# Estimating Error Rates for Firearm Evidence Identifications in Forensic Science

John Song[1], Theodore V. Vorburger[1], Wei Chu[1], James Yen[2],
Johannes A. Soons[1], Daniel B. Ott[1], Nien Fan Zhang[2]
[1]Engineering Physics Division
[2]Staistical Engineering Division
National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, U.S.A.

**Abstract:**
Estimating error rates for firearm evidence identification is a fundamental challenge in forensic science. This paper describes the recently developed Congruent Matching Cells (CMC) method for image comparisons, its application to firearm evidence identification, and its usage and initial tests for error rate estimation. The CMC method divides compared topography images into correlation cells. Four identification parameters are defined for quantifying both the topography similarity of the correlated cell pairs and the pattern congruency of the registered cell locations. An identification (declared match) requires a significant number of CMCs, i.e., cell pairs that meet all similarity and congruency requirements. Initial testing on breech face impressions of a set of 40 cartridge cases fired with consecutively manufactured pistol slides showed wide separation between the distributions of CMC numbers observed for known matching and known non-matching image pairs. Another test on 95 cartridge cases from a different set of slides manufactured using the same process also yielded widely separated distributions. The test results were used to develop two statistical models for the probability mass function of CMC comparison scores. The models were applied to develop a framework for estimating cumulative false positive and false negative error rates and individual error rates of identifications and exclusions for this population of breech face impressions. The CMC method can provide a statistical foundation for estimating error rates in firearm evidence identifications, thus emulating methods used for forensic identification of DNA evidence.

**Keywords:** Forensics, ballistics identification, error rate, congruent matching cell, CMC.

## 1. Introduction

Tool marks are permanent changes in the topography of a surface created by forced contact with a harder object (the tool). When bullets and cartridge cases are fired or ejected from a firearm, the parts of the firearm that make forcible contact with them create characteristic tool marks called "ballistic signatures" [1]. By examining these ballistic signatures side-by-side in a comparison microscope, firearm examiners can determine whether a pair of bullets or cartridge cases was fired or ejected from the same firearm. Firearm examiners can then connect a recovered firearm or other firearm evidence to criminal acts.

Successful identification requires that the relevant firearm surfaces have individuality and that the tool marks are reproducible [1]. In general, tool marks have so-called "class characteristics" that are common to certain firearm designs and manufacturing methods, and "individual

characteristics" arising from random variations in firearm manufacturing and wear [1]. While class characteristics can be used to exclude a firearm as a source of a recovered cartridge case or bullet, only the individual characteristics, which are unique for individual firearms, can form the basis for identification [1].

Side-by-side tool mark image comparisons for firearm identification have more than a hundred-year history [1]. However, the scientific foundation of firearm and tool mark identification has been challenged by recent reports and court decisions. As stated in a 2008 National Academies Report [2], "*The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related tool marks has not yet been fully demonstrated.*" and "*Since the basis of all forensic identification is probability theory, examiners can never really assert a conclusion of an 'identification to exclusion of all others in the world,' but at best can only assert a very small (objective or subjective) probability of a coincidental match.*"

The legal standard for the acceptance of scientific evidence contained in the U.S. Supreme Court decision, called the *Daubert* standard [2], "*places high probative weight on quantifiable evidence that can be tested empirically and for which known or potential error rates may be estimated, such as identification using DNA markers*"[2]. However, as stated in a 2009 National Academies Report [3], "*But even with more training and experience using newer techniques, the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates.*"

Since the 1980's, estimates of Coincidental Match Probability (CMP) have been used for specifying uncertainty of DNA identifications [3]. "*The courts already have proven their ability to deal with some degree of uncertainty in individualizations, as demonstrated by the successful use of DNA analysis (with its small, but nonzero, error rate)*" [3]. It is therefore a fundamental challenge in forensic science to establish a scientific foundation and statistical procedures providing quantitative error rate reports to support firearm identifications, in the same way that reporting procedures have been established for forensic identification of DNA evidence [3]. A number of experimental and theoretical efforts have been pursued along this line including the computer learning approach of Petraco et al. [4,5], the work on likelihood ratio by Riva and Champod [6], the study of examiner error rates by Baldwin et al. [7], the feature-based matching algorithm of Lilien et al. [8,9], and the work on image cross correlation and congruent matching cells (CMC) of Song et al. [10-15].

In this paper we apply the CMC method of Song et al [12-14] to estimations of error rates for false identifications and exclusions for two sets of topography image data of breech face impressions from fired cartridge cases. We review the CMC method in Section 2, then describe validation tests, error rate estimation procedures and initial results in Sections 3 to 5, and provide some concluding observations in Section 6.

## 2. Congruent Matching Cells (CMC) Method

We begin with pairs of measured 3D topography images of breech face impressions whose similarity we wish to quantify (see Fig. 1). A common approach would be to calculate the value of the normalized cross-correlation function (Pearson's correlation coefficient) for the pair of images as a whole [10,11] when they are registered at a position of maximum correlation. Instead,

the CMC method divides the reference image into a rectangular array of cells as shown in Fig. 2. For each cell on the reference image, a search is made on a compared image for a matching cell-sized area. The cell-by-cell analysis is done because a firearm often produces characteristic marks, or individual characteristics [1], on only a portion of the bullet or cartridge case surface. Carrying over the terminology from our previous research in firearms identification [12,13], a region of the surface topography is termed a "valid correlation region" if it contains individual characteristics of the ballistic signature that can be used effectively for firearm identification. Conversely, a region of the surface topography that does not contain individual characteristics of the firearm's ballistic signature is termed an "invalid correlation region" and should be eliminated from consideration for firearm identification. Invalid correlation areas can occur, for example, due to insufficient contact between the firearm surface and the bullet or cartridge case during firing.

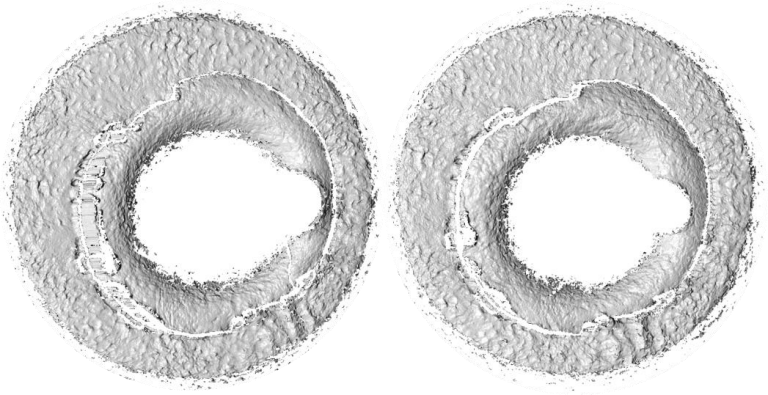

**Fig. 1.** Topography images of breech face impressions obtained from a pair of cartridge cases ejected from slide 3 in the Fadul data set [16] discussed here. The images have a number of features in common. The diameter of each image is about 3.5 mm. The topography contrast is rendered with a virtual light source from the left.

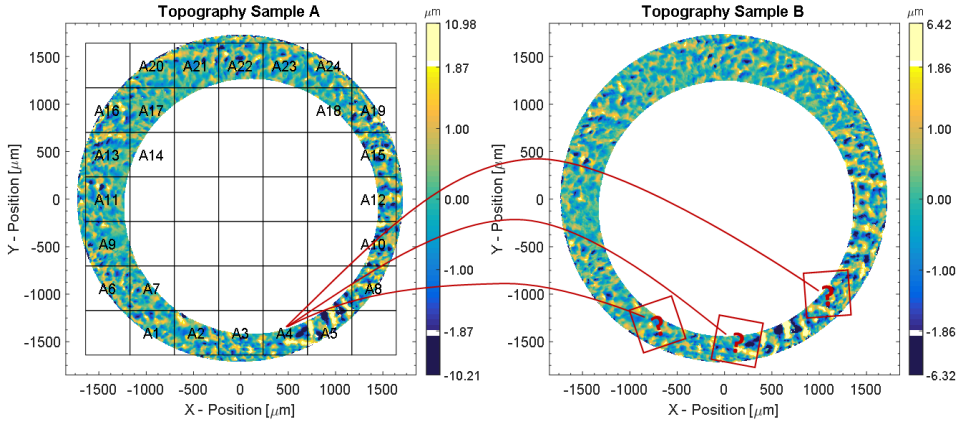


**Fig. 2.** Conceptual diagram of a topography image from Fig. 1 overlaid by a 7 ×7 grid, dividing the image into cells. The drag mark at the 3 o'clock position in Fig. 1 and the central hole and surrounding bulge from the firing pin impression are masked out of the images before the cell division step. Only cells with a sufficient fraction of measured pixels are used for the correlation analysis. Also shown is a schematic diagram of the search procedure to find an area in the compared (right) image that has a strong correlation with one of the cells in the reference (left) image. Here the topography is represented by a color map.

Assuming that two ballistic topographies A and B originate from the same firearm, both will likely contain valid and invalid correlation regions. When A and B are compared with each other, their common valid correlation region is the overlap of the individual valid correlation regions of A and B, which comprise only part, sometimes even a small part, of the entire areas of A and B. If a quantitative measure of correlation is obtained from the entire images of A and B, the correlation accuracy may be relatively low because large invalid regions may be included in the correlation. If the correlation areas are divided into cells, the valid correlation regions can be more likely identified and the invalid correlation regions can be eliminated from correlations. This procedure can significantly increase the correlation effectiveness and accuracy. Furthermore, the use of a statistically large number of congruently matched cells identified by multiple parameters can facilitate the estimation of an error rate from a well characterized population [13].

A correlation cell is a rectangular sub-region of the surface topography image that contains a sufficient quantity of distinguishing peaks, valleys, and other topographic features so that an assessment of topography similarity can be made. If topographies A and B originating from the same firearm are registered at their position of maximum correlation (Fig. 3), the cell pairs located in their common valid correlation regions can be identified, as shown by the solid cell pairs located in ($A_1$, $B_1$), ($A_2$, $B_2$), and ($A_3$, $B_3$). These cell pairs are necessarily characterized by [12,13]:

1) High pairwise topography similarity as quantified by a high value of the normalized cross correlation function maximum $CCF_{max}$;
2) Similar registration angles $\theta$ for all correlated cell pairs in regions A and B; and
3) "Congruent" *x*-*y* spatial distribution patterns for the correlated cell arrays ($A_1$, $A_2$, $A_3$…) and ($B_1$, $B_2$, $B_3$…) or nearly so.

On the other hand, if the registered cell pairs are located in the invalid correlation regions of A and B, such as the dotted cells (*a'*, *a''*, *a'''*) and (*b'*, *b''*, *b'''*) in Fig. 3, or if they originate from different firearms, their maximum cross correlation value $CCF_{max}$ would be relatively low, and their cell arrays would show significant variation in their *x*-*y* distribution patterns and registration angles $\theta$.

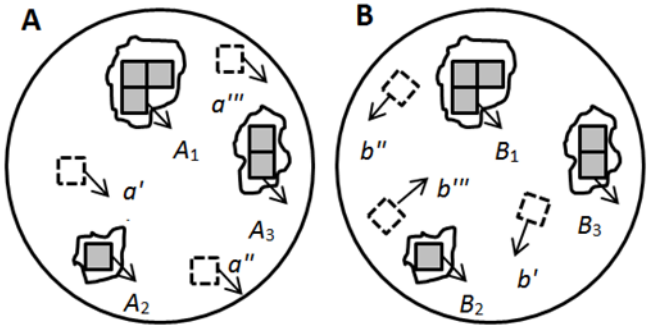


**Fig. 3.** Schematic diagram of topographies A and B originating from the same firearm and registered at the position of maximum correlation. The six solid cell pairs in each image are located in three valid correlated regions ($A_1$, $B_1$), ($A_2$, $B_2$), and ($A_3$, $B_3$). The dotted cell pairs (*a'*, *b'*), (*a''*, *b''*), and (*a'''*, *b'''*) are located in the invalid correlation region.

