

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

Yuhang Lin^{1,2,*}, Heike Hofmann^{2,3}, Curtis Mosher⁴, Eden Amin⁵, Jeff Salyards², and Alicia Carriquiry^{1,2}

¹Iowa State University, Department of Statistics, Ames, IA, 50011, USA

²Iowa State University, Center for Statistics and Applications in Forensic Evidence (CSAFE), Ames, IA, 50011, USA

³University of Nebraska-Lincoln, Department of Statistics, Lincoln, NE, 68583, USA

⁴Iowa State University, Roy J Carver High Resolution Microscopy Facility (HRMF), Ames, IA, 50011, USA

⁵University of Central Oklahoma, Criminal Justice and Forensic Science, 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 $\times 3p$ 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?](#). We distinguish between a 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 problem is on whether two marks were left by the same source.

In this paper we are interested in the second of these problems: in particular, we want to determine whether two marks on cut wires were made by the same tool.

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 identify the source of a wire cutter based on the striations left on the surface. 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. Although the Association of Firearm and Toolmark Examiners (AFTE) developed the Theory of Identification ¹, which outlines the process of comparing tool marks, it is still subjective and relies on the examiner's experience, making results hard to reproduce and validate. Reports by the National Research Council ² and the President's Council of Advisors on Science and Technology ³ criticized this lack of objectivity and called for more automatic and reproducible methods to analyze tool marks that also allow an assessment of error rates in the process.

Earlier research by Ma et al. ⁴ and Zheng et al. ⁵ 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. ⁶ and Vorburger et al. ⁷ have demonstrated the efficacy of using numerical methods to improve accuracy and consistency in tool mark analysis. Hare et al. ⁸ and Ju et al. ⁹ have explored methods for quantifying the similarity between representative signals, but alignment remains a major hurdle.

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 pipeline to analyze the data and draw conclusions based on numerical measures. Here, we provide a data set containing multiple files, as described in Table 1.

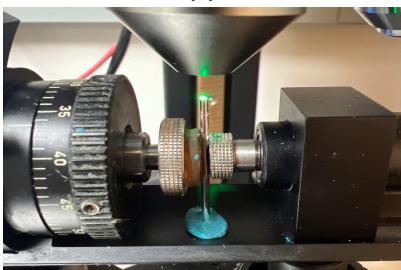
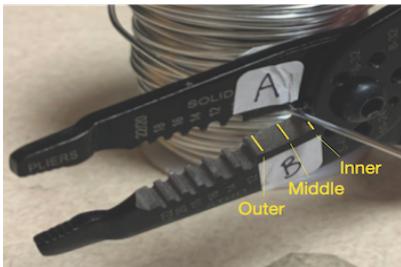
For the reproducibility of all our data and alignment results, we introduce in detail in [Cutting Wires](#) hyperlink location



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

Table 1. Structure of available data and files.

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



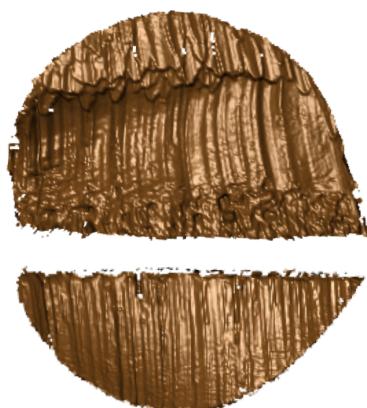
(a)

(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>

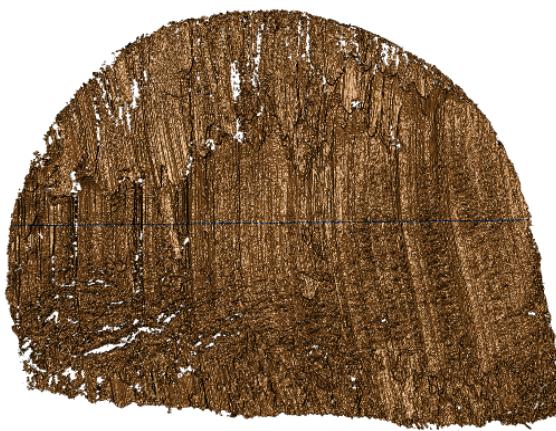
42 incorrect for unnumbered sections it seems that the hyperlinks make sure that the section is on the page how we cut the wire
 43 and collect the 120 scans with 5 tools, in Extract Profiles how we extract profiles from the scans, in Filtered Signals how we
 44 filter signals from the profiles, in Align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

