KNM)
4. Calculate a score-based likelihood ratio(Lund & Iyer 2017) for a known match:

$$LR(KM) = \frac{\text{Probability of observing } x | \text{ known match}}{\text{Probability of observing } x | \text{ known non match}}$$

Derivation of this likelihood ratio depends heavily on the reference distributions derived from known matches and known non matches. These reference distributions are assembled from collected data, which may be sparse for a given set of class characteristics (firearm, barrel manufacturing process, ammunition type).

In order to present evidence in court, it is particularly important to assess the false positive rate (the probability of declaring a match when the data are from different sources). For score-based problems, the false positive rate is tied to the quality of the estimate of the denominator above, that is, the probability of observing  $x$  when it is known that the two pieces of evidence are from different sources. It is generally impractical to derive or even bound this probability experimentally, as that requires the collection and comparison of very large amounts of data from different sources stemming from the population of interest. For bullet comparisons, for instance, it would be necessary to gather bullets fired from a large number of barrels with class characteristics similar to those referenced, compare each fired bullet to each other bullet, and determine the number of bullet matches which have a score greater than  $x$  observed score.

It is much more reasonable to statistically derive the distribution of scores under known match or known non-match conditions by using a bootstrap or resampling technique to gen-

erate the numeric data produced in step 1 above. [Bachrach et al. \(2013\)](#) generated bullet signatures using wavelets and fractals to mimic the identified features of land engraved area signatures. Other data generation and augmentation techniques have been used for handwriting data ([Sesa-Nogueras & Faundez-Zanuy 2013](#), [Rabasse et al. 2008](#)), iris data([Wecker et al. 2010](#)), and facial images ([Masi et al. 2016](#)). With generated data, we can be reasonably certain that any correspondence between two generated patterns is entirely random. As long as important characteristics of the data are preserved, these randomly assembled pattern data provide a much quicker method to assess the distribution of known non-match and known match pattern scores.

## 1.1 Statistical Analysis of Bullet Land Engraved Areas

Comparison of bullet land engraved areas (LEAs) typically involves distilling the 3D scan of the area of interest into a two-dimensional “signature” that removes artifacts originating from the grooves as well as the curvature of the bullet. [Hare et al. \(2017\)](#) provides one method for distilling bullet scans into signatures; alternate methods used in [Vorburger et al. \(2011\)](#) should also be compatible with the approach to data generation described in this paper. [cite CMPS paper, Will’s implementation](#) Important characteristics of the LEA signature depend in part on the method used to transform the collected 3D image into a signature. We would not expect that different filtering and noise removal methods would produce signatures with exactly the same shapes or properties. By working with comparisons at the signature level, rather than the full image level, we also do not have to simulate the entire 3D image, as in [Bachrach et al. \(2013\)](#).

It is important that the presence of problematic features, such as tank rash and missing values due to microscope error, occur with approximately the same frequency in the generated sequences as in sequences from real data. This ensures that the reference distributions for known matches and known non-matches account for real-world constraints. The method for synthetic signature generation proposed in [Bachrach et al. \(2013\)](#) produces sequences and 3D images which do not mimic the problems seen in collected data. Examples of problematic lands in the Hamby 44 test set are shown in [Figure 1](#). To address this issue, this paper proposes a method for generating sequences from examiner test kits and

persistence studies using resampling techniques designed for data with temporal or spatial autocorrelation.

## 1.2 Threshold Bootstrap

The threshold bootstrap was proposed by [Park & Willemain \(1999\)](#) as an alternative to the moving block bootstrap that preserves continuity at block endpoints. Typically used in time series analysis to preserve complex dependency structures, it can be repurposed here because of our desire to maintain spatial continuity. Using the threshold bootstrap ensures that we get a continuous sequence that, while not guaranteed to be differentiable, adequately mimics the smoothness of the source data.

As described in [Park & Willemain \(1999\)](#), the threshold bootstrap divides a weakly dependent, stationary series  $\mathbf{X} = \{X_1, \dots, X_n\}$  with a threshold  $\bar{X}$  that divides the sequence into  $R$  chunks  $C_i = \{X_{i,1}, \dots, X_{i,n_i}\}$ , where the  $i$ th chunk has size  $n_i$ ,  $i = 1, \dots, R$  and  $\sum_{i=1}^R n_i = n$ . Samples are constructed at random with replacement from the set  $\{C_1, \dots, C_R\}$  to produce a sequence of the desired length.