Congruent matching cell pairs, or CMCs, are therefore determined by three sets of identification parameters for quantifying both the topography similarity of the correlated cell pairs and the pattern congruency of the cell distributions. The former is quantified by the normalized cross correlation function maximum $CCF_{max}$ with threshold $T_{CCF}$; the latter is quantified by the registration angle $\theta$ and translation distances in $x$ and $y$ with corresponding thresholds $T_{\theta}$, $T_{x}$, and $T_{y}$. A correlated cell pair is considered a CMC—that is, part of a congruent matching cell pattern—when its correlation value $CCF_{max}$ is greater than a chosen $T_{CCF}$ and the registration angle $\theta$ and $x$, $y$ registration positions are within the chosen $T_{\theta}$, $T_{x}$ and $T_{y}$.

How many CMC pairs are required so that the two surface topographies can be identified as matching? Ideally, this would be determined after carefully designed experiments and error rate estimations. As a starting point, we use a numerical identification criterion $C$ equal to 6, taking a lead from the method of Consecutively Matching Striae (CMS) developed by Biasotti and Murdock [17] for identification of bullet striation signatures. This numerical criterion and objective method have been used internationally for bullet and striated tool mark identifications since 1984. Thus, when the number of CMC pairs of the compared topographies A and B is equal to or greater than $C = 6$, A and B are concluded to be a match. We demonstrate that this criterion works well for the tests that we present in the following sections. However, when applying similar algorithms to other tool marks, such as firing pin impressions [18], the identification criterion will depend on a number of factors such as the size of the tool mark, the total number of evaluated cells, and whether the tool marks are striated or impressed. For each type of population, $C$ would be determined from the population statistics and from estimated targets for error rates. Calculation of error rates for a specific data set is discussed below.

## 3. Validation Tests

An initial validation test of the CMC method was conducted using a set of cartridge cases originally created by Fadul et al. at the Miami-Dade Crime Laboratory [16] for a study of visual firearm identifications by ballistics examiners. The set contains 40 cartridge cases ejected from handguns with ten consecutively manufactured Ruger 9 mm pistol slides. Three slides were used to fire three cartridge cases each, four slides were used to fire four cartridge cases each, and three slides were used to fire five cartridge cases each. The slide is a component of a semi-automatic pistol firing mechanism that absorbs the recoil impact of the cartridge case on its breech face. As a result, the surface topography of the slide's breech face is impressed on the soft primer of the cartridge case upon impact.

Comparisons involving a population of consecutively manufactured firearm parts represent a challenging scenario for accurately identifying bullets or cartridge cases as being fired or ejected from the same firearm. Consecutively manufactured parts can have similar topographic features arising from temporary imperfections in the manufacturing process, such as a worn tool. The presence of these *sub-class characteristics* can lead to false identifications [1]. For this studied set, the breech face was machined using a straight pull step broach [16]. However, the manufacturer finished the surfaces of the slides by sand and bead blasting, a process that should produce random surface topographies [19] with individual characteristics and mitigate the effect of sub-class characteristics. The task then for topography measurement and analysis is to distinguish the individual characteristics of the surface impressions from any underlying similarities in consecutively manufactured slides resulting from earlier phases of the manufacturing process. The

objective is to draw a correct conclusion of identification or exclusion with error rate estimation for any pair of topography images drawn from the 40 cartridge cases that were measured.

The breech face impression topographies on the cartridge cases were measured by a disk scanning confocal microscope [20] operating with a 10× objective having a numerical aperture of 0.3, a nominal working distance of approximately 10.1 mm, and a single field of view of approximately 1.6 mm × 1.6 mm, comprising 512 × 512 pixels. The topography images of the entire breech face impressions were achieved by stitching 3×3 fields of view and were approximately 3.9 mm × 3.9 mm with approximately 1240 × 1240 pixels and a nominal pixel spacing of 3.125 µm. The images were down sampled to a pixel spacing of 6.25 µm to improve the speed of the subsequent image correlations. The root mean square instrument noise was approximately 7 nm as tested by measuring an optical flat at 10× with a long wavelength cutoff of 250 µm on a number of occasions.

Before correlating, the images were trimmed to extract the breech face impression of interest, yielding, on average, an image size of 3.5 mm × 3.5 mm. The images were then bandpass filtered to attenuate noise with short spatial wavelengths and attenuate surface form and waviness with long wavelengths thus highlighting individual characteristics. The short wavelength cutoff of the Gaussian filter was 16 µm, and the long wavelength cutoff was 250 µm. Figure 4 shows a topography image of a breech face impression before and after trimming and filtering.

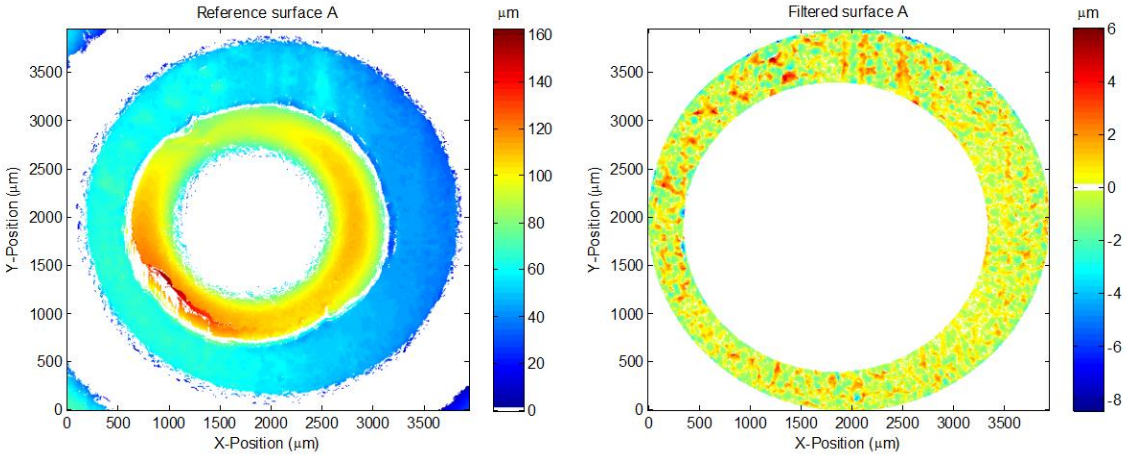


**Fig. 4.** Color coded topography image of one of the breech face impressions before (left) and after (right) trimming, leveling, and filtering. The prominent annular ridge on the left hand image is due to flow back into the firing pin aperture. This feature is trimmed away in the right hand image.

The topography images were correlated using the CMC method. A total of 780 (= 40×39/2) image correlations were performed, comprising 63 (= 3×3 + 4×6 + 3×10) known matching (KM) and 717 (= 780 – 63) known non-matching (KNM) image pair comparisons. The topography images were divided into arrays of 49 (= 7×7) cells. Each cell size for the set of correlation tests was chosen to be 75×75 pixels (nominally 468.75 µm × 468.75 µm), and the range of cell registration angles was restricted to ± 30° with respect to their initial orientation.

Although the nominal number, $N_{\text{nom}}$, of compared cell pairs for each topography correlation equals 49, the actual number $N$ of effective cell pairs for each correlation depends on the number of cells in the reference image that contain a sufficient number of measured pixels for effective correlation.

For example, the empty center portion of the surface shown in Fig. 4, corresponding to the firing pin impression, leads to fewer effective correlation cells than $N_{nom}$. A cell was not used unless at least 10 % of its pixels represented measured data. For this study, the number of evaluated cell pairs in a comparison ranged from 24 to 30, with an average of 26.

One set of validation test results is shown in Fig. 5 [14]. The cell size $a$, the pixel spacing, and the thresholds $T_{CCF}$, $T_\theta$, $T_x$ and $T_y$ are shown on the upper right side. The number of congruent matching cell pairs (CMCs) for the 63 KM topography pairs ranges from 9 to 26; while the number of CMCs for the 717 KNM topography pairs ranges from 0 to 2.

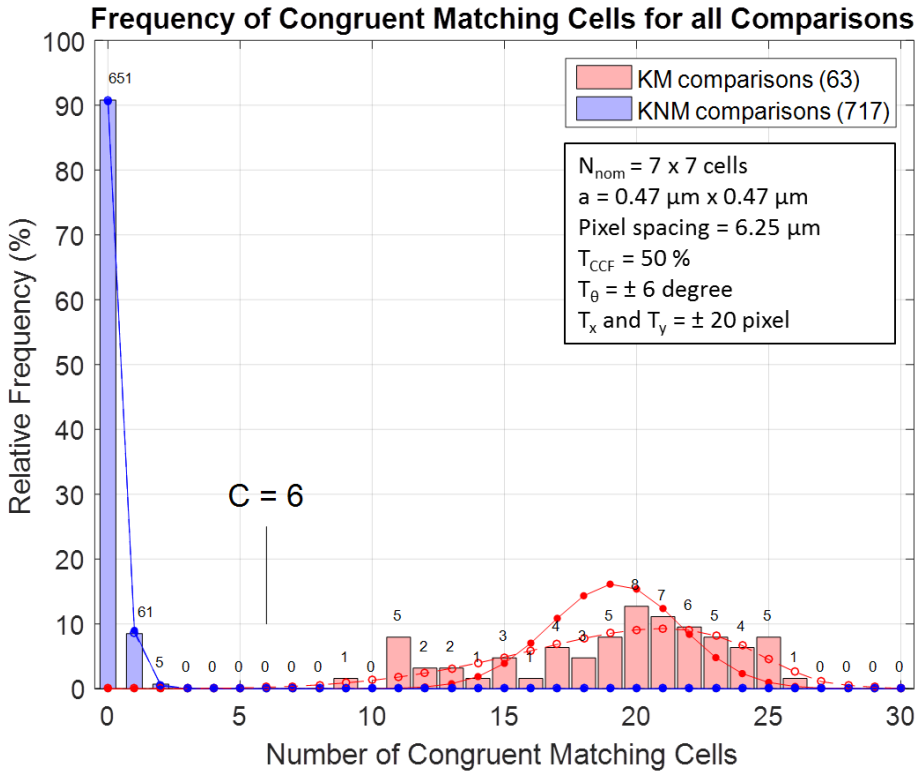


**Fig. 5.** Relative frequency distribution of image pairs vs. CMC number for 63 KM and 717 KNM image pairs. The KM and KNM distributions are each scaled to their particular sample size. The solid blue and solid red curves represent binomial distribution models for the KNM and KM data, respectively, estimated from the data by Eq. 8 and Eq. 10, respectively. The dashed blue and red curves, the first of which is indistinguishable from the solid blue curve, represent the respective beta-binomial distribution models (see Section 5.1). Note that the distribution models are discrete histograms, with the connecting lines drawn for visualization. The number of image pairs having a particular CMC value is shown just above each bar in the histograms.

Of the 717 KNM topography pairs, 651 pairs have CMC = 0 (no congruent matching cells). There are only five non-matching topography pairs that have as many as two congruent matching cells, i.e. CMC = 2 (Fig. 5); one of them is shown in Fig. 6A. For the 63 KM topography pairs, only one topography pair has a CMC number as low as 9. This topography pair is shown in Fig. 6B. All the other KM topography pairs have a CMC number ranging from 11 to 26 (Fig. 5). A close-up of one

pair of matching cells from Fig. 6B, cell A1 vs. cell B1, is shown in Fig. 7. Their topography similarity is quantified by the maximum value of the normalized cross-correlation function $CCF_{max}$ = 67.6 %.

