

# Three-dimensional data of wirecut 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>4,5</sup>, Alicia Carriquiry<sup>1,2</sup>, Eden Amin<sup>2</sup>, Jeff Salyards<sup>2</sup>, and order? Jeff the last? check details? affiliations?<sup>1</sup>

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

<sup>2</sup>Iowa State University, Center for Statistics and Applications in Forensic Evidence (CSAFE), Ames,

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

<sup>4</sup>Iowa State University, Department of Genetics, Development and Cell Biology (GDCB), Ames,

<sup>5</sup>Iowa State University, Roy J Carver High Resolution Microscopy Facility (HRMF), Ames,

\*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  $x \times 3p$  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

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.

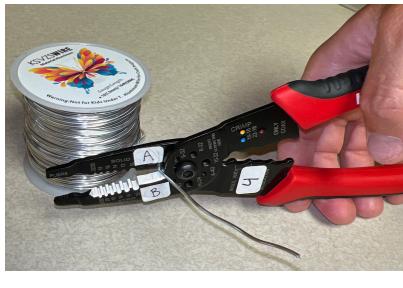
However, there is a lack of a standardized method to analyze it except for visual comparison. Although the Association of Firearm and Toolmark Examiners (AFTE) developed the Theory of Identification<sup>1</sup>, which outlines the process of comparing tool marks, it is still subjective and relies on the examiner's experience, making results hard to reproduce and validate. Several reports, such as those from the National Research Council<sup>2</sup> and the President's Council of Advisors on Science and Technology<sup>3</sup>, pointed out the abovementioned issues and called for more objective and reproducible methods to analyze tool marks. Early research by Ma et al.<sup>4</sup> and Zheng et al.<sup>5</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>6</sup> and Vorburger et al.<sup>7</sup> have demonstrated the efficacy of using numerical methods to improve accuracy and consistency in tool mark analysis. Hare et al.<sup>8</sup> and Ju et al.<sup>9</sup> 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.

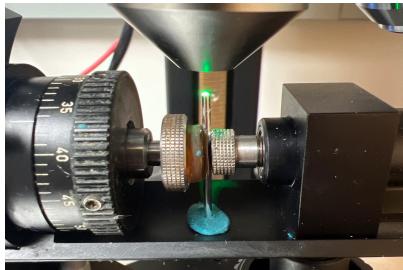
For the reproducibility of all our data and alignment results, we introduce in detail in [Cutting wires](#) [hyperlink location incorrect for unnumbered sections](#) how we cut the wire and collect the 120 scans with 5 tools on 3 locations, in [Extract profiles](#) how we extract profiles from the scans, in [Filtered signals](#) how we filter signals from the profiles, in [Align signals](#) how we align signals from different scans and optimize the alignment with the cross-correlation function (CCF) values. In [Data Records](#), we discuss where our data is held. Then, in [Technical Validation](#), a technical validation was conducted to further compare signals from different sources also match our assumption, together with visual aids for drawing conclusions. Finally, in [Usage Notes](#), we provide available codes for creating the data set and conducting technical validation, as discussed in [Methods](#) and [Technical Validation](#). [Code availability](#) discusses where these codes can be found online. We hope this pipeline developed using this data set can be further generalized and applied to real crime scenes to help investigators draw conclusions based on real wire cut data.

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

	Description	Section
<b>Raw data</b>		
scans /	folder containing 120 topographic 3d scans of aluminum wire cuts in 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</b>		
signals	folder of signals, processed from corresponding profile (csv format)	<a href="#">Filtered signals</a>
CCF values	CCF values of all pairwise aligned signals in 1 CSV	<a href="#">Align signals</a>
<b>Image files</b>		
aligned-signals	pictures of pairwise aligned signals from same sources in PNGs	<a href="#">Align signals</a>

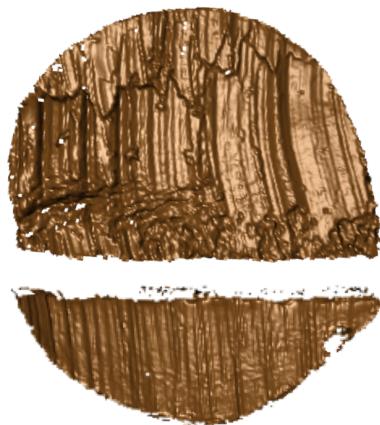


(a)

