

# Adapting the Chumbley Score to Bullet Striations

Ganesh Krishnan, Heike Hofmann

April 21, 2018

# Objective and Motivation

- ▶ Same Source Matching of Bullet lands
- ▶ Evaluate performance of Chumbley Score method when used for Bullet Striations
- ▶ Bullet striations have curvature, not present in toolmarks
- ▶ Identify Error rates and effect of different parameters on them  
(In short finding the best error rates possible )

## Structure

- ▶ Error rates in toolmarks
- ▶ Data being used
- ▶ What is the Chumbley Score Method?
- ▶ Identifying Best parameter Settings for Bullets
- ▶ Modifications to the Algorithm
- ▶ Results

# Variations of Chumbley score method and Error Rates for toolmarks

Research paper	Method	Data Source	False Positives	False Negatives
<b>Faden et al. [2007]</b>	Maximized Correlation	Screwdrivers	-	-
<b>Chumbley et al. [2010]</b> (Same-Surface Same-Angle)	Randomized Chumbley Score	Screwdrivers	2.3%	8.9%
<b>Grieve et al. [2014]</b>	Randomized Chumbley Score	Slip-joint	-	-
<b>Hadler and Morris [2017]</b> (Same-Surface Same-Angle)	Deterministic Chumbley Score	Screwdrivers	0%	6%

Table 1: Error Rates for Toolmarks using variations of the chumbley score method

# Digitized Striation Marks

- ▶ Data
  - ▶ Ruger P85s Bullet Lands, or Hamby scans (Hamby et al. [2009]) provided by NIST (85,491 comparisons)
  - ▶ Bullet striation marks  $\approx 2\text{mm}$
  - ▶ Screwdriver marks  $\approx 7\text{mm}$  (all chumbley score papers)
- ▶ Let  $x(t_1)$ ,  $t_1 = 1, 2, \dots T_1$  and  $y(t_2)$ ,  $t_2 = 1, 2, \dots T_2$  be two digitized marks (where  $T_1$  and  $T_2$  are not necessarily equal).
- ▶  $T_1$  and  $T_2$  are the final pixel indexes of each marking. Therefore give the respective lengths of the markings.
- ▶ Signatures/ Profiles (NIST- Hamby)  $\approx 1200$  pixels (2 mm)  
Screwdriver toolmarks (Chumbley Papers)  $\approx 9000$  pixels (7 mm)

# Chumbley Score

## Step 0 : Defining a coarseness parameter

- ▶ Used to remove drift and (sub)class characteristics from individual markings
- ▶ Lowess or Loess fit residuals = Signatures
- ▶ Removes topographic structure (curvature)
- ▶ Improve the signal to noise Ratio

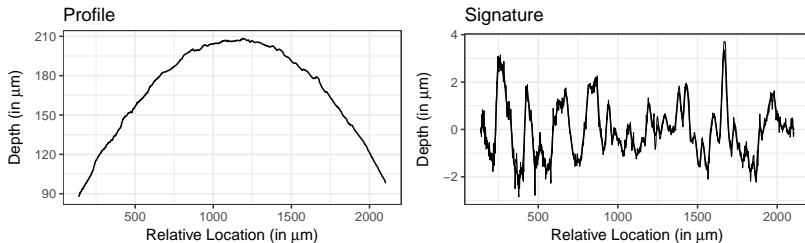


Figure 1: Bullet land profile (left) and the corresponding signature (right).

# Algorithm

- ▶ Two steps: Optimization (1<sup>st</sup>) and Validation (2<sup>nd</sup>).
- ▶ Windows  $\implies$  short segments of the markings
  - ▶ Have *predefined sizes*. ( $T_1$  or  $T_2 \gg w_o$  &  $w_v$ )
    1.  $w_o$  used in the Optimization step
    2.  $w_v$  used in the Validation step

## Optimization step

- ▶ **Goal** :Align markings horizontally as best as possible
- ▶ Correlation Matrix of all possible windows of size  $w_o$  between  $x(t_1)$  and  $y(t_2)$  computed
- ▶ *Identify lag for horizontal alignment*

Window Pair with maximized correlation  $\implies$

Optimal vertical (in-phase) shift of  $t_1^o - t_2^o$

- ▶ For aligning the two markings.

$$(t_1^o, t_2^o) = \arg \max_{1 \leq t_1 \leq T_1, 1 \leq t_2 \leq T_2} \text{cor}(x^{w_o}(t_1), y^{w_o}(t_2))$$

where  $t_1^o, t_2^o$  are the respective starting points of  $w_o$  in  $x(t_1)$  and  $y(t_2)$

