

# Interactive Visualization Framework for Forensic Bullet Comparisons

Nathan Rethwisch\*

Department of Statistics, Iowa State University  
and

Heike Hofmann

Department of Statistics, University of Nebraska Lincoln

April 28, 2025

## Abstract

The current method for forensic analysis of bullet comparison relies on manual examination by forensic examiners to determine if bullets were discharged from the same firearm. This process is highly subjective, prompting the development of algorithmic methods to provide objective statistical support for comparisons. However, a gap exists between the technical understanding of these algorithms and the typical background of many forensic examiners. We present a visualization tool designed to bridge this gap, allowing for the presentation of statistical information in a more familiar format to forensic professionals. The forensic bullet comparison visualizer (FBCV) features a variety of plots that will enable the user to examine every step of the algorithmic comparison process. We demonstrate the utility of the FBCV by applying it to data from the Houston Science Lab, where it helped identify an error in the comparison process caused by mislabeling. This tool can be used for future investigations, such as examining how distance between shots affects scores. The FBCV offers a user-friendly way to convey complex statistical information to forensic examiners, aiding their understanding and utilization of algorithmic comparison methods.

*Keywords:* data visualization, interactive forensic modeling, cross-correlation function, land engraved area, forensic pattern analysis, forensic statistics

---

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

# 1 Introduction and Background

Identifying the firearm used in a crime is a critical component of forensic examination and plays a pivotal role in criminal investigations. Ensuring evidence is appropriately identified is crucial in upholding the integrity of the criminal justice system.

Current forensic practices rely on examiners to visually inspect bullets under a comparison microscope for similarities of marks on the bullets' surfaces. As a bullet is discharged from the firearm, the rifling in the barrel forces the bullet to follow the groove pattern - like rails. Micro-imperfections in the barrel leave scratches (called *striations*) on the bullet's surface. Striations on land engraved areas (LEA; the area between two grooves) are assumed to be unique to the individual firearm. This allows forensic examiners to determine whether two bullets originate from the same source by seeing if these LEAs match ([AFTE 1998](#)). However, this process is highly subjective, relying heavily on an examiner's expertise ([NRC 2009](#), [President's Council of Advisors on Science and Technology 2016](#)). These criticisms triggered the development of algorithmic comparisons ([Carriquiry et al. 2019](#), [Chen et al. 2019](#), [Chu et al. 2013](#), [Ju & Hofmann 2022](#), [Vorburger et al. 2011, 2015](#)) to provide objective measures with the goal of augmenting an examiner's testimony.

Algorithmic comparison methods have demonstrated considerable potential to quantify the similarity between pairwise pieces of evidence. However, current approaches have created a gap between the statistical metrics and the practical understanding of these metrics by practitioners. This gap highlights the need for a more effective method to assist forensic practitioners in assessing and understanding the algorithm's performance.

Here, we are proposing an interactive interface designed to visualize the statistical metrics embedded in the context of the data ([Wickham et al. 2015](#)) in a manner that is intuitive

and accessible to forensic examiners. We introduce the forensic bullet comparison visualizer (FBCV). It leverages the power of interactive graphics by combining a set of interactive visualizations, allowing forensic examiners to engage with the complex algorithmic data at each stage of the process, thereby bridging the gap between statistical analysis and practical forensic application.

Interactive graphics are a powerful method of analysis techniques that allow the user to dynamically explore and manipulate graphical representations of data, usually through clicking, dragging, or rescaling ( [n.d.](#), [?](#), [?](#)). A typical strategy for visualizing complex data is to use high-dimensional plots, which can incorporate multiple variables and levels of aggregation. However, these plots often pose challenges in interpretability, especially among those unfamiliar with the data ([Wills 2008](#)). To improve readability, a cogent method is to use linked data views - a type of interactive graphics that allows data views to communicate with other views. This linkage system allows for the display of highly interpretable graphs while preserving all relevant information. Users are able to select key features (1) and hide details that are less pertinent at the time (0), allowing for intuitive data exploration ([Wills 2008](#)). The FBCV leverages linked views to present data at multiple levels of the bullet analysis process.

This paper presents a short review of the algorithmic comparison process. We then discuss the role of interactive visualizations to support diagnostics at each stage of the process. We also showcase the diagnostic capabilities of these visuals by presenting a real-world use case where we successfully applied the FBCV to identify an error in the data-cleaning process. Finally, we highlight other ways in which the FBCV can be utilized for further investigatory analysis.

The data used for illustrating the process is a dataset of scans provided by a collaboration

of CSAFE (Center for Statistics and Applications of Forensic Evidence) and the Houston Forensic Science Center (HFSC). The data consists of scans from 40 test fires of each of 13 Ruger LCP barrels. Ten barrels (labeled ‘A’ through ‘J’) were consecutively manufactured, while the remaining three (labeled 1-3) come from HFSC’s reference library of firearms. Here, we are analyzing 40 sequential shots from each of the barrel. The lettered barrels were only fired ten times each before this study. For ease of notation, we refer to these shots as 11 through 50. LCP barrels are traditionally rifled barrels with 6 lands and 6 groove areas. These barrels mark well, i.e., striation marks are almost visible to the naked eye, making them well-suited for a forensic analysis. Scans of the bullets were obtained by the High-Resolution Microscopy Lab at Iowa State University using a Sensofar confocal light microscope. For each bullet, 3d topographic images of each of the six land engraved areas (LEA) were acquired at 20x magnification (corresponding to 0.645 micron/pixel), resulting in a total of 3,120 LEA scans (13 barrels x 40 bullets x 6 lands).

These scans provide the basis for algorithmic comparisons. For the processing of scans and comparison of signals, we follow the steps outlined in [Hare et al. \(2017\)](#), implemented in the `bulletxtrctr` package in R ([Hofmann et al. 2022](#)). We apply the following steps to each of the 3,120 scans [steps 1-4] and each pair of scans [step 5]. The results from these comparisons were then rendered in our visualization framework [Section 2].