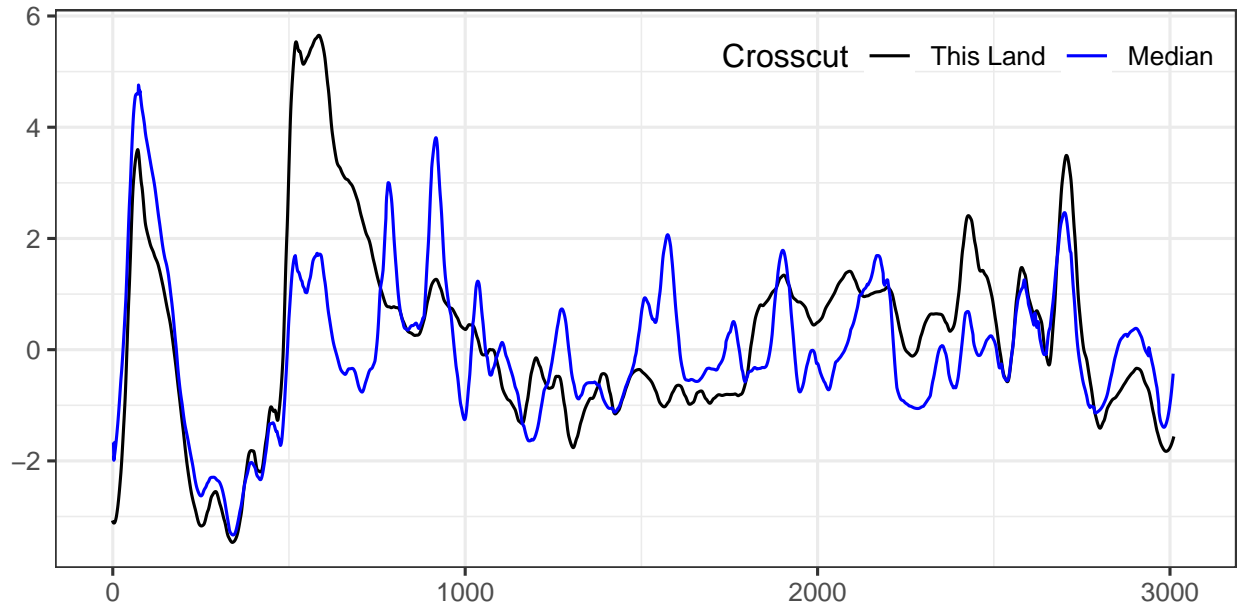
Figure 2 shows a bullet signature, with the signature mean value marked; the bottom two images shows the blocks produced for use in the threshold bootstrap, where each block contains 1 or 3 cycles, respectively, as defined by [Park & Willemain \(1999\)](#).

This paper demonstrates the use of the threshold bootstrap to create simulated LEA signatures under known match and known non-match conditions, using several different sets of bullet signatures derived from examiner test kits. We describe first the creation of known non-match sequences, and then the creation of known match sequences, by applying the threshold bootstrap process systematically to a library of LEA signatures.

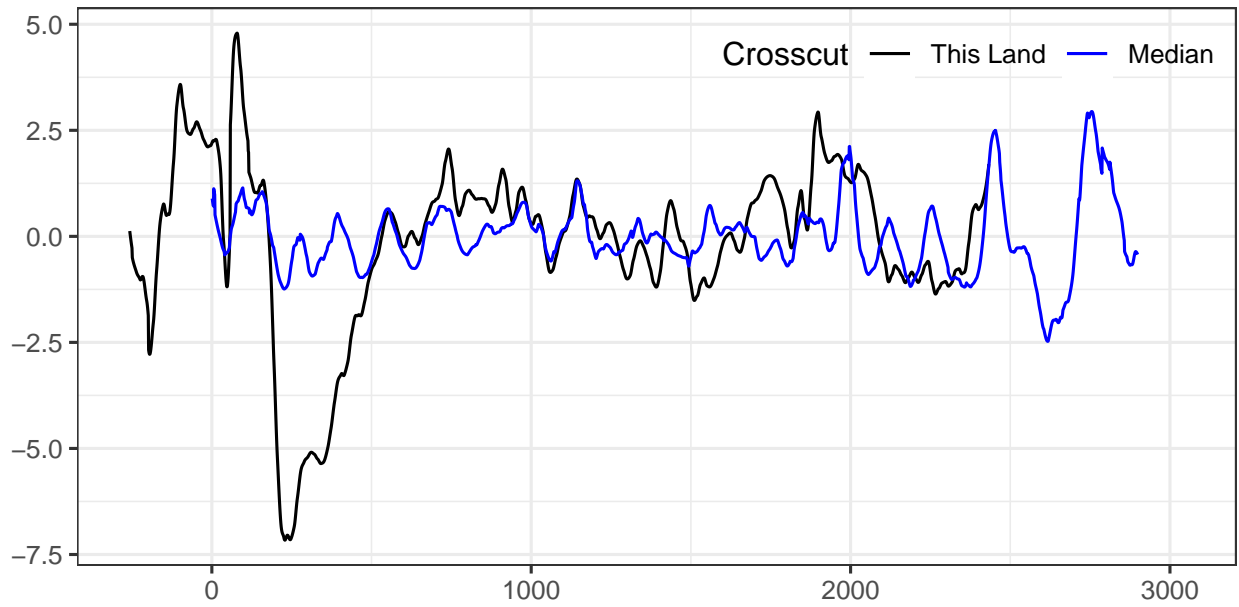
## 2 Simulating Bullet LEA Signatures

### 2.1 Notation

Let us define notation to describe the data used in this study. We have test sets  $h = 1, \dots, H$  which may be developed under different conditions and which may use different types of firearms and ammunition. Each study uses firearms (or more specifically, barrels)



(a) Barrel 2, Bullet 2, Land 5, with tank rash



(b) Barrel 8, Bullet 2, Land 2, with extreme pitting

Figure 1: Overview of bullet lands with prominent deficiency such as tank rash or extreme pitting in the Hamby-44 study, with corresponding crosscuts. Median crosscuts were calculated from the joint alignment of known bullets in Hamby 36, Hamby 44, and Hamby 224.