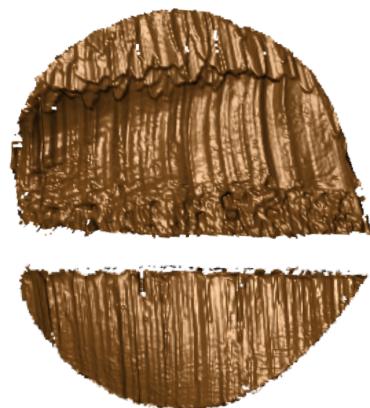


(b)

Blade A



Blade C



Blade B

(c)

Blade D

**Figure 1.** (a) A Kaiweets wire cutter of model KWS-105 was used to cut the wire. we don't have another high resolution pic showing wirecutter cutting wire (b) A confocal microscope was used to scan the wire surfaces. need an extra pic of the tip (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. full requirements see <https://www.nature.com/sdata/publish/submission-guidelines#figures>

## 41 Methods

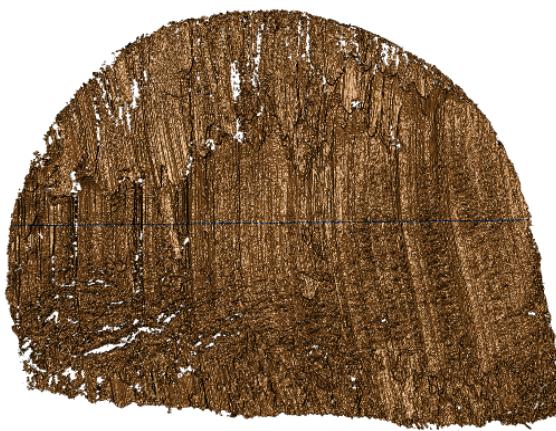
42 In this study, aluminum wire was used to create an optimal scenario where the most amount of information could be transferred  
 43 from the tool to the substrate, despite the wire in some real cases being made of lead. The physical property of aluminium  
 44 wire make it an excellent candidate for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

## 45 Cutting wires

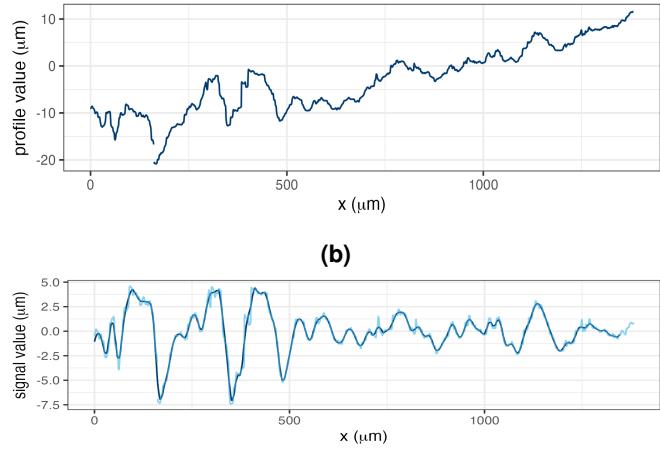
46 The aluminum wire used was 16 Gauge/1.5 mm, anodized. In order to cut the wire, 4-inch pieces were unspooled and cut using  
 47 Kaiweets wire cutters, model KWS-105, as shown in Figure 1a, for 1 blade location, either inner, middle, or outer, which gives  
 48 us 1 replicate. Each piece was then cut into half to create 2-inch pieces for each side, AB and CD, with a sharpie line marking  
 49 the cut ends, giving us 4 samples. Then, we can use the standard scanning protocols for the confocal microscope, shown in  
 50 Figure Figure 1b, to scan the wire tip surfaces. The scanned surfaces are saved in a resolution of  $0.645\mu\text{m} \times 0.645\mu\text{m}$  per  
 51 square pixel in an x3p file format. Here, we are showing AB and CD sides in Figure 1c, with the back of A being C and the  
 52 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  
 53 tent into 2 pieces along the bending position, resulting in 8 scans. We repeated this process for all 3 locations along the blade  
 54 and 5 wire cutters, with 2 replicates for each tool-edge-location combination, resulting in 120 scans. Each piece was labeled  
 55 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)  
 56 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.

