NCJRS
National Criminal Justice Reference Service

Office of Justice Programs
BJA BJS NIJ OJJDP OVC SMART

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**A Quantitative Understanding of Uniqueness and Reproducibility of Firearm Toolmark Surfaces**
Author(s):	**Michael T. Stocker**
Document Number:	**310904**
Date Received:	**November 2025**
Award Number:	**DJO-NIJ-19-RO-0009**

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Federal Award Number: DJO-NIJ-19-RO-0009

Project Title: A Quantitative Understanding of Uniqueness and Reproducibility of Firearm Toolmark Surfaces

Principal Investigator: Michael T. Stocker

National Institute of Standards and Technology

100 Bureau Drive Mailstop 8212

Gaithersburg, MD 20899

Phone: 301-975-5102

Email: stocker@nist.gov

Award Recipient Organization:

National Institute of Standards and Technology

100 Bureau Drive Mailstop 8212

Gaithersburg, MD 20899

Project Period:

January 1, 2020 – December 31, 2023 (including approved one year no cost extension)

Award Amount:

$458,108

Disclaimer: Any mention of commercial products within this document is for information purposes only; it does not imply recommendation or endorsement by NIST.

Table of Contents

Collaborators: Stephano Chatham[2], Ronald G. Dixson[1], Rachael Gominsky[2], Maritoni Litorja[1], Thomas B. Renegar[1], Erich Smith[2], Johannes A. Soons[1], Robert M. Thompson[1], James Yen[1]

1 – National Institute of Standards and Technology

100 Bureau Drive

Gaithersburg, MD 20899

2 – Federal Bureau of Investigation Laboratory

Firearms/Toolmarks Unit

2501 Investigation Parkway

Quantico, VA 22556

Abbreviations

ACCF – Areal Cross Correlation Function

AFTE – Association of Firearm and Toolmark Examiners

AFM – Atomic Force Microscope

BFCM – Breechface Casting Mold

CD – Critical Dimension

FAJ – Fiducial Alignment Jig

FATM – Firearm and Toolmark

FOV – Field of View

FPS – Firing Pin Stop

FTU – Firearms and Toolmark Unit

KM – Known Matching

KNM – Known Non-Matching

NIST – National Institute of Standards and Technology

NA – Numerical Aperture

OSAC – Organization of Scientific Area Committees

SAAMI – Sporting Arms and Ammunition Manufacturers Institute

SEM – Scanning Electron Microscope

SSMD – Strictly Standardized Mean Difference

S&W – Smith and Wesson

Keywords: 3D surface topography measurement, firearm and toolmark analysis, toolmark transfer, firearm toolmark uniqueness, firearm toolmark reproducibility

1.0 Project Background

The Association of Firearms and Tool Mark Examiners (AFTE) defines a tool as "An object used to gain mechanical advantage. Also thought of as the harder of two objects which when brought into contact with each other, results in the softer one being marked" [1]. Tools can be anything from a screwdriver to a pair of fingernail clippers. Firearms are also considered tools in that their machined surfaces impart toolmarks onto the cartridge case and bullet during firing. Striated toolmarks are left on the fired bullet as it passes through and contacts the lands and grooves of a rifled barrel. Impressed toolmarks are imparted on the soft primer surface of the fired cartridge case as it contacts the firing pin and breech face. The random nature of the tool's marking surface during manufacturing forms the basis of forensic firearm identification. The AFTE theory posits that all toolmarks transferred to the cartridge case and bullet are unique (individuality) to the tool (firearm surfaces) that created them.

In 2009, the National Academy issued a report [2] emphasizing the lack of objectivity in various forensic disciplines, including firearm and toolmark analysis. As a result, various studies have been done to demonstrate the scientific validity of the examiner-based forensic discipline, reporting low examiner false-positive and false-negative decision error rates [3,4]. In addition, the reports motivated extensive global research regarding the application of 3D surface topography measurements to objective firearm identification. Significant progress has been made in this area, including ballistic reference databases [5-7], identification algorithms [8-12], error rates [13], quality control [14], and instrumentation. However, a largely unexplored area to leverage and promote advances in objective 3D firearm identification exists in developing a comprehensive and quantitative understanding of the firearm toolmark surfaces themselves. There is no quantitative information in the literature regarding 1) the range of spatial wavelengths of the individualizing features (uniqueness) found on firearm toolmark surfaces, nor 2) how consistently firearms generate these individualizing marks (reproducibility). A lack of this foundational knowledge impedes advancements in several ways.

Without a quantitative evaluation of the range of unique spatial wavelengths, practitioners and 3D instrument manufacturers have no authoritative guidance on the appropriate measurement resolution for 3D surface topography measurements of forensic toolmark samples. The level of 3D measurement resolution required in the application of these instruments is an issue specifically identified by the Organization of Scientific Area Committees (OSAC) [15]. OSAC indicates this as a major gap in current knowledge. The range of sizes of these unique spatial wavelengths remains an unanswered question. For example, practitioners have utilized higher resolution instrumentation, such as scanning electron microscopes (SEM), in the past [16,17]. However, there is no research indicating that this level of resolution is advantageous to effectively compare forensic toolmark samples. Arbitrarily measuring at fine resolutions involves expensive equipment, can lead to wasted data acquisition and analysis times, and may even degrade similarity score/error rate evaluations. On the other hand, the presence of individualizing features at small scales (sub-micrometer to tens of micrometers) offers the possibility to tailor the technology and algorithms for the features at these scales. Quantitative knowledge of uniqueness will facilitate improved and optimized measurement and analysis of

firearm evidence with respect to the implementation of 3D surface topography measurement in the forensic laboratory.

A key area in objective firearm identification research is in determining error rates. Part of this effort requires an understanding of how various sources of uncertainty may influence a given error rate calculation. One source of uncertainty lies in the level of toolmark reproducibility from one test-fire to the next for a given firearm. Typically, a firearm's machined surfaces are not compared to or measured directly in a forensic examination. Access to these unique markings is achieved through "toolmark transfer" of the machined surface marks on the firearm (from manufacturing) to the various portions of the fired cartridge case and bullet. A variety of factors, including manufacturing/assembly tolerances, surface detail persistence [18], firearm contamination, ammunition pressures, and primer material, affect the fidelity and reproducibility of the toolmarks transferred from the firearm to the test-fired cartridge component. A quantitative understanding of this reproducibility is essential to fully understand how it may affect error rates.

2.0 Project Summary

2.1 Project Objectives and Goals

The two-fold objective of the project was to generate quantitative information regarding 1) the range of unique spatial wavelengths found on firearm tool marking surfaces and 2) how well firearm toolmarks reproduced from firing to firing considering the effects of manufacturing method, primer materials, and ammunition pressures. The goal of the project was to use the quantitative uniqueness and reproducibility data to facilitate advances in objective firearm identification based on 3D optical surface topography technologies. The project also aimed to quantify toolmark transfer fidelity using the data and analysis generated for the uniqueness and reproducibility studies. The scope of this project focused on analyzing toolmarks from firearm breech face impressions on 9 mm Luger caliber cartridge cases. The selected firearm and ammunition sources were typical of those examined in casework.

Uniqueness was evaluated in this study on a matrix of test-fires (described in 2.3.3) that were generated during the project and measured both optically and with an AFM. Toolmark reproducibility was also evaluated on a matrix of test-fires measured optically. Replica casts of the firearm breech faces were generated and served as ground truth to understand the spatial wavelengths present on the firearm versus on the test-fire. Measurement and evaluation of these casts, along with measurement and evaluation of the test-fires, provided a way to estimate toolmark transfer efficacy.

2.2 Research Questions

The project intended to answer the following research questions:

1. What is the spatial wavelength range of individualizing features present on the primer areas of fired cartridge cases as a function of different breechface manufacturing methods and ammunition parameters?
2. How well do the individualizing features present on the primer areas of fired cartridge components reproduce as a function of different breechface manufacturing methods and ammunition parameters?
3. What is the toolmark transfer fidelity for the parameters studied? (i.e., what sizes of features present on the firearm breechface transfer to the test-fired primer surface?)

2.3 Project Design

The completion of this project required the generation of test-fires and forensic casts representing different firearm manufacturing methods and ammunition parameters. These test-fires and forensic casts were measured on two different high-resolution measurement instruments. Analysis techniques were developed that enabled the determination of the spatial wavelengths present on the physical breechface of the firearms and the primer surface area of the fired cartridges. These materials, instruments, and methods are described below.

2.3.1 Firearm Models and Ammunition

Three firearm models were chosen to represent different manufacturing, roughing, and finishing operations of the firearm's breechface surface. Glock 19, Hi-Point C9, and Smith and Wesson (S&W) MP9 consecutively manufactured sets were selected. Glock 19 firearms are classified as having "parallel" breechface markings, likely from the broaching operation Hi-Point C9 firearms are also classified as having "parallel" breech face markings. However, their markings are much more prominent, due to a belt sanding finishing operation. S&W breechface surfaces have "granular" breechface markings. Each set contains 10 consecutively manufactured firearms.

Two ammunition parameters were selected to study their effect on toolmark formation and transfer: primer material and ammunition pressure. Brass and nickel-plated brass were chosen as the ammunition primer materials. Standard and +P+ were chosen as the ammunition pressures. For ammunition availability reasons, we limited our ammunition to those containing 3.5 mm diameter primers. While hand loaded ammunition is certainly a factor in some casework, we did not evaluate it in this study. We were aware of the effect of lacquer coatings on toolmark transfer [19], but it also was not addressed in this study. Table 1 lists the ammunition brands selected, covering the two primer materials and two pressures.

Primer Material	Ammunition Pressure	Ammunition Brand
Brass	Standard	TULAMMO, 9 mm Luger, 115 grain, FMJ, Steel Case, Berdan Primed
Brass	+P+	Winchester Ranger, 9 mm Luger, 115 grain, JHP, Brass Case
Nickel	Standard	PMC, 9 mm Luger, 115 grain, JHP, Brass Case
Nickel	+P+	Winchester, 9 mm Luger, 115 grain, JHP, Brass Case

Table 1 – List of specific ammunition used in the study.

2.3.2 Instrumentation

Two well-characterized calibrated 3D microscope systems at NIST were used to acquire high-resolution data of the acquired test-fires; 1) an optical Zygo Nexview NX2 coherence scanning interferometric (CSI) microscope and 2) a Veeco Critical Dimension (CD) atomic force microscope (AFM). The Zygo microscope was operated at a magnification of 50X, producing a field-of-view (FOV) of approximately 350 µm x 350 µm and an image effective pixel spacing on the order of 350 nm. This particular objective had a numerical aperture (NA) of 0.55, providing an optical resolution of 0.52 µm. This optical resolution specification is based on the Sparrow criterion at a wavelength of 570 nm. Optical test-fire measurements employed a 6 x 11 stitched measurement, producing an approximate stitched FOV of 1.40 mm by 2.45 mm.