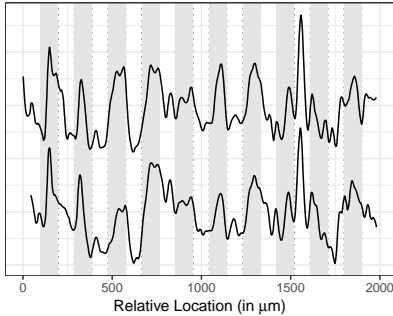
- ▶ Let  $t_1^*$  and  $t_2^*$  be relative optimal locations, where  $t_i^* = t_i^o / (T_i - w_o)$  for  $i = 1, 2$ , such that  $t_1^*, t_2^* \in [0, 1]$ .
- ▶ Once (sub-)class characteristics are removed, these locations have uniform distribution in  $[0, 1]$

## Validation Step

- ▶ Two sets of windows of size  $w_v$  chosen from both markings (see Figure 2)
- ▶ First set or **Same Shift**
  - ▶ pairs of windows are extracted from the two markings using the optimal vertical shift.  $t_1^o - t_2^o$
- ▶ Second set or **Different Shift**
  - ▶ the windows are extracted using a different (out-of-phase) shift.

# In-phase and Out-of-phase

In-phase sample



Out-of-phase sample

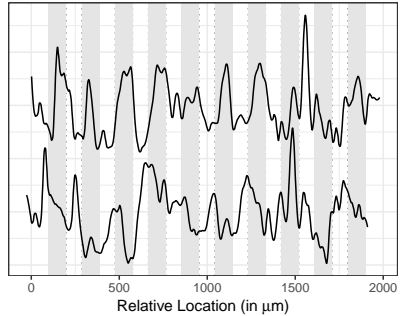


Figure 2: Two markings made by the same source. For convenience, the markings are moved into phase on the left and out-of phase on the right. In-phase (left) and out-of-phase (right) samples are shown by the light grey background. The Chumbley-score is based on a Mann-Whitney U test of the correlations derived from these two sets of samples.



- ▶ Both same- and different-shift pairs correlations between the markings are calculated.
- ▶ For Same-Source markings, correlations
  - ▶ for the in-phase shift should be high
  - ▶ for out-of-phase shift should be low.
    - ▶ Provide a measure for the base-level correlation to which in-phase shift correlations can be compared.
- ▶ The Chumbley score is the Mann Whitney U statistic computed by comparing between in-phase sample and out-of-phase sample.

# Block Diagram

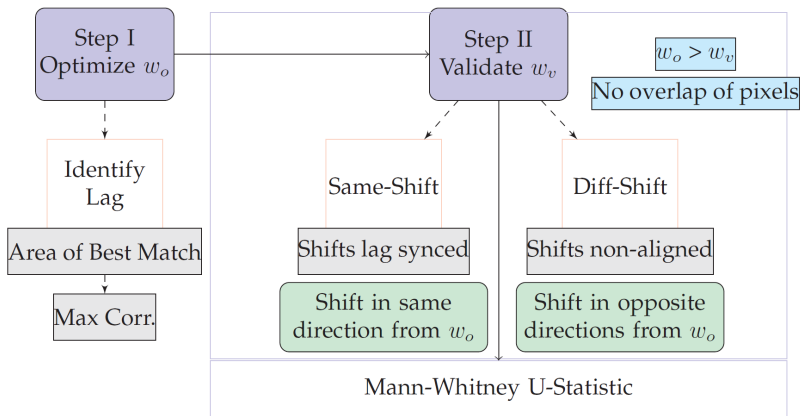


Figure 3: An overview of the adjusted chumley score method as given by Hadler and Morris [2017]

## Starting Points

More precisely, let us define starting points of the windows of validation  $s_i^{(k)}$  for each marking  $k = 1, 2$  as

$$s_i^{(k)} = \begin{cases} t_k^o + iw_v & \text{for } i < 0 \\ t_k^o + w_o + iw_v & \text{for } i \geq 0, \end{cases} \quad (1)$$

for integer values of  $i$  with  $0 < s_i^{(k)} \leq T_k - w_v$  where  $s \in \mathbb{Z}$

# The Hadler and Morris [2017] method (CS1)

- Same-shift pairs of length  $w_v$  are all pairs that start in:

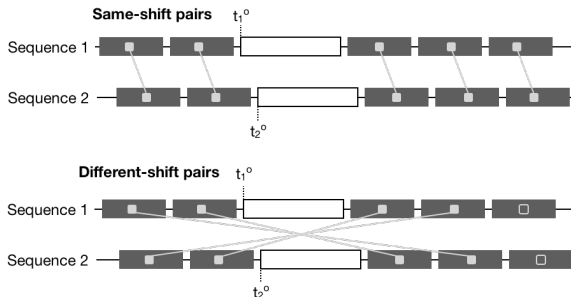
$$(s_i^{(1)}, s_i^{(2)}) \quad \forall i \in \mathbb{Z}$$

for which both  $s_i^{(1)}$  and  $s_i^{(2)}$  are defined.

- Different-shift pairs are defined as

$$(s_i^{(1)}, s_{-i-1}^{(2)}) \quad \forall i \in \mathbb{Z}$$

where both  $s_i^{(1)}$  and  $s_{-i-1}^{(2)}$  are defined (see fig. 4).



# Failed Tests

- ▶ By definition (equation 1), some number of tests fail to produce a result
- ▶ Either because the number of eligible same-shift pairs is 0, or the number of different-shift pairs is 0.
- ▶  $t_1^o, t_2^o$  not necessarily independent
  - ▶ **same-source:** Assume high dependence,  $\text{corr}(t_1^o, t_2^o) \approx 1$ 
    - ▶ Example:  $w_o = 120$ , coarseness ( $c$ ) = 0.3,  $\text{corr}(t_1^o, t_2^o) = 0.85$
  - ▶ **diff-source:** Assume independence of  $t_1^o, t_2^o$ 
    - ▶ Example:  $w_o = 120$ , coarseness ( $c$ ) = 0.3,  $\text{corr}(t_1^o, t_2^o) = 0.12$

## Failure Rate

$$P\left(t_1^o < w_v \cap t_2^o > T_2 - w_o - w_v\right) + \\ P\left(t_1^o < T_1 - w_o - w_v \cap t_2^o < w_v\right).$$

# Same-shift failure

- ▶ Same-source  $\approx 0$
- ▶ Different-source  $\approx 2 P(t_i < w_o)^2 = \frac{2w_v^2}{(T_1 - w_o)(T_2 - w_o)}$

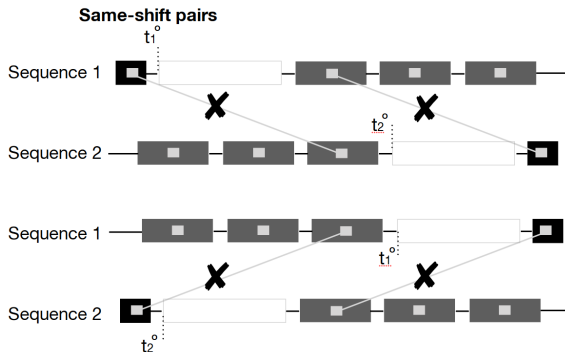


Figure 5: Sketch of same-shift pairings (top) when the lag is too large to accomodate a a vaildation window in either of the two signatures

# Different-Shift Failure

► Same-source (Assuming  $t_1^o \approx t_2^o \approx 2w_v/(T_i - w_o)$ )

$$P(t_1^o < w_v \cap t_2^o < w_v) + P(t_1^o < w_v \cap t_2^o < w_v) = 2P(t_1^o < w_v)$$

► Different-source  $\approx 2P(t_i < w_o)^2 = \frac{2w_v^2}{(T_1 - w_o)(T_2 - w_o)}$

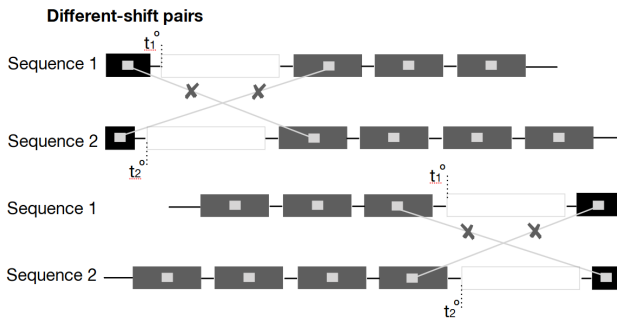


Figure 6: Sketch of diff-shift pairings (top) when the number of diff-shift computations is likely to 0

# Proposed Modification

- ▶ Failures due to missing Same-shift pairs unavoidable
- ▶ Failures due to missing different-shift pairs preventable

Define same-shift pairs identical to Hadler and Morris [2017] as pairs

$$(s_i^{(1)}, s_i^{(2)}) \quad \forall i \in \mathbb{Z}$$

where the boundary conditions of both sequences are met simultaneously.

