

# Three-dimensional data of wire-cut surface scans under the confocal microscope (110 character maximum, inc. spaces)

Yuhang Lin<sup>1,2,\*</sup>, Heike Hofmann<sup>2,3</sup>, Curtis Mosher<sup>4</sup>, Eden Amin<sup>5</sup>, Jeff Salyards<sup>2</sup>, and Alicia Carriquiry<sup>1,2</sup>

<sup>1</sup>Iowa State University, Department of Statistics, Ames, IA, 50011, USA

<sup>2</sup>Iowa State University, Center for Statistics and Applications in Forensic Evidence, Ames, IA, 50011, USA

<sup>3</sup>University of Nebraska-Lincoln, Department of Statistics, Lincoln, NE, 68583, USA

<sup>4</sup>Iowa State University, Roy J Carver High Resolution Microscopy Facility, Ames, IA, 50011, USA

<sup>5</sup>University of Central Oklahoma, Forensic Science Institute, Edmond, OK, 73034, USA

\*corresponding author(s): Yuhang Lin (yhlin@iastate.edu)

## ABSTRACT

Update later: max of 170 words: describe the study, the assay(s) performed, resulting data, and reuse potential

Wire cut data is important in forensic investigations but lacks a systematic way of analyzing the data. We created a data set of 120 scans of aluminum wire cut in  $\times 3\mu$  format, using 5 wire cutters and 3 locations along the 4 blades, with 2 replicates for each combination. A systematic pipeline with multiple analysis plots was developed to analyze the data and draw conclusions based on numerical measures.

## Background & Summary

An important part of a forensic analysis is the investigation of marks left at a crime scene. Forensic examiners are in particular interested in the origin of those marks, ie. in the investigation of their source. This is known as the Source Identification Problem in Forensic Science [citation?](#). The forensic science community generally distinguishes between the **specific source problem**, where the examiner is interested in whether a mark was left by a specific tool, and the **common source problem**, where the focus of the investigation is on whether two marks were left by the same tool. Current accepted practice in both of these situations is based on a visual inspection of the items under a comparison microscope and results, according to the Theory of Identification<sup>1</sup> developed by the Association of Firearm and Toolmark Examiners (AFTE), in a conclusion of the form *identification* (the marks are believed to have been made by the same source), *elimination* (the marks on the two items are believed to have been made by different tools), and *inconclusive* (there are not enough similarities or dissimilarities between the marks to allow either an identification or an elimination). At its core, this assessment is subjective in its nature and has been criticized in reports by the National Research Council<sup>2</sup> and the President's Council of Advisors on Science and Technology<sup>3</sup> for its lack of objectivity and the absence of error rates.

Both of these issues rely on data with known ground-truth.

Biedermann<sup>4</sup> distinguishes between internal and external perspectives in the forensic literature. The external perspective only allows general statements based on (black-box) studies relating examiners' conclusions to ground truth without considering any evidence of a particular case. To allow an internal perspective, there is a need to quantitatively capture the basis for an evaluation based on specific evidence.

When a bladed tool cuts a wire, striation marks are left on the cut surface of the wire, as shown in Figure 1.

Wire cut data is a type of forensic tool mark data used to assess similarity of striations left on the surface by a wire cutter. There have been cases where the evidence and testimony on wire cut evidence played a crucial role in the criminal investigation and conviction of a defendant.

However, there is a lack of a standardized method to analyze it, except for visual comparison.

Work on striations made by screwdrivers<sup>5</sup> identified angle of attack<sup>6</sup>, rotational axis<sup>7</sup>, and direction<sup>8</sup> as the main factors for comparisons.

Earlier research by Ma et al.<sup>9</sup> and Zheng et al.<sup>10</sup> has focused on collecting and distributing datasets for this purpose and providing a foundation for future advancements in tool mark analysis. Studies such as those by Chu et al.<sup>11</sup> and Vorburger et al.<sup>12</sup> have demonstrated the efficacy of using numerical methods to improve accuracy and consistency in tool mark analysis.



**Figure 1.** Microscopic close-up of striations left by a blade on the cut end of a wire.

42 Hare et al.<sup>13</sup> and Ju et al.<sup>14</sup> have explored methods for quantifying the similarity between representative signals, but alignment  
43 remains a major hurdle.

44 In this study, we would like to follow the same path and provide a data set of wire cut scans, and also discuss a systematic  
45 pipeline to analyze the data and draw conclusions based on numerical measures. Here, we provide a data set containing  
46 multiple files, as described in Table 1.