51 Methods

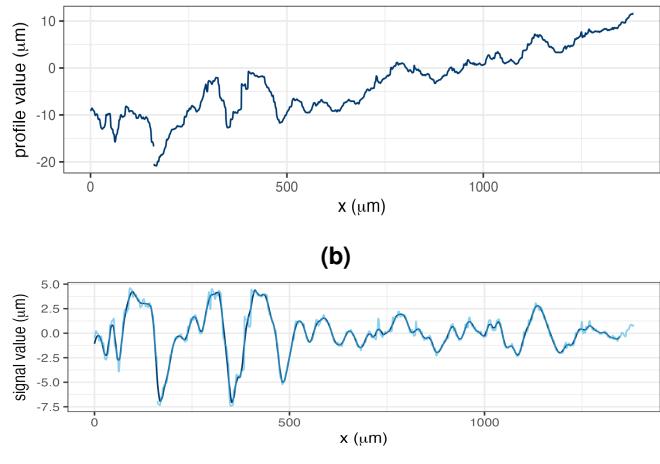
52 In this study, aluminum wire was used to create an optimal scenario where the most amount of information could be transferred
 53 from the tool to the substrate despite the wire, in some real cases, being made of lead. The physical property of aluminum
 54 wire makes it an excellent candidate for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

55 Cutting Wires

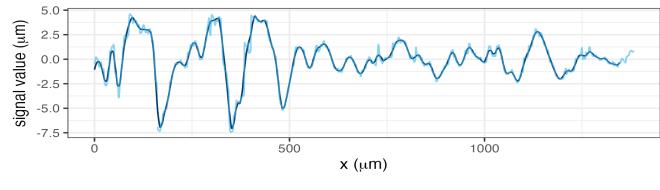
56 The aluminum wire used was 16 Gauge/1.5 mm, anodized. In order to cut the wire, 4-inch pieces were unspooled and cut using
 57 Kaiweets wire cutters, model KWS-105, as shown in Figure 2a, for 1 blade location, either inner, middle, or outer, which gives
 58 us 1 replicate. Each piece was then cut into half to create 2-inch pieces for each side, AB and CD, with a sharpie line marking
 59 the cut ends, giving us 4 samples. Then, we can use the standard scanning protocols for the confocal microscope, shown in
 60 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
 61 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
 62 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
 63 tent into 2 pieces along the bending position, resulting in 8 scans. We repeated this process for all 3 locations along the blade
 64 and 5 wire cutters, with 2 replicates for each tool-edge-location combination, resulting in 120 scans. Each piece was labeled
 65 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)
 66 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.



(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.

67 Extract Profiles

68 Numerical comparisons between 2 replicates cannot be done directly on the $\times 3p$ files. We need to extract representative
 69 functions from the scans first. A representative function with the most information is considered as a signal for one scan,
 70 which can be used later for comparison. To obtain this function, we first need a profile of the scan, which is a sequence of
 71 values along a user-drawn line on the surface. The profile should capture most features of the scan and be orthogonal to the
 72 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
 73 the scan to maximize the feature captured, as shown in dark blue in Figure 3a. We can then investigate the values under this
 74 profile line. The profile function along the line is plotted in Figure 3b.