The AFM is an ultra-high-resolution scanning probe instrument that raster scans an area to be measured. For this study, the AFM was used to acquire areal measurements on the order of 40 µm x 40 µm in size. The AFM was set to sample the measured surface at 100 nm in both the scan axis and the non-scan axis directions. Both instruments are shown in Fig. 1.

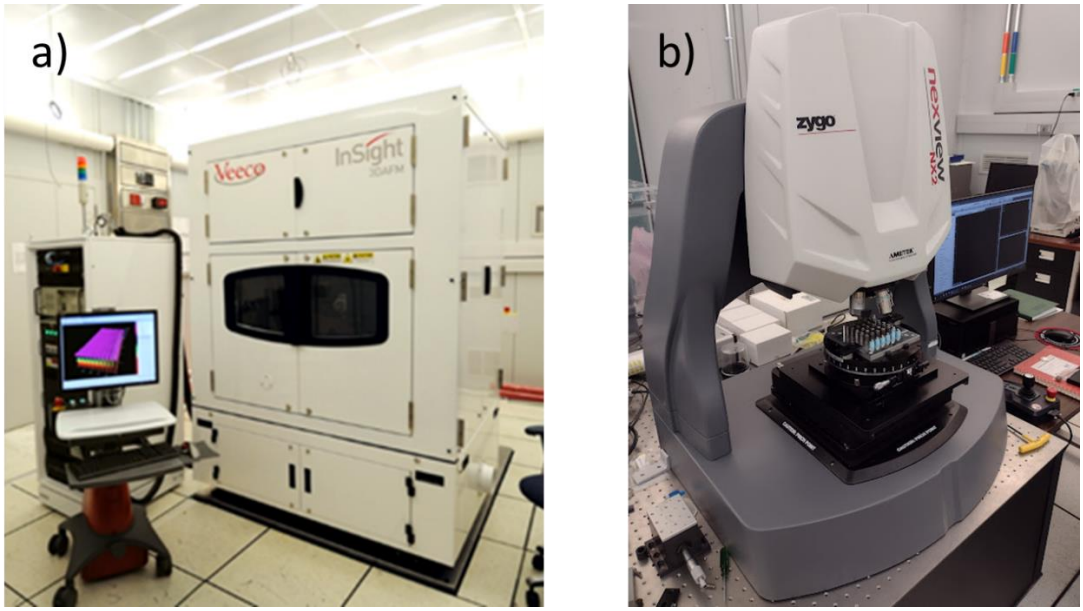


Fig. 1. a) Veeco CD-AFM, b) Zygo Nexview NX2 CSI microscope.

2.3.3 Experiment Design

2.3.3.1 Uniqueness Study

Test-fire generation for the uniqueness portion of the study was a full factorial of the experimental variables, including all 10 slides of each firearm model set. Three test-fires were produced for each experiment variable combination, resulting in a total of 360 test-fires to evaluate for uniqueness across all three firearm models. This design is summarized in Table 2(a). These test fires were all measured optically.

Uniqueness	
Experiment Variable	**Samples**
Consecutively Manufactured Sets (Each set a different manufacturing method)	3
Number of slides utilized from each set	10
Primer materials	2
Ammunition pressures	2
Test fires per combination	3
TEST FIRES =	360

a)

Reproducibility	
Experiment Variable	**Samples**
Consecutively Manufactured Sets (Each set a different manufacturing method)	3
Number of slides utilized from each set	3
Primer materials	2
Ammunition pressures	2
Test fires per combination	10
TEST FIRES =	360

b)

Table 2 – a) Experiment design for uniqueness portion of study, b) Experiment design for reproducibility portion of study.

In addition, for evaluating uniqueness at higher resolutions than possible with optics, test-fires were produced and measured on the AFM. The AFM test-fires were limited to brass-primer standard-pressure ammunition. Additional test-fire modifications were necessary to enable them to fit in the AFM. The cartridges required cutting them to a specific height and adding fiducial marks to facilitate reproducible measurements in the same location. Because of the difficulties in sample preparation and increased measurement time, only 3 test-fires per firearm model were generated for AFM measurement. To increase the number of KM toolmarks, three measurement sites were chosen on each cartridge.

2.3.3.2 Reproducibility Study

For the reproducibility portion, a limited design was employed, reducing the number of slides to 3 from each set and increasing the number of test-fires per combination to 10. This also resulted in 360 test-fires. A refined experiment design allowed us to generate fewer test fires for reproducibility, wherein some of the test fires were reused from the uniqueness study for the reproducibility study. A total of 252 new test-fires were generated for the reproducibility study. 108 test-fires from the uniqueness study were used to provide the required 360 test-fires. This design is shown in Table 2(b).

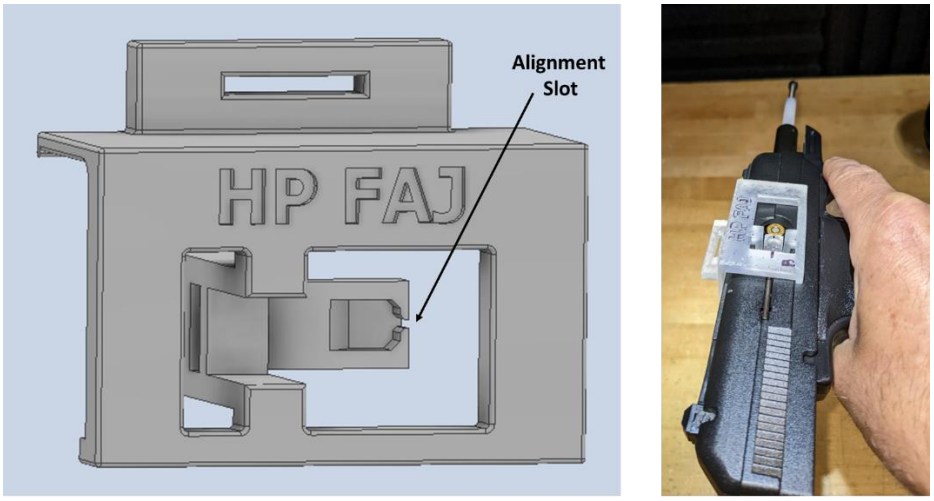
2.3.3.3 Toolmark Transfer Study

Casts of the firearm breech face surfaces were created to serve as ground truth for understanding what features and spatial wavelengths were physically present on the firearm's breechface. Casts of the slides were made before and after the uniqueness and reproducibility portions of the study. In total, two casts per each of the 30 firearms were made, totaling 60 casts. Before and after casts for the first three Hi-Point firearms did not cure properly. As such, only seven before and after sets were used in the analysis for each firearm model, reducing the number of casts measured and used to 42. These casts were used to characterize the toolmark transfer fidelity of each of the firearm and ammunition combinations.

2.3.4 Test-fire and Cast Preparations

2.3.4.1 Test-fire Generation and Measurement

Optical high-resolution measurements can be very time-consuming with very large file sizes. This motivated a decision to measure only a fraction of the available primer area on each of the test-fire breechface impressions. This, in turn, complicated the test-fire generation process. It created the need to precisely position each cartridge in the chamber before firing. This ensures that the same area of interest on the firearm's breechface can be repeatably captured. Without this effort, we would have difficulty in generating KM samples. To guarantee we were capturing and measuring the same area of the breech face for each firearm, a jig for each firearm model was designed, fabricated, and utilized during test-firing. An example of the fiducial alignment jig (FAJ) for the Hi-Point C9 firearms is shown in Fig. 2. For each test-fire, the jig was inserted in the chamber, and the cartridge was rotated to align a fiducial on the cartridge headstamp with the jig alignment slot.



a) b)

Fig. 2. a) CAD design of the FAJ for the Hi-Point C9 firearm. The alignment slot used to orient the cartridge case is noted with the arrow. b) Picture of the Hi-Point FAJ inserted in the chamber while aligning a cartridge.

A measurement automation scheme was devised to measure the hundreds of test-fires generated in the study. The scheme involved loading a pallet with 25 cartridges at a time for measurement. Each cartridge required precise alignment on the pallet, using the same alignment fiducial utilized for aligning the cartridge in the chamber when it was test-fired. Each pallet was measured at low magnification on a 3D optical surface topography microscope to determine the center position of each cartridge case. The center position information was then used to program a measurement automation routine on the Zygo Nexview NX2 microscope, enabling 50X measurements of each test-fire on the pallet. This measurement automation workflow is shown in Fig. 3(a). A total of twenty-five pallets were required to measure the complete set of test-fires. The process of preparing and measuring each pallet took approximately 18 to 20 hours. Fig. 3(b) shows an example of a fully loaded pallet.

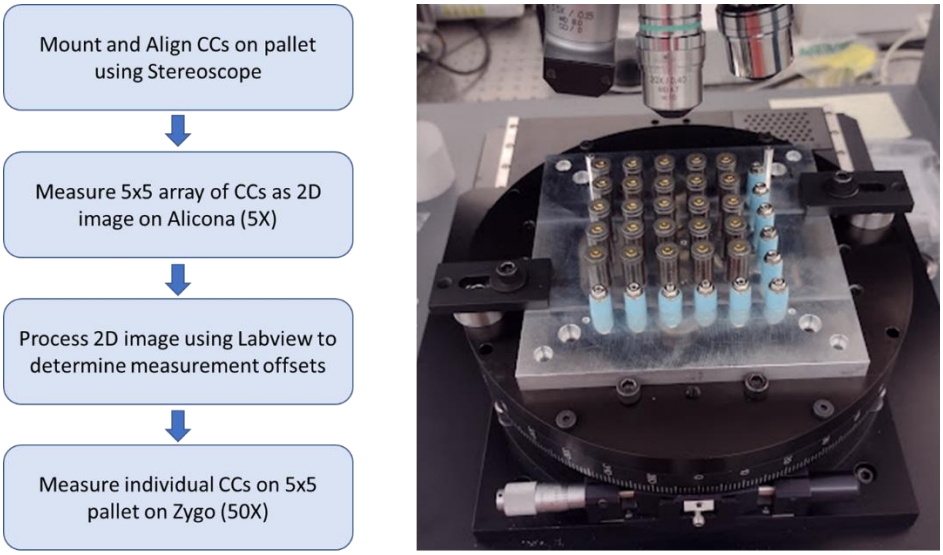


Fig. 3. a) Test-fire measurement automation workflow, b) Pallet loaded on Zygo Nexview NX2 microscope

2.3.4.2 Cast Generation and Measurement

Breechface casting molds (BFCM) were designed and fabricated for the three different firearms used in the study. The BFCM made the casting of the breechface surface easier and produced a cast that was more conducive to measurement on the microscopes. When a breechface cast is made, it also captures the firing pin and firing pin aperture. Studying the firing pin was not part

of this study, but the cast and test-fire data could be used later to study the size and reproducibility of individualizing features present on firing pins. For that reason, a firing pin stop (FPS) was designed and fabricated to ensure the same amount of the firing pin was captured on each cast for a given firearm model. Fig. 4 shows the BFCM and FPS inserted in a Hi-Point firearm slide in preparation for making a cast. Cast measurements were also made on the Zygo Nexview NX2 microscope, at the same magnification and resolution as the test-fires.