47 For the reproducibility of all our data and alignment results, we introduce in detail in [Cutting Wires](#) [hyperlink location](#)  
48 [incorrect for unnumbered sections](#) it seems that the hyperlinks make sure that the section is on the page how we cut the wire  
49 and collect the 120 scans with 5 tools, in [Extract Profiles](#) how we extract profiles from the scans, in [Filtered Signals](#) how we  
50 filter signals from the profiles, in [Align Signals](#) how we align signals from different scans and optimize the alignment with the  
51 cross-correlation function (CCF) values. In [Data Records](#), we discuss where our data is held. Then, in [Technical Validation](#),  
52 a technical validation was conducted to further compare signals from different sources also match our assumption, together  
53 with visual aids for drawing conclusions. Finally, in [Usage Notes](#), we provide available codes for creating the data set and  
54 conducting technical validation, as discussed in [Methods](#) and [Technical Validation](#). [Code availability](#) discusses where these  
55 codes can be found online. We hope this pipeline developed using this data set can be further generalized and applied to real  
56 crime scenes to help investigators draw conclusions based on real wire cut data.

## 57 **Methods**

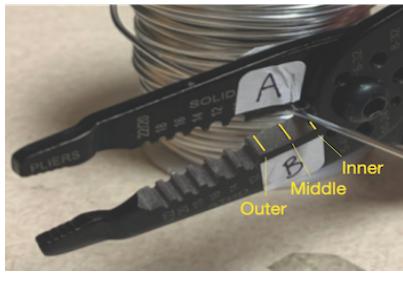
58 In this study, aluminum wire was used to create an optimal scenario where the most amount of information could be transferred  
59 from the tool to the substrate despite the wire, in some real cases, being made of lead. The physical property of aluminum  
60 wire makes it an excellent candidate for keeping marks while being relatively easy to bend and non-toxic.

### 61 **Cutting Wires**

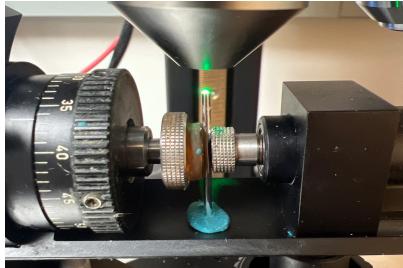
62 The aluminum wire used was 16 Gauge/1.5 mm, anodized. In order to cut the wire, 4-inch pieces were unspooled and cut using  
63 Kaiweets wire cutters, model KWS-105, as shown in Figure 2a, for 1 blade location, either inner, middle, or outer, which gives  
64 us 1 replicate. Each piece was then cut into half to create 2-inch pieces for each side, AB and CD, with a sharpie line marking  
65 the cut ends, giving us 4 samples. Then, we can use the standard scanning protocols for the confocal microscope, shown in  
66 Figure Figure 2b, to scan the wire tip surfaces. The scanned surfaces are saved in a resolution of  $0.645\mu m \times 0.645\mu m$  per  
67 square pixel in an x3p file format. Here, we are showing AB and CD sides in Figure 2c, with the back of A being C and the  
68 back of B being D. Both AB and CD sides form tent structures on the tips of the wire, and we can separate each side of the  
69 tent into 2 pieces along the bending position, resulting in 8 scans. We repeated this process for all 3 locations along the blade  
70 and 5 wire cutters, with 2 replicates for each tool-edge-location combination, resulting in 120 scans. Each piece was labeled  
71 with the naming conventions, T(ool) 1/2/3/4/5 (Edge) A/B/C/D W(ire) - L(ocation) I(nner)/M(iddle)/O(uter) - R(epetition)  
72 1/2, with T1AW-LI-R1 being the piece cut by tool 1 on the A edge at the inner location for the first repetition.

**Table 1.** Structure of available data and files.

	Description	Section
<b>Raw data</b>		
scans/	folder containing 120 topographic 3d scans corresponding to 30 aluminum wire cuts (x3p format)	<a href="#">Cutting Wires</a>
meta.csv	meta information for each cut with tool, blade, and location information (CSV format)	<a href="#">Cutting Wires</a>
<b>Manual derivatives</b>		
profiles/	folder of files with manually extracted profiles (CSV format)	<a href="#">Extract Profiles</a>
<b>Computational derivatives in folder 'data-derived/'</b>		
wire-signals	signals processed from corresponding profile (zipped CSV format)	<a href="#">Filtered Signals</a>
wire_pairwise_ccf	CCF values of all pairwise aligned signals (zipped CSV format)	<a href="#">Align Signals</a>
<b>Image files</b>		
pngs/	folder containing pictures of 3d scans of wire cuts (PNG format)	<a href="#">Cutting Wires</a>
profile-images/	folder containing pictures of profile extracted from wire cuts (PNG format)	<a href="#">Extract Profiles</a>
<b>Visual Inventory in folder 'assessment/analysis-manual/'</b>		
processing-wires	display of pairwise aligned signals from the same sources (HTML format)	<a href="#">Align Signals</a>



(a)