The KM and KNM distributions of Fig. 5 show a significant separation. Additional tests using slightly different versions of the correlation software and different parameter values show similar results without an overlap [14]. Tests performed with optical intensity images of the breech face impressions, instead of 3D topography images, also show similar results without any false identifications or false exclusions [15]. In standard binary classifier terms, these results indicate both high sensitivity and specificity [21] and suggest the feasibility of using the CMC method for firearm identifications. The identification accuracy can likely be improved further by designed experiments to optimize the image processing, cell size, parameter threshold values, and registration intervals.

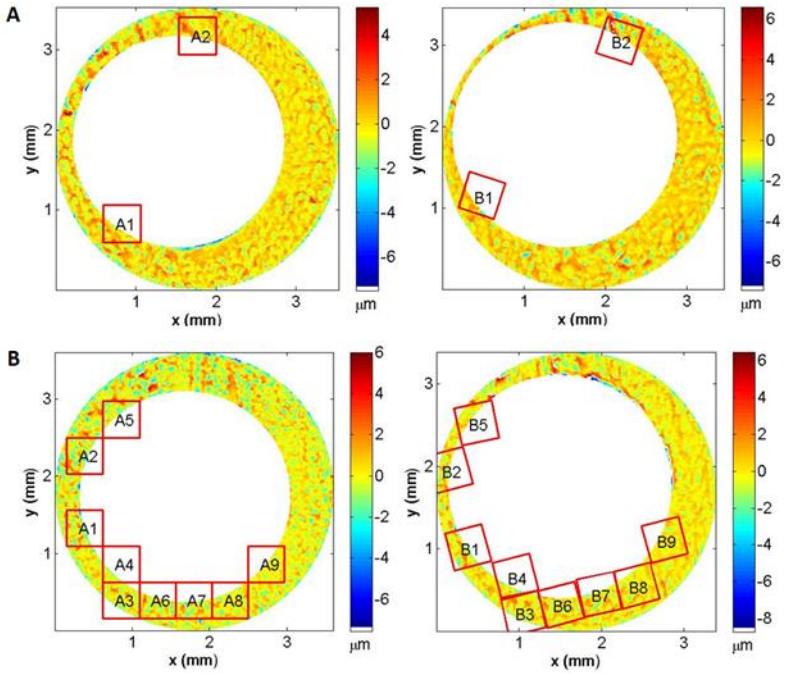


**Fig. 6.** Depiction of congruent matching cells for two sets of correlated topography pairs. For the 717 KNM topography pairs, only five pairs have a CMC value as high as 2; one of these image pairs is shown in (A). For the 63 KM topography pairs, only one has a CMC value as low as 9; that pair is shown in (B). The cell pattern A1-A9 on the left of Fig. 6B is congruent with the cell pattern B1-B9 on the right. The filtered surface topographies of the breech face impressions are depicted by the color scale of the diagram.

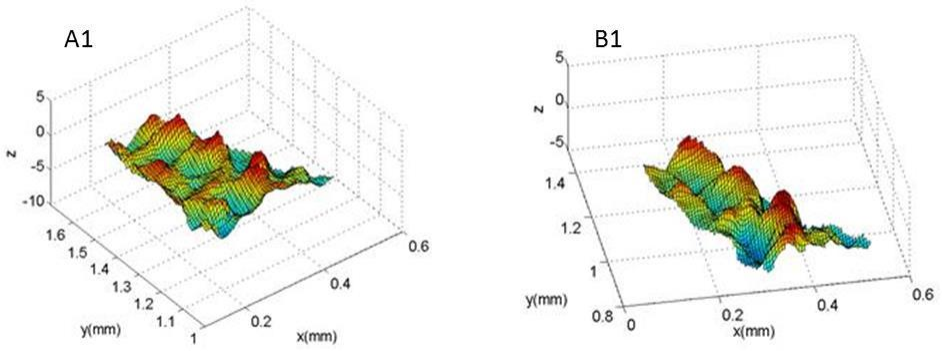


**Fig. 7.** Topography comparison of KM cell pair A1 and B1 from the KM image pair of Fig. 6B. Common topography features are apparent. The normalized correlation value, $CCF_{max}$, is 67.6 %.

## 4. Error Rate Analysis

### 4.1 A Statistical Framework

We seek to develop an approach for estimating the expected error rates of ballistic identifications based on the CMC method. Generally speaking, error rates can be considered from two points of view [22,23]. The first point of view (or class) addresses the reliability of the identification procedure. This reliability is specified by the false positive and false negative error rates for a given set of KM and KNM samples. The false positive error rate represents the expected frequency or probability of obtaining an erroneous result of identification (declared match) when comparing samples from different sources (KNM). The false negative error rate represents the probability of obtaining an erroneous result of exclusion (declared non-match) when comparing samples from the same source (KM). The false positive and false negative error rates can be used as a measure of the reliability of the identification procedure.

The second point of view addresses the probability of an incorrect conclusion for an identification (declared match) or exclusion (declared non-match). It represents the expected frequency or error rate that a result of either identification or exclusion is false. This way of describing error rate is, in general, of great interest during legal proceedings. For example, when a firearms examiner concludes that the evidence and reference items are from the same source, the judge may ask: "What is the probability that these two items are actually from different sources?" However, error rates in this class depend not only on the reliability of the identification procedure, but also on the ratio of same-source image pairs to different-source image pairs in the population of comparisons relevant to the case [6] or (for this paper) relevant to the validation test (see Sec. 4.5).

Another way to describe this second point of view is with a Bayesian approach, where the ratio of same-source to different-source populations is cast as *prior odds*. Multiplying this factor by the *likelihood ratio* [24,25] yields *posterior odds*, say, for an identification being correct. The likelihood ratio is the ratio of the probabilities of obtaining a specific comparison result, say a declared match, under the competing hypotheses of same source and different source samples. Thus, the likelihood ratio expresses the strength of the obtained evidence irrespective of the prior odds. It can be calculated from data and models such as those in Fig. 5.

Aside from the points of view discussed above, error rate estimations can also be defined either for the cumulative range of comparison scores associated with a conclusion of identification or exclusion, or defined for individual comparison scores of interest. In this paper we calculate both types of results from the distributions obtained with the CMC method. Thus, for example, the cumulative false positive error rate represents the probability of obtaining a CMC score larger than or equal to $C$, the identification criterion, when comparing samples from different sources (KNM). Alternatively, for a specific CMC comparison score, we calculate individual error rates of identifications and exclusions. Thus, for example, when CMC = 15, the individual identification error rate represents the probability that an identification based on a CMC score of 15 is a falsely declared match.

The large number of cell correlations associated with the CMC method using multiple identification parameters facilitates a statistical approach to modeling error rates. The CMC method is based on pass-or-fail tests of individual cell pairs comprising an image pair of breech face impressions. In this section we develop statistical models for the probability distribution of the number of successful tests in a comparison, i.e., the CMC numbers of KM and KNM comparisons. After estimating model parameters from experimental results for KM and KNM comparisons, the models are applied to estimate potential error rates.

### 4.2 A Binomial Probability Model for the Distribution of CMCs

For a pair of images, $N$ represents the number of correlated cell pairs. If, for example, there are 49 cells in the array of the reference image, but nine of those have an insufficient fraction of pixels with measurement values, then $N$ is reduced to 40. For a given correlated cell pair, a random variable $X$ represents the outcome of the CMC method for that cell pair. When the CMC method determines that the cell pair is part of the set of congruent matching cells, i.e. when its correlation value $CCF_{\max}$ is greater than a chosen threshold $T_{CCF}$ and the registration angle $\theta$ and $x$, $y$ registration positions are within the chosen threshold limits $T_\theta$, $T_x$ and $T_y$, then $X = 1$; otherwise $X = 0$. We use the symbol $P$ to represent probability in general and the symbol $p$ to represent the probability that $X = 1$. That is, $P(X=1) = p$, and $P(X=0) = 1 - p$.

We now make two key assumptions that will be revisited in later sections: 1) the comparisons between cell pairs are independent from each other, and 2) each cell pair comparison for the KNM images has the same probability $p = p_{KNM}$ to qualify as a CMC, and each cell pair comparison for the KM images has the same probability $p = p_{KM}$ to qualify as a CMC (Note: this notation represents a change from that of a previous article [13]). Thus, for the first image pair with $N_1$ correlated cell pairs, we now have a sequence of Bernoulli trials [26], $X_{11},...,X_{1N_1}$, which are independent from each other but have a common probability, $p_{KNM}$ or $p_{KM}$. We denote the number of successful trials, the CMC number, for the first image pair by $Y_1$. That is, $Y_1 = \sum_{i=1}^{N_1} X_{1i}$. Under the stated assumptions, $Y_1$ is a binomially distributed random variable [26], namely, $Y_1 \sim Bin(N_1, p)$. The functional form of $Bin$ is shown in Eq. 7. Similarly, for $M$ KNM or KM image pairs, we have $Y_1,\ldots,Y_M$. Assuming $\{Y_j, j = 1,\ldots, M\}$ are independent from each other, we have a

sequence of binomially distributed random variables, $Y_j = \sum_{i=1}^{N_j} X_{ji}$ , for $j = 1,...,M$ and $Y_j \sim Bin\ (N_j, p)$. In addition, we can state

$$\sum_{j=1}^{M} Y_j \sim Bin(\sum_{j=1}^{M} N_j, p) \cdot \tag{1}$$

For observed values of $\{Y_j ; j = 1,\ldots, M\}$, the maximum likelihood estimator of $p$ is given by [27]:

$$\hat{p} = \frac{\sum_{j=1}^{M} Y_j}{\sum_{j=1}^{M} N_j} = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N_j} X_{ji}}{\sum_{j=1}^{M} N_j} \tag{2}$$

Therefore, for the sub-population of KNM image comparisons, the false positive cell probability, denoted by $p_{\mathrm{KNM}}$, is the probability that a KNM cell pair comparison results in a CMC. Likewise, for the sub-population consisting of KM image pairs, the false negative cell probability is denoted by $(1 - p_{\mathrm{KM}})$.

To estimate $p_{\mathrm{KNM}}$ and $p_{\mathrm{KM}}$ from the data, we apply Eq. (2) to the sub-population of 717 KNM image pairs and the sub-population of 63 KM image pairs. For each sub-population, we estimate $p$ by counting all the CMC cell pairs that pass the four threshold criteria for a match:

$$\hat{p}_{\mathrm{KNM}} = \frac{\text{Number of KNM CMC cell pairs}}{\text{Total number of evaluated KNM cell pairs}} , \tag{3a}$$

$$\hat{p}_{\mathrm{KM}} = \frac{\text{Number of KM CMC cell pairs}}{\text{Total number of evaluated KM cell pairs}} . \tag{3b}$$

For the test results depicted in Fig. 5, the estimates obtained from Eq. 3 are:

$$\hat{p}_{\mathrm{KNM}} = 71/18859 = \mathbf{0.003765},$$
$$\hat{p}_{\mathrm{KM}} = 1207/1628 = \mathbf{0.7414}. \tag{4}$$

To evaluate the model, we compare the observed frequency of comparisons with a particular CMC number $h$ to the respective modeled frequency. For KM comparisons, the observed frequency is obtained as:

$$f_{\mathrm{KM}}\,(CMC = h) = \frac{\text{Number of KM image pair comparisons with } CMC = h}{\text{Total number of KM image pair comparisons}}. \tag{5}$$

In Fig. 5, the resulting frequency distribution is depicted by the red histogram. The respective modeled frequency distribution, depicted by the solid red curve, is obtained as:

$$\hat{f}_{\mathrm{KM}}(CMC = h) = \frac{\sum_{j=1}^{M} P(Y_j = h \mid \mathrm{KM})}{\text{Total number of KM comparisons}} = \frac{\sum_{j=1}^{M} Bin(h \mid N_j, \hat{p}_{\mathrm{KM}})}{\text{Total number of KM comparisons}} \quad (6)$$

where the summation of the binomial probability values is performed over all KM comparisons. If all comparisons have the same number of evaluated cells $N$, Eq. (6) would simplify to:

$$\hat{f}_{\mathrm{KM}}(CMC = h) = Bin(h \mid N,\ \hat{p}_{\mathrm{KM}}) = C_N^h \cdot \hat{p}_{\mathrm{KM}}^h \cdot (1 - \hat{p}_{\mathrm{KM}})^{N-h}, \quad (7)$$

where the binomial coefficient $C_N^h$ is the number of possible combinations of $h$ out of $N$ elements.