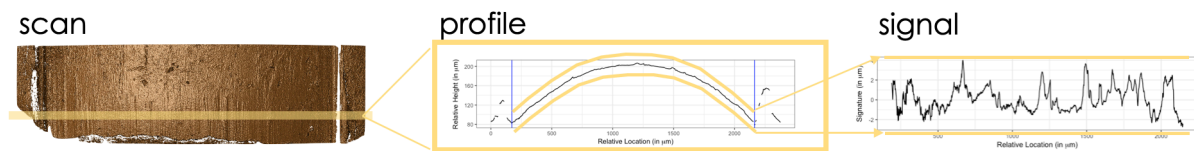


Figure 1: Signal extraction from a 3d topographic scan. From left to right we see a rendering of a 3d topographic scan of a land-engraved area, the profile corresponding to the horizontal yellow line, and the signal resulting from removing the bullet curvature from the profile.

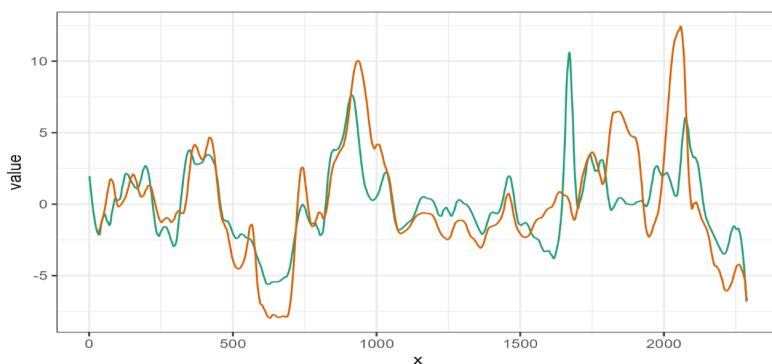


Figure 2: Aligned Signals from lands of two separate bullets.

1. A 3d LEA scan (left in Figure 1) is inspected for its suitability for comparisons, scans of low quality or of damaged lands (due to tank ‘rash’, pitting, or cracks) are removed from the analysis.
2. A crosscut is chosen orthogonal to the direction of well-marked striations (marked as a yellow line in the rendering of the scan left in Figure 1).
3. The topographical measurements corresponding to this crosscut are extracted from the scan. The middle plot in Figure 1 shows the profile of these height measurements along the crosscut. The spike in height at either end of the profile indicates the start of the neighboring groove areas and need to be excluded from the comparison (values

outside the vertical blue lines).

4. The signal for a LEA (shown on the right of Figure 1) is created by removing the bullet curvature from the profile using a non-parametric smooth (Cleveland 1979).
5. Finally, signals are aligned pairwise, as shown in Figure 2, and metrics assessing their similarity –such as the number of matching peaks, height of matching peaks, number of consecutively matching peaks, and, more statistically, cross-correlation– are extracted.

These metrics provide the basis of a quantifiable comparison of the strength of similarities with statistical models and algorithms. Common examples of such algorithms include random forests (Hofmann et al. 2022) and congruent matching profile segments (Chen et al. 2019, Ju & Hofmann 2022). A large number of these algorithms are based on the maximized cross-correlation function between pairs of signals. This is the metric which we will use in this paper. However, this is not an actual restriction, any other similarity metric would work similarly well, with its usability only restricted by the metric’s diagnostic ability.

Assume that  $X = \{X_t\}_{1 \leq t \leq N_X}$  and  $Y = \{X_s\}_{1 \leq s \leq N_Y}$  are the observed surface measurements (signals) of two land engraved areas (with  $N_X, N_Y$  the number of the respective observations). The correlation between  $X$  and  $Y$  is defined as the ratio of their covariance scaled by their respective variances:

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

$Y^{(k)}$  defines the  $k$ th lag of  $Y$  with  $Y_s^{(k)} = Y_{k+s}$  with  $k \in [-M, M]$  and  $0 \leq M < N_Y$ .

The choice of  $M$  depends on the minimal number of values  $N_Y - M$  used as a basis for an evaluation of the similarity of the two signals. With that we define the maximized

cross-correlation function  $CCF_{\max}(X, Y)$  as

$$CCF_{\max}(X, Y) = \operatorname{argmax}_{k \in [-M, M]} \operatorname{corr}(X, Y^{(k)}).$$

Here, we use  $M = 500$  for the alignment of signals. This corresponds to a horizontal shift of  $\pm 500$  values (equal to  $\pm 500 \times 0.645\mu = .323mm$ ) corresponding to about a quarter of a scan's width.

When assessing the similarity of one bullet to another, a common approach is to assemble all scores from comparing pairs of lands in form of a square matrix and visualize it in form of a tile plot, see Figure 3. The fill color encodes the score (here, the ccf) between a pair of LEAs. Higher values indicate higher similarity, shown in shades of orange. Tiles filled with grey values indicate less similarity. The two bullets shown in the example are known to have been fired through the same barrel. In this case, we expect six pairs of lands with high similarities (in-phase), while all other pairs (out-of phase) should result in low scores. This is exactly the pattern that can be seen in Figure 3.

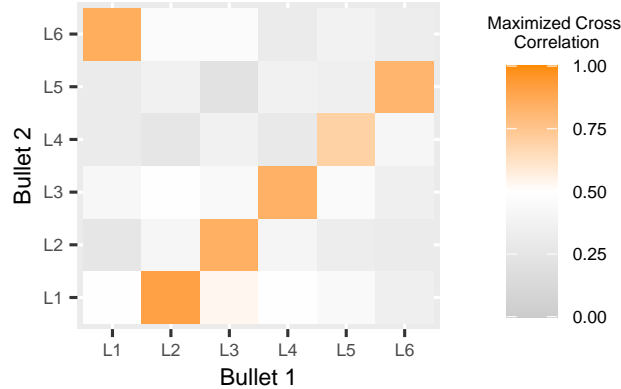


Figure 3: Tile plot of all pairwise comparisons of LEA signals from two different bullets

A summary statistic is computed using the average of in-phase scores, as these reflect bullet similarity while also being highly indicative of model performance. This is given as

$$\overline{\text{CCF}}_{\max}(B_1, B_2) = \underbrace{\left[ \frac{1}{n} \sum_{(i,j) \in \mathcal{P}} c_{ij} \right]}_{\text{in-phase average}},$$

where  $c_{ij}$  is the score between land  $i$  on bullet 1 and land  $j$  on bullet 2, with  $1 \leq i, j \leq n = 6$ , where  $\mathcal{P}$  denotes the pairs of lands that capture the best alignment between bullets  $B_1$  and  $B_2$ .