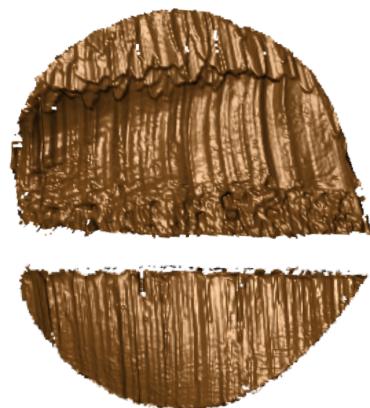


(b)

Blade A



Blade C



Blade B

(c)

Blade D

**Figure 2.** (a) A Kaiweets wire cutter of model KWS-105 was used to cut the wire, with inner, middle and outer locations marked. (b) A confocal microscope was used to scan the wire surfaces. (c) After separating 2 tent structures by the connecting position, we obtained 4 samples - 2 samples from blade A and B, and others from blade C and D. width and height are tuned manually | full requirements see <https://www.nature.com/sdata/publish/submission-guidelines#figures>

### 73 Extract Profiles

74 Numerical comparisons between 2 replicates cannot be done directly on the `x3p` files. We need to extract representative  
 75 functions from the scans first. A representative function with the most information is considered as a signal for one scan,  
 76 which can be used later for comparison. To obtain this function, we first need a profile of the scan, which is a sequence of  
 77 values along a user-drawn line on the surface. The profile should capture most features of the scan and be orthogonal to the  
 78 striation marks of the scan, which are formed by the ups and downs of grooves. So, we draw the line across the wide region of  
 79 the scan to maximize the feature captured, as shown in dark blue in Figure 3a. We can then investigate the values under this  
 80 profile line. The profile function along the line is plotted in Figure 3b.

### 81 Filtered Signals

82 With the profile extracted, we can then obtain the signal. Two Gaussian filters, as discussed in Cleveland et al.<sup>15</sup>, are applied  
 83 to these resulting profiles. In particular, we first used a large low-pass filter with bandwidths of 400 microns to remove the  
 84 large trend, as it can overwhelm the signals, and then used a small high-pass filter of 40 microns to average across noise and  
 85 remove spikes, as shown in Figure 3c. (add reference: W. S. Cleveland, E. Grosse and W. M. Shyu (1992) Local regression  
 86 models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). Finally, the  
 87 extreme tail values are removed.

### 88 Align Signals

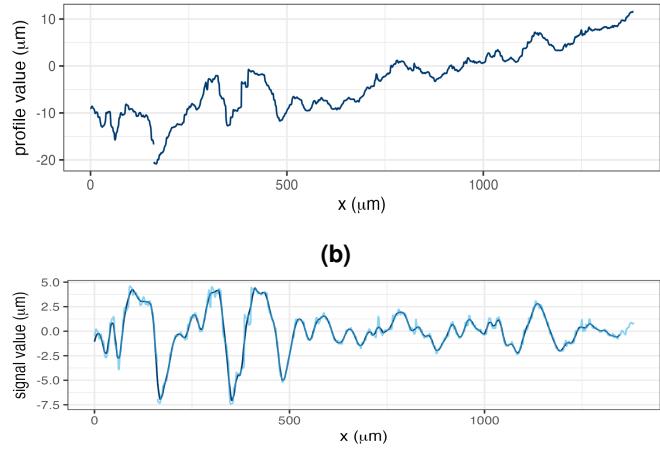
89 Signals extracted from different scans can be put together for comparison, and we maximize the cross-correlation function  
 90 (CCF) values between the signals to find the best alignment numerically. For example, we compare T1AW-LI-R1 to T1AW-  
 91 LI-R2, T1CW-LI-R1 to T1CW-LI-R2, and so on. That is comparing each row in Figure 4. We know that signals from two  
 92 replicates with the same tool-edge-location combination should yield similar signals as in the first and second columns of  
 93 Figure 5, which will give alignments of massive overlapping and high CCF values close to 1. The alignments and values we  
 94 got in the rightmost column of Figure 5 fulfill our expectations.

### 95 Data Records

96 The complete data set is available on the ISU DataShare repository at <https://iastate.figshare.com/>, which is public and open  
 97 access for every interested researcher. The structure of the data set is described before in Table 1.



(a)



(b)

(c)

**Figure 3.** (a) A profile line in dark blue was drawn across the striations of the scan. (b) The profile function extracted along the profile line in (a). (c) The raw signal in light blue is obtained by using the low-pass filter on the profile function in (b) and the filtered signal is obtained by using the high-pass filter on the raw signal.

## 98 Technical Validation

99 For the data collection process, two team members did the cutting and labeling together, then one person did the scanning and  
100 named according to the naming convention introduced in [Cutting Wires](#). The scanning was done in a specific order to ensure  
101 consistency across all scans. The data was saved in a consistent format to ensure they could be easily accessed and analyzed.  
102 A third person then checked the data to ensure that the data was consistent in naming and accurate.