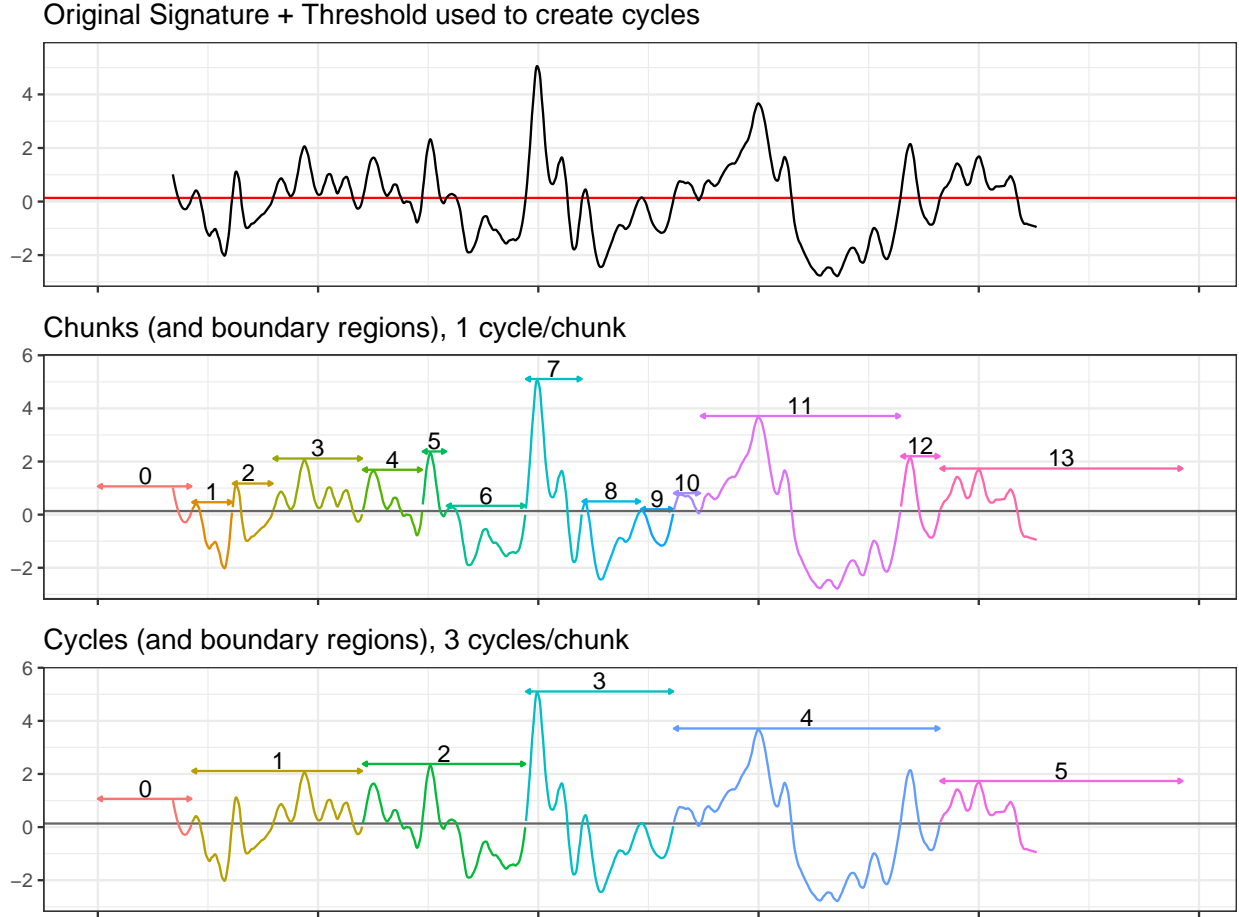


Figure 2: Threshold Bootstrap Cycle Construction. In the second graph, chunks are composed of a single cycle, but any multiple of  $n$  cycles could be used, depending on the dependency structure in the data. The third graph shows the same signature, divided into chunks consisting of 3 cycles. The first and last chunks are boundary chunks, and do not contain a full cycle consisting of one region above the median and one region below the median. Boundary chunks are recycled to construct the ends of a bootstrapped sequence. The LEA signature shown is from Barrel 1, Bullet 1, Land 1 in the Hamby 252 data set downloaded from NIST Ballistics Toolmark Research Database (Zheng et al. 2016); the signature was created using the `bulletxtctr` package and the method described in Hare et al. (2017).