A pairwise comparison of  $K$  number of bullets results in a set of scores of size  ${}^K C_2 = \frac{1}{2}K(K-1)$  or  ${}^K C_2 + K$ , if we also consider to allow a comparison of a bullet to itself (done to achieve an empirical assessment of the range of scores we can expect to see for a particular type of ammunition and firearm). Different types of visualizations of these set of scores are discussed in the next section.

## 2 Visualization Framework

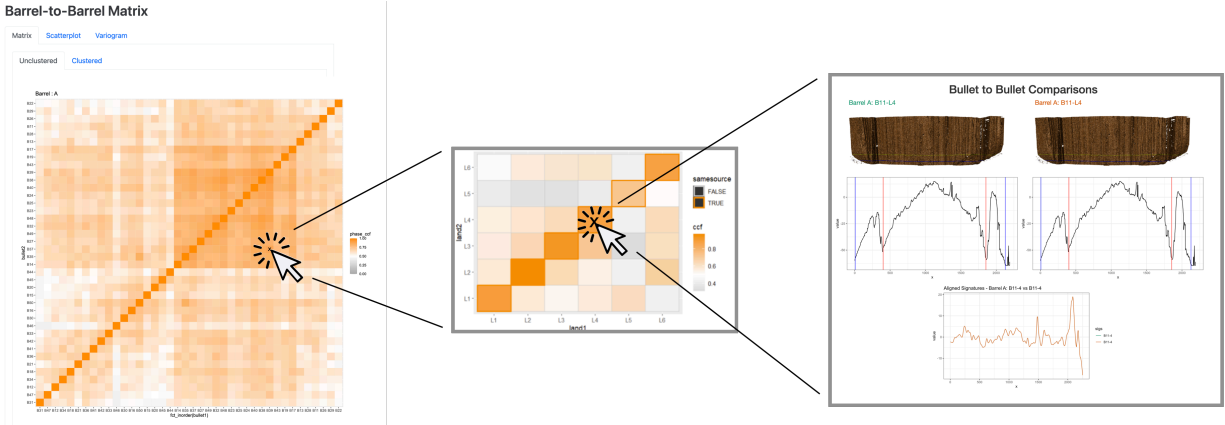


Figure 4: Three connected levels of information. From left to right, there is a tile plot of scores from all bullets in one barrel, a tile plot of scores at the land-level for one pair of bullets, and a set of diagnostic plots for comparing a single pair of lands.



As seen in the previous section, there are similarity scores at the bullet-to-bullet level, there are scores at the land-to-land level, and there are important diagnostics for individual pairs of lands. For any given comparison, we have pertinent information at each of these levels (see Figure 4). The statistical perspective focuses on scores within the distribution of other, comparable scores, while the focus in a forensic examination is on the individual. The idea of this visualization tool is to connect these different levels and perspectives for a seamless exploration.

The forensic bullet comparison visualizer (FBCV) is created in HTML using a combination of Javascript and R code. This allows us to leverage the everyday familiarity of links for implementing connections across levels of information. An implementation of the FBCV showing all comparisons involving bullets from barrel A can be found at <https://tinyurl.com/y53n3mkm>.

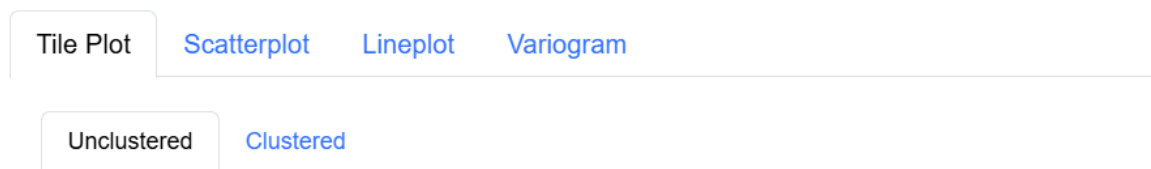


Figure 5: Tabs in the Interface framework

The main interface of the FBCV consists of a set of tabs with choices for the visualization of the set of bullet scores, see Figure 5. Each tab shows an interactive visualization that focuses on a different aspect of the data, which is discussed in more detail next.

**Tile plots** are our default choice for visualizing all bullet comparisons. The left side of Figure 6 shows the 40x40 matrix of all pairwise bullet comparisons in for barrel A and same-bullet scores on the diagonal. Each row and column corresponds to the bullet involved in

the comparison, each cell represents the maximum phase correlation score for the respective comparison.