75 Filtered Signals

76 With the profile extracted, we can then obtain the signal. Two Gaussian filters, as discussed in Cleveland et al.¹⁰, are applied
 77 to these resulting profiles. In particular, we first used a large low-pass filter with bandwidths of 400 microns to remove the
 78 large trend, as it can overwhelm the signals, and then used a small high-pass filter of 40 microns to average across noise and
 79 remove spikes, as shown in Figure 3c. (add reference: W. S. Cleveland, E. Grosse and W. M. Shyu (1992) Local regression
 80 models. Chapter 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). Finally, the
 81 extreme tail values are removed.

82 Align Signals

83 Signals extracted from different scans can be put together for comparison, and we maximize the cross-correlation function
 84 (CCF) values between the signals to find the best alignment numerically. For example, we compare T1AW-LI-R1 to T1AW-
 85 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
 86 replicates with the same tool-edge-location combination should yield similar signals as in the first and second columns of
 87 Figure 5, which will give alignments of massive overlapping and high CCF values close to 1. The alignments and values we
 88 got in the rightmost column of Figure 5 fulfill our expectations.

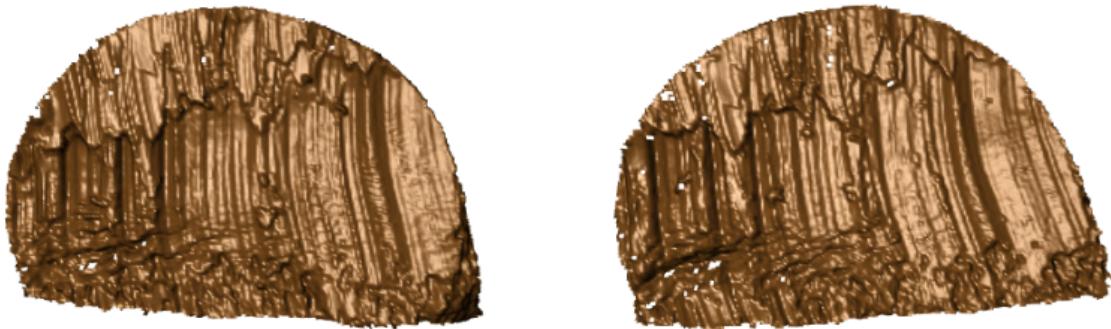
89 Data Records

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

92 Technical Validation

93 For the data collection process, two team members did the cutting and labeling together, then one person did the scanning and
 94 named according to the naming convention above. The scanning was done in a specific order to ensure consistency across all