$i = 1, \dots, I_h$ , with bullets  $j = 1, \dots, J_{hi}$  fired from each barrel in the study. Each bullet has  $k = 1, \dots, K_{hi}$  land engraved areas.

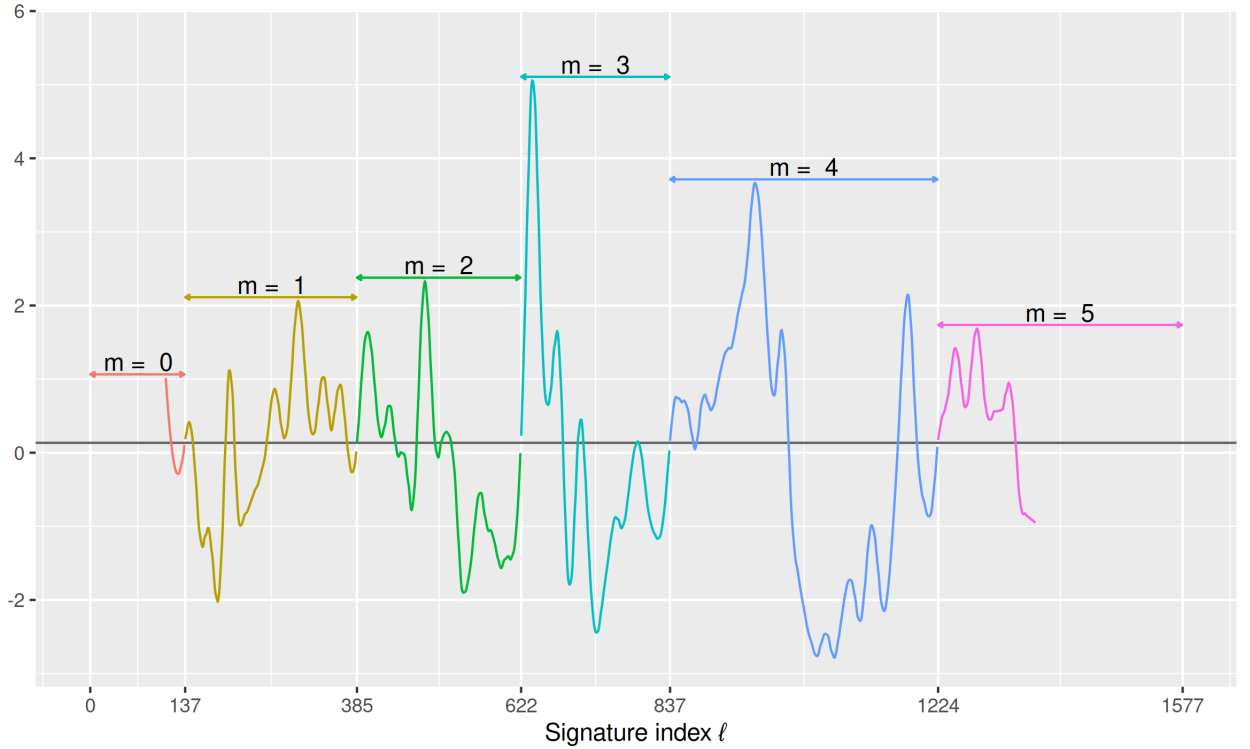
We will define the index  $k$  such that any signatures  $S_{hijk}$  with the same  $h$  and  $i$  index have the same land index  $k$  (that is, we assume that lands are indexed such that all bullets fired from the same barrel have the same index scheme corresponding to the same land in the rifled barrel). This condition significantly simplifies the notation without loss of generality. In practice, when aligning these sets, we do have to re-index the bullets from the scanned sequence of lands, but this is not a difficult process.

We will split each signature  $S_{hijk}$  into  $m$  chunks

$$\{\{S_{hijk1}, \dots, S_{hijk(\ell_1-1)}\}, \{S_{hijk\ell_1}, \dots, S_{hijk(\ell_2-1)}\}, \dots, \{S_{hijk\ell_m}, \dots, S_{hijkL}\}\}.$$

That is, let  $m = 1, \dots, M$  index the chunks, and  $\ell = 1, \dots, L$  index the sequence S. By indexing the chunks this way, we implicitly get both the chunk length and the chunk order.

Indexing signatures and chunks in  $\ell$  and  $m$



## 2.2 Source Data

This method requires signatures from known match and known non-match bullets; that is, from bullets fired from the same barrel as well as bullets fired from different barrels.