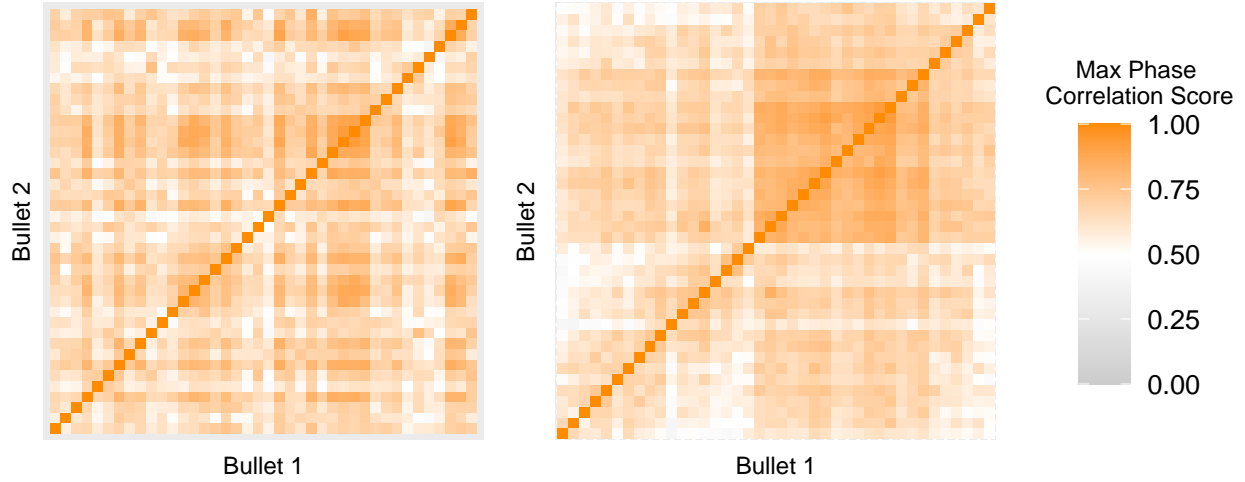


Figure 6: Tile plots for all pairwise comparisons within a bullet

Note that this matrix is not static: the spatial area of each tile maps interactively to the corresponding 6x6 land-to-land tile plot, such as the one in Figure 3. When a user clicks on a square in the 40x40 matrix, the FBCV retrieves the corresponding 6x6 matrix of land-to-land scores, giving users more details on the land-to-land comparisons that contribute to the score of the selected square.

The interactivity extends beyond the 6x6 matrix. By clicking on an individual square within this matrix, additional information is provided for the two LEAs and their comparison: renderings of the two LEA scans with marked crosscut locations, plots of their profiles, and the aligned signals. These web links directly map the different levels of comparisons as shown in Figure 4 to individual comparisons and allow the user to move naturally between abstraction levels.

Ordering rows and columns in tile plots has a large impact on the visualization. By clicking the *Clustered* subtab, the user can view an altered version of the original 40x40 tile plot.

The ordering of the rows and columns is based on a complete-linkage hierarchical clustering of the score matrix. The right side of Figure 6 displays this ordering. Here, we see two fairly distinct clusters. This version of the tile plot groups bullets by their similarity, which helps to identify any significant performance discrepancies in the data. Also note that the interactivity of the clustered tile plot is the same as for the original plot.

**Scatterplots** provide an alternative representation of the scores: Figure 7 shows an example of the default scatterplot. The first bullet in the comparison is represented on the x-axis, while the associated maximum phase correlation score is displayed on the y-axis. Additionally, we use color to represent the shot number of the second bullet. When the user hovers over a point, all other points containing the second bullet in the comparison are highlighted. This enables the user to identify bullets with poor scores across the dataset or those exhibiting similar patterns across all comparisons.

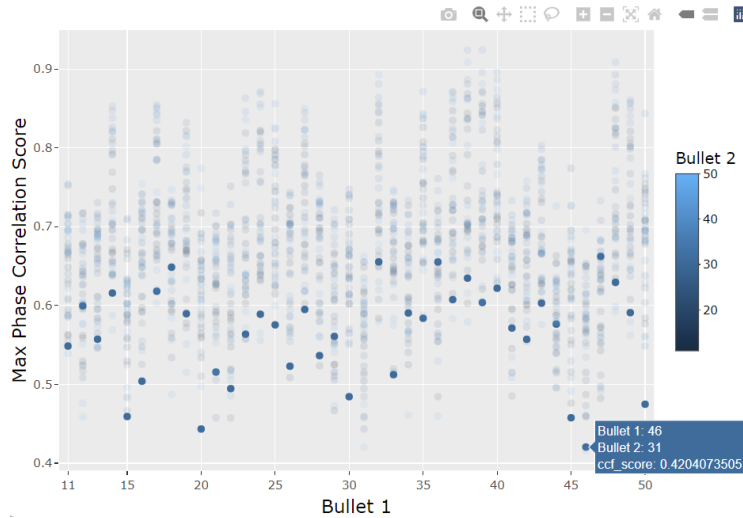


Figure 7: An interactive scatterplot for bullet comparisons

Clicking on a point again brings up the 6x6 matrix of land-to-land comparisons resulting in the selected point's score.

Clicking on the *lineplot* tab brings up the same scatterplot, with the key distinction that points representing the same second bullet in the comparison are connected by a line. This visualization helps to emphasize the relationships and trends between the points that share the same bullet.

**Variograms** are used to represent values as a function of the distance. In this context, the variogram illustrates how the similarity of bullets is affected by the number of bullets fired between them, shown in Figure 8. The x-axis represents the numerical distance between shots (11 vs. 12 corresponds to a distance of 1, while 11 vs. 50 is a distance of 39). The y-axis represents the algorithmic score between the bullets. The blue line shows a loess fit to capture the main trend. Clicking on any point within the variogram leads to the same interactive pipeline as the scatterplots and other visualizations.

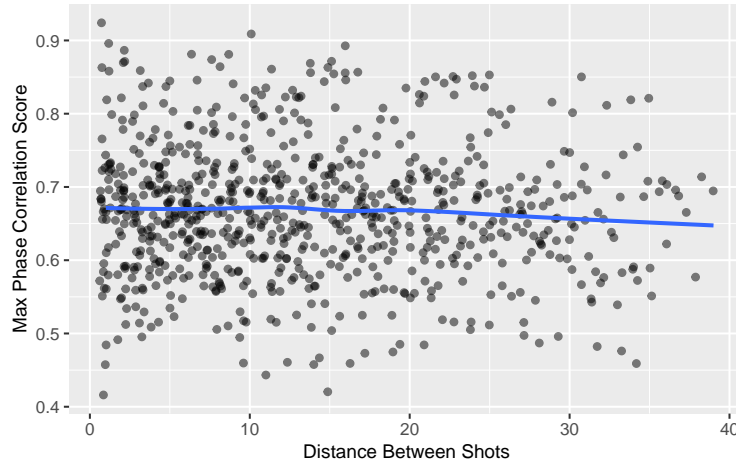


Figure 8: The variogram included in the visualization framework

These visualizations are integrated into a single HTML webpage, providing a comprehensive view of the data and offering accessible insights into scores.