103 For the validation of the scans and their processing, we investigate the correlation scores of pairwise aligned signals. Large  
104 scores between signals are indicative of being made by the same tool. As shown previously in Figure 5 in [Align Signals](#), we  
105 would expect a high correlation score between signals from scans of wires cut with the same tool. For signals from scans of  
106 wires cut with a different tool, we would expect a low correlation score. For example, we have two scans from different tools,  
107 T1AW-LI-R1 and T2AW-LI-R1, as shown in Figure 6a and Figure 6b. The alignment is shown in Figure 6c with a 0.2 CCF  
108 value, which is low, as expected.

109 We also put resulting CCFs for all pairwise comparisons in the boxplot, together with the receiver operating characteristic  
110 (ROC) curve, as in Figure 7 and Figure 8. We can see that the CCF values for the same sources are close to 1, while the CCF  
111 values for different sources are much lower than expected. This is consistent with our expectations and validates our data  
112 processing pipeline.

## 113 Usage Notes

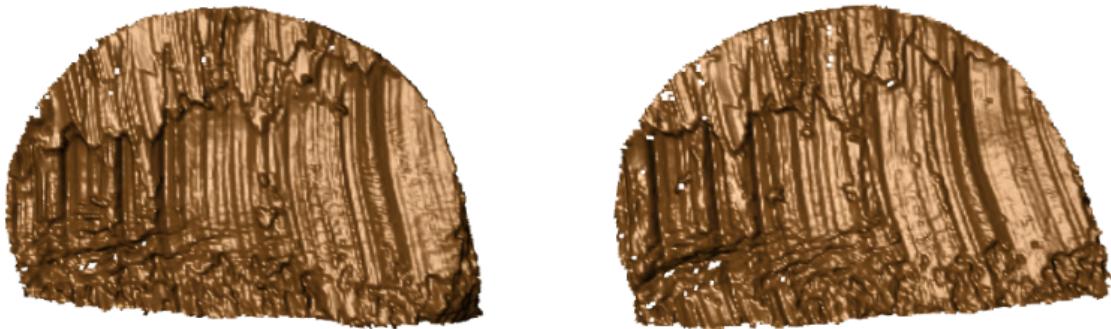
114 The R package `x3ptools`<sup>16</sup> available from CRAN supports working with files in `x3p` format. The sample scripts in R  
115 for processing scans from `x3p` format to their signal and alignment are available on GitHub [heike/wirecuts-data](#) in the  
116 assessment/code folder, as described in Table 2.

117 We already conduct pairwise comparisons and visualize some of the comparison results in [Align Signals](#) and [Technical  
Validation](#), and we can also produce other analysis plots.

118 Suppose we put the CCF values in a tilemap with different tools, locations and edge combinations. In that case, we expect  
119 only the diagonal to have high CCF values, close to 1 and marked as orange in the tilemap, as the diagonal represents the  
120 same source, and the rest of the matrix to have low CCF values, close to 0 and marked as gray. In Figure 9, the behavior is  
121 consistent with our expectation overall, except for some rare cases with tool 5 edge D. The density plot in Figure 10 shows the  
122 distribution of the CCF values with the same sources and different sources. The overlapping points between the tails of these  
123 two distributions can be a rough threshold.

124 Furthermore, the ROC curve in Figure 8 shows the sensitivity / true positive rate against the false positive rate (FPR) (1 -  
125 specificity). The curve is very close to the upper left corner, which is excellent for classification and drawing conclusions. It  
126 gives us a true threshold of 0.589 to control the FPR to be less than 0.05 with a false negative rate (FNR) to be 0, (false positive  
127