One of the strengths of utilizing the threshold bootstrap for sequence generation is that we can be confident that features of the LEA sequences which are attributable to the pre-processing method used are preserved in the data. This is, of course, a trade-off: the feature distributions which are the end result of the bootstrapping process will not generalize to other methods for creating LEA signatures. Similarly, if there are differences in the features of signatures produced from different ammunition or barrel manufacturing methods, it may be unwise to use distributions generated from different source materials, particularly because it is known that different combinations of firearms and ammunition mark differently (Bonfanti & De Kinder 1999). Ideally, the distributional results would be relatively robust to these changes, as there are several levels of abstraction between the data and the resulting match score, but that is left to some future study.

Provided that relevant source data exist, the threshold bootstrap is applicable to any automatic processing method which reduces a 3D scan of a land to a single sequence of points describing a cross-section of the land engraved area. The choice to demonstrate the method using one specific processing method is primarily one of convenience.

In this paper, we will use data collected from 3 Hamby sets (Hamby et al. 2009): Hamby 36, Hamby 44, and Hamby 224. Each set was produced by firing the same 10 barrels to create a set of 20 known and 15 unknown bullets (thus, we are working with between 60 and 105 bullets, depending on whether we include the unknown bullets in the alignments; each barrel will have at least 6 scans). These sets were scanned using a Sensofar confocal light microscope, and processed using the method described in Hare et al. (2017). Once scans were reduced to signatures, bullets fired from each barrel were aligned by maximizing the cross-correlation over the 6-land sequence.

## 2.3 Threshold Bootstrap Procedure

### 2.3.1 Chunk Creation

Each bullet signature  $S_{hijk}$  is split into  $m$  chunks

$$\{\{S_{hijk1}, \dots, S_{hijk(\ell_1-1)}\}, \{S_{hijk\ell_1}, \dots, S_{hijk(\ell_2-1)}\}, \dots, \{S_{hijk\ell_m}, \dots, S_{hijkL}\}\}$$



around the sequence mean or median. Each chunk consists of  $t$  complete cycles around the middle line value. We can reference these chunks (rather than the underlying sequence) as  $C_{hijk m}$ , if it is so desired. Boundary chunks  $C_{hijk1} := (S_{hijk1}, \dots, S_{hijk(\ell_1-1)})$  and  $C_{hijk m} := (S_{hijk\ell_m}, \dots, S_{hijkL})$  may not contain the full  $t$  cycles; these chunks are handled separately and  $C_{hijk m}$  is reversed so that it has the same orientation as  $C_{hijk1}$  - this allows for a bit more flexibility in generating the sequence ends using the cycle bootstrap.

### 2.3.2 Bootstrap Signature Creation

Then a signature generated from the threshold bootstrap,  $S^*$ , of target length  $\mathcal{T}_t^*$  is assembled from  $Z^*$  chunks as follows:

1. The sign of the sequence,  $a$ , is selected randomly from  $-1, +1$ , where  $-1$  means that full chunks will initially be negative and the last value will be positive.
2. Boundary chunks  $C_0^*, C_Z^*$  are sampled from the collection of boundary chunks. These chunks have length  $\mathcal{L}_0$  and  $\mathcal{L}_Z$  respectively.
3. The total number of missing values at the beginning of the boundary chunks,  $m^*$  is calculated.
4.  $C_0^*$  is arranged so as to be compatible with the chosen sign  $a$ .
5.  $C_Z^*$  is reversed and arranged to be compatible with the chosen sign  $a$ .
6. The list of all complete chunks  $\{C\}_{0 < l < L}$  is permuted and truncated such that the cumulative length of the chunks is less than the remaining length excluding missing values,  $\mathcal{L}_0 + \mathcal{L}_Z - m^*$ , or the first chunk in the case that the first chunk selected is longer than the remaining length.
7. The chunks are assembled by index
8. The sequence is either truncated or padded with NAs so that it has target length  $\mathcal{T}_t^*$