- ▶ Let us assume that this results in  $I$  pairs.
- ▶ Let  $s_{(j)}^{(k)}$  to be the  $j$ th starting location in sequence  $k = 1, 2$ ,  
i.e.  $s_{(1)}^{(k)} < s_{(2)}^{(k)} < \dots < s_{(I)}^{(k)}$ .



We then define the pairs for different-shifts as

$$\left(s_{(j)}^{(1)}, s_{(l-j+1)}^{(1)}\right) \text{ for } j = \begin{cases} 1, \dots, l & \text{for even } l \\ 1, \dots, (l-1)/2, (l-1)/2 + 2, \dots, l & \text{for odd } l \end{cases} \quad (2)$$

- ▶ For an odd number of same-shift correlations
  - ▶ We skip the middle pair for the different-shift correlations (see fig. 7).
- ▶ This pairing ensures that the number of different-shift pairings is the same or at most one less than the number of same-shift pairings in all tests.

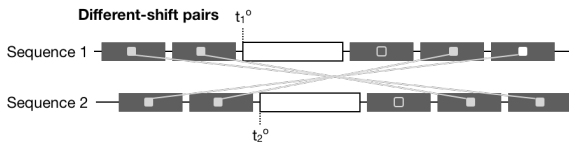


Figure 7: Sketch of adjusted different-shift pairings. At most one of the same-shift pairings can not be matched with a different-shift pair.

# Case where CS1 fails but CS2 does not fail

**CS1** Hadler and Morris [2017] algorithm

**CS2** the suggested modified algorithm

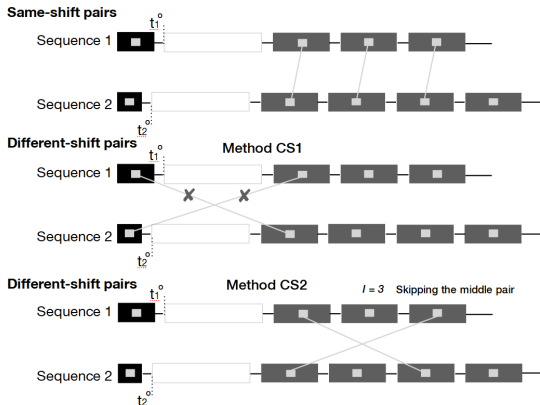


Figure 8: Sketch of a case where CS1 fails but CS2 does not fail

# Results Failed Tests

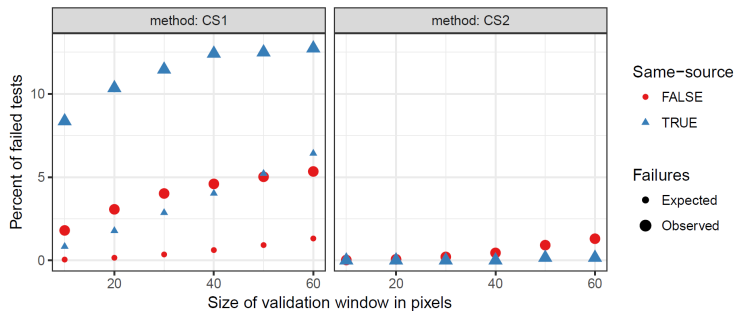


Figure 9: Percent of failed land-to-land comparisons for  $w_o = 120$  and coarseness  $c = 0.25$

# Conclusions Failed Tests

- ▶ With an increase in the  $w_v$  higher percent of tests fail under both CS1 and CS2
- ▶ Number is highly dependent on the comparison window sizes
- ▶ Correlated to the ground truth,
- ▶ **Higher for known same-sourced lands (CS1)** than for known different sourced lands.
- ▶ **CS1** fails to conduct a test about 8 to 13 % of the time for **known same-source lands**, and 2 to 6% of the time for **known different source lands**.
- ▶ Number for CS1 always higher than the corresponding theoretical number of failed tests.
- ▶ Using CS2, the case with **largest** # of failed tests is still **lower** than the case where CS1 gives the **lowest** # of failed tests
- ▶ Even for high coarseness, CS2 will have lower number of failed tests than CS1, Making it more robust.
- ▶ CS2 performs better for **both** same and different-sources
- ▶ Solves a **critical issue** of CS1 known same-source matching, by having a **negligible** number of Known same-source failed tests