### 4.3 Distributions for Individual Identification Parameters

It is instructive to plot the experimental frequency distributions of correlated KM and KNM cell pairs with respect to each CMC identification parameter and to estimate the respective cell trial success probabilities. Figure 8 shows the experimental frequency distributions of correlated cell pairs for KM (red) and KNM (blue) comparisons with respect to the four identification parameters: $CCF_{\max}$ (Fig. 8A), registration angle $\theta$ (Fig. 8B), and $x$-, $y$- registration distances (Figs. 8C and 8D). The thresholds $T_{\mathrm{CCF}}$, $T_{\theta}$, $T_{\mathrm{x}}$, and $T_{\mathrm{y}}$ are also shown.

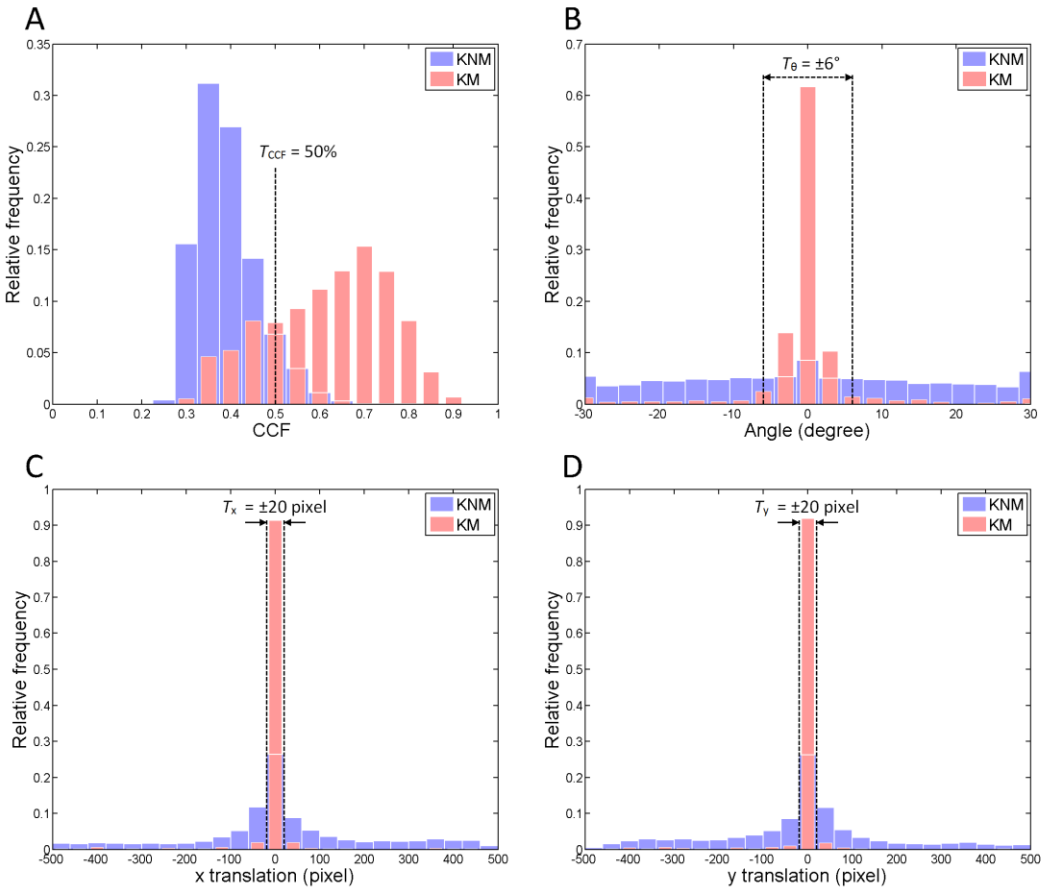


**Fig. 8.** Experimental relative frequency distributions of the correlated cell pairs for the KM (red) and KNM (blue) comparisons with respect to the identification parameters: A) $CCF_{\max}$ with a threshold $T_{CCF}$ = 50%; B) $\theta$ with $T_\theta$ = ± 6°; C) $x$ with $T_x$ = ± 20 pixels (or ± 0.125 mm); and D) $y$ with $T_y$ = ± 20 pixels (or ± 0.125 mm). The KM (63 pairs) and KNM (717 pairs) distributions are each scaled to their particular sample size.

Although there are large overlaps between the KM and KNM cell distributions for each parameter, combining all four parameters yields the significant separation between the CMC distributions of KM and KNM image pairs shown in Fig. 5. The combined false positive and false negative probabilities, $p_{KNM}$ and $(1 - p_{KM})$, for each correlated cell pair can likewise be estimated by combining the estimated false positive and false negative frequencies associated with each of the four identification parameters ($CCF_{\max}$, $\theta$, $x$ and $y$) taking into account any correlation between them. First, the number of KNM cell pairs that pass the $T_{CCF}$ threshold ($CCF_{\max} \geqslant 50\%$ for the test case) are counted and compared with the total number of cell pairs to derive an estimation for the individual probability $p_{KNM}(CCF)$ (see Table 1). Then only the cell pairs passing the $T_{CCF}$ test are included in the conditional frequency distribution for the next parameter $\theta$, from which the conditional probability $p_{KNM}(\theta|CCF)$is estimated, and so on [13]. The false positive and false negative cell trial probabilities associated with all thresholds estimated in this way are the same as those calculated by Eqs. 2 and 3 and are shown in Table 1.

### 4.4 Error Rate Estimation with the Binomial Model

The estimated cell trial success probabilities, $\hat{p}_{\mathrm{KNM}}$ and $\hat{p}_{\mathrm{KM}}$, are inserted into the binomial model to estimate potential error rates for cartridge cases fired from different source (KNM) and same source (KM) firearms under similar conditions. Figure 9 shows a conceptual diagram for two CMC probability mass functions, $\Phi_{\mathrm{CMC}}$ and $\Psi_{\mathrm{CMC}}$, for KM and KNM topography pairs, respectively. As discussed in Section 4.2, the probability mass function $\Psi_{\mathrm{CMC}}$ for KNM comparisons is modeled as:

$$\Psi(CMC = h \mid N, \hat{p}_{\mathrm{KNM}}) = Bin(h \mid N, \hat{p}_{\mathrm{KNM}}) = C_N^h \cdot \hat{p}_{\mathrm{KNM}}^h \cdot (1 - \hat{p}_{\mathrm{KNM}})^{N-h}. \qquad (8)$$

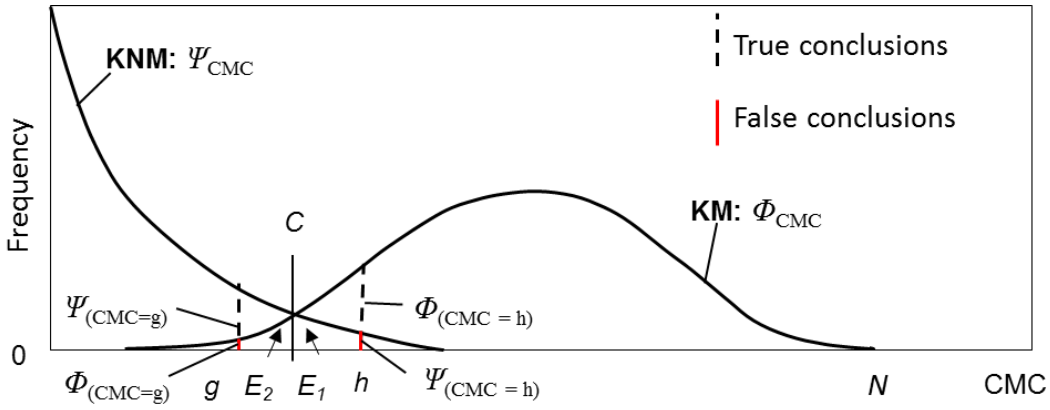


**Fig. 9.** Conceptual diagram of the CMC probability mass functions for KM and KNM comparisons, $\Phi_{\mathrm{CMC}}$ and $\Psi_{\mathrm{CMC}}$. To illustrate clearly the listed quantities, the schematic depicts the discrete probability distributions as continuous density functions that overlap much more than they would be expected to in practice. The regions $E_1$ and $E_2$ under the curves represent cumulative false positive and false negative error rates. For each individual "matching" conclusion, $h \geqslant C$, and "non-matching" conclusion, $g < C$, there are probabilities for both "True" and "False" conclusions as demonstrated by the black dashed bars and the solid red bars, respectively.

The cumulative false positive error rate $E_1$ is given by the sum of the probability mass function values $\Psi_{CMC}$ for CMC values between $C$ and $N$:

$$E_1 = \sum_{CMC\,=\,C}^{CMC\,=\,N} \Psi_{(CMC)} = \Psi_{(CMC=C)} + \Psi_{(CMC=C+1)} + \cdots + \Psi_{(CMC=N)}$$
$$= 1 - \left(\Psi_{(CMC=0)} + \Psi_{(CMC=1)} + \cdots + \Psi_{(CMC=C-1)}\right). \qquad (9)$$

The cumulative false positive error rate $E_1$ is determined by three factors: the number of correlation cell pairs $N$ in a comparison, the numerical identification criterion $C$ of the CMC method ($C$ = 6), and the combined false cell trial success probability $\hat{p}_{\mathrm{KNM}}$ of each correlated cell pair estimated from Eq. 3a.

Similarly, the probability mass function $\Phi_{\mathrm{CMC}}$ for KM correlations (Fig. 9) is modeled as:

$$\Phi(CMC = g \mid N, \hat{p}_{KM}) = \; Bin(g \mid N, \hat{p}_{KM}) = C_N^g \cdot \hat{p}_{KM}^g \cdot (1 - \hat{p}_{KM})^{N-g}. \tag{10}$$

The cumulative false negative error rate $E_2$ is given by the sum of the probability mass function values $\Phi_{CMC}$ for CMC values between 0 and ($C$ – 1):

$$E_2 = \sum_{CMC\,=\,0}^{CMC\,=\,C-1} \Phi_{(CMC)} = \Phi_{(CMC=0)} + \Phi_{(CMC=1)} + \cdots \Phi_{(CMC=C-1).} \tag{11}$$

We note again the assumptions underlying the binomial distribution model:

(1) All cell pair evaluations of each compared image pair represent independent Bernoulli trials with the same success probability, and
(2) The image pair comparisons are independent from each other but have the same cell trial success probability for cell pair comparisons within each image pair.