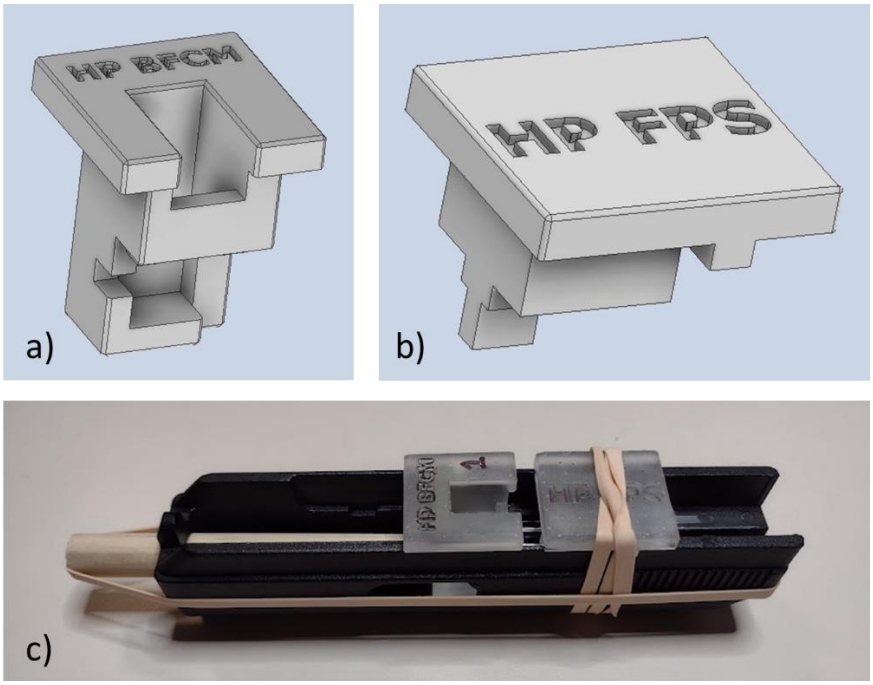


Fig. 4. a) BFCM for Hi-Point firearm, b) FPS for Hi-Point firearm, c) BFCM and FPS inserted in a Hi-Point C9 slide in preparation to make a breechface cast.

2.4 Data Analysis Methods

The normalized areal cross-correlation function ($ACCF_{MAX}$) [8] is one of many analytical comparison methods that have been applied to the determination of similarity between surface topography measurements of breechface impressions. Comparison methods, including $ACCF_{MAX}$, are validated by running reference data sets through them. The separation between known matching (KM) and known non-matching (KNM) reference distributions indicates the effectiveness of the method's discriminating power, in addition to providing a means to estimate error rates. This is shown schematically in Fig. 5, with reasonable separation displayed between KM and KNM distributions and the false negative and false positive error rates indicated in the tails of those distributions.

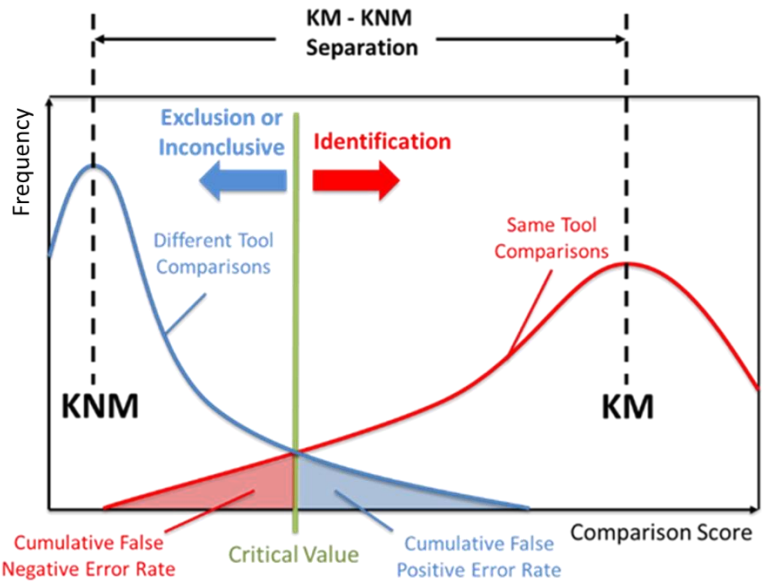


Fig. 5. Schematic showing frequency distributions of known matching and known non-matching comparison scores. The cumulative false negative and false positive error rates are highlighted in the tails of those distributions.

In this study, $ACCF_{MAX}$, in combination with KM and KNM distributions, were used to construct a metric to evaluate firearm toolmark uniqueness and reproducibility. They were used in conjunction with a bandpass analysis to quantify the range and reproducibility of individualizing spatial wavelengths present in the breechface impressions of the test-fires and casts produced. Because of the high-resolution instrumentation used, the measured topography data included accurate high-resolution spatial wavelength information ranging from nanometers to millimeters. A bandpass analysis was devised, shown schematically in Fig. 6, creating bands that spanned different portions of the roughness and waviness [ASME B46 ref] segments of the surface's spatial wavelength spectrum. λ_S and λ_C are the short cutoff and long-cutoff spatial frequency limits that define the bandpass range. A given set of measured breechface impression and cast topographies are filtered to form the different bandpassed data sets. The $ACCF_{MAX}$ is applied to each data set to build the KM and KNM distributions at each bandpass.

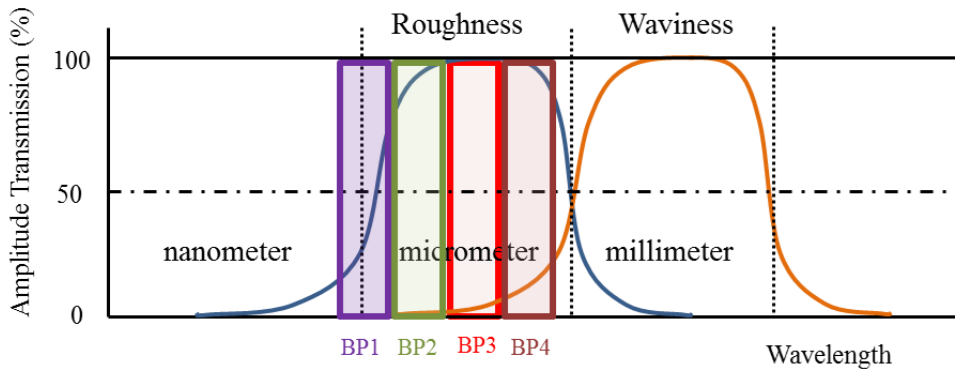


Fig. 6. Schematic of bandpass analysis that breaks up the surface spatial wavelength spectrum into appropriate wavelength ranges.

A summary statistic, the strictly standardized mean difference (SSMD), was used to represent the statistical separation between and the KM and KNM distributions. SSMD is defined as

$$SSMD = \frac{\mu_{KM} - \mu_{KNM}}{\sqrt{(\sigma_{KM})^2 + (\sigma_{KNM})^2}} \quad (1)$$

where the difference between the known matching distribution mean (μ_{KM}), and the known non-matching distribution mean (μ_{KNM}), is divided by the root sum square of the known matching distribution standard deviation (σ_{KM}), and the known non-matching standard deviation (σ_{KNM}). The SSMD has the benefit of normalizing the separation by the combined standard deviations of the two distributions (i.e., the separation alone as a metric is not sufficient). A plot of SSMD values is generated as a function of spatial wavelength bandpass values. A SSMD value greater than zero indicates the presence of individualizing features at those wavelengths. A higher SSMD value is interpreted to mean a stronger presence of features at that wavelength range than features at a different wavelength range with a lower SSMD value.

2.5 Expected Applicability of the Research

The foundational knowledge (quantitative uniqueness and reproducibility information) generated will give necessary guidance, enable advancements, and provide immediate benefits to the discipline:

1) Provide practitioners with guidance on required lateral measurement resolution
2) Enable 3D instrument manufacturers to produce tailored instruments with targeted resolution capabilities for toolmark evaluation
3) Define and demonstrate optimized data analysis parameters (filtering cutoffs) for the most accurate and efficient toolmark analysis
4) Inform error rate uncertainty estimates with respect to toolmark reproducibility
5) Contribute quantitative knowledge of firearm toolmark transfer fidelity
6) Provide practitioners with a validated method for examining inoperable or dismantled firearms based on casting.

3.0 Participants and Collaboration Organizations

3.1 NIST participants

Michael Stocker was the principal investigator. He provided general oversight for the project in addition to involvement in mechanical design, sample generation, and measurement and analysis of the firearm toolmark samples. Development of new analysis and data acquisition routines for processing the higher-resolution data were performed by Dr. Johannes Soons. Dr. Maritoni Litorja created image processing techniques used in the required measure automation routines. Thomas Renegar provided support in the form of mechanical design and measurement of firearm toolmark samples. AFM measurement and analysis was performed by Dr. Ronald Dixson. Dr. James Yen provided statistical analysis and support. Robert Thompson served as a subject matter expert, providing guidance for the project as a whole.

3.2 Federal Bureau of Investigation Firearms and Toolmarks Unit (FTU) Collaboration

Erich Smith is a Physical Scientist and senior firearms examiner at the FTU. He served as a subject matter expert for the project. He provided the ammunition and access to the consecutively manufactured firearm sets used in this study. Stephano Cheswick is a Physical Scientist at the FTU. He provided support through test-fire generation for the study.

4.0 Changes in approach

4.1 50X Magnification Measurements

The original proposal specified using 100X magnification measurements from a confocal microscope. We chose to use a newer interferometric-based coherence scanning surface topography microscope: the Zygo Nexview NX2. To achieve reasonable measurement times and manageable file sizes, without sacrificing much in the spatial sampling of the measured surfaces, we chose to use 50X magnification.

4.2 +P+ High-pressure Ammunition

Due to ammunition shortages in 2020 and 2021 and the need to have certain primer material and pressure ammunition combinations, we had to select a different high-pressure ammunition. We chose +P+ ammunition. While this is not a SAAMI standardized ammunition, there is a reasonable expectation that +P+ ammunition runs about 10% to 15% higher in pressure than standard ammunition, which was still suitable for the purposes of this study.

5.0 Outcomes

5.1 Results

High-resolution and ultra-high-resolution breechface impression data was acquired, processed, and analyzed to evaluate uniqueness, reproducibility, and toolmark transfer for three different

firearm models: Glock 19, Hi-Point C9, and S&W MP9. The results of these analyses are described below.

