Quantifying the similarity between images for forensic analysis

Alicia Carriquiry

February 26, 2024

forensicstats.org

csafe
Center for Statistics and Applications in Forensic Evidence

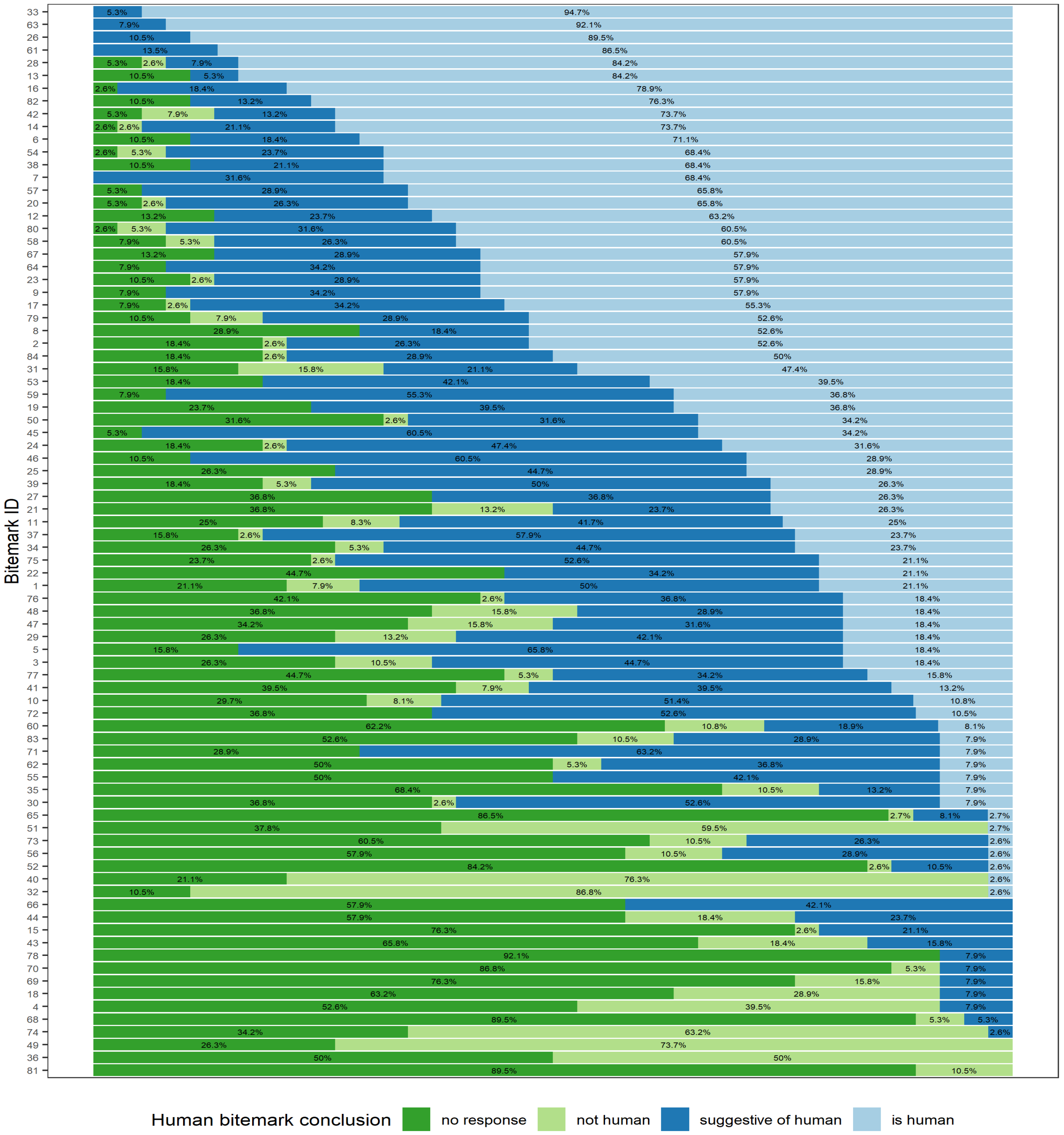
- Motivation
- Same or different source?
- Bayes' Rule and the likelihood ratio
- When data are images
  - Bullets
  - Handwriting
- Parting thoughts