Note that Condition 1 implies a binomial distribution for the number of CMCs in an image pair comparison and Condition 2 implies that the whole set of KM or KNM comparisons consists of a sequence of independent binomial random variables corresponding to image pair comparisons with the same $p$ value, but allowing for variable $N$. The error rates estimated from the value of $\hat{p}$ are random variables themselves and their uncertainties should also be assessed [4,5].

The cumulative false positive and false negative error rates $E_1$ and $E_2$ may be estimated from Eqs. 8 and 10 using the known number of effective cells $N$, the CMC identification criterion $C$ = 6 and the estimated false positive and false negative cell trial success probabilities $\hat{p}_{\mathrm{KNM}}$ and $(1 - \hat{p}_{\mathrm{KM}})$ shown in Eq. 3. The results are shown in Table 2 for $N$ = 26, the average number of evaluated cells in the image comparisons. For the 717 KNM comparisons, the cumulative false positive error rate is $E_1 = 6.2\times10^{-10}$, which represents the sum of the $\Psi_{\mathrm{CMC}}$ probabilities between 6 and $N$. The cumulative false negative error rate is $E_2 = 7.4\times10^{-09}$ which represents the sum of the $\Phi_{\mathrm{CMC}}$ probabilities between 0 and 5. These error rates will vary depending on the specific data population, the distribution models, and all the parameters chosen for the calculation, such as the cell size. Error rates for different models are discussed briefly below.

### 4.5 Individual Error Rates for Identifications and Exclusions

Figure 9 also illustrates the probabilities of true and false conclusions by the dashed black bars and the solid red bars for specific CMC values $h$ and $g$. These probabilities can be used to calculate likelihood ratios for various conclusions, that is, the ratio of the likelihoods of obtaining the conclusion under two competing hypotheses (matching or non-matching samples) [23-25]. Alternatively, we define here the individual probability $R_1$ that an identification is false as the probability that an image pair is non-matching when its CMC value appears in the matching region with a specific score $h$ ($h \geq C$) and conversely the individual probability $R_2$ that an exclusion is false as the probability that an image pair is matching when its CMC value appears in the non-matching region with a specific score $g$ ($g < C$). For our experiment, $R_1$ can be estimated as:

$$R_{1(CMC=h)} = \frac{K \times \Psi_{(CMC=h)}}{K \times \Psi_{(CMC=h)} + \Phi_{(CMC=h)}} , \qquad (h \geq 6) , \tag{12}$$

where $K$ is the ratio of the sample sizes of KNM and KM topography image pairs. In a Bayesian approach, $K$ represents the prior odds against obtaining a match in the current population of 40 breech face images before conducting the forensic test. For this study, $K$ is equal to 717/63 under the condition that we randomly select two of the cartridge cases from our set before comparing their topographies to determine whether they are matching.

Conversely $R_2$ can be estimated by

$$R_{2(CMC=g)} = \frac{\Phi_{(CMC=g)}}{\Phi_{(CMC=g)} + K \times \Psi_{(CMC=g)}} , (g < 6). \tag{13}$$

The parameters, $R_1$ and $R_2$, could be useful when addressing questions, such as "given the conclusion of identification based on a CMC score $h$ ($h \geq C$), what is the probability that the cartridge cases were actually ejected from different firearms (individual false identification probability $R_1$)?", or "given the conclusion of exclusion based on a CMC comparison score $g$ ($g < C$), what is the probability that the two cartridge cases were actually ejected from the same firearm (individual false exclusion probability $R_2$)?" In Bayesian terms, $R_1$ and $R_2$ represent posterior probabilities of erroneous identifications and exclusions, respectively.

Table 2 shows values of $R_1$ for $h$ = 9 and 26 and values of $R_2$ for $g$ = 0 and 2. The largest false identification probability for this set is $R_{1(\mathrm{CMC}=9)} = 2.3\times10^{-10}$; the largest false exclusion probability is $R_{2(\mathrm{CMC}=2)} = 3.0\times10^{-11}$. The small estimated value for the individual false identification probability is largely due to the rapid decline in the modeled probability mass function curve $\Psi_{\mathrm{CMC}}$ for KNM comparisons, which matches well with the experimental data (Fig. 5). However, the very small value for the individual false exclusion probability is questionable, as the observed CMC frequency distribution for KM comparisons is wider than the binomial distribution model (solid red line in Fig. 5). For realistic databases with a large number of entries of firearms and ammunition, even when compartmented according to model and manufacturer, the overlap of KM and KNM distributions can become significant and the error rates will likely increase significantly.

It is both insightful and misleading to illustrate the calculation of $R_1$ and $R_2$ using the red and black bars in Fig. 9, which are values for the normalized distributions, $\Psi_{\mathrm{CMC}}$ and $\Phi_{\mathrm{CMC}}$. One can conveniently imagine $R_1$ for a particular value of $h$ to be given by the length of the red bar in Fig. 8, i.e., the probability $\Psi_{\mathrm{CMC}}(h)$ of obtaining a CMC score $h$ for two non-matching samples, divided by the total length of the red bar and dashed black bar together, i.e., the probability $\Psi_{\mathrm{CMC}}(h)$ + $\Phi_{\mathrm{CMC}}(h)$ of obtaining a CMC score of $h$. This would be neglecting the ratio $K$ between the size of the KNM and KM populations, because the distributions, $\Phi_{\mathrm{CMC}}$ and $\Psi_{\mathrm{CMC}}$, are conventionally shown normalized to unity. In fact, there are many more possible pairs of non-matching comparisons than matching comparisons, as taken into account by the factor $K$, which increases the probability of a wrong conclusion of identification. One could visualize this by plotting the $\Psi_{\mathrm{CMC}}$ distribution with a total area of $K$ instead of unity. The red bar at CMC = $h$ would then be $K$ times as long as that shown in Fig. 9, as would the black dotted bar at $g$.

## 5. Re-assessing the Binomial Model

The binomial model described above contains the assumption that a single value ($p_{\mathrm{KNM}}$) characterizes the probability that a pair of cells from KNM images will pass all criteria and qualify as a false positive CMC cell pair. The resulting model fits the KNM data quite well (see Fig. 5, blue line), and theoretically, we expect the use of a single false positive cell trial success probability $p_{\mathrm{KNM}}$ to be a good approximation for KNM data. If two cells were from images of breech face impressions from different firearms, the fact that they appear to qualify as a CMC cell pair is likely driven by random, non-selective factors as long as subclass characteristics, the carry-over of pre-fire tool marks, and systematic measurement errors are not significant factors in the evaluation.

The situation is more complicated for cell pairs of KM images. Variations in firing conditions, firearm wear, and contaminants cause variations in the tool marks imparted on the cartridge case and the domain of the breech face impression area. These effects and others cause variations in the size and quality of the common valid correlation area of a KM image pair comparison, which, in turn, may cause variations in the probability $p_{\mathrm{KM}}$ of a cell pair to be qualified as a CMC. For comparisons of KNM samples, these effects simply add additional random factors to a comparison result, which is already largely driven by random factors and which is unlikely to cause major variations in the cell trial success probability $p_{\mathrm{KNM}}$. Variations in the cell trial success probability $p_{\mathrm{KM}}$ would, to some extent, explain the higher dispersion of the observed CMC numbers for KM comparisons than predicted by the binomial model (solid red curve in Fig. 5), which is based on the assumption of a single value of $p_{\mathrm{KM}}$ for all KM comparisons. To account for these observations, we relax the assumption of the same cell trial success probability $p_{\mathrm{KM}}$ for all KM comparisons.

### 5.1 A Beta-binomial Probability Model for the Distribution of CMCs

In this approach, we still assume that a CMC image comparison can be modeled as a set of independent Bernoulli trials characterized by the same cell trial success probability. However, we now allow the cell trial success probability $p$ to vary from comparison to comparison. Here we assume that the parameter $p$ can be modeled as a random variable with a beta distribution. The choice of the beta distribution has several advantages. The beta distribution is defined by two parameters, $\alpha$ and $\beta$, which allow for a wide range of distribution shapes. The domain of the beta distribution is restricted to the interval [0, 1], which makes it a convenient distribution to model probabilities. In a Bayesian framework, the beta distribution is a conjugate distribution of the binomial distribution, yielding an analytical expression for the resulting compound beta-binomial distribution [28]. Finally, the resulting beta-binomial distribution can approximate the binomial distribution to arbitrary precision when needed [29].

Similar to Section 4.1, we model an image comparison $j$ with $N_j$ evaluated cell pairs as a sequence of Bernoulli trials, $X_{j1},...,X_{jN_j}$, which are independent from each other and have a common success probability $p = p_j$. The CMC number of the comparison, i.e., the sum of the trial outcomes $X_{ji}$, is $Y_j$, which for a given $p = p_j$ has a binomial distribution $Y_j|p_j \sim Bin(N_j, p_j)$. For $M$ image comparisons, we have $Y_j|p_j \sim Bin(N_j, p_j)$, for $j$ = 1 to $M$, where $p_j$ now has a beta distribution, i.e., $p_j \sim Beta(\alpha, \beta)$ with positive $\alpha$ and $\beta$. The probability mass function of the resulting beta-binomial random variable $Y$ for given values of $N$, $\alpha$ and $\beta$ is given by [29]:

$$P(Y = k|N, \alpha, \beta) = C_N^k \frac{B(k + \alpha, N - k + \beta)}{B(\alpha, \beta)}, \quad (14)$$

where $B(\alpha, \beta)$ is a beta function with parameters $\alpha$ and $\beta$, and $k$ is a CMC value.

For the KM and KNM comparison results discussed in Section 3, we obtained maximum likelihood estimates of the parameters $\alpha$ and $\beta$ using the algorithm described by Smith [29]. The respective values are: $\hat{\alpha}_{\mathrm{KNM}} = 2.15$ and $\hat{\beta}_{\mathrm{KNM}} = 569.1$ for the KNM comparisons and $\hat{\alpha}_{\mathrm{KM}} = 6.55$ and $\hat{\beta}_{\mathrm{KM}} = 2.29$ for the KM comparisons. The modeled frequency distributions for the KM and KNM CMC results are depicted by the dashed curves in Fig. 5. As expected, the beta-binomial model for the KNM comparisons is nearly indistinguishable from the respective binomial model. For the KM comparisons, on the other hand, the beta-binomial model shows a significant improvement in the ability to model the dispersion of the experimental results. The result is a significant increase in the estimated cumulative false negative error rate from $E_2 = 7.4\times10^{-09}$ for the binomial model to $E_2 = 2.1\times10^{-3}$ for the beta-binomial model. For the KNM comparisons, the estimated cumulative false positive cumulative error rate increases from $E_1 = 6.2\times10^{-10}$ to $E_1 = 3.2\times10^{-8}$.

## 5.2 Software Development

We repeated the analysis with several changes to the algorithms and correlation parameters. First, the images are no longer down sampled, resulting in an average pixel spacing of 3.125 μm instead of 6.25 μm. Second, the low-pass and high-pass filters are now, respectively, zeroth order and second order Gaussian regression filters [30] to attenuate filtering edge effects at the image domain boundaries. Their cutoffs are now, respectively, 25 μm and 250 μm. Third, cell registration was improved through a combination of Fourier-based and direct optimization of the normalized cross correlation value at overlapping image areas as a function of sample translation and rotation. Fourth, the effect of spurious local registration optima was reduced by increasing requirements for the minimum percentage of measured pixels in a cell from 10 % to 25 % and by decreasing the size of the search domain to ± 0.75 mm for sample translations. Finally, we optimized the initial placement of the cells on the donut shape of the reference image to ensure maximum coverage of the respective sample domain, resulting in an increase of the average number of evaluated cells.