5.1.1 Uniqueness

SSMD bandpass plots were generated from the uniqueness data acquired during the study. Figs. 7(a), 7(b), and 7(c) are the SSMD bandpass plots for the high-resolution optically-measured test-fire data from the Glock 19, Hi-Point C9, and S&W MP9 respectively. All four experimental variable combinations are represented on each plot: brass-primer and regular-pressure shown in the dashed line with circle markers, brass-primer and high-pressure (noted as "Plus" in the legend) shown in the solid line with circle markers, nickel-primer and regular-pressure shown in the dashed line with square markers, and nickel-primer and high-pressure shown in the solid line with square markers.

Two different bandpass sets were defined for the optical and AFM data. Optical data were analyzed using the following bandpasses: <5 µm, 5 µm to 10 µm, 10 µm to 20 µm, 20 µm to 40 µm, 40 µm to 80 µm, 80 µm to 160 µm, 160 µm to 320 µm, 320 µm to 640 µm, 640 µm to 1280 µm, and >1280 µm. Because the AFM data contained smaller spatial wavelengths, smaller bandpass regions were added to the optical bandpasses: <1.25 µm, 1.25 µm to 2.5 µm, 2.5 µm to 5 µm, 5 µm to 10 µm, 10 µm to 20 µm, 20 µm to 40 µm, and >40 µm.

Due to long data acquisition times, AFM data was only acquired on brass-primer regular-pressure test-fires. Fig. 7(d) shows the SSMD bandpass plots for the AFM data: Glock 19 data is shown in the dashed line with circle markers and S&W data is shown in the dashed line with square markers. Attempts were made to acquire AFM data on Hi-Point C9 test-fires, but these primer topographies frequently exceeded the AFM scan range maximum of 5 µm.

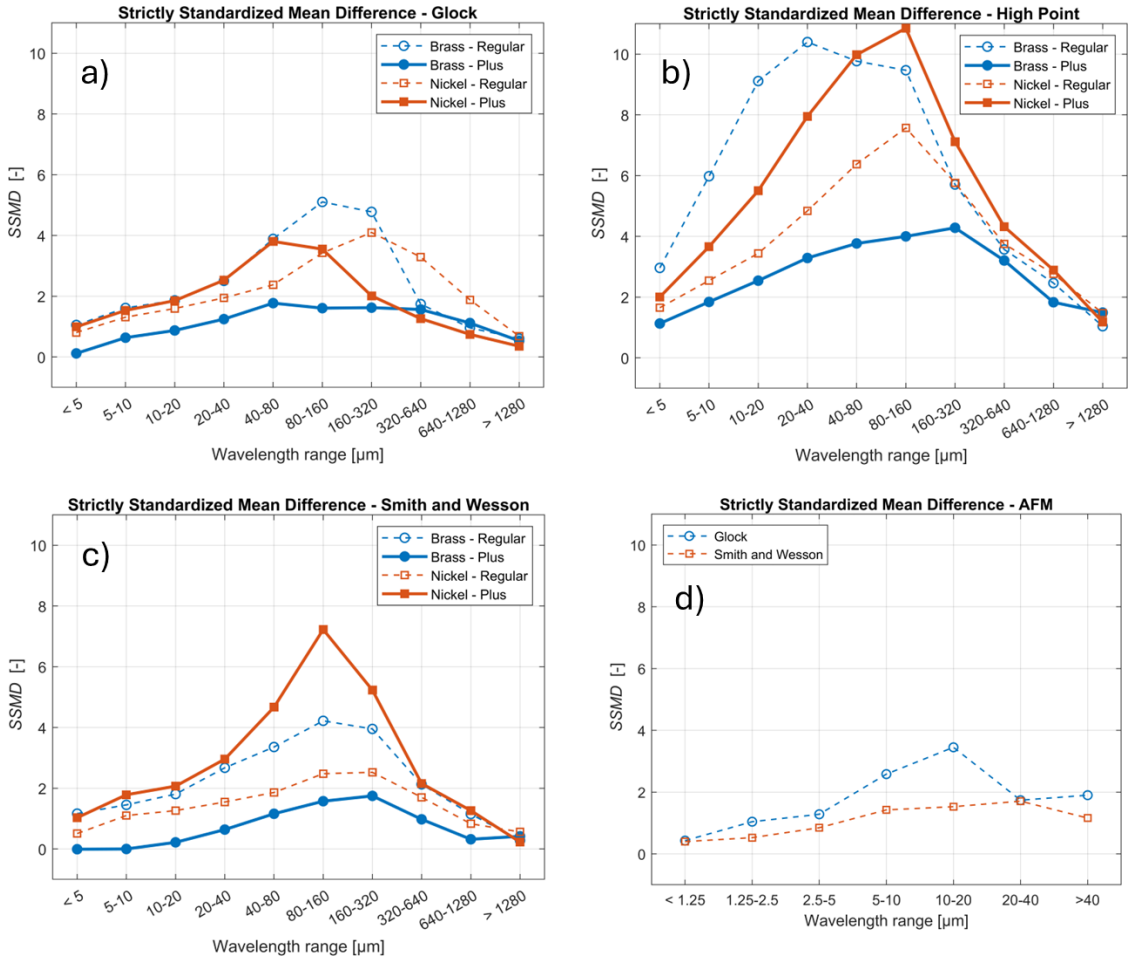


Fig. 7. Graphs of SSMD uniqueness of test-fires a) optical data for Glock 19 test-fires, b) optical data for Hi-Point C9 test-fires, c) optical data for S&W MP9 test-fires, and d) AFM ultra-high-resolution data for Glock and S&W test-fires .

5.1.2 Reproducibility

SSMD bandpass plots were also generated from the reproducibility data acquired during the study. Figs. 8(a), 8(b), and 8(c) are the SSMD bandpass plots for the high-resolution optically-measured reproducibility test-fire data from the Glock 19, Hi-Point C9, and S&W MP9 respectively. All four experimental variable combinations are again represented on each plot: brass-primer and regular-pressure shown in the dashed line with circle markers, brass-primer and high-pressure shown in the solid line with circle markers, nickel-primer and regular-pressure shown in the dashed line with square markers, and nickel-primer and high-pressure shown in the solid line with square markers. Due to long data acquisition times, no AFM reproducibility data was taken. The bandpasses defined for the high-resolution optical surface topography reproducibility data were the same as those defined above for the high-resolution optically-measured test-fire uniqueness data.

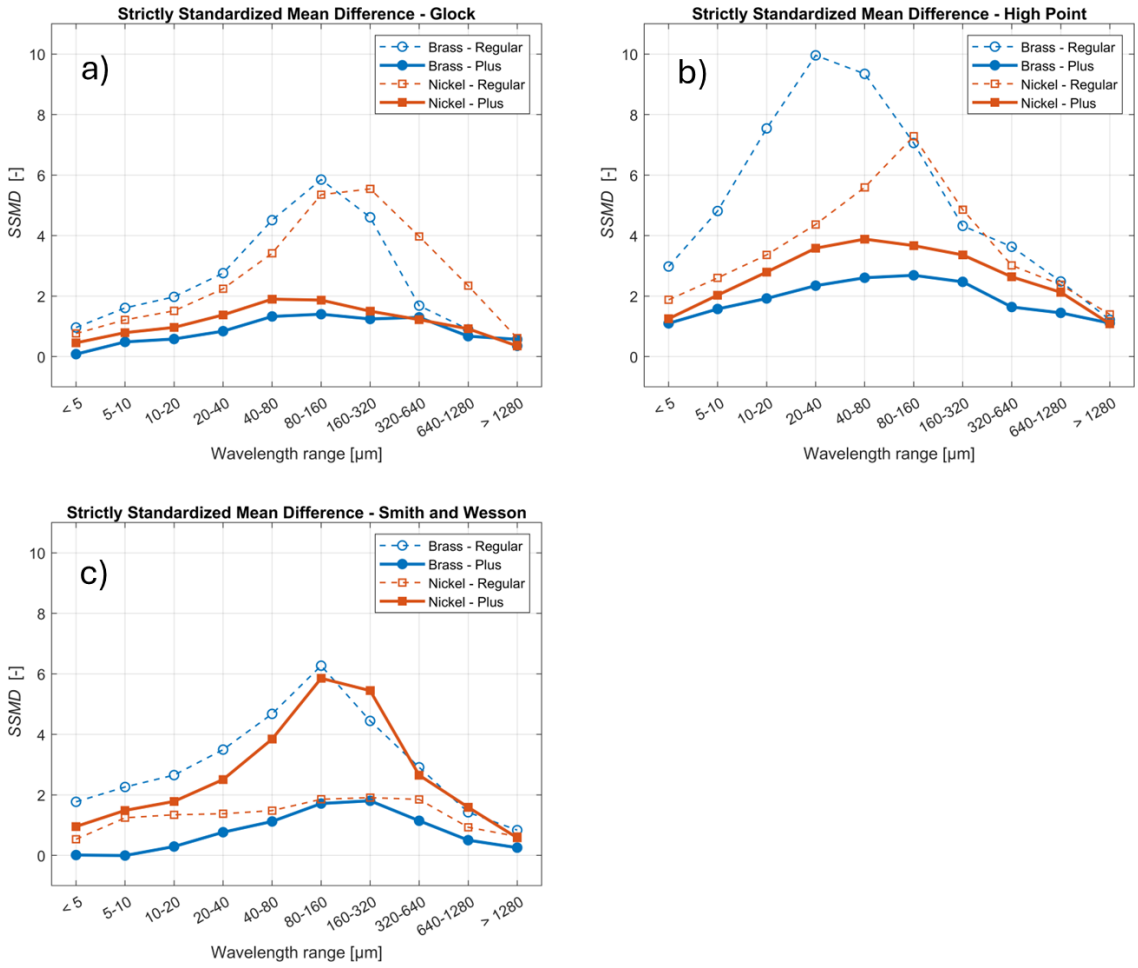


Fig. 8. a) SSMD plots for Glock 19 optically-measured test-fire reproducibility data, b) SSMD plots for Hi-Point C9 optically-measured test-fire reproducibility data, and c) SSMD plots for S&W MP9 optically-measured test-fire reproducibility data.

5.1.3 Toolmark Transfer Fidelity

Evaluating firearm toolmark transfer requires an understanding of what toolmarks exist on the physical firearm breechface and what survives the firing process and is present on the fired cartridge components, in this case, the primer breechface impression. As described above (2.3.3.3), casts were made of each slide for all three firearm models used in the study. Toolmark transfer fidelity is characterized by comparing test-fire SSMD values to cast SSMD values for a given firearm model. Direct subtraction or division of the cast and test-fire SSMD data was not mathematically intuitive. Instead, we observe the transfer fidelity through visual comparison of the cast and test-fire SSMD values from the same firearm. Figs. 9(a), 10(a), and 11(a) all show the same cast SSMD data, but in each plot a different firearm cast profile is indicated by the red arrow for comparison. For example, in Fig. 9(a), the Glock cast SSMD profile is indicated by the red arrow and can be visually compared to four Glock test-fire SSMD plots in Fig. 9(b). Similarly,

visual comparisons of the cast and test-fire SSMD data for Hi-Point and S&W in Figs. 10 and 11 can be made.