Edge A



Edge C



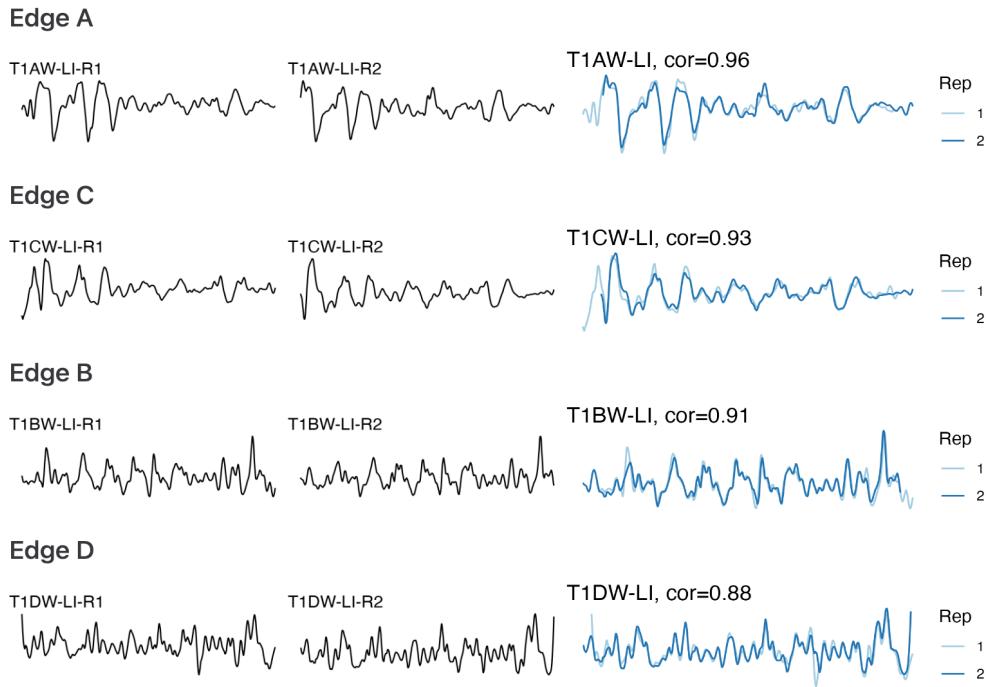
Edge B



Edge D



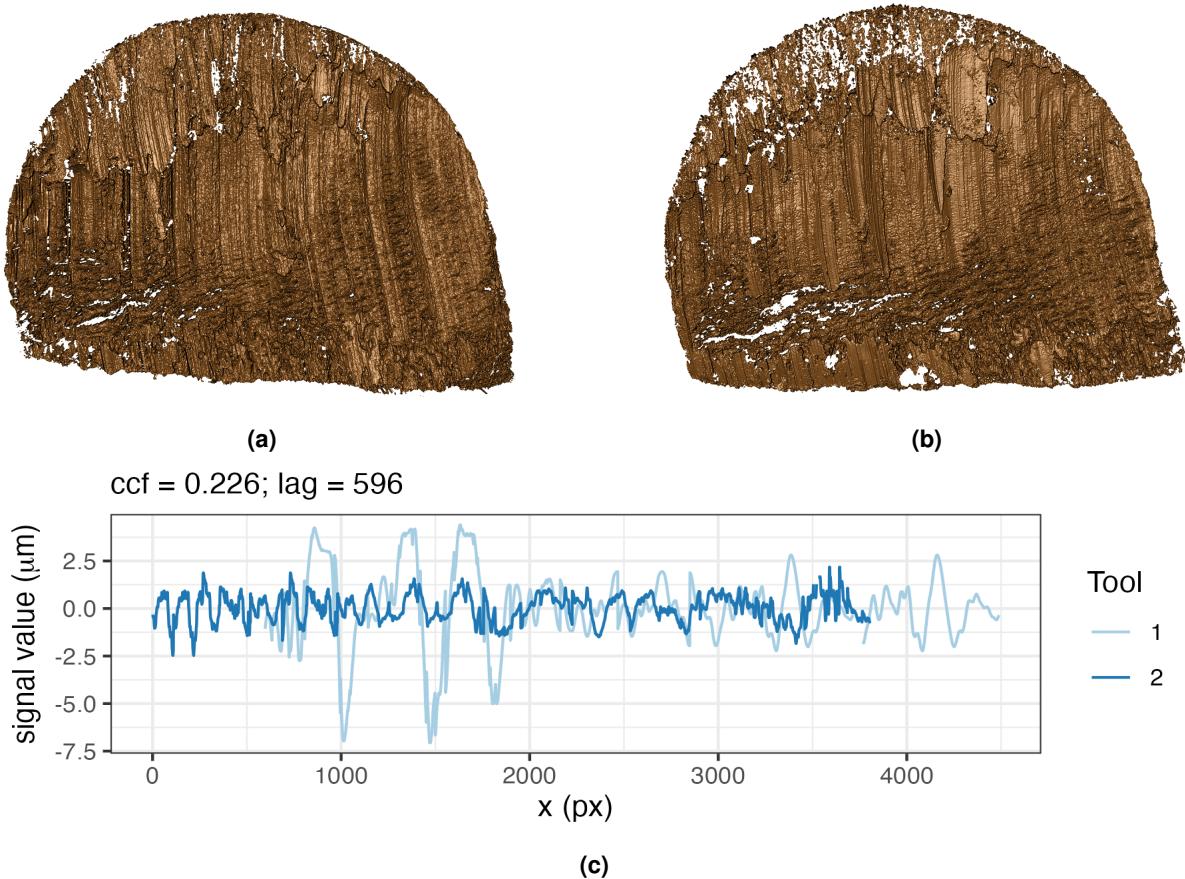
**Figure 4.** Scans from different sides of tool 1 at the inner location.