### Cycle Breaks for Hamby Barrel 1 Aligned LEAs

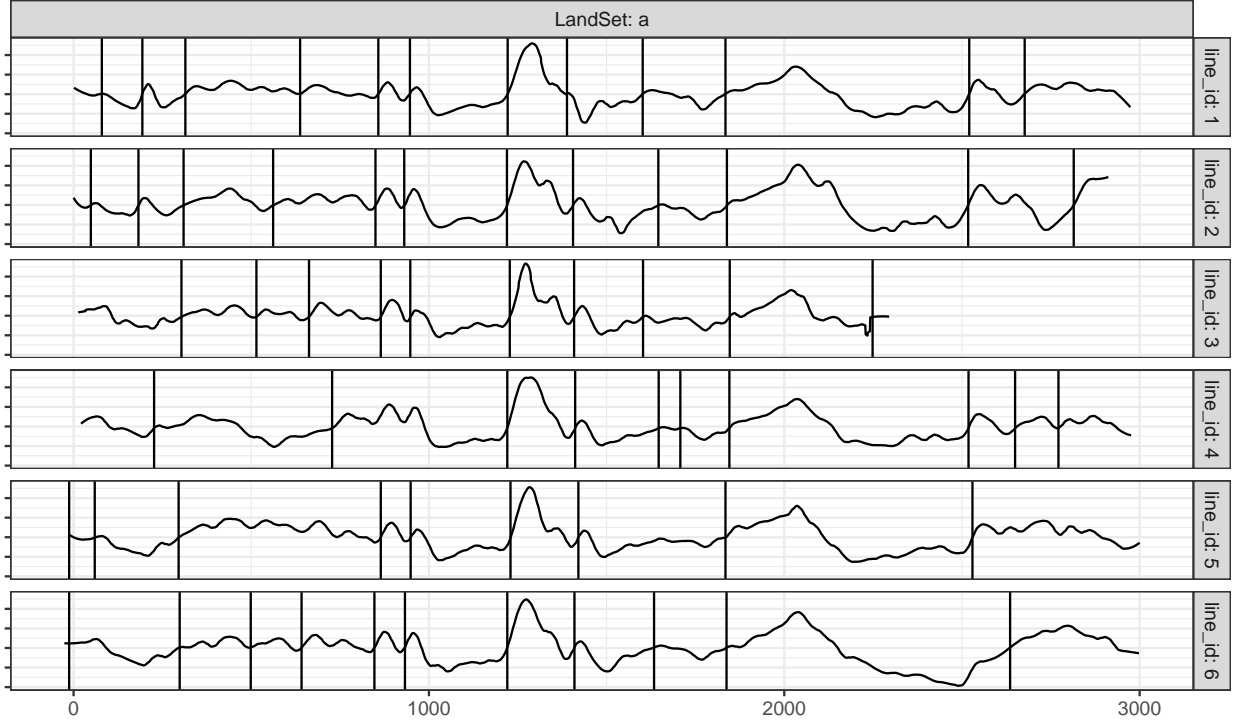


Figure 3: There is significant correspondence between the breaks of aligned sequences. We must avoid using sequences in different rows when comparing two generated sequences, because the incidental probability of correspondence is higher than it would be for chunks created from completely independent sequences resampled with replacement.

#### 2.3.3 Generation of Known Non-Match Sequences

For known non-matching sequences, the input sequences  $S$  are the computed signatures from each of the lands in the collection of imaged data and are indexed by  $i$ , the barrel number,  $j$ , the sequence number for each barrel  $i$ , and  $k$ , the land number.

As signatures from the same physical land in the same barrel have similar features, we need to ensure that the bootstrap sampling scheme used does not allow for incidental overlap due to the fact that our signatures are not independent. As shown in Figure 3, the chunk boundaries of aligned sequences tend to be aligned, though the correspondence is not exact due to slight variations in how the signature median intersects the signature. Even with these variations, simple random sampling from the entire set of available signature chunks would introduce the potential for unacceptable correspondence between sequences

which are intended to be non-matching. Thus, a more restrictive sampling method which does not allow for the possibility of sampling from signatures originating from the same source barrel is necessary.

The simplest way to ensure non-matching sequences is to compare signatures sampled from a collection of barrels  $G_1$  to signatures sampled from a collection of barrels  $G_2$ , where  $G_1 \cap G_2 = \emptyset$ . This ensures that incidental matches from different signatures originating from the same land of the same barrel do not occur, because same-barrel comparisons are not possible under this sampling strategy. Any arrangement such that generated sequences  $S_1^*$  and  $S_2^*$  are do not contain chunks from the same barrel is permissible, however, it is much simpler to consider a sampling scheme where  $S_1^*$  is generated from chunks originating from barrels in  $G_1$  and  $S_2^*$  is generated from chunks originating from barrels in  $G_2$  than to consider the more general case.

#### 2.3.4 Generation of Known Match Sequences

Generation of known match sequences is somewhat more complicated, in part because it is necessary to generate two separate sequences: the pattern which might be thought of as originating from the physical object (in this case, toolmarks on the lands of the barrel), which is consistent across sequences, and the variability around this pattern which reflects the differences in the firing process, any damage to the bullet, and any measurement error. To generate these two separate sequences, we must first have a set of co-aligned signatures from the same land of the same barrel. These sets of lands will be used to generate a set of consensus sequences which will be resampled to generate the underlying structure corresponding to the physical object. Deviations from the consensus sequence will be resampled separately to produce individual signature differences from the consensus sequence. In both cases, resampling will take place using the general threshold bootstrap framework described in Section 2.3.2.