Three important qualities are important to note when visually comparing these data in Figs. 9, 10, and 11: 1) Test-fire SSMD data equal in magnitude to cast SSMD data indicates high transfer fidelity, 2) Test-fire SSMD data at any given wavelength should not be higher than its corresponding cast SSMD value at the same wavelength, and 3) A lower test-fire SSMD value at a given wavelength, compared to its corresponding cast SSMD value at the same wavelength, indicates a loss in toolmark transfer at that wavelength.

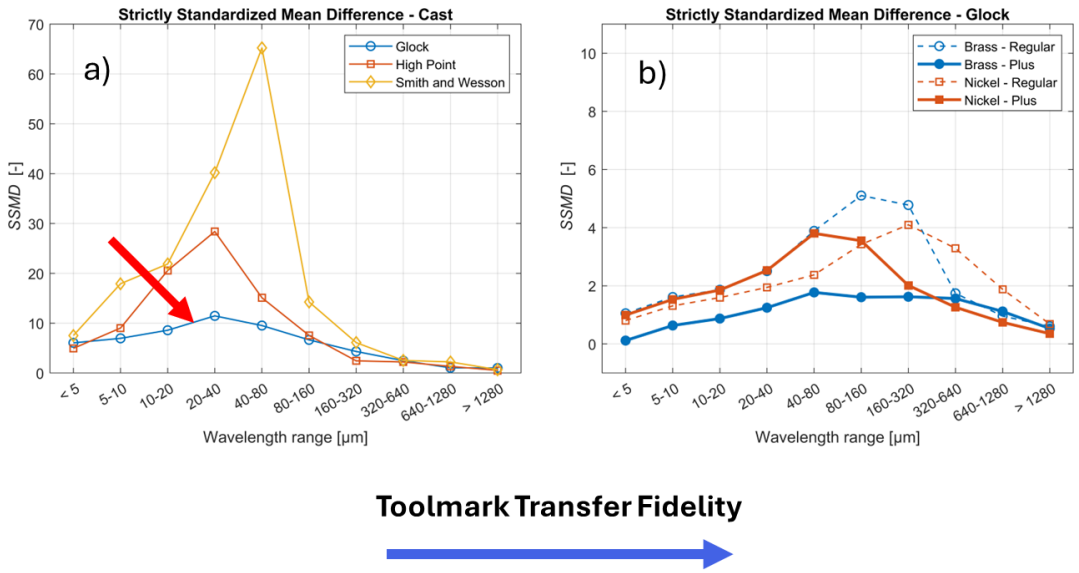


Fig. 9. Glock toolmark transfer fidelity. a) SSMD plots for casts of Glock 19, Hi-Point C9, and S&W MP9 firearms. The Glock 19 cast SSMD curve is indicated with the red arrow., b) SSMD plots for Glock 19 optically-measured test-fire uniqueness data. Transfer fidelity can be observed by visually comparing the magnitude of the Glock cast SSMD values to the test-fire SSMD values.

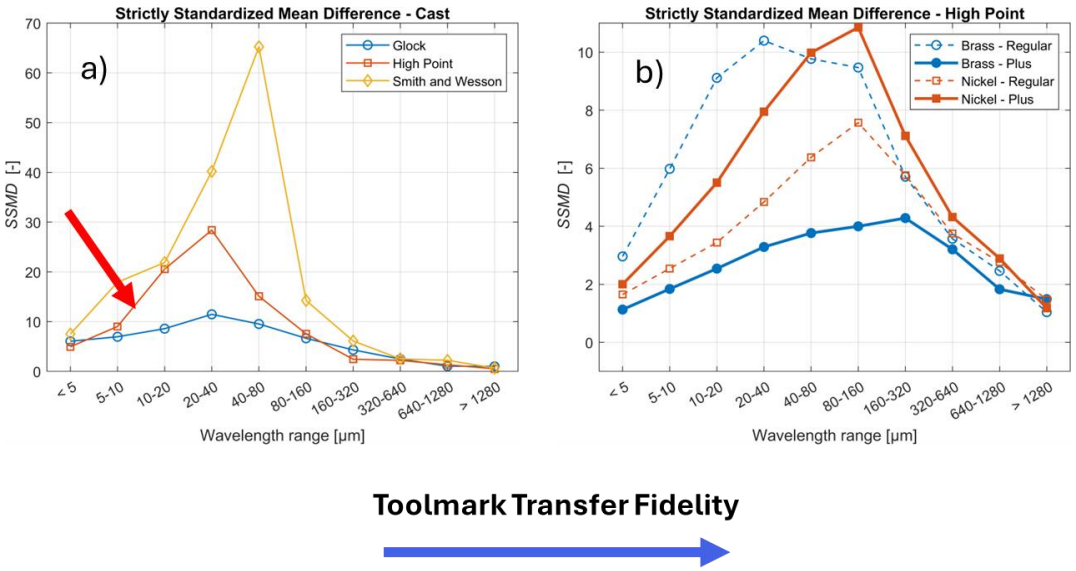


Fig. 10. High Point toolmark transfer fidelity a) SSMD plots for casts of Glock 19, Hi-Point C9, and S&W MP9 firearms. The Hi-Point C9 cast SSMD curve is indicated with the red arrow., b) SSMD plots for Hi-Point C9 optically-measured test-fire uniqueness data. Transfer fidelity can be observed by visually comparing the magnitude of the Hi-Point cast SSMD values to the test-fire SSMD values.

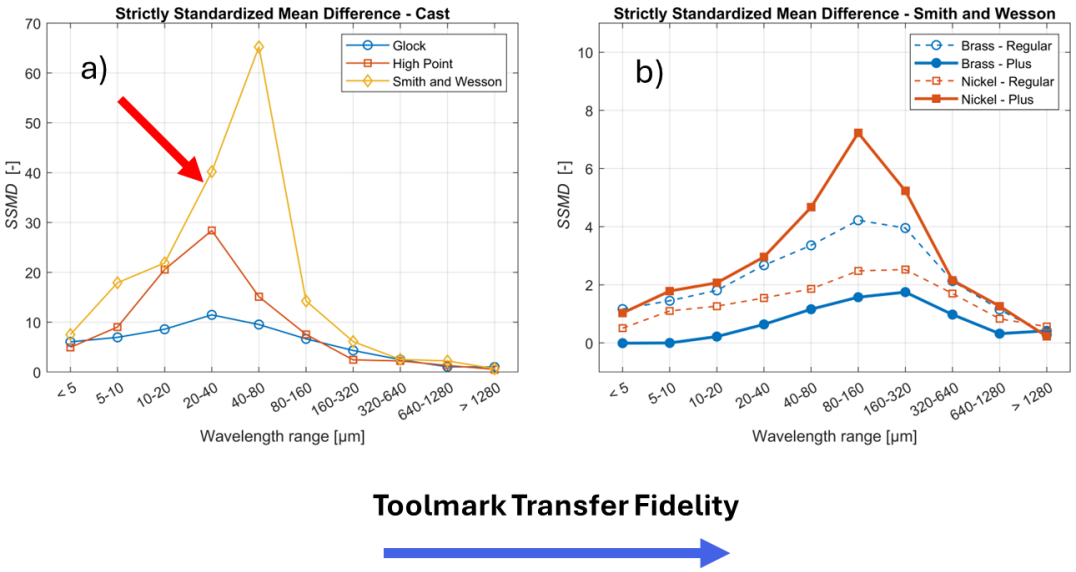


Fig. 11. S&W toolmark transfer fidelity a) SSMD plots for casts of Glock 19, Hi-Point C9, and S&W MP9 firearms. The S&W MP9 cast SSMD curve is indicated with the red arrow., b) SSMD plots for S&W MP9 optically-measured test-fire uniqueness data. Transfer fidelity can be observed by visually comparing the magnitude of the S&W cast SSMD values to the test-fire SSMD values.

5.2 Discussion

Successful firearm identification requires toolmarks on the firearm to be unique and the firing process (at some level) to impart those marks to the fired cartridge components (in this case the primer) in a reproducible manner. In the absence of these conditions, the nature of the KM and KNM distributions will not enable discrimination of samples from same or specific sources. Several generalities can be stated. If there are few or no unique features on the firearm, there will be little to no separation between the KM and KNM distributions. High reproducibility (regardless of the presence of unique features on the firearm) will guarantee high KM scores. High reproducibility without the presence of unique features (subclass characteristics present) will cause the KNM scores to be high as well. Low reproducibility (regardless of the presence of unique features on the firearm) will result in low KM and KNM scores. A continuum of these effects manifests in the data presented above.

The experiment design was intended to separate the effects of uniqueness and reproducibility on KM and KNM distributions. Because many slides and few repeats were used for the uniqueness study and few slides and many repeats were used for reproducibility study, the uniqueness data presented is affected more by the presence of unique features and the reproducibility data presented is affected more by the existence of reproducibility of the firing process. However, they are not completely separable. Changes in reproducibility will have an effect on the uniqueness data and vice versa. Even so, the data is useful, exhibiting interesting trends and effects that are discussed below.

5.2.1 Uniqueness Trends

SSMD values for high-resolution optically-measured test-fire uniqueness data for Glock and Hi-Point (Figs. 7(a) and 7(b)) all exceeded zero at all wavelength bandpasses tested. This indicates the presence of unique firearm breechface marks that are transferred to the primer, even at the extreme bandpasses defined in this study (< 5 μm and > 1280 μm). With the exception of the two smallest bandpasses (< 5 μm and 5 μm to 10 μm), the S&W exhibited the same trend. Hi-Point showed the highest SSMD values, and therefore the strongest presence of individualizing features, on all four of the ammunition combinations, compared to Glock and S&W. This is perhaps not surprising, given that Hi-Point primer breechface marks contain deep striated marks generated from a sanding finishing operation.

We note that the 1) brass-primer regular-pressure ammunition and 2) nickel-primer high-pressure ammunition combinations scored the highest SSMD values across all three firearm models (Fig. 7), with the exception of the Glock, where the nickel-primer regular-pressure ammunition was on par with the nickel-primer high-pressure ammunition. In contrast, the brass-primer high-pressure ammunition consistently produced the lowest SSMD values across all three firearm models.

Regarding the high-pressure ammunition, it is possible to imagine two toolmark transfer mechanisms at play. First, the high-pressure ammunition could cause deep well-defined, reproducible impression toolmarks on the primer during firing (yielding higher SSMD values). Alternatively, the high-pressure ammunition could cause a more chaotic firing event, resulting in more motion (and likely smearing) between the firearm breechface and the primer (yielding lower SSMD values). Also, it may simply be that certain firearm-ammunition combinations mark better than other combinations.