## 57 Extract profiles

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



(a)



(b)

(c)

**Figure 2.** (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.

## 65 Filtered signals

66 With the profile extracted, we can then obtain the signal. Two Gaussian filters are applied to these resulting profiles. In  
 67 particular, we first used a large low-pass filter with bandwidths of 400 microns to remove large trend, as it can overwhelm the  
 68 signals, and then used a small high-pass filter of 40 microns to average across noise and remove spikes, as shown in Figure 2c.  
 69 Cleveland et al.<sup>10</sup>??? (add reference: W. S. Cleveland, E. Grosse and W. M. Shyu (1992) Local regression models. Chapter  
 70 8 of Statistical Models in S eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.). Finally, the extreme tail values  
 71 are removed.

## 72 Align signals

73 Signals extracted from different scans can be put together for comparison, and we maximize the cross-correlation function  
 74 (CCF) values between the signals to numerically find the best alignment. For example, we compare T1AW-LI-R1 to T1AW-  
 75 LI-R2, T1CW-LI-R1 to T1CW-LI-R2, and so on. That is comparing each row in Figure 3. We know that signals from two  
 76 replicates with the same tool-edge-location combination should yield similar signals as in the first and second column of  
 77 Figure 4, which will give alignments of massive overlapping and high CCF values close to 1. The alignments and values we  
 78 got in the rightmost column of Figure 4 fulfill our expectations.

## 79 Data Records

80 The complete data set is available on the ISU DataShare repository at <https://iastate.figshare.com/>, which is public and open  
 81 access for every interested researcher. The data set consists of 120 scans in the x3p file format with the naming convention  
 82 as described before. (Explain the x3p header info?)

## 83 Technical Validation

84 For the data collection process, two team members did the cutting and labeling together, then one person did the scanning and  
 85 named according to the naming convention above. The scanning was done in a specific order to ensure consistency across all  
 86 scans. The data was saved in a consistent format to ensure they could be easily accessed and analyzed. A third person then  
 87 checked the data to ensure that the data was consistent in naming and accurate.

88 For the validation of the scans and their processing we investigate the correlation scores of pair-wise aligned signals. Large  
 89 scores between signals are indicative of being made by the same tool. As showed previously in Figure 4 in Align signals, we  
 90 would expect a high correlation score between signals from scans of wires cut with the same tool. For signals from scans of  
 91 wires cut with a different tool, we would expect a low correlation score. For example, we have two scans from different tools,  
 92 T1AW-LI-R1 and T2AW-LI-R1, as shown in Figure 5a and Figure 5b. The alignment is shown in Figure 5c with 0.2 CCF  
 93 value, which is low, as expected.

We also put resulting CCFs for all pairwise comparisons in the boxplot, together with the receiver operating characteristic (ROC) curve, as in Figure 6 and Figure 7. We can see that the CCF values for the same sources are close to 1, while the CCF values for different sources are much lower than expected. This is consistent with our expectations and validates our data processing pipeline.

## Usage Notes

The R package `x3ptools` (available from CRAN) supports working with files in x3p format.  
Sample scripts in R for processing scans from x3p format to their signal are available from ... github.

Further analysis can be conducted with the GitHub R package `wire` and the GitHub R shiny app `wireShiny` ([citation?](#)). We already conduct between-replicate comparisons in the technical validation section, and we can also conduct across-replicate comparisons to establish error rates threshold and produce other analysis plots.

Suppose we put the CCF values in a tilemap with different tool, location and edge combinations. In that case, we expect only the diagonal to have high CCF values, close to 1 and marked as orange in the tilemap, as the diagonal represents the same source, and the rest of the matrix to have low CCF values, close to 0 and marked as gray. In Figure 8, the behavior is consistent with our expectation overall, except for some rare cases with tool 5 edge D, which is caused by [?????](#). The density plot in Figure 9 shows the distribution of the CCF values with the same sources and different sources. The overlapping points between the tails of these two distributions can be a rough threshold.

Furthermore, the ROC curve in Figure 7 shows the sensitivity / true positive rate against the false positive rate (FPR) (1 - specificity). The curve is very close to the upper left corner, which is excellent for classification and drawing conclusion. It gives us a true threshold of 0.589 to control the FPR to be less than 0.05 with false negative rate (FNR) to be 0, ([false positive rate \(FPR\) / false discovery rate \(FDR\)](#) -> define the H0 or call it false identification rate (FIR)[???](#)), and 0.658 to control the FPR to be less than 0.01, with FNR to be 0.02.