Edge A



Edge C



Edge B



Edge D



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

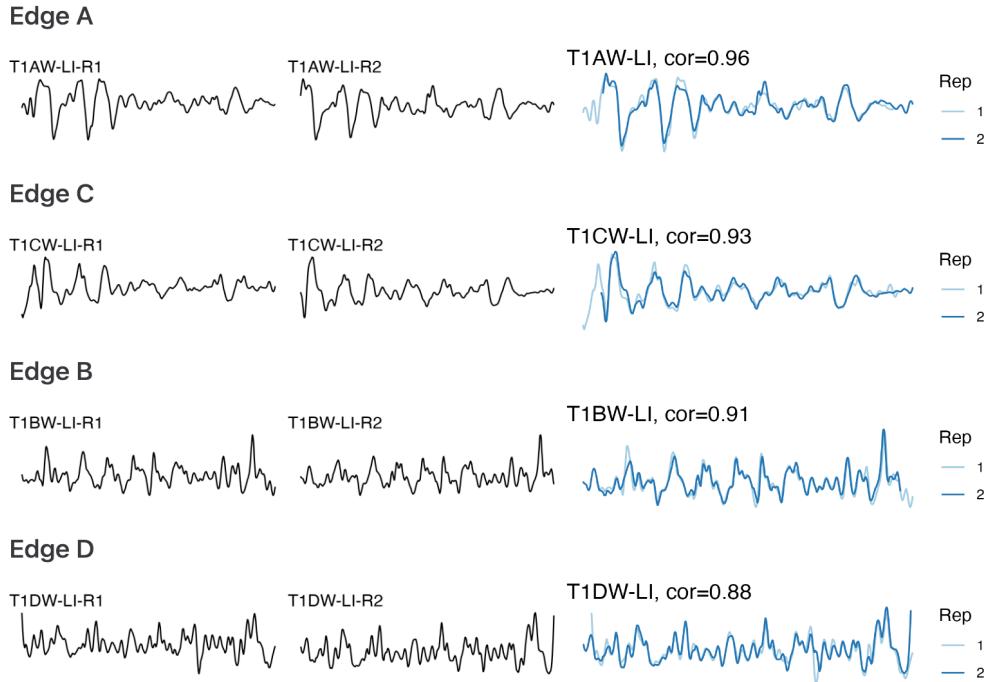


Figure 5. The first and second columns show the signals extracted from Figure 4, and the third column shows the alignments and CCF values between pairs of signals.

95 scans. The data was saved in a consistent format to ensure they could be easily accessed and analyzed. A third person then
96 checked the data to ensure that the data was consistent in naming and accurate.

97 For the validation of the scans and their processing, we investigate the correlation scores of pairwise aligned signals. Large
98 scores between signals are indicative of being made by the same tool. As shown previously in Figure 5 in Align Signals, we
99 would expect a high correlation score between signals from scans of wires cut with the same tool. For signals from scans of
100 wires cut with a different tool, we would expect a low correlation score. For example, we have two scans from different tools,
101 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
102 value, which is low, as expected.

103 We also put resulting CCFs for all pairwise comparisons in the boxplot, together with the receiver operating characteristic
104 (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
105 values for different sources are much lower than expected. This is consistent with our expectations and validates our data
106 processing pipeline.

107 Usage Notes

108 The R package `x3ptools`¹¹ available from CRAN supports working with files in `x3p` format. The sample scripts in R
109 for processing scans from `x3p` format to their signal and alignment are available on GitHub [heike/wirecuts-data](#) in the
110 assessment/code folder, as described in Table 2.

111 We already conduct pairwise comparisons and visualize some of the comparison results in Align Signals and Technical
112 Validation, and we can also produce other analysis plots.

113 Suppose we put the CCF values in a tilemap with different tools, locations and edge combinations. In that case, we expect
114 only the diagonal to have high CCF values, close to 1 and marked as orange in the tilemap, as the diagonal represents the
115 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
116 consistent with our expectation overall, except for some rare cases with tool 5 edge D. The density plot in Figure 10 shows the
117 distribution of the CCF values with the same sources and different sources. The overlapping points between the tails of these
118 two distributions can be a rough threshold.

119 Furthermore, the ROC curve in Figure 8 shows the sensitivity / true positive rate against the false positive rate (FPR) (1 -
120 specificity). The curve is very close to the upper left corner, which is excellent for classification and drawing conclusions. It
121 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