# Coarseness

- ▶ Remove (sub-)class characteristics from profiles before comparisons for matching.
- ▶ Hadler and Morris [2017] suggest a coarseness parameter of 0.25 for toolmark comparisons.
- ▶ For bullet lands, coarseness might need to be adjusted because of the strong effect bullet curvature has on profiles.
- ▶ Optimal locations are distributed uniformly once (sub-)class characteristics are removed.
- ▶ Distinct boundary effects:  $c > 0.20$  optimal locations  $t^*$  are found at the very extreme ends of a profile more often than one would expect based on a uniform distribution. -smaller coarseness value of  $c = 0.15$  to be suitable

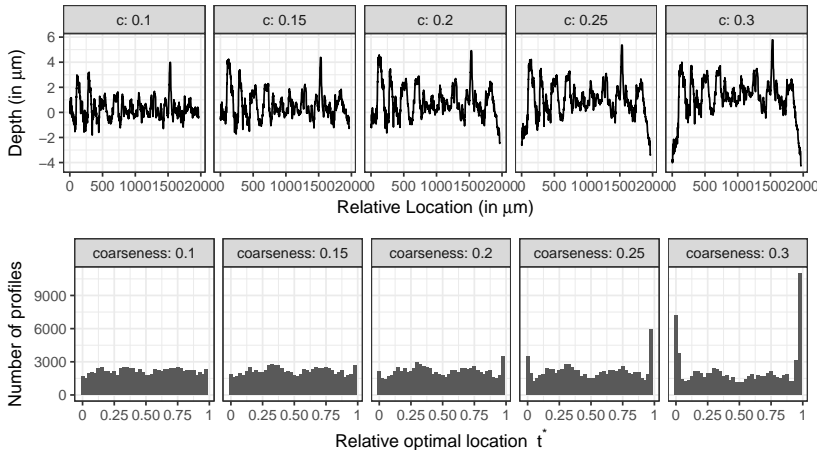


Figure 10: Overview of the effect of different coarseness parameters  $c$  on the profile shown in Figure 1 (top). The bottom row shows histograms of the (relative) optimal locations  $t^o$  identified in the optimization step for different values of the coarseness parameter  $c$ .

# Type II error rates

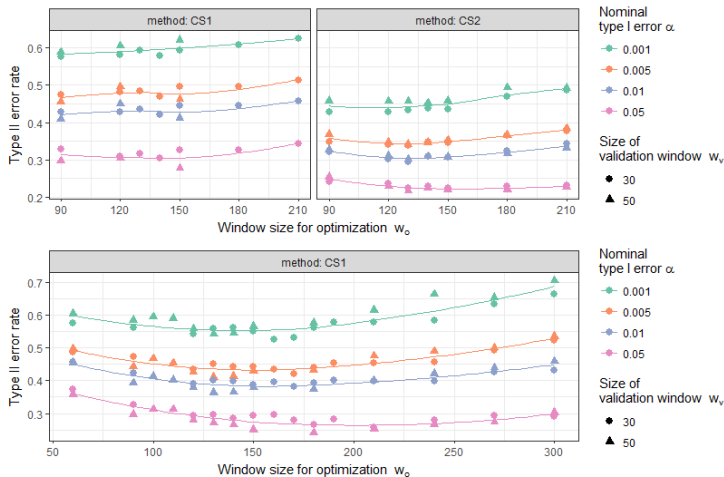


Figure 11: Type 2 error of methods CS1 and CS2 across a range of different optimization windows  $w_o$ . Top two figures are for a **coarseness**  $\approx 0.15$  and the bottom one is for  $0.3$



# Conclusions for Type II Error

## CS1

- ▶ Best works best for  $w_o$  of  $\approx 130$  to 160 and  $w_v$  50 when the smoothing is  $c \approx 0.3$ .
- ▶ The Type II rate is lowest for a nominal  $\alpha$  of 5%, with type I error rate of 6.2% and the Type II error rate of 24%.
- ▶ For lower nominal alpha levels of 1%, 0.5% and 0.1% the lowest type II error rate increases to about 36.4%, 41% and 52.5% respectively.
- ▶ Gets worse for coarseness 0.15

## CS2

- ▶ Significantly reduced over CS1
- ▶ For a window size of  $w_o = 130$  we see a minimum in type II error rate across all type I rates considered. - Smaller validation sizes  $w_v$  are typically associated with a smaller type II error.
- ▶ CS2 shows an increase in the power of the test.
- ▶ Type II CS2, still much higher for bullet lands than for toolmarks.
- ▶ Fix in CS2 will also improve power for matching toolmarks than CS1
- ▶ Bullet-to-bullet comparison using CS2  $\approx$  more power out of the test.