## Code availability

table of codelist files with short explanation, similar to table 1  
no, we can't use the website as a place for more detailed procedures. This paper is the detailed procedure.no more README, scanning procedures in another HTML

We put together the cutting and the standard scanning procedures mentioned in Section [cross-ref not working](#) with more pictures for each step into a [README of the GitHub repository heike/Wirecuts](#) ([High-res pics needed in the README](#)).

The data set can be easily accessed with the CRAN R package `x3ptools`. Further analysis can be conducted with the GitHub R package `wire` and the GitHub R shiny app `wireShiny` ([citation](#)).

## References

- AFTE. The association of firearm and tool mark examiners: Theory of identification as it relates to toolmarks. *AFTE J.* **30**, 86–88 (1998).
- NRC. *National Research Council: Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009).
- 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 the President of the United States, President's Council, Washington, D.C., 2016).
- 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).
- Zheng, X. A., Soons, J. A. & Thompson, R. M. NIST Ballistics Toolmark Research Database | NIST (2016).
- 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).
- 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).
- 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).

- 139    9. Ju, W. & Hofmann, H. The R Journal: An Open-Source Implementation of the CMPS Algorithm for Assessing Similarity  
140    of Bullets. *The R J.* **14**, 267–285, [10.32614/RJ-2022-035](https://doi.org/10.32614/RJ-2022-035) (2022).
- 141    10. Cleveland, W. S., Grosse, E. & Shyu, W. M. Local Regression Models. In *Statistical Models in S* (Routledge, 1992).

142    **Acknowledgements**

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

147    **Author contributions statement**

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

153    **Competing interests**

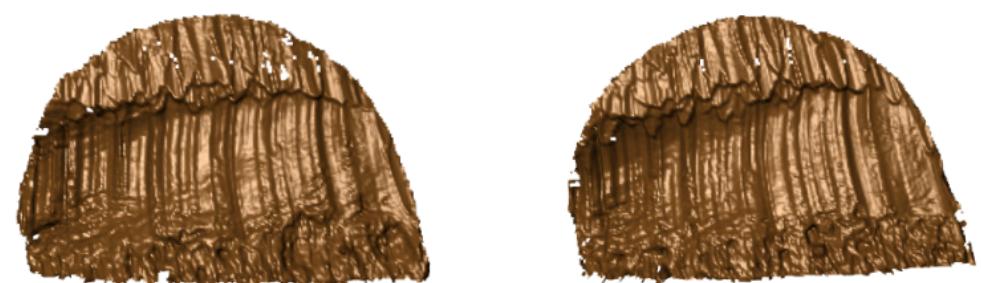
154    (mandatory statement)