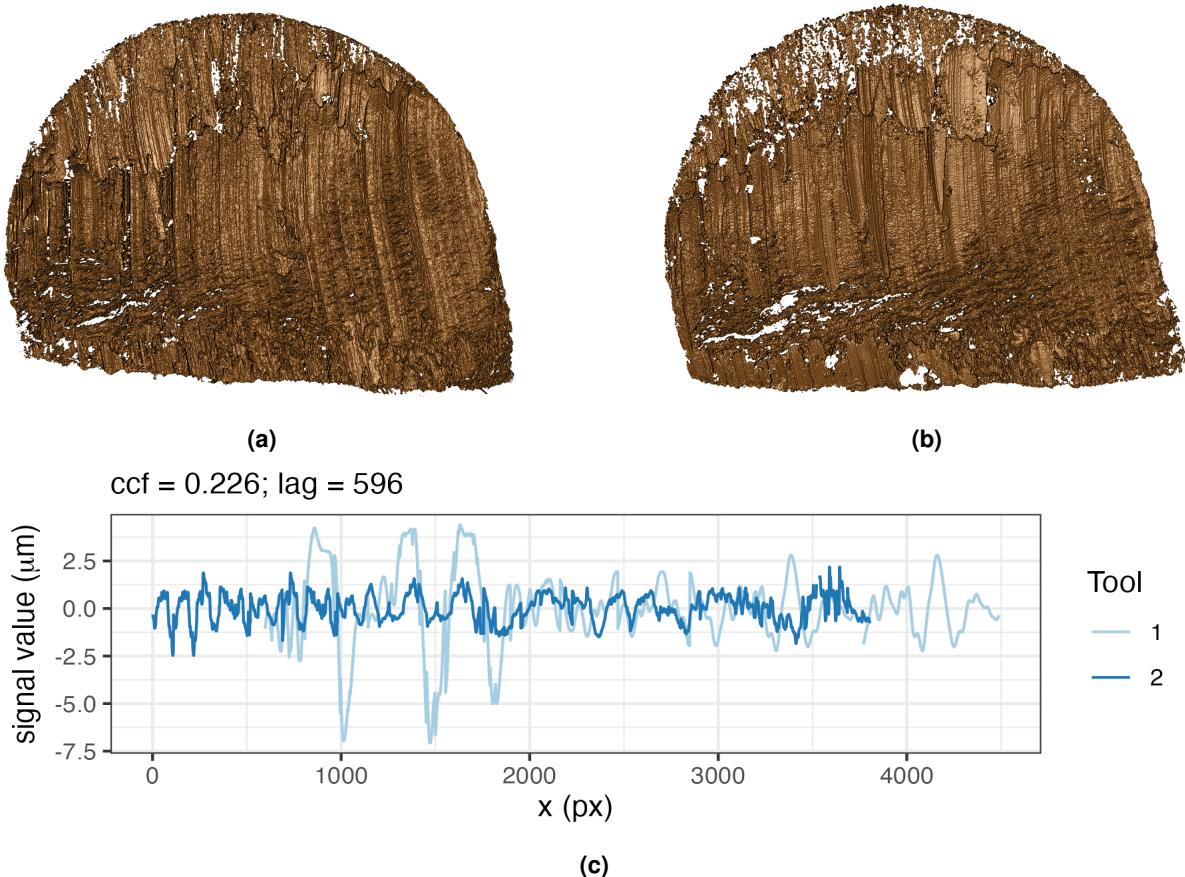


Figure 6. (a) Scan T1AW-LI-R1 cut by tool 1. (b) Scan T2AW-LI-R1 cut by tool 2. (c) Alignment of signals from T1AW-LI-R1 and T2AW-LI-R1.