Junk science

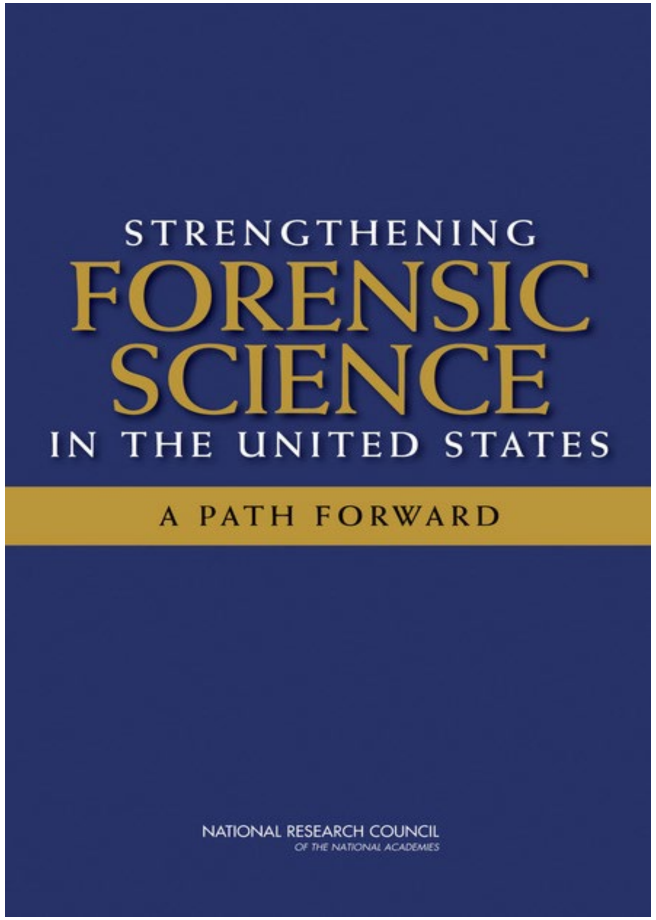
Forensic bitemark analysis has no scientific or empirical justification.

Little agreement

- Images of injuries from real cases.
- Reasonable agreement on 3 or 4 samples.



Questionable science



- Most forensic disciplines stand on shaky science.
- Worst are *pattern comparison* disciplines:
  - Latent prints
  - Ballistics
  - Shoe and tire treads
  - Blood stain patterns
  - Handwriting
  - Bitemarks

- Motivation
- Same or different source?
- Bayes' Rule and the likelihood ratio
- When data are images
  - Footwear analysis
- Parting thoughts

A crime is committed

- Crime scene investigators collect *evidence* from the scene.

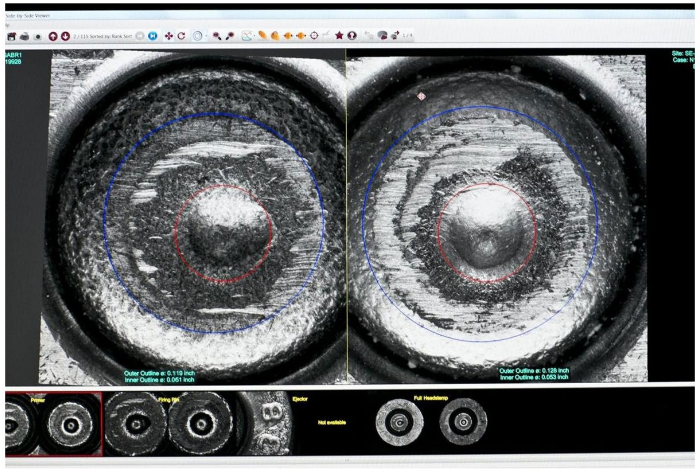
- Different types of evidence:
  - Biological (blood, saliva…)
  - Digital
  - Physical (glass, fibers…)
  - Patterns (fingerprints, shoeprints,…)