Dividing the sequences shown in Figure 4 into the mean structure and variability around the mean produces two separate sets of information which can be resampled separately. [Cite Basu's theorem saying these are point-wise independent?](#)

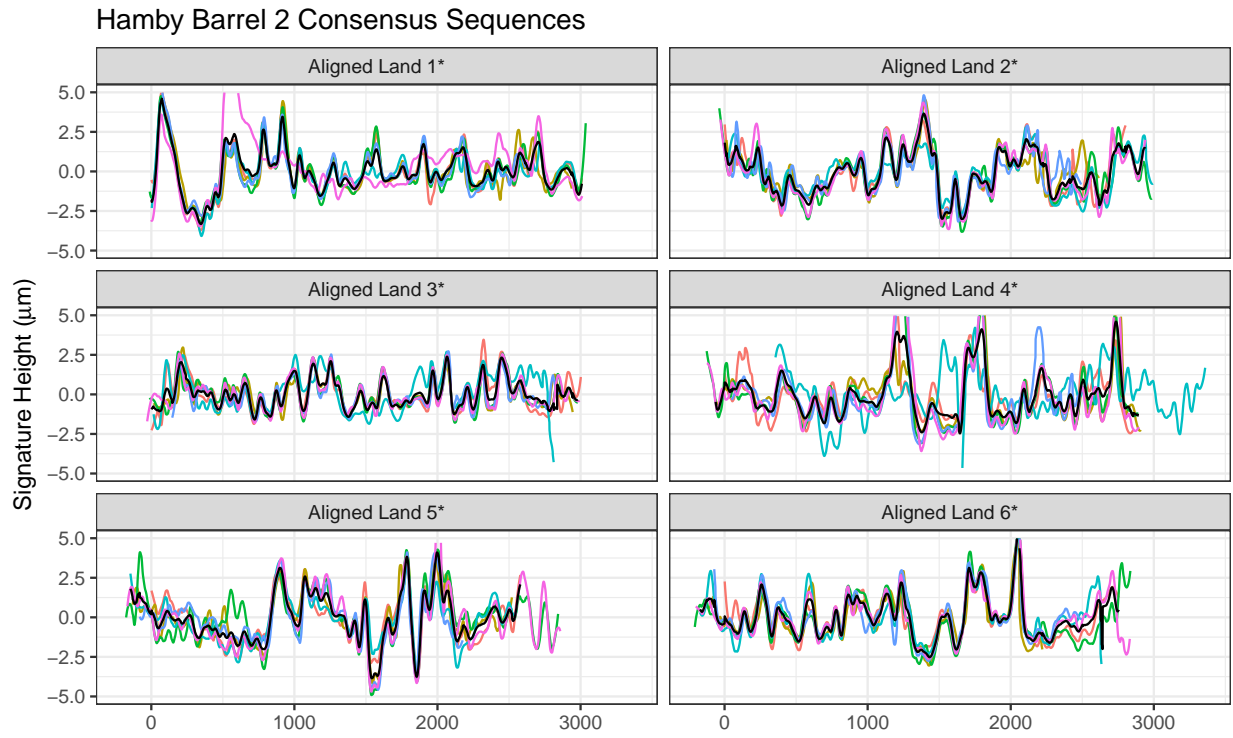


Figure 4: Aligned signatures from Barrel 2 of the Hamby studies. The mean (where at least 3 sequences are defined) is shown in black.