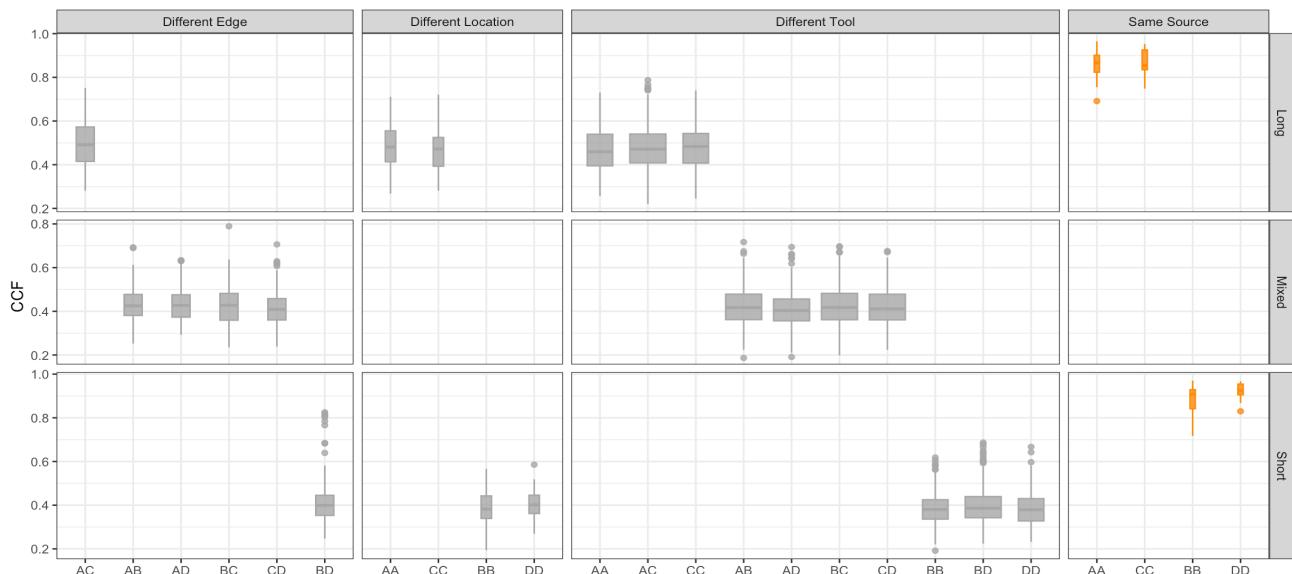


Figure 7. The boxplot shows that signals from the same sources have higher CCFs than those from different sources.

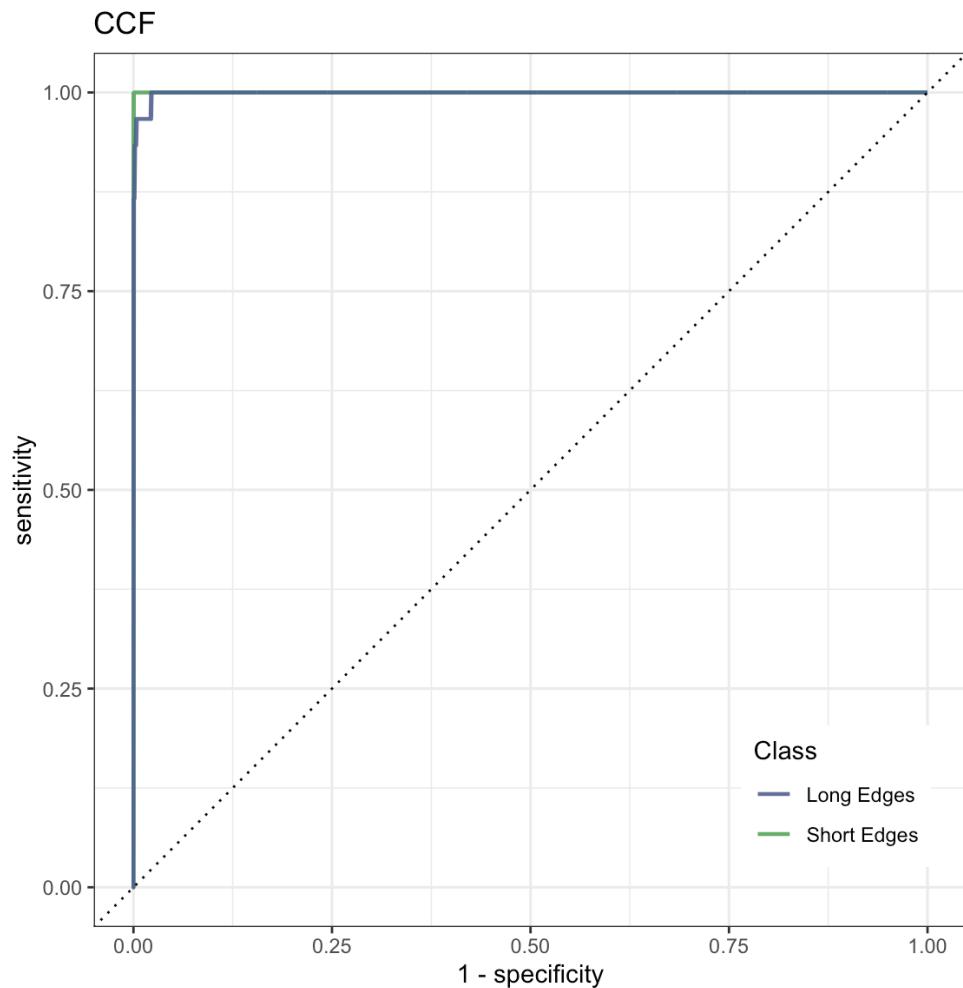


Figure 8. The ROC curve is bending very close to the upper left corner, which means excellent in classification and drawing conclusions.