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

Edge A



Edge C



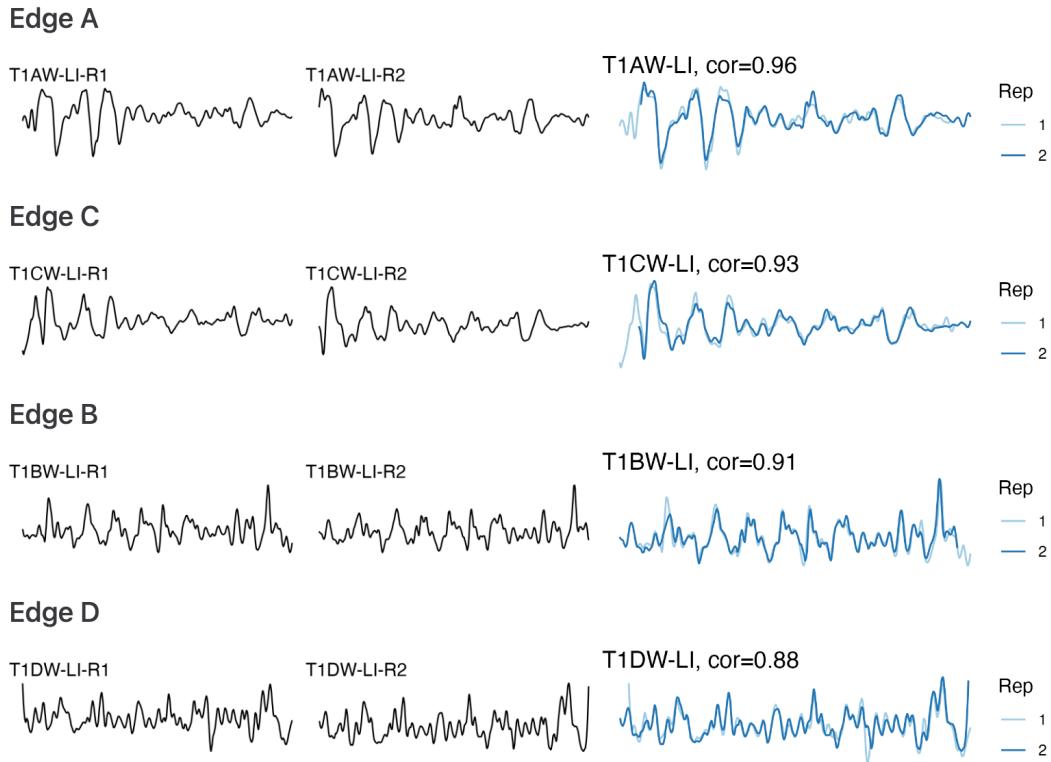
Edge B



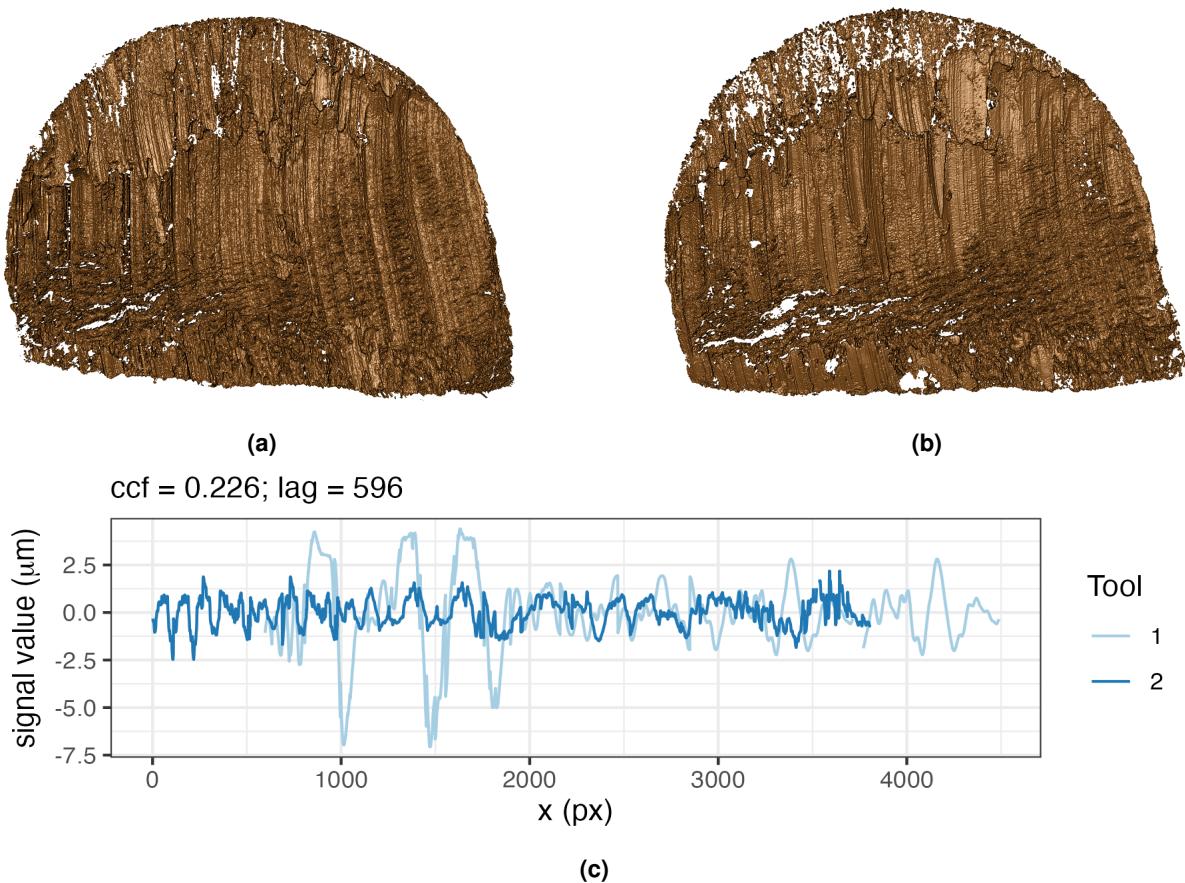
Edge D



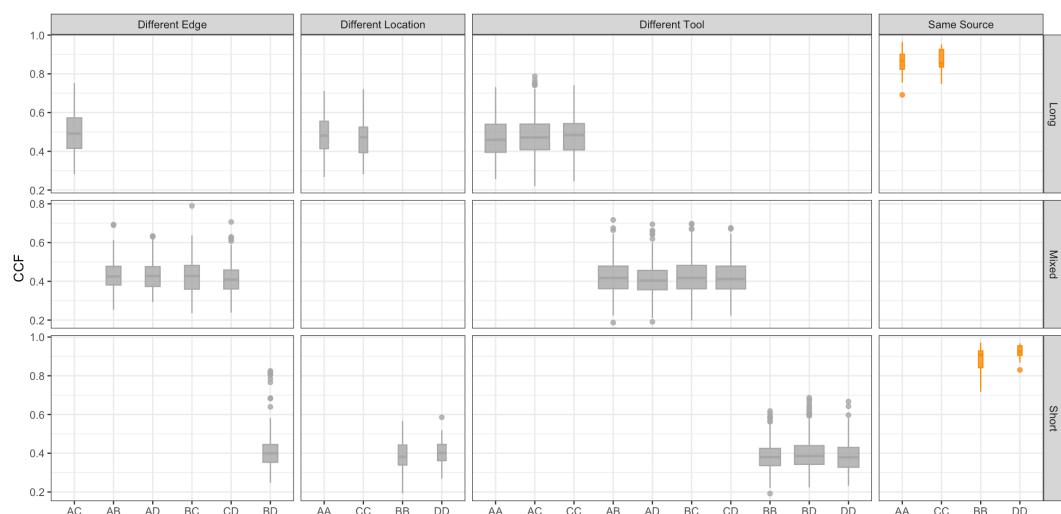