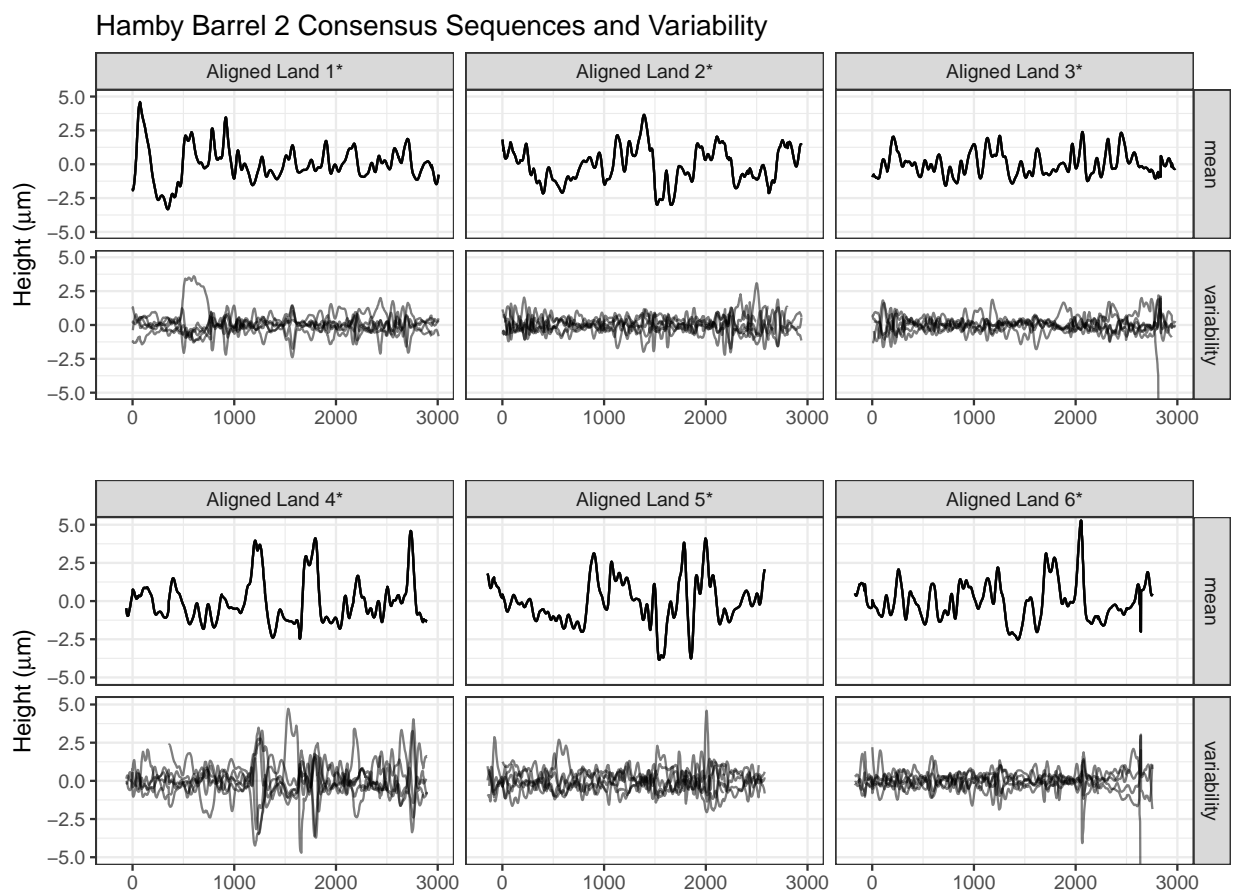


Figure 5: Aligned signatures (left) and variability (right) for Barrel 2 of the Hamby studies.

## 2.4 Consecutive Matching Striae and resampling

## 2.5 Theoretical probability of signature overlap due to same-piece resampling

## 2.6 Computational assessment

# 3 Case study: Resampling Hamby bullet signatures

## References

- Bachrach, B., Gao, P., Xu, R., Wang, W., Mishra, A., Tang, K. & Zhang, G. (2013), ‘Development of Synthetically Generated LEA Signatures to Generalize Probability of False Positive Identification Estimates’, p. 72.
- Bonfanti, M. & De Kinder, J. (1999), ‘The Influence of the Use of Firearms on their Characteristic Marks.’, *Association of Firearms & Tool Mark Examiners (AFTE) Journal* **31**, 318–323.
- Hamby, J. E., Brundage, D. J. & Thorpe, J. W. (2009), ‘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.
- Hare, E., Hofmann, H. & Carriquiry, A. (2017), ‘Automatic matching of bullet land impressions’, *Ann. Appl. Stat.* **11**(4), 2332–2356.  
**URL:** <https://doi.org/10.1214/17-AOAS1080>
- Lund, S. P. & Iyer, H. (2017), ‘Likelihood Ratio as Weight of Forensic Evidence: A Closer Look’, *Journal of Research of the National Institute of Standards and Technology* **122**.  
**URL:** <https://nvlpubs.nist.gov/nistpubs/jres/122/jres.122.027.pdf>
- Masi, I., Trần, A. T., Hassner, T., Leksut, J. T. & Medioni, G. (2016), Do We Really Need to Collect Millions of Faces for Effective Face Recognition?, in B. Leibe, J. Matas, N. Sebe & M. Welling, eds, ‘Computer Vision – ECCV 2016’, Vol. 9909, Springer International

Publishing, Cham, pp. 579–596.

**URL:** [http://link.springer.com/10.1007/978-3-319-46454-1\\_35](http://link.springer.com/10.1007/978-3-319-46454-1_35)

Park, D. & Willemain, T. R. (1999), ‘The threshold bootstrap and threshold jackknife’, *Computational Statistics & Data Analysis* **31**(2), 187–202.

**URL:** <http://linkinghub.elsevier.com/retrieve/pii/S0167947399000080>

Rabasse, C., Guest, R. M. & Fairhurst, M. C. (2008), ‘A New Method for the Synthesis of Signature Data With Natural Variability’, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **38**(3), 691–699.

Sesa-Nogueras, E. & Faundez-Zanuy, M. (2013), ‘Writer recognition enhancement by means of synthetically generated handwritten text’, *Engineering Applications of Artificial Intelligence* **26**(1), 609–624.

**URL:** <http://linkinghub.elsevier.com/retrieve/pii/S0952197612000668>

Vorburger, T., Song, J.-F., Chu, W., Ma, L., Bui, S., Zheng, A. & Renegar, T. (2011), ‘Applications of cross-correlation functions’, *Wear* **271**(3-4), 529–533.

**URL:** <http://linkinghub.elsevier.com/retrieve/pii/S0043164810001407>

Wecker, L., Samavati, F. & Gavrilova, M. (2010), ‘A multiresolution approach to iris synthesis’, *Computers & Graphics* **34**(4), 468–478.

**URL:** <http://linkinghub.elsevier.com/retrieve/pii/S0097849310000798>

Zheng, X. A., Soons, J. A. & Thompson, R. M. (2016), ‘NIST Ballistics Toolmark Research Database | NIST’.

**URL:** <https://www.nist.gov/publications/nist-ballistics-toolmark-research-database>