Table 2. Overview of available codes.

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

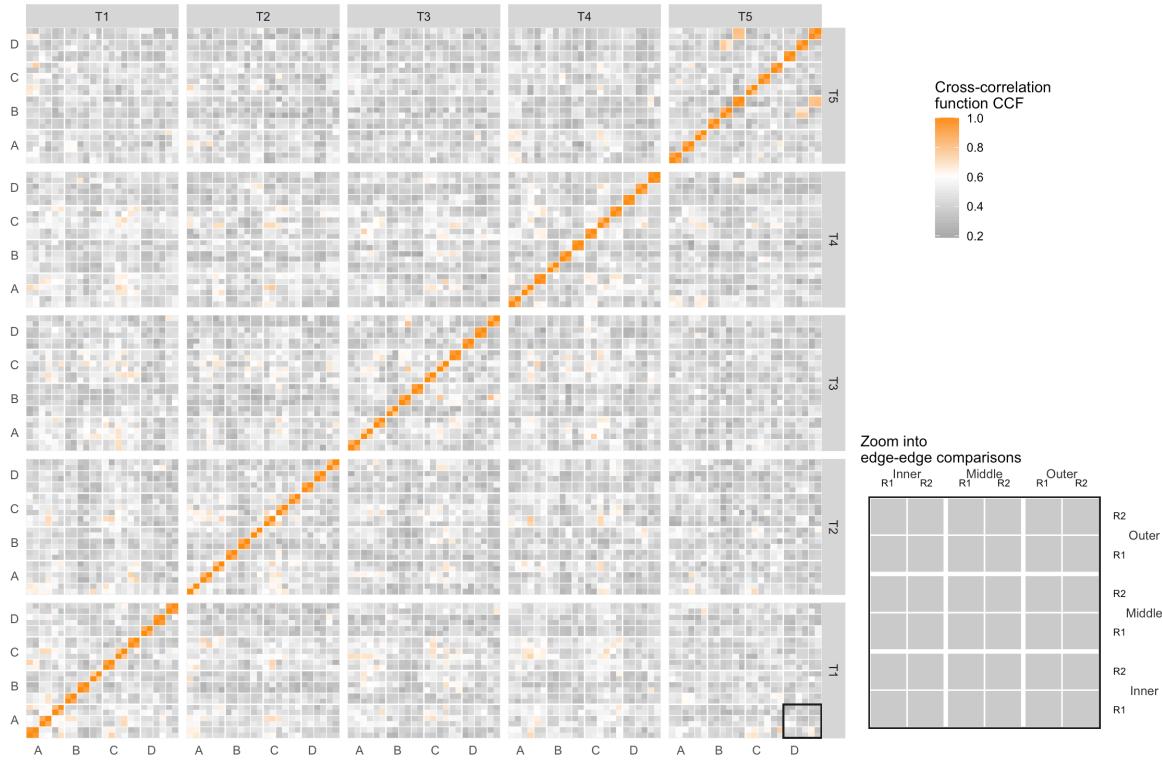


Figure 9. The tilemap shows signals from the same source have CCFs close to 1.