5.2.2 Filtering Parameters

Typical pre-processing of breechface impression data involves filtering the surface to highlight individualizing features. For comparing surfaces using the $ACCF_{MAX}$, 15 µm and 400 µm are typical low and high filter cutoffs, respectively. The peaks and majority of ranges of optically-measured test-fire uniqueness SSMD values (Fig. 7) for all three firearm models and all four ammunition combinations fell within this range, validating the use of these cutoff values. These cutoff values demonstrate a sweet spot for the most accurate and efficient toolmark analysis. Current optical systems utilized (at low magnification) in forensic toolmark analysis are easily capable of resolving features sub-10 micrometers. The successful application of a 15 µm low cutoff filter implies that higher resolution instrumentation may not be necessary or advantageous.

5.2.3 Reproducibility Trends

SSMD values for the high-resolution optically-measured test-fire reproducibility data (Fig. 8) looked, not surprisingly, quite similar to the optically-measured test-fire uniqueness SSMD values (Fig. 7). One exception to this was the nickel-primer high-pressure ammunition. This ammunition produced much lower SSMD values for the Glock and Hi-Point firearms relative to the uniqueness data.

The SSMD values shown in Fig. 8 are directly affected by increases and decreases in reproducibility, but are also affected by increases and decreases in individualizing features on the primer surface of the test-fires. High KM $ACCF_{MAX}$ comparison scores are significant indicators of the reproducibility of individualizing features. To observe reproducibility levels across the firearms and ammunition combinations, box plots containing the KM $ACCF_{MAX}$ score distributions are provided in Figs. 12, 13, and 14. For this discussion, the KNM $ACCF_{MAX}$ scores are ignored.

For the Glock firearms, the regular-pressure ammunitions (Figs. 12(a) and 12(c)) demonstrated high reproducibility across the various wavelength bands, as evidenced by the smaller red boxes for the KM distributions. High-pressure ammunition (Figs. 12(b) and 12(d)) resulted in lower KM $ACCF_{MAX}$ scores and significant drops in reproducibility (larger red boxes). A similar trend is seen in the Hi-Point firearms (Fig. 13), but for only one of the regular-pressure ammunitions: the brass-primer one (Fig. 13(a)). There was very little difference in the

reproducibility and KM $ACCF_{MAX}$ scores for the nickel-primer regular- and high-pressure ammunitions (Figs. 13(c) and 13(d)), apart from a slight decrease in scores across the wavelength bands. Reproducibility for the S&W firearms were approximately the same for three of the four ammunitions: 1) brass-primer regular-pressure (Fig. 14(a)), 2) brass-primer high-pressure (Fig. 14(b)), and 3) nickel-primer high-pressure (Fig. 14(d)). While the brass-primer high-pressure reproducibility was on par with the other two, the KM $ACCF_{MAX}$ scores dropped measurably, indicating a decrease in the presence of individualizing features for this ammunition in this firearm. The nickel-primer regular pressure produced the worse reproducibility (Fig. 14(c)) when fired through the S&W firearms.

A near-universal behavior lies in the reproducibility at the smallest and largest wavelengths. With the exception of the largest wavelength KM $ACCF_{MAX}$ scores in the Hi-Point test-fires seen in Figs 13(b), 13(c), and 13(d), the best reproducibility always occurred at the extreme low and high wavelength portions of the spectrum. The better reproducibility at higher wavelengths is perhaps easier to understand, essentially that large-form individualizing features survive. Less intuitive is the notion that the smallest individualizing features on the firearms survive the violent firing process and reproduce better than the larger wavelength ones.

Finally, it is observed that the brass-primer regular-pressure ammunition seemed to enable the best reproducibility results.

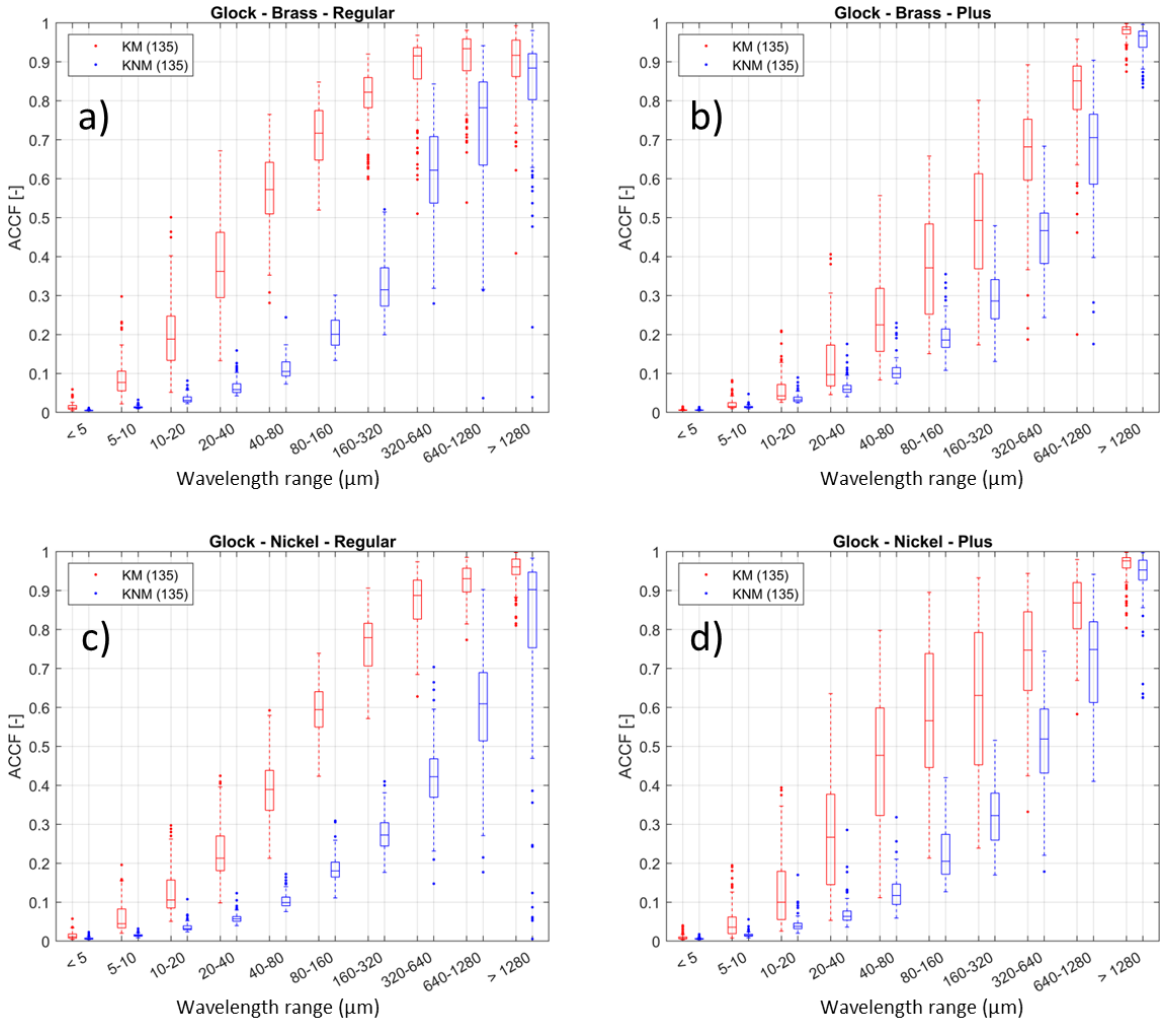


Fig. 12. Box plots of KM and KNM $ACCF_{MAX}$ scores for Glock 19 optically-measured test-fire reproducibility data. a) Brass-primer regular-pressure ammunition test-fire data, b) Brass-primer high-pressure ammunition test-fire data, c) Nickel-primer regular-pressure ammunition test-fire data, and d) Nickel-primer high-pressure ammunition test-fire.

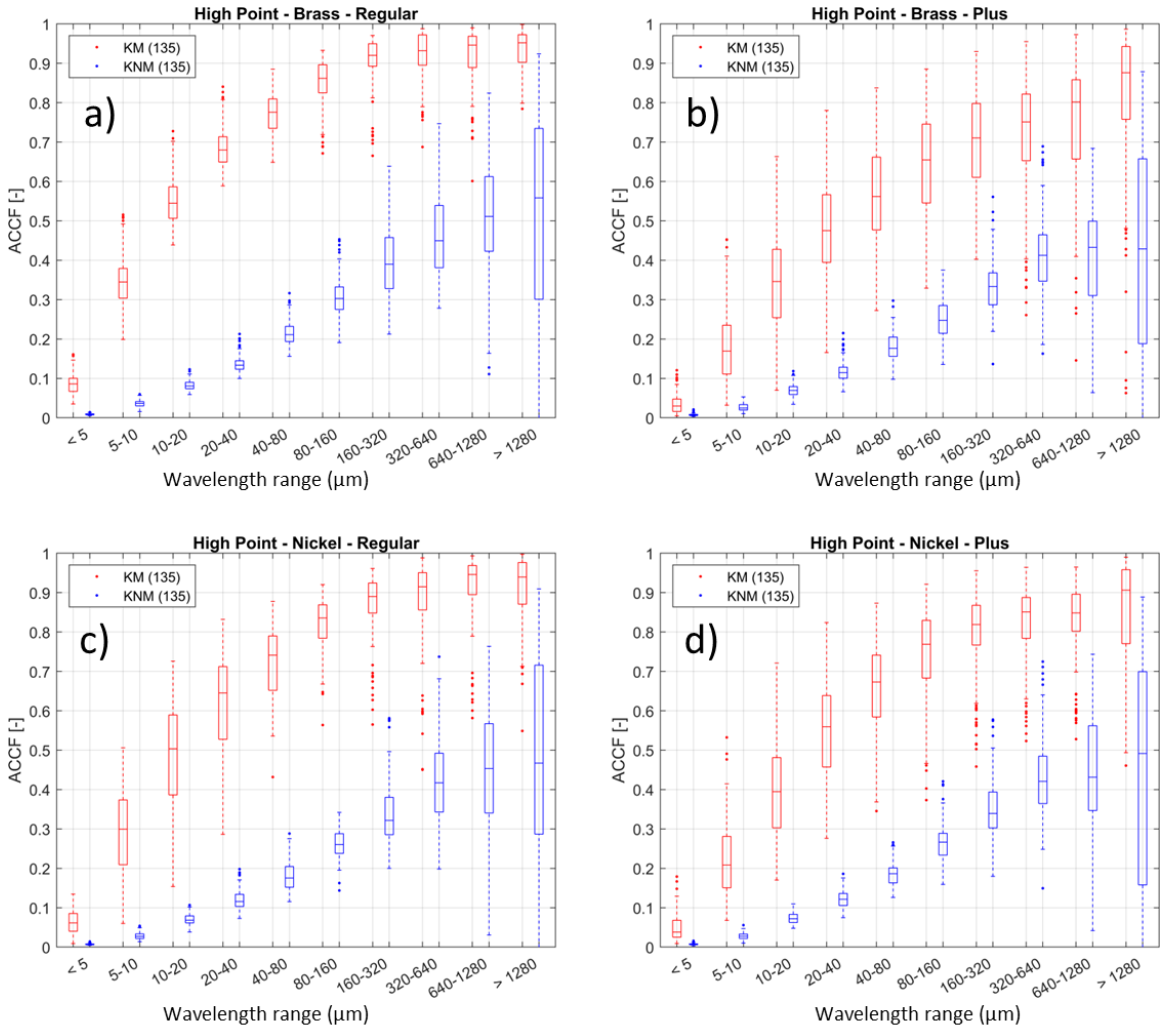


Fig. 13. Box plots of KM and KNM $ACCF_{MAX}$ scores for Hi-Point C9 optically-measured test-fire reproducibility data. a) Brass-primer regular-pressure ammunition test-fire data, b) Brass-primer high-pressure ammunition test-fire data, c) Nickel-primer regular-pressure ammunition test-fire data, and d) Nickel-primer high-pressure ammunition test-fire data.