Figure 10 shows the results of the revised analysis. The chosen cell size remains the same, but now comprises 150 × 150 pixels. The x-y registration thresholds remain at ± 20 pixels (± 62.5 μm considering the pixel spacing is 3.125 μm). The search range of registration angles and x-y displacements was limited to ± 45° and ± 750 μm, respectively. The number of evaluated cells per comparison varies between 26 and 35 cells, with an average of 31 cells. For the 63 KM cartridge pairs, the number of CMCs ranges from 18 to 32, and for the 717 KNM cartridge pairs, the number of CMCs has the same range (0 to 2) as that of Fig. 5. For the beta-binomial model with 31 cells, the estimated cumulative false positive error rate decreases to $E_1 = 5.5\times10^{-9}$ and the cumulative false negative error rate to $E_2 = 1.1\times10^{-4}$.

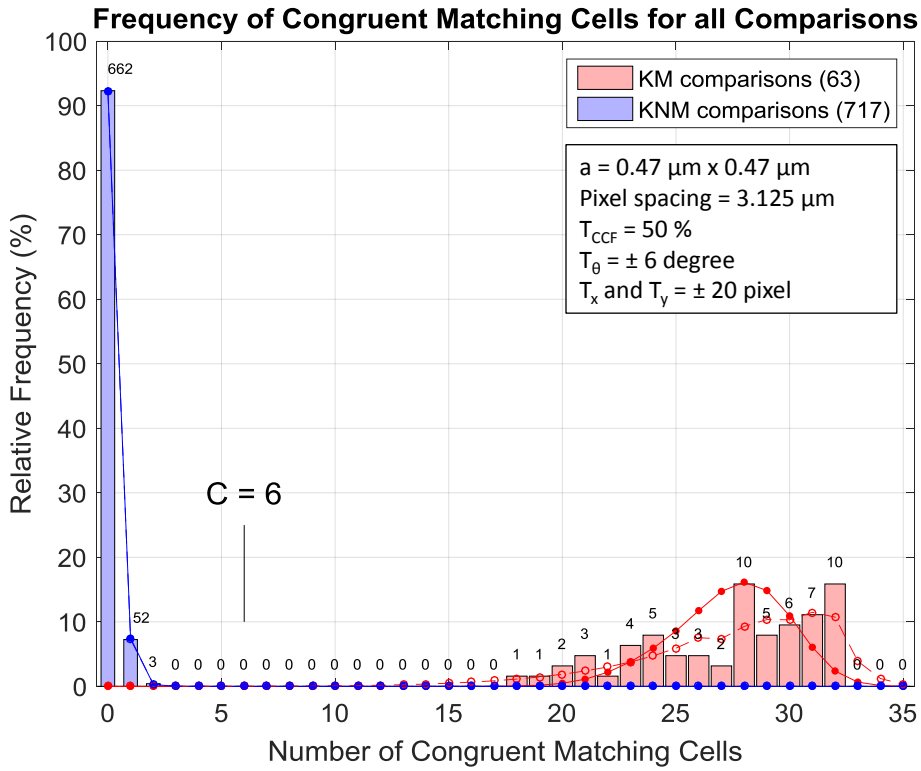


**Fig. 10.** Relative frequency distribution of CMC numbers for KM and KNM image pairs obtained with modified software and correlation parameters. The solid red and dashed red curves represent the binomial and beta-binomial distribution models for the KM data. The overlapping solid blue and dashed blue curves represent the two respective models for the KNM data. The models are estimated from the histogram data. Note that the distribution models are discrete, with the connecting lines drawn for visualization.

### 5.3 Clustering

It is important to note that the image comparisons in our experiments are not independent for two reasons. First, each firearm slide was used to mark more than one cartridge case. Second, each cartridge case was used in more than one comparison. This lack of independence can lead to clustering effects in the experimental frequency distributions. For example, if one cartridge case is poorly marked, several comparisons with this cartridge case are affected in a similar fashion. Of the five comparisons yielding a CMC number of 11 in Fig. 5, three involved the same cartridge case and four involved the same firearm slide. This clustering is currently not addressed by our models, which assume independence of the various comparisons in our set.

The KM data of Fig. 10 are plotted again in Fig. 11 with the identity of the ten slides indicated by different colors. In Fig. 11, the histogram abscissa represents the percentage of evaluated cells in a comparison that were classified as CMC cells. There is clearly a difference between images from different slides. For example, all KM image comparisons involving samples from slides 1 or 3 yield CMC numbers exceeding 90 % of the number of evaluated cells. For slide 7, on the other hand, half the comparisons yielded CMC numbers between 55 % and 70 % of evaluated cells.

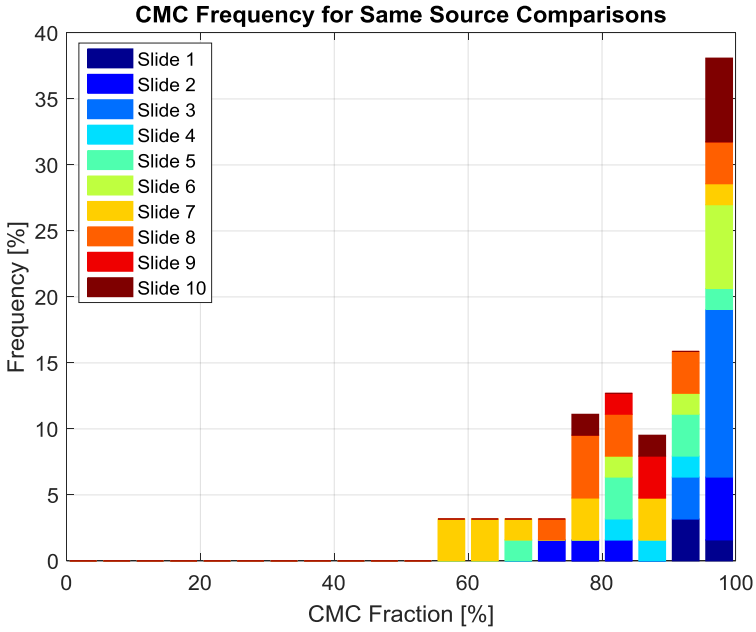


**Fig. 11.** Frequency distribution of KM image comparisons for the fraction of evaluated cells that were classified as CMC cells.

The differences are illustrated further in Fig. 12 by a scatter plot of the KM images' CMC fractions for each slide. The differences in the spread of the CMC fractions from one slide to another suggest differences in the capacity of the slides to impress similar topographies on the breech faces. Clearly, the five images of slide 3 when paired off to yield the ten KM images' CMC fractions on the right hand side of Fig. 12 have consistently higher similarity than the five images of slide 7 whose ten KM CMC fractions are shown on the left hand side.

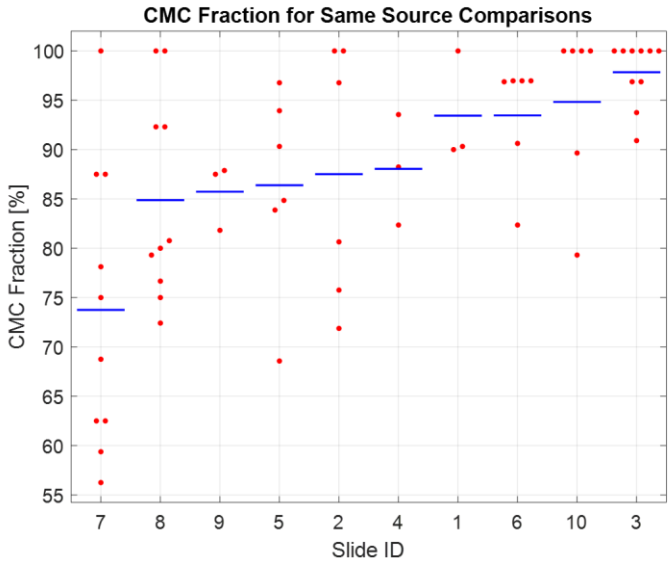


**Fig. 12.** Scatter plot of CMC fractions for different slides using the data shown in Fig. 11. The blue line represents the mean value for each slide.

Figure 13 shows the comparison of one breech face impression topography, image A, with two other cartridge case topographies, B and C, all fired from a firearm using slide 7. In the first comparison, A vs. B, all 32 evaluated cells were classified as CMC cells. In the second comparison, A vs. C, only 19 cells were classified as CMCs. Some of the failed cells can be explained by insufficient trimming of the firing pin impression area on the inside edge in the upper right quadrant of image C. However, overall, stronger matching features are present in comparison A-B than A-C, as reflected by the difference in the average normalized CCF values of the CMC cells of 86 % vs. 66 %.

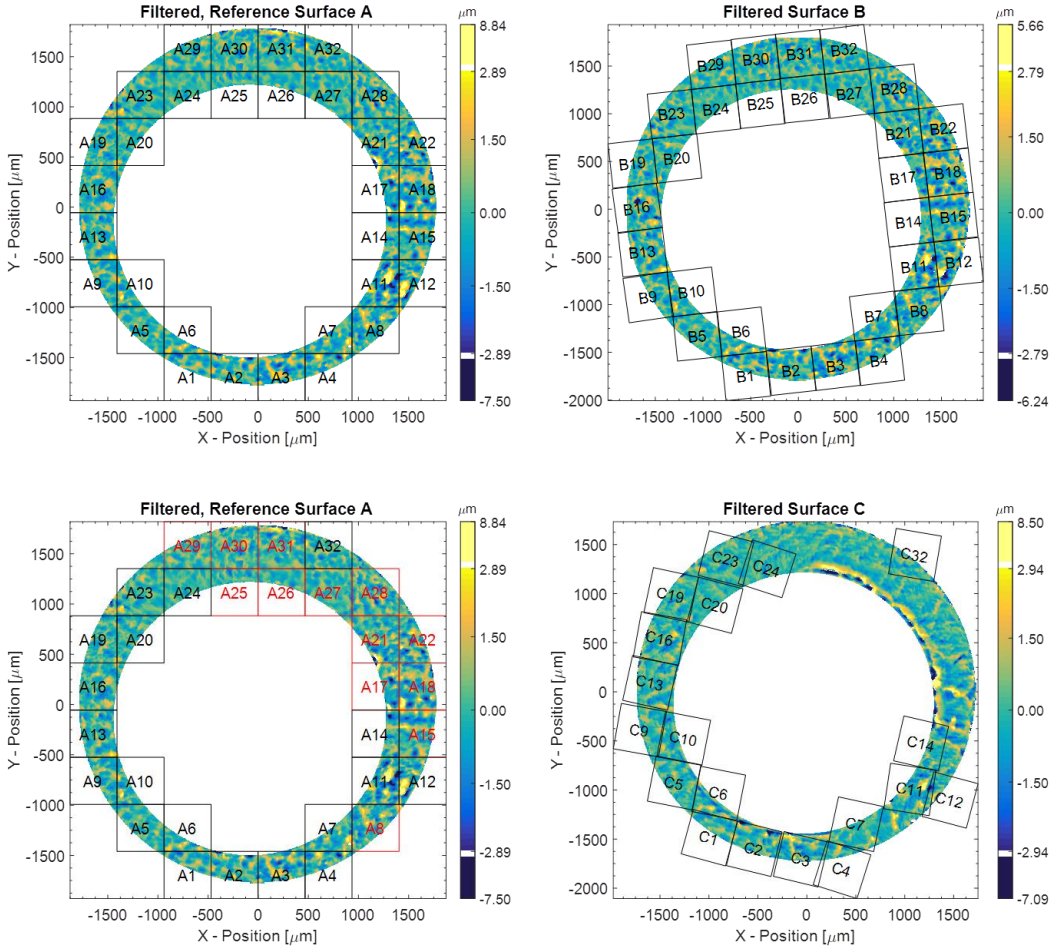


**Fig. 13.** Top--comparison of the breech face impression of cartridge cases A and B fired from a firearm using slide 7. Bottom--comparison of the breech face impression of cartridge cases A and C fired also from a firearm using slide 7.