### 3 Use Cases

#### 3.1 The Case of Barrel D

The provided visualizations have shown scores for barrel A, but not all firearms displayed such straight-forward results. One such case was the scores for barrel D. After running the pairwise comparisons and analyzing the visualization framework, it became apparent that there was an error in the analysis for bullets 35-40. These bullets performed well when compared to each other but did not score highly compared to the other bullets shot from this firearm, as shown in Figure 9. In the next section, we highlight different strategies to deal with this data discrepancy, comparing two control bullets (33-34) and the affected bullets (35-40). The bullets primarily used in this analysis are highlighted in Figure 9.

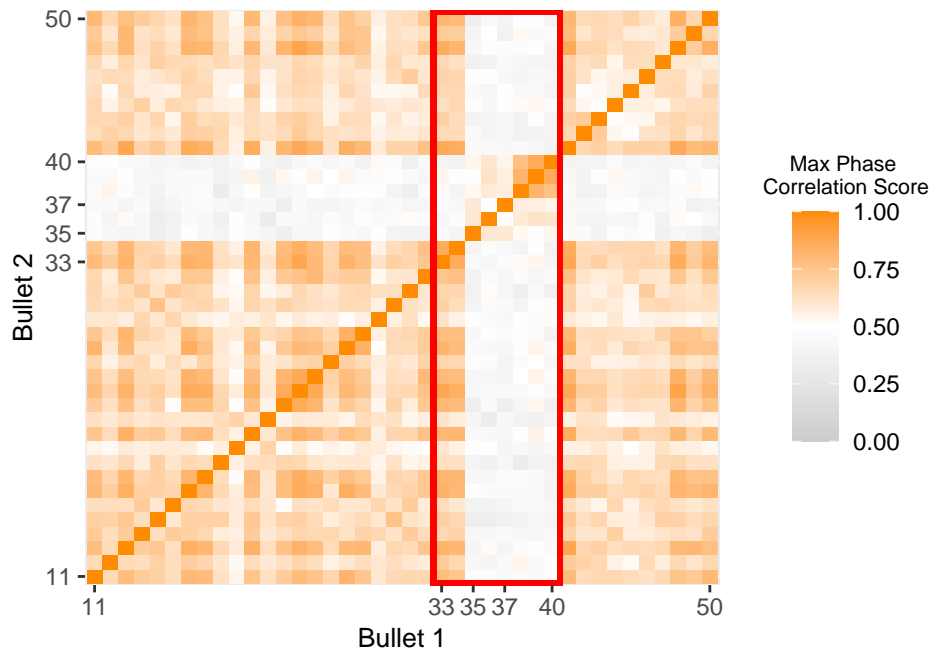


Figure 9: Tile plot of pairwise comparisons for barrel D

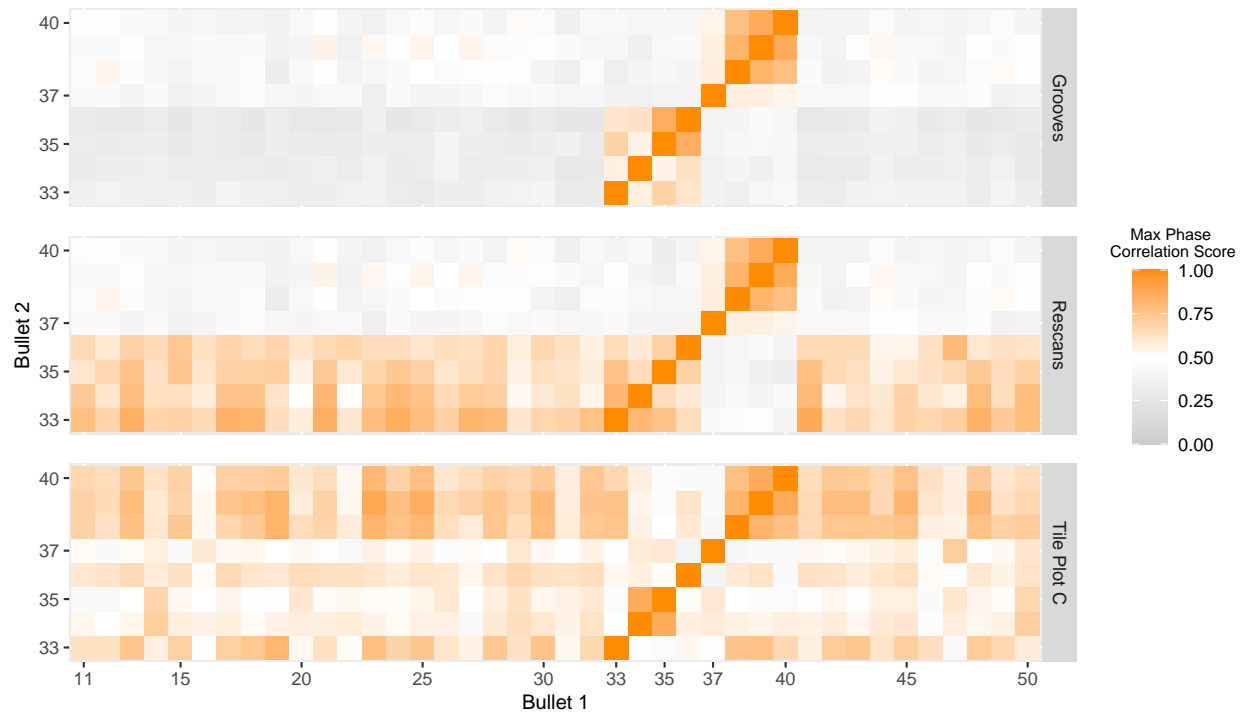
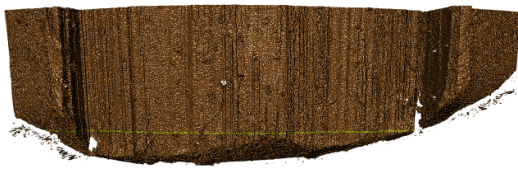