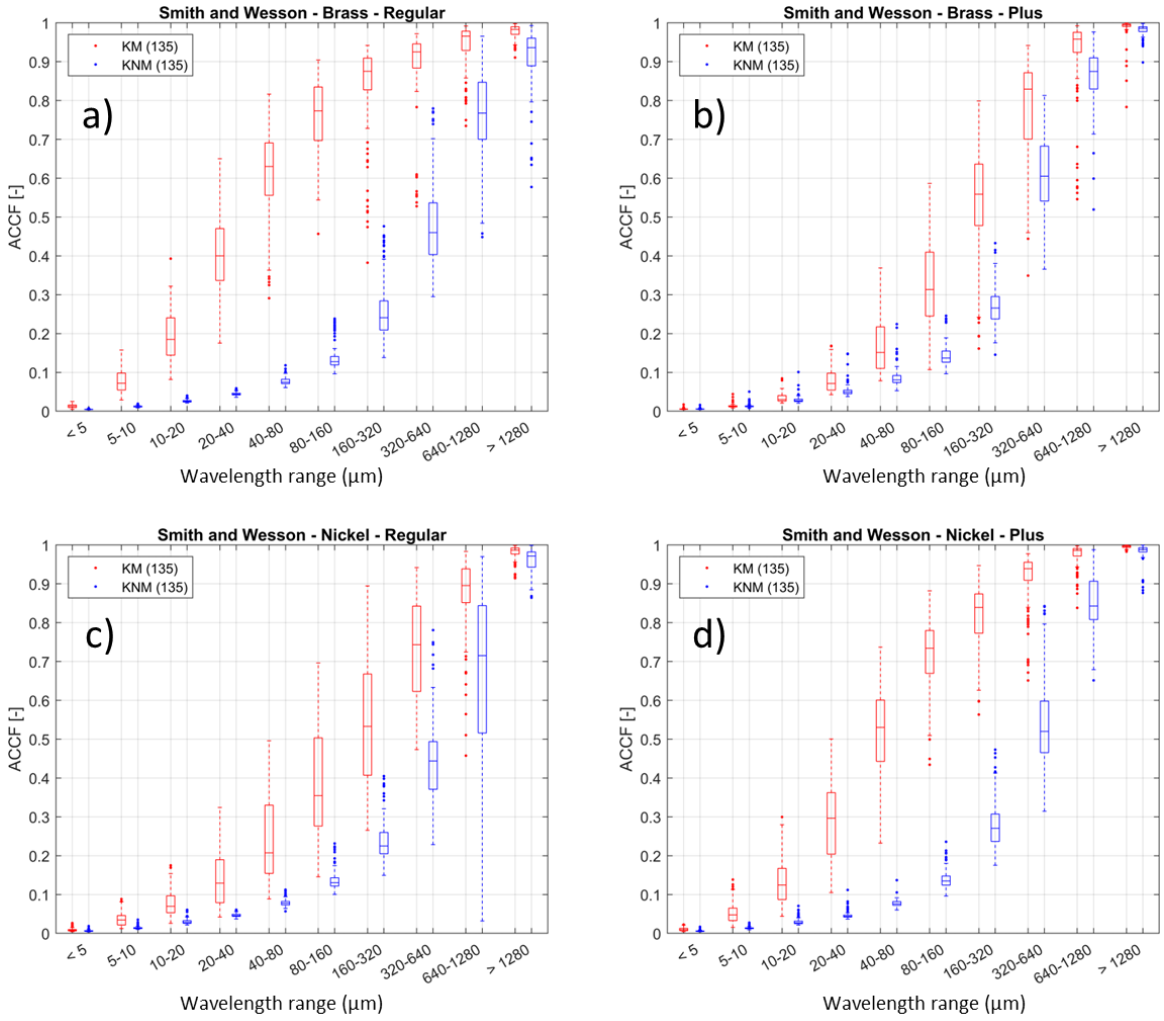


Fig. 14. Box plots of KM and KNM $ACCF_{MAX}$ scores for S&W MP9 optically-measured test-fire reproducibility data. a) Brass-primer regular-pressure ammunition test-fire data, b) Brass-primer high-pressure ammunition test-fire data, c) Nickel-primer regular-pressure ammunition test-fire data, and d) Nickel-primer high-pressure ammunition test-fire data.

5.2.4 KNM Scores at Higher Wavelengths

For the reproducibility discussion above, we focused on the KM scores in the box plots shown in Figs. 12, 13, and 14. The KNM distributions, however, exhibited a somewhat expected, but not previously measured trend. The KNM score distributions increased as a function of increasing wavelength without exception on all firearms and ammunition types. Typically, in pre-processing surface topography data for forensic toolmark evaluations, these large features are removed. As stated above, larger KNM scores imply the presence of sub-class features. What is interesting, and clearly demonstrated in the data, is that sub-class features are nearly non-existent at the smallest wavelengths, but then increase monotonically as wavelength increases.

5.2.5 AFM SSMD Data

AFM was employed to probe smaller wavelengths (Fig. 7(d)) than was possible with the optical microscope system. For the AFM analyses, three bandpasses were added, extending down to less than 1.25 µm. Surprisingly, the AFM test-fire SSMD values stayed above zero, even to the smallest bandpass, indicating the existence of individualizing features at these extremely small wavelength scales. Fig. 15 is a side-by-side combination of the AFM SSMD data from Fig. 7(d) and the Glock optically-measured test-fire uniqueness SSMD data from Fig. 7(a). Fig. 16 is the same, but for S&W. The side-by-side view enables visual comparison of the AFM and optically-measured test-fire uniqueness SSMD data. In general, the AFM data in Figs. 15 and 16 complemented the optical test-fire SSMD well.

In Fig. 15 (Glock SSMD data), the common bandpass (20 µm to 40 µm) is circled in red on both plots. The AFM SSMD value at that bandpass was a little under 2 while the optical SSMD value was a little over 2. From there, the Glock AFM SSMD data trends downward (similar to the Glock optical SSMD data), eventually stopping at a value between 0 and 1 at the smallest bandpass. The peak in the Glock AFM SSMD data (at the 10 µm to 20 µm bandpass) is peculiar and did not present in the Glock optical SSMD data. A possible explanation is that the optical system failed to measure the individualizing features at these wavelengths due to optical properties of the features (i.e., the topographic features weren't observable at visible illumination wavelengths). The AFM would not be sensitive to this type of issue, as it is a contact measurement system.

The agreement between the S&W AFM and optical SSMD data (Fig. 16) was equally as good as the Glock data. The same common bandpasses were circled in red. As with the Glock data, the S&W AFM SSMD value at that bandpass is a little under 2 while the S&W optical SSMD value is a little over 2. The S&W AFM data also trended downwards to a value somewhere between 0 and 1 at the smallest bandpass, but without the presence of a local peak.

The AFM SSMD data confirms the presence of individualizing features at wavelengths smaller than currently measured or analyzed in typical successful forensic toolmark evaluations based on 3D surface topography measurements. It's not to say they would not add value, but there is a two-fold burden to incorporate features of this scale in forensic toolmark evaluations: 1) the cost of equipment capable of resolving these features is much more expensive and more difficult to use than 3D optical measurement systems, and 2) processing data at much higher spatial resolutions are currently time-prohibitive.

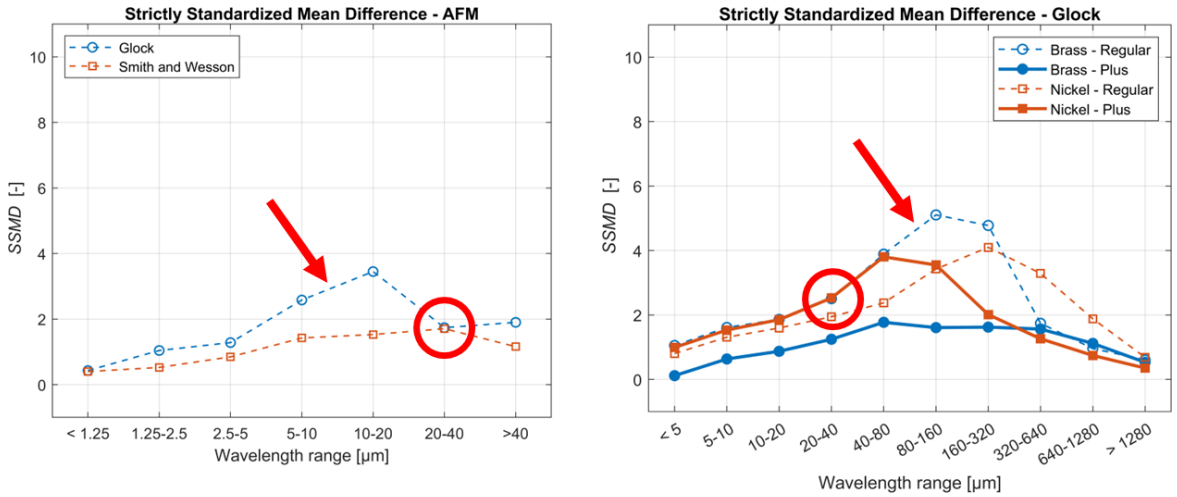


Fig. 15. Glock SSMD AFM-measured test-fire vs optically-measured test-fire comparison. a) AFM SSMD uniqueness data copied directly from Fig. 7(d) and b) Glock SSMD uniqueness data copied directly from Fig. 7(a). The common bandpass (20 µm to 40 µm) is circled in red. The red arrows indicate the brass-primer regular-pressure data to compare between AFM and optical.