## 5.4 Testing the Models

We evaluated the models derived in Sections 4 and 5 on a different set of cartridge cases created by Weller et al. [31]. The cartridge cases were obtained from another set of eleven firearm slides produced by the same manufacturer using the same manufacturing process as that of the Fadul set described in Section 3. The Weller set consists of 95 cartridge cases, 9 cartridge cases each for 10 consecutively manufactured slides and 5 cartridge cases for one extra slide that was not

manufactured consecutively with the others. The measurement procedure and image processing parameters were the same as those discussed previously, and the CMC parameters were the same as those discussed in Section 5.2. For this dataset, the domain of the trimmed breech face impressions typically consists of thicker "donut" areas. This resulted in a higher number of evaluated cells per comparison, ranging from 28 to 49 cells with an average of 42 cells.

Figure 14 shows the relative frequency distribution of the observed CMC numbers for the 370 KM and 4095 KNM image comparisons. Once again, there are no false positive or false negative results, and the KNM and KM data are widely separated. For the 370 KM cartridge pairs, the number of CMCs ranges from 21 to 47; for the 4095 KNM cartridge pairs, the number of CMCs again ranges from 0 to 2. Also shown in Fig. 14 are the modeled frequency distributions for the CMC results. For the KM data, the binomial and beta binomial probability models are different, but less so than for the KM Fadul data, with the beta-binomial model providing a slightly better description of the data dispersion. For the average number of evaluated cells, 42, the cumulative false positive and false negative error rates obtained from the beta-binomial model are, respectively, $E_1 = 5.9\times10^{-11}$ and $E_2 = 3.8\times10^{-11}$.

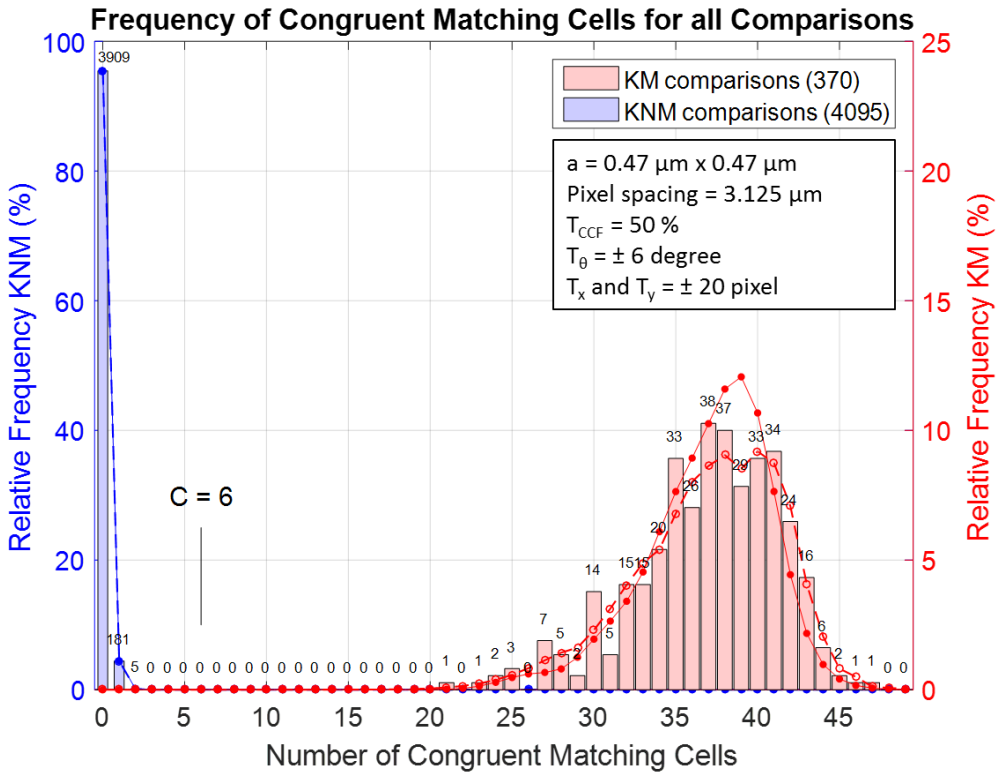


**Fig. 14.** Relative frequency distribution of CMC numbers for KM and KNM image pairs of the Weller dataset [31]. The solid red and dashed red curves represent the binomial and beta-binomial distribution models for the KM data. The nearly indistinguishable solid blue and dashed blue curves represent the two respective models for the KNM data. Note that the distribution models are discrete, with the connecting lines drawn for visualization. The right hand scale for the KNM data is magnified by a factor of four to show differences more clearly.

We have also compared the Weller data [31] with models using the $p_{KNM}$, $p_{KM}$, $\alpha$, and $\beta$ parameters estimated from the Fadul data set discussed in Section 5.2. Figure 15 shows good agreement between both models and the KNM data and reasonable agreement between the beta binomial model and the KM data. For both models, the predicted expected mean value for the number of CMCs in the Weller KM comparisons is approximately one CMC lower than that of the respective experimental data and models shown in Figure 14. This evaluation shows the consistency of the results and suggests the general applicability of the beta-binomial model for describing KM data and the applicability of both models for describing KNM data from these types of manufactured slides. For the average number of evaluated cells, 42, the cumulative false positive and false negative error rates obtained from the beta-binomial model are, respectively, $E_1 = 3.7\times10^{-8}$ and $E_2 = 1.8\times10^{-5}$.

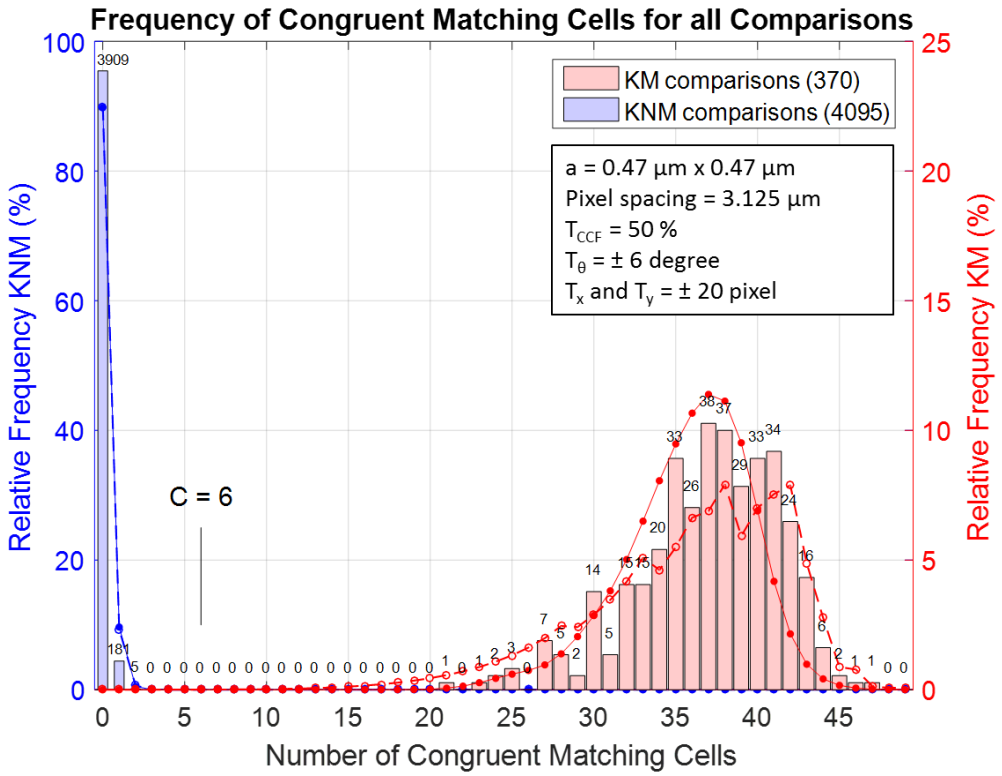


**Fig. 15.** Relative frequency distribution of CMC numbers for KM and KNM image pairs of the Weller dataset. The solid red and dashed red curves represent the binomial and beta-binomial distribution models for the KM data, whose parameters were estimated from the Fadul data. The nearly indistinguishable solid blue and dashed blue curves represent the two respective models for the KNM data. Note that the distribution models are discrete, with the connecting lines drawn for visualization.

## 6. Summary and Discussion

Reporting error rates for firearm identification is a fundamental challenge in forensic science. We have developed a practical statistical approach to estimating error rates based on the CMC identification method. Initial results for comparisons of the breech face impression topography on cartridge cases ejected from two different sets of pistols of the same brand, each with slides that

were consecutively manufactured using the same process, show wide separation between the CMC scores of KM and KNM samples. Models for the frequency distribution of the CMC scores show good agreement with the experimental data and yield very small error rates for erroneous classifications, in particular for false positive identifications.

The error rate estimates are derived from models for the probability distribution of the similarity metric, the CMC score, for KNM and KM image comparisons. The initial binomial models were based on two key approximations: 1) the CMC cell trials in a comparison are statistically independent, and 2) all the KNM and KM cell pairs in our population have, respectively, the same false positive identification probability $p_{KNM}$ and the same true positive identification probability $p_{KM}$. Barring the presence of sub-class characteristics, which is unlikely for sand-blasted breech face surfaces, these approximations seem reasonable for KNM image comparisons. The estimated binomial distribution $\Psi_{CMC}$ for the respective CMC scores matches the experimental data quite well for both the Fadul and Weller datasets (blue lines in Figs. 5, 10, 14, and 15). From a legal perspective, the KNM distribution is critical for ballistics identifications as it describes the probability of false positives (false identifications), which are to be avoided at almost any cost. For these KNM data, the modeled beta-binomial distribution is virtually identical to the modeled binomial distribution.

On the other hand, the binomial distribution for KM image comparisons shows a lower dispersion than the experimental data (solid red lines in Figs. 5, 10, 14, and 15), resulting in error rate estimates that are too low. This is not unexpected, as variations in firing conditions, wear, and contaminants cause variations in the tool marks imparted on the cartridge cases. These effects cause variations in the size and quality of the common valid correlation area of a KM image pair comparison, which, in turn, are likely to cause variations in the average probability $p_{KM}$ of a cell pair to be qualified as a CMC in a comparison. The beta-binomial model described in Section 5 allows the average cell trial success probability $p$ to vary from comparison to comparison. This approach improves agreement between the modeled and experimental KM CMC distributions (dashed red lines in Figs. 5, 10, 14, and 15).

We emphasize that the estimated error rates in this report are specific to the sets of firearms studied here and are not applicable to other firearm scenarios. Furthermore, the presented models do not address correlations between the experimental comparison results due to the use of a sample image or firearm slide in more than one comparison. Some differences between the modeled and experimental results, such as the CMC = 11 values in Fig. 5, can be attributed to this effect.

On the other hand, the small error rates calculated form the models suggest the feasibility of applying the CMC method to a larger number of firearms manufactured under similar conditions and producing correlation results to support identification and exclusion decisions. The probability models for the CMC scores estimated from one firearm set were consistent with the distributions observed for the second set, indicating consistency of the estimated error rates.