**Figure 3.** Scans from different sides of tool 1 at the inner location.



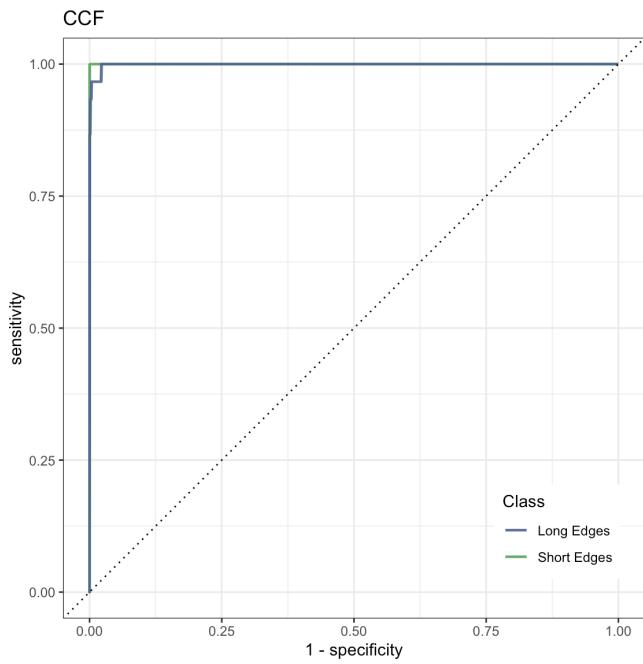
**Figure 4.** The first and second columns show the signals extracted from Figure 3, and the third column shows the alignments and CCF values between pairs of signals.