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

124 **Code availability**

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

128 **References**

- 129 1. AFTE. The association of firearm and tool mark examiners: Theory of identification as it relates to toolmarks. *AFTE J.* **30**, 86–88 (1998).
- 130 2. NRC. *National Research Council: Strengthening Forensic Science in the United States: A Path Forward* (National
 Academies Press, 2009).
- 131 3. President's Council of Advisors on Science and Technology. *President's Council of Advisors on Science and Technology:
 132 Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (Executive Office of
 133 the President of the United States, President's Council, Washington, D.C., 2016).
- 134 4. Ma, L. *et al.* NIST Bullet Signature Measurement System for RM (Reference Material) 8240 Standard Bullets. *J. Forensic
 135 Sci.* **49**, 1–11, [10.1520/JFS2003384](https://doi.org/10.1520/JFS2003384) (2004).
- 136 5. Zheng, X. A., Soons, J. A. & Thompson, R. M. NIST Ballistics Toolmark Research Database | NIST (2016).
- 137 6. 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).
- 138 7. 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).

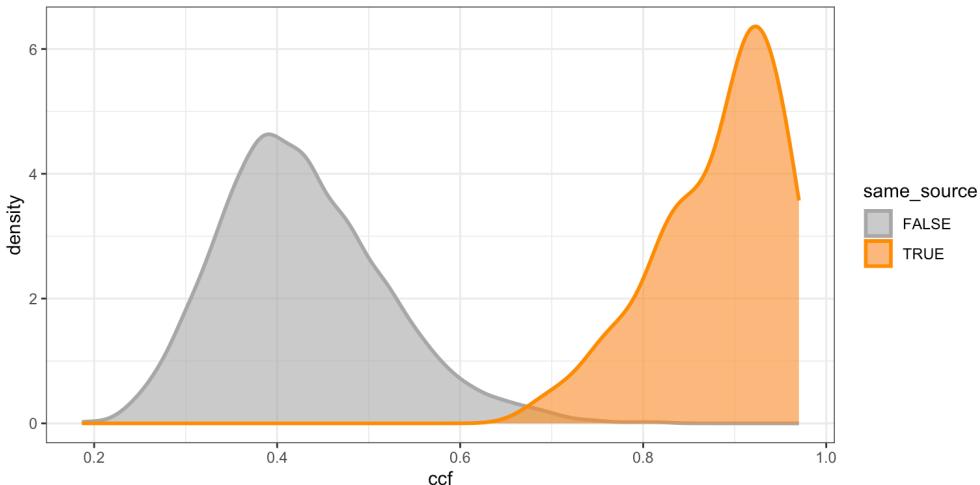


Figure 10. The density plot shows tails of distributions overlap, which can be used as a rough threshold for drawing conclusions.

- 142 8. Hare, E., Hofmann, H., Carriquiry, A. *et al.* Automatic matching of bullet land impressions. *The Annals Appl. Stat.* **11**,
143 2332–2356, [10.1214/17-AOAS1080](https://doi.org/10.1214/17-AOAS1080) (2017).
- 144 9. Ju, W. & Hofmann, H. The R Journal: An Open-Source Implementation of the CMPS Algorithm for Assessing Similarity
145 of Bullets. *The R J.* **14**, 267–285, [10.32614/RJ-2022-035](https://doi.org/10.32614/RJ-2022-035) (2022).
- 146 10. Cleveland, W. S., Grosse, E. & Shyu, W. M. Local Regression Models. In *Statistical Models in S* (Routledge, 1992).
- 147 11. Hofmann, H., Vanderplas, S., Krishnan, G. & Hare, E. X3ptools: Tools for Working with 3D Surface Measurements,
148 [10.32614/CRAN.package.x3ptools](https://doi.org/10.32614/CRAN.package.x3ptools) (2018).

149 Acknowledgements

150 This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooper-
151 ative Agreement 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie
152 Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, Univer-
153 sity of Pennsylvania, Swarthmore College and the University of Nebraska-Lincoln.

154 Author contributions statement

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

160 Competing interests

161 (mandatory statement)
162 H.H. is a technical advisor to AFTE (Association of Firearms and Toolmarks Examiners), fellow of the ASA (American
163 Statistical Association), and committee member of the ASA Forensic Science Committee. H.H. has testified as court witness
164 on behalf of judge April Neubauer, NY State Supreme Court Criminal Term in New York City. other competing interests -
165 Alicia?