Figure 10: Overview of the adjustments made throughout the data cleaning process and their impact on the tile plots

### Comparing Groove-Engraved Areas

One potential reason for this suboptimal performance may stem from the groove-engraved areas (GEAs) being scanned rather than the land-engraved areas. Note that these areas are usually not used for examinations because grooves preserve marks from the tool they are made. When firearms are manufactured, a broaching tool is used to create the grooves for the rifling. This incorporates marks specific to the tool on the surface of the barrel. Because the same broaching tool is used for multiple barrels, marks on grooves are not specific to the firearm, limiting the ability to conclusively link striations on groove-engraved areas of a bullet to a particular firearm.



(a) Original land from bullet 35



(b) Scan of a groove from bullet 35

Figure 11: Comparing the land-engraved area and the groove-engraved area of bullet 35 from barrel D.

When analyzing the scans of groove-engraved areas (GEAs), it became apparent that the usage of grooves was not the root of the problem. Figure 11a below shows the original scan from Bullet 35, while Figure 11b shows a rescanned groove for that same firearm. Notably, the scan from GEAs has fewer topographical protrusions than the original image, indicating a smoother surface profile.

We then conducted an analysis where bullets 33-36 in the original data were replaced with comparisons using the grooves. For this test, bullets 33 and 34 act as a control - they represent scans that were already producing expected results in the dataset. Bullets 35 and 36 represent two bullets performing poorly in barrel D. The matrix of affected bullets was recreated using our interactive framework, shown in the top tile plot of Figure 10.

This visualization reinforces the conclusion that the discrepancy in the max phase correlation score was not due to grooves being scanned instead of lands. Not only are bullets 35 and 36 performing worse than before, but the scores on our control bullets - 33 and 34, also dropped significantly. Thus, we can conclude that the inaccuracy of the original data is not because groove areas were analyzed instead of lands.

## **Rescanning Land-Engraved Areas**

Our following action was to rescan and process the 3D topographical imaging on the bullets in question. The same control and test groups were used as the groove comparisons. Bullets 33 - 36 were rescanned, with 33 and 34 as the control. Replacing those comparisons in the interactive framework showed a drastic difference in results. In the matrix, the max phase correlation scores of the rescanned bullets aligned closely with those the other bullets fired from firearm D, shown in the middle plot of Figure 10. Furthermore, no significant change was found in the scores of the control group. When utilizing other parts of the FBCV, such as the 6x6 matrix and looking at the raw scans, these comparative results were reinforced. The alignment of signals for the rescans showed significant improvement compared to the previous alignment among barrel D.

Thus, the observed discrepancy in performance stems not from the algorithmic process itself but from inconsistencies in the raw scans utilized in the initial data processing. This could be due to various factors, including mislabeling or inadequate scanning. However, if this is a case of mislabeling scans, it raises questions regarding the provenance of the original scans.

## **Closing the Loop**

To answer whether the data was mislabeled, we compared the original scans of bullet D to data obtained from all of the 12 other firearms in the Houston dataset. We selected one bullet from each firearm that performed exceptionally compared to the other bullets from that firearm (12 bullets in total). Those bullets were then compared to both each other and bullet 39 from firearm D, one of the originally poor-performing bullets.

We found that 11 of the 12 bullets showed poor performance when compared to the selected bullet from firearm D. The exception was the bullet selected from firearm C. When



comparing that bullet to bullet 39 of barrel D, the algorithmic results showed that the two bullets were likely fired from the same weapon. Figure 12 shows the 6x6 matrix from our framework tool when comparing the selected bullet from firearm C to bullet 39 of barrel D. The alignment between these two bullets is extremely strong, even stronger than many other bullet comparisons where both bullets originated from firearm C.

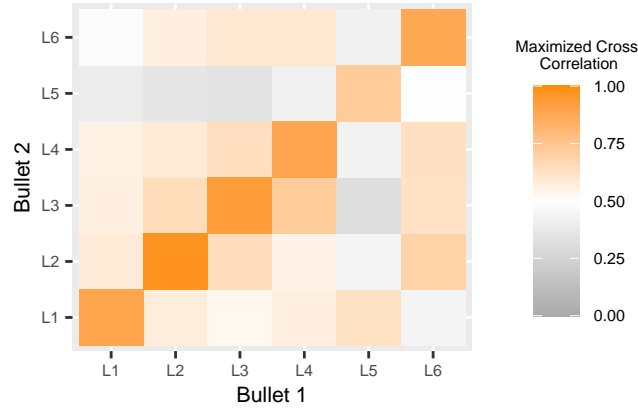


Figure 12: Replacing 6x6 tile plot from bullet C with bullet D

It is also important to note that the selected bullets did not perform well when compared with each other (i.e., results from the chosen bullet from firearm A did not match the bullet from firearm B). This implies the similarity between the selected bullet from firearm C and the poor-performing bullet from firearm D cannot be attributed to well-performing bullets being accurate among comparisons with other weapons. Thus, we have strong evidence that the bullets were mislabeled, with the poorly performing bullets originating from barrel C.

This notion is strengthened by our visualization tool. When the comparisons for bullets 35-40 are substituted with those initially labeled as bullets 35-40 fired by barrel D, there is no significant deviation in performance according to the matrix shown in the bottom plot of Figure 10. While there may appear to be a dip in performance around this area, the differentiation happens in bullets 34-37, which is not the complete scope of the substituted

bullets. Overall, these scans perform very well compared to other bullets shot by firearm C. Thus, we can conclude that bullets 35-40 were mislabeled and shot by firearm C, not D.