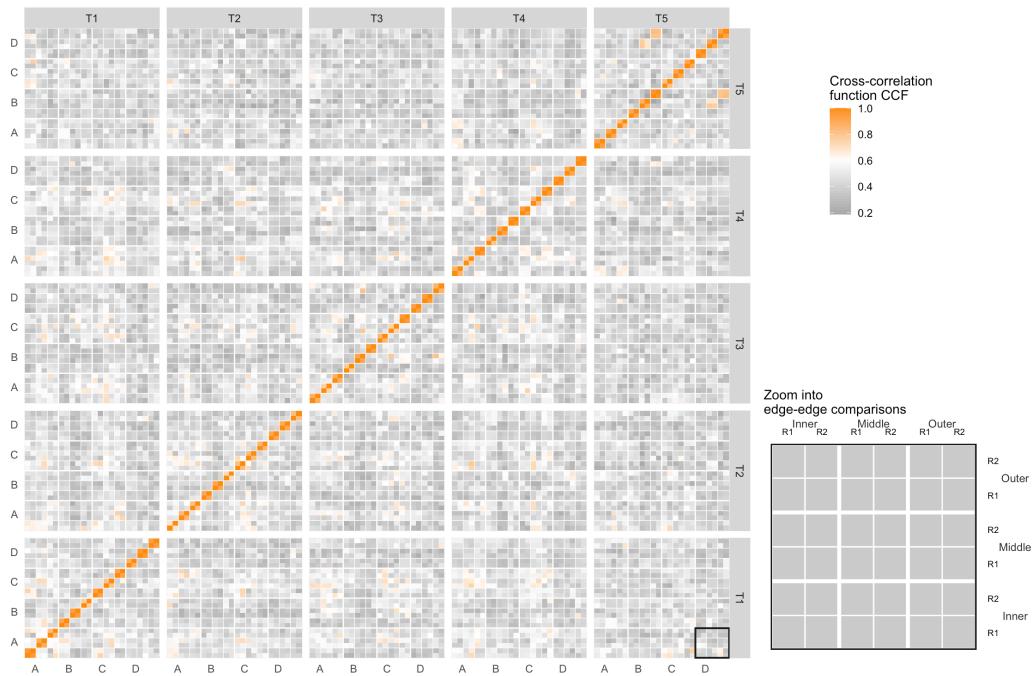
**Figure 5.** (a) Scan T1AW-LI-R1 cut by tool 1. (b) Scan T2AW-LI-R1 cut by tool 2. (c) Alignment of signals from T1AW-LI-R1 and T2AW-LI-R1.



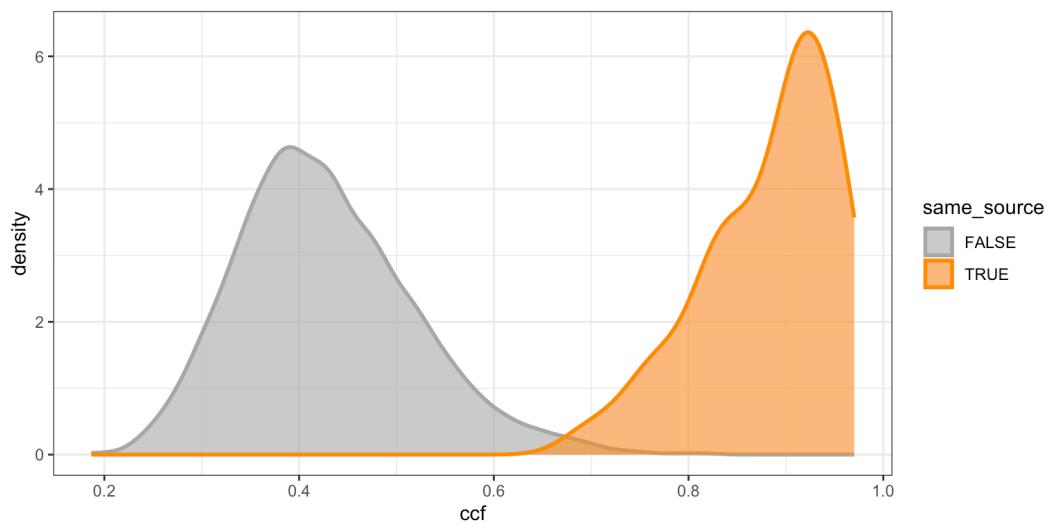
**Figure 6.** The boxplot shows that signals from the same sources have higher CCFs than those from different sources.



**Figure 7.** The ROC curve is bending very close to the upper left corner, which means excellent in classification and drawing conclusions.



**Figure 8.** The tilemap shows signals from the same source have CCFs close to 1.



**Figure 9.** The density plot shows tails of distributions overlap, which can be used as a rough threshold for drawing conclusions.