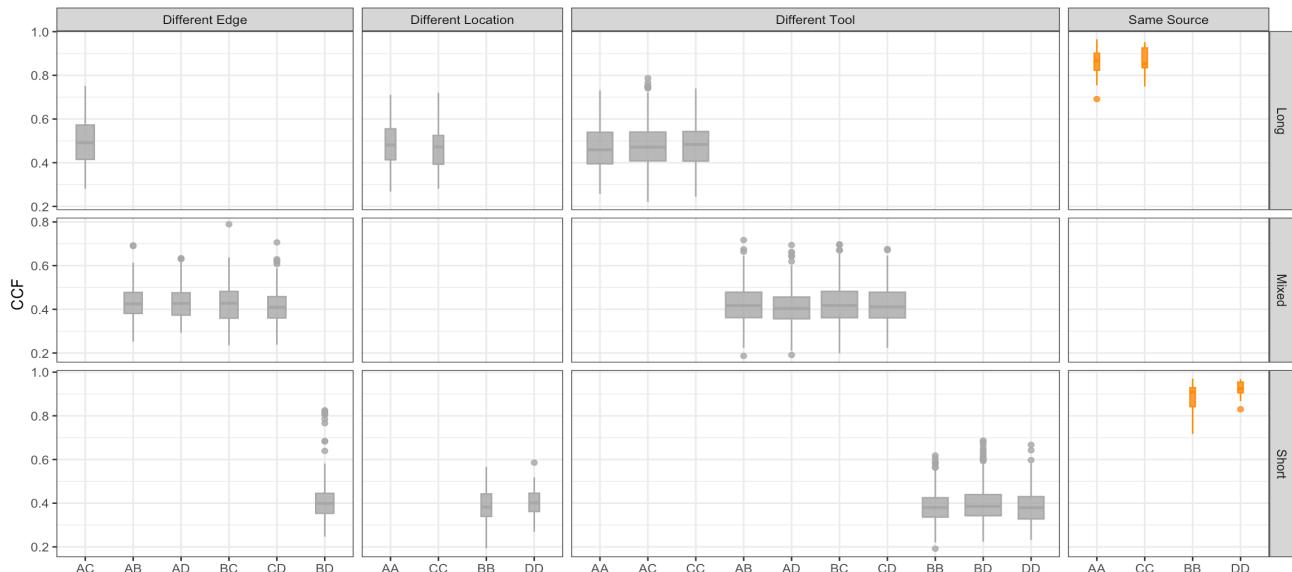
**Figure 5.** The first and second columns show the signals extracted from Figure 4, and the third column shows the alignments and CCF values between pairs of signals.

**Table 2.** Overview of available codes.

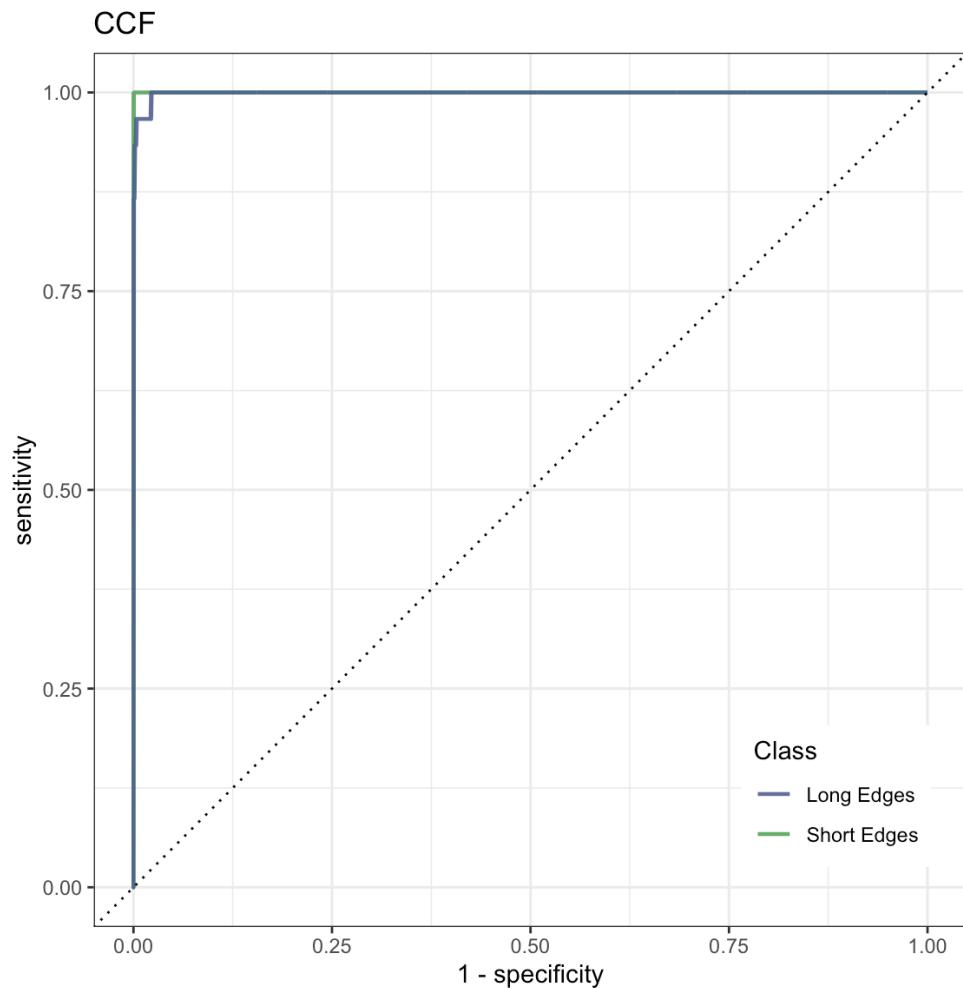
	Description	Section
Inspect raw scans		
1-create_pngs_from_x3p.R	obtain images of x3ps in scans /	<a href="#">Cutting Wires</a>
Extract profiles		
2-create_profiles_from_x3p.R	manually extract profiles from each scan	<a href="#">Extract Profiles</a>
3-create-single-profile-file.R	create meta profile information	<a href="#">Extract Profiles</a>
Derive signals		
4-create_signals_from_profiles.R	derive signals from each profile	<a href="#">Filtered Signals</a>
Align signals		
5-create-images.R	create images for pairwise alignment	<a href="#">Align Signals</a>
6-align-pairwise.R	compute pairwise alignment CCF values	<a href="#">Align Signals</a>
7-all-comparison-results.R	visualize comparison results	<a href="#">Align Signals</a>