### **3.2 Bullet Distance Analysis**

A key focus of this tool was evaluating whether model performance changes as more bullets are fired from a weapon. Using variograms, we provided a visual representation of this performance. A variogram is a geostatistical tool used to visualize spatial dependence. In our context, spatial dependence refers to the relationship between model performance and distance (i.e., number of shots) between selected bullets. Figure 13 illustrates variograms for all firearms in the Houston dataset. Note that poor-performing bullets from barrel D were removed in this analysis, as well as two bullets from barrel J that displaying similar performance issues. Future analysis will investigate the underlying cause of these discrepancies in bullet J.

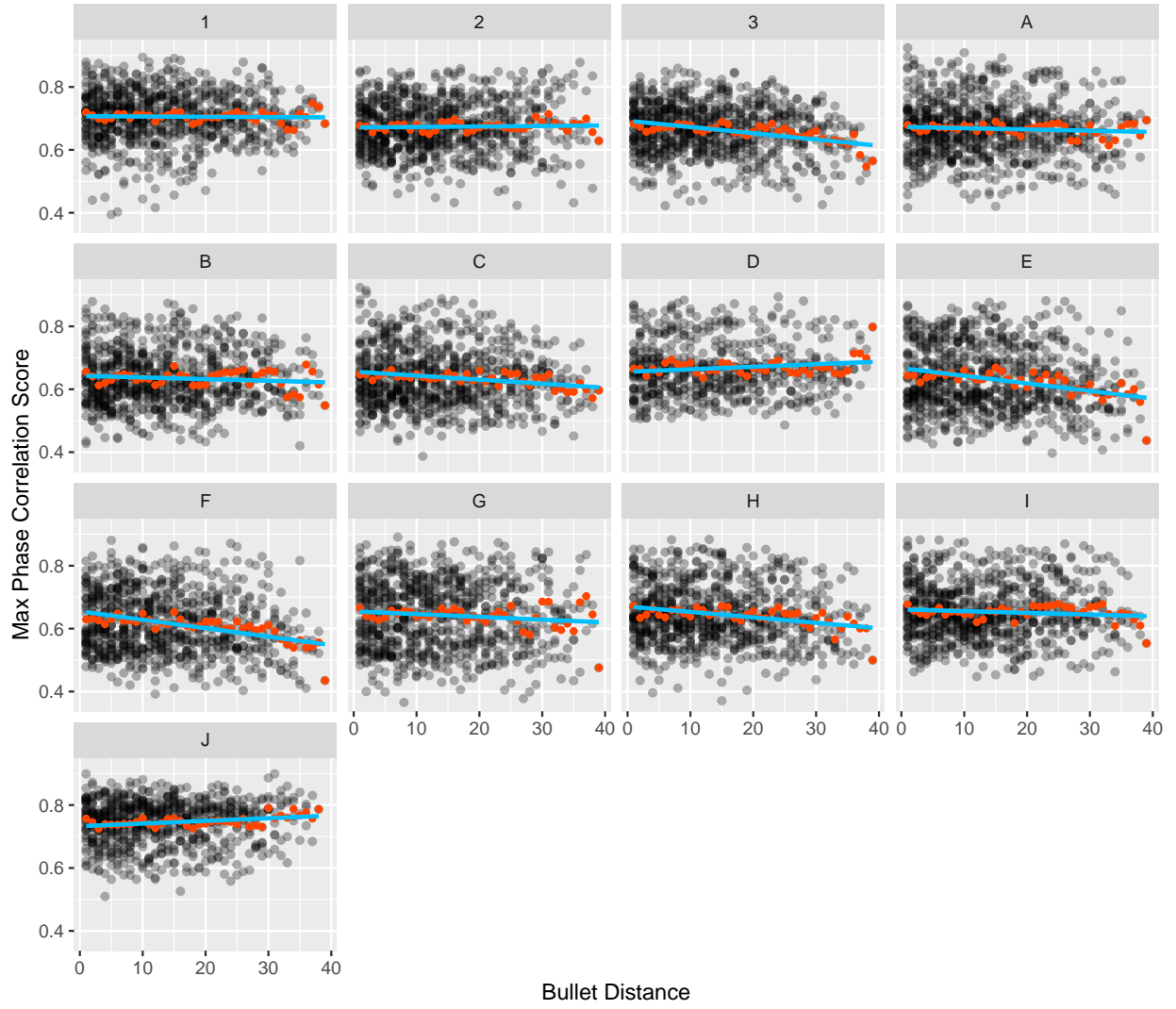


Figure 13: Variograms across all firearms

A linear model was fit to each of the individual bullets. The formula for the linear model can be defined as:

$$\widehat{CCF}_{\max}(B_1, B_2) = \theta_0 + \theta_1 \times \frac{1}{N(d)} \sum_{i=1}^{N(d)} CCF_{\max}^i(B_1, B_2)$$

where  $d = \|B_2 - B_1\|$ ,  $N(d)$  is the number of observations at a given distance, and  $i$  is the  $CCF_{\max}$  between  $B_2$  and  $B_1$  at a given  $d$ .  $\theta_0$  gives a baseline score for each individual barrel, while  $\theta_1$  represent the coefficient associated with the change in mean  $CCF_{\max}$  score

of bullets at a given  $d$ .

It is apparent that there are differences in baseline model performance between different firearms.  $\theta_0$  ranges from 0.64291 for barrel B to 0.73287 for barrel J, with a mean of 0.67228.

We hypothesized that model performance would decline as the distance between shots fired increases, leading  $\theta_1$  to be negative. This is true in most firearms, but  $\theta_1$  is positive for firearms 2, D, and J. Overall, however,  $\theta_1$  is relatively small among all barrels, falling between -0.00268 and 0.00084, with a mean of -0.00083. Combined with the visual analysis of the variogram, we conclude that bullet distance has minimal effect on model accuracy relative to other causes of variability. This is evident when  $d < 30$ , whereas beyond  $d > 30$ , there is insufficient data to make definitive claims.

Furthermore, we acknowledge that there is significant variability among the in-phase scores across the firearms we examined. This variability suggests that additional factors beyond bullet distance have a substantial effect on model performance and accuracy.