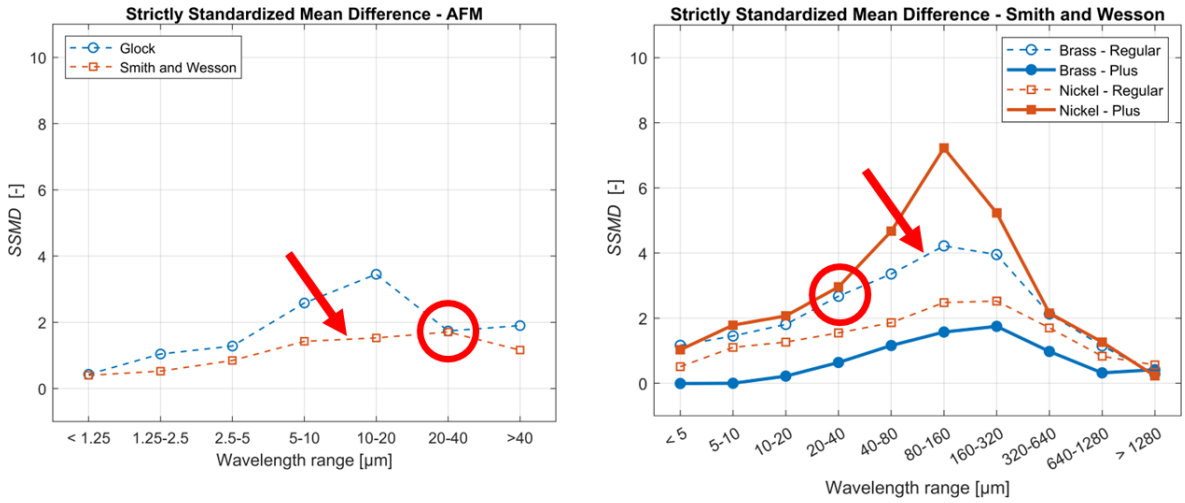


Fig. 16. S&W SSMD AFM-measured test-fire vs optically-measured test-fire comparison. a) AFM SSMD uniqueness data copied directly from Fig. 7(d) and b) S&W SSMD uniqueness data copied directly from Fig. 7(c). The common bandpass (20 µm to 40 µm) is circled in red. The red arrows indicate the brass-primer regular-pressure data to compare between AFM and optical.

5.2.5 Toolmark Transfer Fidelity

The cast SSMD data in Figs. 9, 10, and 11 demonstrated that all three firearm models contained individualizing features that peaked between 20 µm and 80 µm. The cast SSMD peak values were approximately 11, 29, and 65 for the Glock, Hi-Point, and S&W respectively. The peaks for the optically-measured test-fire SSMD data were shifted to the right (relative to the cast data), with most of the peaks occurring between 40 µm and 320 µm, revealing that many of the smaller wavelength features physically present on the firearm breechface do not survive the firing process. By looking for the cast SSMD scores that matched as closely as possible with the test-fire SSMD data (in a given bandpass), one can generally state the size of features that have the highest probability of surviving the firing process.

For the Glock firearm, the bandpass that matched most closely was the 320 µm to 640 µm. A larger band of wavelengths seem to survive for the Hi-Point, ranging from 80 µm to 1280 µm. The shorter band (80 µm to 160 µm) matched closely, as did the larger band (640 µm to 1280 µm). However, the Hi-Point cast data showed lower SSMD values than the optically-measured test-fire SSMD data for the two in-between bands (160 µm to 320 µm, 320 µm to 640 µm). The mechanical properties of the cured casting material could explain this effect. When cured, the material is still flexible, possibly allowing these larger features to be distorted from cast to cast, while the test-fires might reconstruct these larger features more faithfully. The S&W showed the strongest presence of smaller individualizing features (ranging from 10 µm to 160 µm). The features that survived the S&W firing process were again in the 320 µm to 640 µm band.

5.3 Limitations

This study focused on three common firearm systems and four ammunition types frequently seen in casework. Conclusions reached regarding the firearms and ammunition used here cannot predict how other firearm systems and ammunitions will behave. Future work will involve conducting the same type of study on other firearm systems and ammunition types.

Ideally, all ammunition used in this project would have been from the same manufacturer. Due to ammunition shortages at the beginning of the project, the four different ammunition types were from different vendors. Observations made and conclusions reached should consider this factor.

Image pair correlations were not statistically independent. One image was always used in at least two image correlations, sometimes more. Therefore, one bad image could produce several bad data points in the KM and KNM distributions. These bad data points could skew the SSMD values derived from these distributions.

Hardness is a material property that describes its resistance to deformation, scratches, and indentation. Hardness of the selected primer materials (brass and nickel-plated brass) have the potential to influence the analysis of uniqueness, reproducibility, and toolmark transfer reported above. While hardness measurements of the primer materials were initially planned,

it was decided to not include them in this study due to the very thin primer surface thickness that complicated accurate measurement of the property.

5.4 Conclusion

A comprehensive study of breechface impression toolmark uniqueness, reproducibility, and transfer fidelity was conducted. We demonstrated the existence of small individualizing breechface impression features (less than 15 µm) that are routinely filtered out as noise. Brass-primer regular-pressure ammunition provided the best reproducibility. Current surface topography measurement pre-processing cutoff filters encompass the majority of individualizing features found on these three firearm models. Increasing KNM was observed with increasing wavelength, indicating the presence of subclass features increases with wavelength.

6.0 Artifacts

6.1 List of Products

The project required the full two years to generate, measure, and analyze toolmark samples. It did not lend itself to presenting or publishing small accomplishments midstream. Now that we have finalized the data and analyses, we will be presenting and publishing on the findings at multiple forensics conferences and journals.

6.2 Data sets generated

6.2.1 Optical surface topography data

Raw Zygo Nexview NX2 data (.datx) was generated for all test-fires and casts. These were also converted to x3p format and used in the various analyses. Both the datx and x3p files are archived.

6.2.2 AFM Data

Raw AFM data (.dt2) was generated for the AFM-specific test-fires. These were also converted to x3p format and used in various analyses. Both the raw AFM and x3p files are archived.

6.2.3 Test-fires and casts

All physical test-fires and casts generated will be stored and protected from degradation for possible future measurement and evaluation.

6.2.4 Custom correlation software

A custom version of our correlation software was developed and used to process the high-resolution data produced in the study. The similarity of compared images for each wavelength band was evaluated using the $ACCF_{MAX}$. A critical step in the calculation of this similarity score is the registration of the compared images, i.e., the estimation of their relative lateral position

and orientation. For this study, we registered the images by maximizing the $ACCF_{MAX}$ value of the compared band-pass filtered images. Here the band-pass filter was re-applied to the unfiltered image data after registration, using only the overlapping image areas. This approach avoids image values at non-overlapping areas affecting the similarity score. A major challenge was the registration of large, high-resolution, images at small wavelength bands based only on $ACCF_{MAX}$ optimization. We developed modifications to existing image registration algorithms to address: 1) the need for high registration resolution in both position and orientation, 2) the presence of many local optima in the $ACCF_{MAX}$ similarity score, in particular for known non-matching images, and 3) the desired independence of registration solutions for low-frequency image data and high-frequency image data; this to reduce artificial correlations in registration solutions between wavelength bands. The latter two issues limited, to some extent, the application of coarse-to-fine registration approaches, in particular for known non-matching images. We developed and refined a registration procedure that was independently applied to each wavelength band. This code is archived for future use.

References

1. Association of Firearm and Toolmark Examiners, Glossary, 6th edition, version 6.101613, 2021
2. The National Research Council, Strengthening Forensic Science in the United States — A Path Forward, NRC, Washington, DC, 2009 pp. 153–154, 184, 155.
3. Baldwin, D.P., Bajic, S.J., Morris, M., and D. Zamzow. "A study of false-positive and false-negative error rates in cartridge case comparisons." Ames Laboratory, USDOE, Technical Report #IS-5207 (2014) afte.org/uploads/documents/swggun-false-postive-false-negative-usdoe.pdf
4. Fadul, T.G., Hernandez, G.A., Stoiloff, S., and S. Gulati. "An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing consecutively manufactured Glock EBIS barrels with the same EBIS pattern." National Institute of Justice Grant #2010-DN-BX-K269, December 2013. www.ncjrs.gov/pdffiles1/nij/grants/244232.pdf.
5. National Research Council, Ballistic Imaging, *The National Academies Press*, Washington, DC, 2008.
6. Vorburger T.V., et al, "Surface Topography Analysis for a Feasibility Assessment of a National Ballistics Imaging Database," NIST Interagency Report NISTIR 7362, National Institute of Standards and Technology, 2007.
7. De Kinder J, Tulleners F, Thiebaut H, "Reference Ballistic Imaging Database Performance, *Forensic Science International* **140** 207 (2004)
8. Vorburger T.V., "Applications of Cross-Correlation Functions," *Wear*, **271** 529-533 (2010)
9. Chu W, et al, "Pilot Study of Automated Bullet Signature Identification Based on Topography Measurements and Correlations," *Journal of Forensic Sciences*, **55** 341-347 (2010)
10. Song J, et al, "Optimizing gaussian filter long wavelength cutoff for improving 3D ballistics signature correlations," *Proceedings of ASPE*, **23** (2008)
11. W. Chu, M. Tong, J. Song, Validation tests for the congruent matching cells (CMC) method using cartridge cases fired with consecutively manufactured pistol slides, AFTE J. 45 (4) (2013) 361–366.
12. J. Song, Proposed "Congruent matching cells (CMC)" method for ballistic identification and error rate estimation, AFTE J. 47 (3) (2015) 177–185.
13. Song, John, et al. "Estimating error rates for firearm evidence identifications in forensic science." *Forensic science international* 284 (2018): 15-32.

14. Stocker, Michael T., et al. "Addressing Quality Assurance Issues in 3D Firearm and Toolmark Imaging| NIST." *Journal of the Association of Firearms and Toolmarks Examiners*50.Journal of the Association of Firearms and Toolmarks Examiners (2018).

15. OSAC, "Resolution Requirements for 3D Virtual Comparison Microscopy", 2016, PDF file, Accessed April 9, 2019, https://www.nist.gov/sites/default/files/documents/2018/05/08/rd_needs_-_resolution_requirements_for_3d_virtual_comparison_microscopy.pdf

16. H. Katterwe, R. Goebel and K.D. Gross, "The Comparison Scanning Electron Microscope First Experiments in Forensic Application" (Reprint from 1980), *AFTE Journal, 1983 Volume 15, Number 2 (April), Page 47 thru 55,*

17. Michael D. Scanlan and Andrew D. Reinholz, U.S. Fish and Wildlife National Forensics Laboratory, Ashland, Oregon*; "*Scanning Electron Microscopy for Firearm and Toolmark Comparisons", AFTE Journal, Volume 45 Number 1 -- Winter 2013

18. Zheng, A., etal, "Modeling Firearm Toolmark Persistence through 3D Measurements and Analysis", presented at the 47th Annual Training Seminar of the Association of Firearm and Toolmark Examiners, New Orleans, LA, May 2016

19. Hayes, C. S., M. Basoa, and R. Freese. "Reduction of characteristic breechface marks due to primer sealants." *AFTE JOURNAL* 36 (2004): 139-146.