**Figure 6.** (a) Scan T1AW-LI-R1 cut by tool 1. (b) Scan T2AW-LI-R1 cut by tool 2. (c) Alignment of signals from T1AW-LI-R1 and T2AW-LI-R1.



**Figure 7.** The boxplot shows that signals from the same sources have higher CCFs than those from different sources.



**Figure 8.** The ROC curve is bending very close to the upper left corner, which means excellent in classification and drawing conclusions.



**Figure 9.** The tilemap shows signals from the same source have CCFs close to 1.

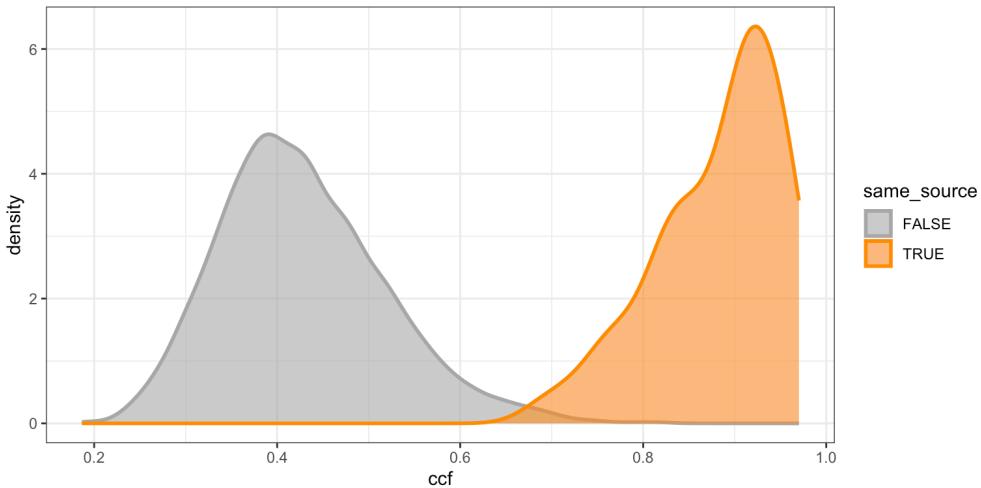
128 rate (FPR) / false discovery rate (FDR) -> define the H0 or call it false identification rate (FIR)???, and 0.658 to control the  
 129 FPR to be less than 0.01, with FNR to be 0.02.

### 130 **Code availability**

131 We made available all codes we used for inspecting raw scans, extracting profiles, deriving signals, aligning signals, and visual-  
 132 izing comparison results discussed in [Methods](#) and [Technical Validation](#), as described in Table 2. All results are reproducible  
 133 using these codes provided.