The work reported here is a demonstration of concept for the objective CMC method. Studies with larger databases, including direct comparisons with manual evaluations, will be required in order to demonstrate feasibility of the CMC method for crime lab casework. We are working to participate in black-box studies of firearms experts similar to that of Baldwin et al. [7], which was favorably reviewed by a recent government report of the President's Council of Advisors on

Science and Technology [32]. This could yield a comparison of error rates of the CMC method and of subjective methods.

We are also working to test the CMC method and error rate procedure on different sets of consecutively manufactured firearms, where the fabrication process leaves stronger common tool marks than the sand blasting process studied here, as well as on a statistically designed set of test fires comprising firearms and ammunitions from different manufacturers. Promising results for the application of the CMC method to conventional optical microscopy image data have been reported by Tong et al. [15]. We are also working on a procedure based on international standards [33] to incorporate uncertainty statements into the reported error rates, and we aim to scale the approach to be usable with databases of forensic samples.

We expect that the estimated probability distributions and error rates are affected by the type of firearm, ammunition, measurement method, correlation parameters, and other factors. However, the error rate procedures based on the CMC method can be adapted, with refinements and model adjustments for specific situations, to provide scientific support and error rate estimates for other firearm identification scenarios, including large scale databases, in a manner similar to the Coincidental Match Probability (CMP) procedures developed for DNA identifications [3]. The CMC method may also serve as a foundation for manufacturers to develop next generation ballistic identification systems characterized by high accuracy and error rate reporting. An error rate procedure could also be used for laboratory assessment and accreditation in accordance with the ISO 17025 standard [34] and ASCLD/LAB procedures [35].

We envision a time when ballistic examiners can input either topographies or optical intensity images into a program that automatically conducts correlations using the CMC method, and displays the correlation conclusion (identification or exclusion) with an error rate estimate. The CMC method and statistical procedure can provide a scientific foundation and a practical method to estimate error rates for supporting ballistic identifications in forensic science.

**Acknowledgements:** The funding for this project is provided by the Forensic Measurement Challenge Program (FMC2012) and the Special Programs Office (SPO) of NIST. The authors are grateful to T. Fadul of the Miami-Dade Crime Lab and to T. Weller for providing test samples, to X. Zheng of NIST for providing the topography images, to R. Silver, R.M. Thompson, and S. Ballou of NIST for their support of this work and their suggestions, and to J. Lu, J. Filliben, J. Butler, and J.M. Libert of NIST for their careful manuscript reviews.

**References:**

1. Scientific Working Group for Firearm and Toolmarks (SWGGUN), The Foundations of Firearm and Toolmark Identification, http://www.swggun.org.
2. The National Research Council, Ballistic Imaging, (NRC, Washington DC 2008) pp. 3, 82, 20.
3. The National Research Council, Strengthening Forensic Science in the United States—A Path Forward, (NRC, Washington DC, 2009) pp. 153-154, 184, 155.
4. N.D.K. Petraco, et al., Application of machine learning to tool marks: statistically based methods for impression pattern comparisons, NIJ Report 239048, National Institute of Justice, Washington, DC (2012).

5. N.D.K. Petraco, L. Kuo, H. Chan, E. Phelps, C. Gambino, P. McLaughlin, F. Kammerman, P. Diaczuk, P. Shenkin, N. Petraco, and J. Hamby, Estimates of striation pattern identification error rates by algorithmic methods, AFTE J. 45 (3) (2013) 235-244.
6. F. Riva and C. Champod, Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases, J Forensic Sci., 59 (3) (2014) 637-647; doi: 10.1111/1556-4029.12382.
7. D.P. Baldwin, S.J. Bajic, M. Morris, and D. Zamzow, A Study of False-positive and False-negative Error Rates in Cartridge Case Comparisons, USDOE Technical Report # IS-5207 (Defense Forensics Science Center, Forest Park, Georgia, April 2014).
8. R. Lilien, Applied Research and Development of a Three-dimensional Topography System for Firearm Identification Using GelSight, NIJ Report 248639, National Institute of Justice, Washington DC, (2016); https://www.ncjrs.gov/pdffiles1/nij/grants/248639.pdf.
9. T. Weller T, N. Brubaker, P. Duez, and R. Lilien, Introduction and initial evaluation of a novel three-dimensional imaging and analysis system for firearms forensics, AFTE J. 47 (2015) 198–208.
10. J. Song, E. Whitenton, D. Kelley, R. Clary, L. Ma, and S. Ballou, SRM 2460/2461 standard bullets and cartridge cases project. J. Res. Natl. Inst. Stand. Technol. 109 (6) (2004) 533-542.
11. NIST SRM 2460/2461 standard bullet and cartridge cases, available at http://www.nist.gov/pml/div683/grp02/sbc.cfm.
12. J. Song, Proposed NIST ballistics identification system (NBIS) using 3D topography measurements on correlation cells. AFTE J. 45 (2) (2013) 184-189.
13. J. Song, Proposed "Congruent matching cells (CMC)" method for ballistic identification and error rate estimation, AFTE J. 47 (3) (2015) 177-185.
14. W. Chu, M. Tong, and 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.
15. M. Tong, J. Song, W. Chu, and R.M. Thompson, Fired cartridge case identification using optical images and the congruent matching cells (CMC) method, J. Res. Natl. Inst. Stand. Technol. 119 (2014) 575-582; http://dx.doi.org/10.6028/jres.119.023.
16. T.G. Fadul Jr., G.A. Hernandez, S. Stoiloff, and S. Gulati, An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing 10 consecutively manufactured slides. NIJ Report No. 237960, National Institute of Justice (2012).
17. A.A. Biasotti and J.E. Murdock, Criteria for identification or state of the art of firearm and toolmark identification. AFTE Journal, 16 (4) (1984) 16-34.
18 H. Zhang, J. Song, M. Tong, and W. Chu, Correlation of firing pin impressions based on the congruent matching cross-sections (CMX) method, Forensic Science International, 263 (2016) 186-193.
19. J.B.P., Williamson, The shape of solid surfaces, in Rough Surfaces, T.R. Thomas, ed., first edition (Longman, Harlow, Essex, UK, 1982).
20. T.V. Vorburger, J. Song, and N. Petraco, Topography measurements and applications in ballistics and tool mark identifications, Surf. Topogr.: Metrol. Prop. 4 (2016) 013002; doi:10.1088/2051-672X/4/1/013002.
21. L.A. Thibodeau, Sensitivity and specificity, in S. Kotz, N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences, Wiley, New York, ed. 1 (1988) 370-372.

22. J.J. Koehler, Fingerprint error rates and proficiency tests: what they are and why they matter, Hastings Law Journal, 59(5) (2008) 1077-1100.
23. C. Aitken, P. Roberts, and G. Jackson, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings, Royal Statistical Society (2010).
24. W. Kerkhoff, R.D. Stoel, E.J.A.T. Mattijssen, and R. Hermsen, The likelihood ratio approach in cartridge case and bullet comparison, AFTE J. 45 (3) (2013) 284-289.
25. S. Bunch and G. Wevers, Application of likelihood ratios for firearm and toolmark analysis, Science and Justice, 53 (2013) 223-229.
26. A. Papoulis, Probability, Random Variables and Stochastic Processes, 2nd ed., McGraw-Hill, New York (1984) 42-43.
27. Ref. 26, p. 75
28. R.R. Wilcox, A review of the beta-binomial model and its extensions, J. Educational Statistics, 6 (1) (1981) 3-32.
29. D.M Smith, Algorithm AS 189: Maximum likelihood estimation of the parameters of the Beta-binomial distribution, J. Royal Statistical Society, Series C (Applied Statistics), 32 (2) (1983) 196-204.
30. S. Brinkman and H. Bodschwinna, Advanced Gaussian filters, in L. Blunt, and X. Jiang, eds. Advanced Techniques for Assessment Surface Topography, Elsevier, London (2003) Chap 4.
31. T.J. Weller, A. Zheng, R. Thompson, and F. Tulleners, Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides, J. Forensic Sci. 57 (4) (2012) 912–917; doi: 10.1111/j.1556-4029.2012.02072.x.
32. "REPORT TO THE PRESIDENT Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods" by the Executive Office of the President, President's Council of Advisors on Science and Technology, September 2016.
33. JCGM 100:2008, Evaluation of measurement data—Guide to the expression of uncertainty in measurement (GUM) (Joint Committee on Guides in Metrology, 2008; http://www.bipm.org/en/publications/guides/gum.html.
34. ISO 17025—General requirements for the competence of testing and calibration laboratories, (ISO, Geneva (2005).
35. The American Society of Crime Laboratory Directors / Laboratory Accreditation Board (ASCLD/LAB), Measurement traceability policy, available at http://www.ascld.org.

**Table 1. The combined false positive and true positive CMC probabilities, $\hat{p}_{KNM}$ and $\hat{p}_{KM}$, for each correlated cell pair are estimated from the individual false positive and true positive identification probabilities of each identification parameter, which are estimated from the experimental distributions of the correlated cell pairs of the KM and KNM topographies by a conditional probability test method (see Fig. 8).**

| **Correlation parameters and thresholds** | | | **Individual false positive and true positive identification probabilities** | | | |
|---|---|---|---|---|---|---|
| $CCF_{max}$, $T_{CCF}$ = | | 50% | $P_{KNM(CCF)}$ = | 0.0782 | $P_{KM(CCF)}$ = | 0.7678 |
| $\theta$, | $T_{\theta}$ = | ±6° | $P_{KNM(\theta\|CCF)}$ = | 0.2673 | $P_{KM(\theta\|CCF)}$ = | 0.9800 |
| $X$, | $T_x$ = | ±125 μm (±20 Pixels) | $P_{KNM(x\|CCF,\theta)}$ = | 0.3604 | $P_{KM(x\|CCF,\theta)}$ = | 0.9902 |
| $Y$, | $T_y$ = | ±125 μm (±20 Pixels) | $P_{KNM(y\|CCF,\theta,x)}$ = | 0.5000 | $P_{KM(y\|CCF,\theta,x)}$ = | 0.9951 |
| Combined false positive and true positive CMC probability $\hat{p}_{KNM}$ and $\hat{p}_{KM}$ | | | $\hat{p}_{KNM}$ = | **0.003765** | $\hat{p}_{KM}$ = | **0.7414** |

**Table 2. Estimated cumulative false positive and false negative error rates $E_1$ and $E_2$ based on the binomial model of Section 4 for the set of 780 cartridge image pairs with average effective cells $N$ = 26 (Fig. 5). Also shown are the individual false identification probabilities $R_1$ at selected CMC values of 9 and 26 and the individual false exclusion probabilities $R_2$ at selected CMC values of 0 and 2.**

| **CMC** | $E_1$ | $E_2$* | $\Phi_{(CMC)}$ | $\Psi_{CMC)}$ | $R_{1(CMC)}$ | $R_{2(CMC)}$ |
|---|---|---|---|---|---|---|
| 0 | | | $5.3\times10^{-16}$ | 0.91 | | $5.2\times10^{-17}$ |
| 2 | | | $1.4\times10^{-12}$ | $4.2\times10^{-03}$ | | $3.0\times10^{-11}$ |
| CMC < 6 | | $7.4\times10^{-09}$ | | | | |
| 9 | | | $2.2\times10^{-05}$ | $4.5\times10^{-16}$ | $2.3\times10^{-10}$ | |
| 26 | | | 0.00042 | $9.3\times10^{-64}$ | $2.5\times10^{-59}$ | |
| CMC ≥ 6 | $6.2\times10^{-10}$ | | | | | |

***Note:** Both $E_1$ and $R_1$ are estimated by the binomial probability model.