# ROC Curves

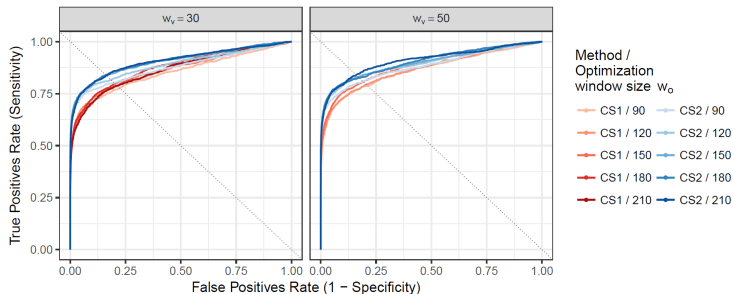


Figure 12: ROC curves of methods CS1 and CS2 for different sizes of optimization window  $w_o$ .

- ▶ Superior performance of CS2 over CS1
- ▶ Best performances wrt ROC curves are reached for  $w_o$  150 and higher.
- ▶ Points of equal error rates (EERs): intersection of the dotted line and the ROC curves.

**THANK YOU. Questions?**



# Appendix

## U Statistic:

This is computed from the joint rank of all correlations of both the same and different shift samples. As given by Hadler and Morris [2017]

Null Hypothesis: If the toolmarks were not match i.e not made by the same tool.

Let  $n_s$  and  $n_d$  be the number of same shift and different shift windows

$$N = n_s + n_d$$

Additionally, let  $R_s(i)$  and  $R_d(j)$  be the ranks associated with the combined vector of correlations for the same-shift and different-shift correlations, for  $i = 1, 2, \dots, n_s$  and  $j = 1, 2, \dots, n_d$ . Then the Mann–Whitney U-statistic is given by

The mann whitney U statistic is given by

$$U = \sum_{i=1}^{n_s} R_s(i)$$

with the standardized version which includes provision for rank ties

$$\overline{U} = \frac{U - M}{\sqrt{V}}$$

where prior to normalization the U-statistic has the mean as

$$M = n_s \left( \frac{N + 1}{2} \right)$$

and variance

$$V = \frac{n_s n_d}{N(N-1)} \left[ \sum^{n_s} R_s(i)^2 + \sum^{n_d} R_d(j)^2 \right] - \frac{n_s n_d (N+1)^2}{4(N-1)}$$

- L. Scott Chumbley, Max D. Morris, M. James Kreiser, Charles Fisher, Jeremy Craft, Lawrence J. Genalo, Stephen Davis, David Faden, and Julie Kidd. Validation of tool mark comparisons obtained using a quantitative, comparative, statistical algorithm. *Journal of Forensic Sciences*, 55(4):953–961, 2010. ISSN 1556-4029. doi: 10.1111/j.1556-4029.2010.01424.x. URL <http://dx.doi.org/10.1111/j.1556-4029.2010.01424.x>.
- David Faden, Julie Kidd, Jeremy Craft, L. Scott Chumbley, Max D. Morris, M. James Genalo, Lawrence J. and Kreiser, and Stephen Davis. Statistical confirmation of empirical observations concerning toolmark striae. *AFTE Journal*, 39(2):205–214, 2007.
- Taylor Grieve, L. Scott Chumbley, Jim Kreiser, Laura Ekstrand, Max Morris, and Song Zhang. Objective comparison of toolmarks from the cutting surfaces of slip-joint pliers. *AFTE Journal*, 46(2): 176–185, 2014.

Jeremy R. Hadler and Max D. Morris. An improved version of a tool mark comparison algorithm. *Journal of Forensic Sciences*, pages n/a–n/a, 2017. ISSN 1556-4029. doi: 10.1111/1556-4029.13640. URL <http://dx.doi.org/10.1111/1556-4029.13640>.

James E. Hamby, David J. Brundage, and James W. Thorpe. The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries. *AFTE Journal*, 41(2):99–110, 2009.