### 134 **References**

- 135 1. AFTE. The association of firearm and tool mark examiners: Theory of identification as it relates to toolmarks. *AFTE J.* **30**, 86–88 (1998).
- 136
- 137 2. NRC. *National Research Council: Strengthening Forensic Science in the United States: A Path Forward* (National  
 Academies Press, 2009).
- 138
- 139 3. President's Council of Advisors on Science and Technology. *President's Council of Advisors on Science and Technology: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (Executive Office of  
 140 the President of the United States, President's Council, Washington, D.C., 2016).
- 141
- 142 4. Biedermann, A. The strange persistence of (source) “identification” claims in forensic literature through descriptivism,  
 143 diagnosticism and machinism. *Forensic Sci. Int. Synerg.* **4**, 100222, [10.1016/j.fsisyn.2022.100222](https://doi.org/10.1016/j.fsisyn.2022.100222) (2022).
- 144
- 145 5. Baiker, M. *et al.* Quantitative comparison of striated toolmarks. *Forensic Sci. Int.* **242**, 186–199, [10.1016/j.forsciint.2014.06.038](https://doi.org/10.1016/j.forsciint.2014.06.038) (2014).
- 146
- 147 6. Baiker, M., Pieterman, R. & Zoon, P. Toolmark variability and quality depending on the fundamental parameters: Angle  
 of attack, toolmark depth and substrate material. *Forensic Sci. Int.* **251**, 40–49, [10.1016/j.forsciint.2015.03.003](https://doi.org/10.1016/j.forsciint.2015.03.003) (2015).



**Figure 10.** The density plot shows tails of distributions overlap, which can be used as a rough threshold for drawing conclusions.

- 148 7. Garcia, D. L., Pieterman, R. & Baiker, M. Influence of the axial rotation angle on tool mark striations. *Forensic Sci. Int.* 279, 203–218, [10.1016/j.forsciint.2017.08.021](https://doi.org/10.1016/j.forsciint.2017.08.021) (2017).
- 149
- 150 8. Cuellar, M., Gao, S. & Hofmann, H. An algorithm for forensic toolmark comparisons. *Forensic Sci. Int. Synerg.* 9, 100543, [10.1016/j.fsisy.2024.100543](https://doi.org/10.1016/j.fsisy.2024.100543) (2024).
- 151
- 152 9. Ma, L. *et al.* NIST Bullet Signature Measurement System for RM (Reference Material) 8240 Standard Bullets. *J. Forensic Sci.* 49, 1–11, [10.1520/JFS2003384](https://doi.org/10.1520/JFS2003384) (2004).
- 153
- 154 10. Zheng, X. A., Soons, J. A. & Thompson, R. M. NIST Ballistics Toolmark Research Database | NIST (2016).
- 155
- 156 11. Chu, W., Thompson, R. M., Song, J. & Vorburger, T. V. Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria. *Forensic Sci. Int.* 231, 137–141, [10/gn65cz](https://doi.org/10/gn65cz) (2013).
- 157
- 158 12. Vorburger, T. *et al.* Applications of cross-correlation functions. *Wear* 271, 529–533, [10.1016/j.wear.2010.03.030](https://doi.org/10.1016/j.wear.2010.03.030) (2011).
- 159
- 160 13. Hare, E., Hofmann, H., Carriquiry, A. *et al.* Automatic matching of bullet land impressions. *The Annals Appl. Stat.* 11, 2332–2356, [10.1214/17-AOAS1080](https://doi.org/10.1214/17-AOAS1080) (2017).
- 161
- 162 14. Ju, W. & Hofmann, H. The R Journal: An Open-Source Implementation of the CMPS Algorithm for Assessing Similarity of Bullets. *The R J.* 14, 267–285, [10.32614/RJ-2022-035](https://doi.org/10.32614/RJ-2022-035) (2022).
- 163
- 164 15. Cleveland, W. S., Grosse, E. & Shyu, W. M. Local Regression Models. In *Statistical Models in S* (Routledge, 1992).
- 165
- 166 16. Hofmann, H., Vanderplas, S., Krishnan, G. & Hare, E. X3ptools: Tools for Working with 3D Surface Measurements, [10.32614/CRAN.package.x3ptools](https://doi.org/10.32614/CRAN.package.x3ptools) (2018).

## 165 Acknowledgements

166 This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative  
167 Agreement 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie  
168 Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University  
169 of Pennsylvania, Swarthmore College and the University of Nebraska-Lincoln.

170 **Author contributions statement**

171 Let's follow the Elsevier definitions: <https://www.elsevier.com/researcher/author/policies-and-guidelines/credit-author-statement>  
172 Y.L.: Methodology, Software, Validation, Data Curation, Writing - Draft; H.H.: Conceptualization, Methodology, Valida-  
173 tion, Writing - Review & Editing; C.M.: Lab supervision; E.A.: Physical Specimen, Scanning; J.S: Forensic advice; A.C.:  
174 Funding acquisition.  
175 All authors reviewed the manuscript.

176 **Competing interests**

177 (mandatory statement)

178 H.H. is a technical advisor to AFTE (Association of Firearms and Toolmarks Examiners), fellow of the ASA (American  
179 Statistical Association), and committee member of the ASA Forensic Science Committee. H.H. has testified as court witness  
180 on behalf of judge April Neubauer, NY State Supreme Court Criminal Term in New York City. [other competing interests -](#)  
181 [Alicia?](#)