While the distance between shots may not be a significant factor in predictive performance, we recognize there is significant amount of variability among in-phase scores between shots. Potential factors may include, but are not limited to, inconsistencies in bullet scan accuracy, slight differences in ammunition characteristics, and differences in the way in which a firearm was discharged, may also play a key role in observed variability. These factors prompt further investigation to determine their role in achieving consistent model predictions and reducing variability among bullets.

## 4 Discussion

### 4.1 Key Findings

The FBCV was proven effective in identifying an error in the highlighted dataset. Through a series of interactive visualizations, we identified the cause of the performance issues, which was attributed to labeling error. Through a series of rescans, we resolved the error, allowing us to move forward with the analysis. This demonstrates the practical use of the FBCV in error identification.

The FBCV also includes variograms, which allow for insightful analysis on how bullet distance affects model accuracy. Using these visualization tools, we examined the impact on 13 different firearms. By fitting a line to the mean in-phase score at each distance among variograms, we observed that model accuracy does not significantly degrade with increasing bullet distance. However, the variability of bullet scores is high, indicating other factors have a substantial effect on prediction performance outside of how many shots were fired between bullets. The plots included in the FBCV offer a promising tool for future analysis on additional datasets, which could contribute to expanding the literature in this area.

### 4.2 Limitations

This study faces a number of limitations, chiefly regarding file storage. Because of the large number of comparisons, a considerable amount of files were rendered for each stage in the FBCV’s process. Each firearm contains 1600 bullet-to-bullet comparisons. Then, for each bullet-to-bullet comparison, there are 36 sub-comparisons made at the land level, corresponding to the six lands of each bullet. Therefore, we must process  $1600 \times 36 =$

57,600 png images at the land level.

At the lowest level of the FBCV, an image is processed for each of the lands being compared, the cutoffs of the LEA for both lands, and the aligned signals. Therefore, 288,000 background PNG files are processed before they are aggregated into HTML format. To mitigate the storage burden, we employed a strategy to avoid creating new background files for each HTML rendering. For the 40x40 matrix, the 6x6 land-to-land matrix, and the bullet comparison informational HTML, we used a single background file for each type of HTML. This background file was reused across different comparisons, enabling more efficient rendering of the FBCV without duplicating files for each individual bullet comparison.

Despite these adjustments, the substantial number of files presents a significant storage challenge, and Github was unable to accommodate more than the number of files associated with one firearm. Consequently, the figures presented in this analysis primarily focus on firearm A. Furthermore, the need to render such a large number of files complicates the process of making rapid updates to the scans, as any modification typically requires re-rendering a large number of figures. Still, as demonstrated in previous examples, the framework does allow for updates as changes are introduced.

### **4.3 Future Work**

Future research in this area should focus on exploring the underlying causes of variability between shots and identifying the factors contributing to this variability. This is crucial to improving model performance and establishing confidence within the forensic community regarding the adoption of such algorithmic models. Additionally, exploring the impact of firearm storage on bullet performance presents a valuable direction for future research.

## 5 Conclusion

This paper presented an interactive framework for analyzing algorithmic comparisons of whether two bullets were fired from the same firearm. The framework includes various visualizations that allow the user to assess algorithmic performance at a broader scope while also diagnosing issues at every level of the comparative analysis process. The FBCV was used to analyze algorithmic performance on the Houston dataset. It successfully identified a problematic error in the comparison process, and investigative steps were taken to discern the cause of the error, which is attributed to mislabeling. In the future, this visualization framework can provide summary overviews of algorithmic performance and diagnose problems in the data processing of forensic bullet analysis. By offering an interface that is intuitive and accessible, it presents an option that can support forensic examiners and lead to more accurate forensic analysis.

## References

(n.d.).

AFTE (1998), ‘The association of firearm and tool mark examiners: Theory of identification as it relates to toolmarks’, *AFTE Journal* **30**(1), 86–88.

Carriquiry, A., Hofmann, H., Tai, X. H. & VanderPlas, S. (2019), ‘Machine learning in forensic applications’, *Significance* **16**(2), 29–35.

Chen, Z., Chu, W., Soons, J. A., Thompson, R. M., Song, J. & Zhao, X. (2019), ‘Fired bullet signature correlation using the Congruent Matching Profile Segments (CMPS) method’, *Forensic Science International* **305**, 109964.

Chu, W., Thompson, R. M., Song, J. & Vorburger, T. V. (2013), ‘Automatic identifica-

- tion of bullet signatures based on consecutive matching striae (CMS) criteria’, *Forensic Science International* **231**(1-3), 137–141.
- Cleveland, W. S. (1979), ‘Robust Locally Weighted Regression and Smoothing Scatterplots’, *Journal of the American Statistical Association* **74**(368), 829–836.
- Hare, E., Hofmann, H., Carriquiry, A. et al. (2017), ‘Automatic matching of bullet land impressions’, *The Annals of Applied Statistics* **11**(4), 2332–2356.
- Hofmann, H., Vanderplas, S., Ju, W. & Krishnan, G. (2022), *Bulletxtrctr: Automatic Matching of Bullet Striae*.  
**URL:** <https://github.com/heike/bulletxtrctr>
- Ju, W. & Hofmann, H. (2022), ‘An Open-Source Implementation of the CMPS Algorithm for Assessing Similarity of Bullets’, *The R Journal* **14**(2), 267–285.
- NRC (2009), *National Research Council: Strengthening Forensic Science in the United States: A Path Forward*, National Academies Press.
- President’s Council of Advisors on Science and Technology (2016), *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.
- 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.
- Vorburger, T. V., Song, J. & Petraco, N. (2015), ‘Topography measurements and applications in ballistics and tool mark identifications’, *Surface Topography: Metrology and Properties* **4**(1), 013002.



Wickham, H., Cook, D. & Hofmann, H. (2015), ‘Visualizing statistical models: Removing the blindfold’, *Statistical Analysis and Data Mining: The ASA Data Science Journal* **8**(4), 203–225.

Wills, G. (2008), ‘Linked data views’.