Questions that may be asked

- Was it a crime?
- Time and manner of death.
- Chemical composition of suspicious substance.
- Location where crime took place.
- Number of assailants.
- Source (or origin) of the evidence.

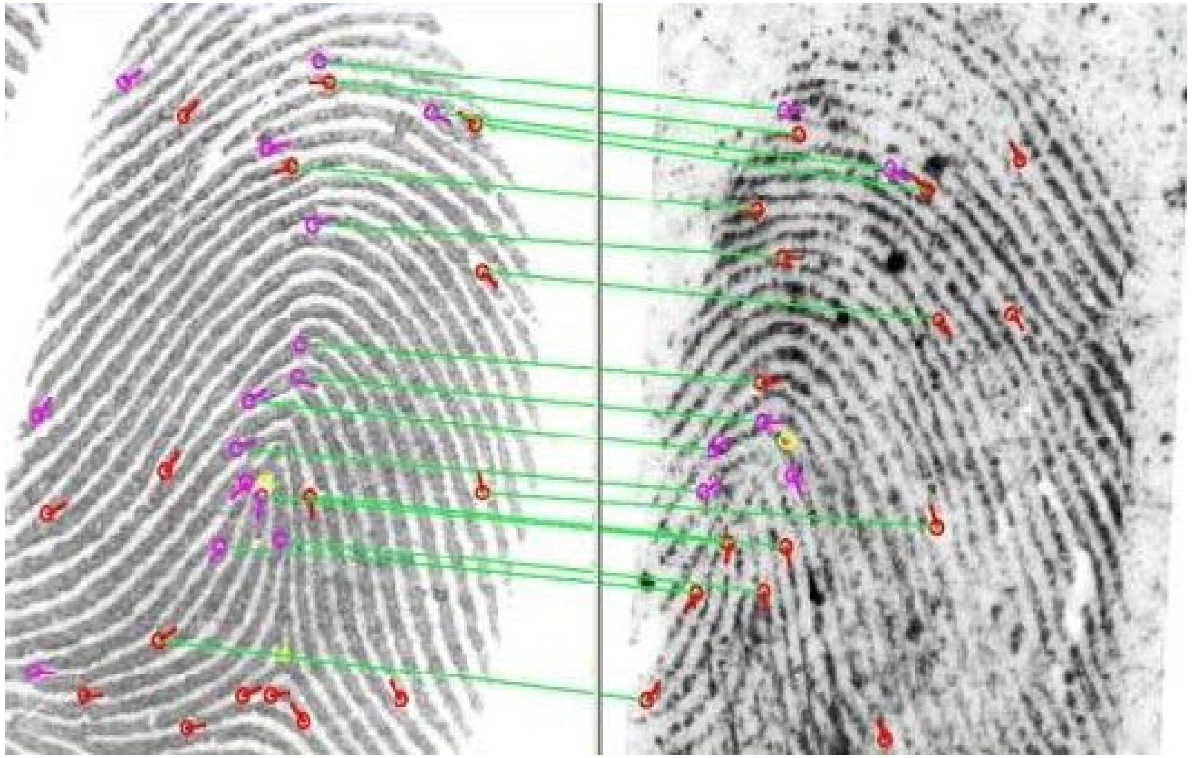
Questions about *source*

- Is the defendant’s finger the source of the latent print?
- Did fibers on the body come from defendant’s carpet?
- Did defendant’s gun fire the bullets?
- Was the shoeprint made by defendant’s shoe?



The status quo

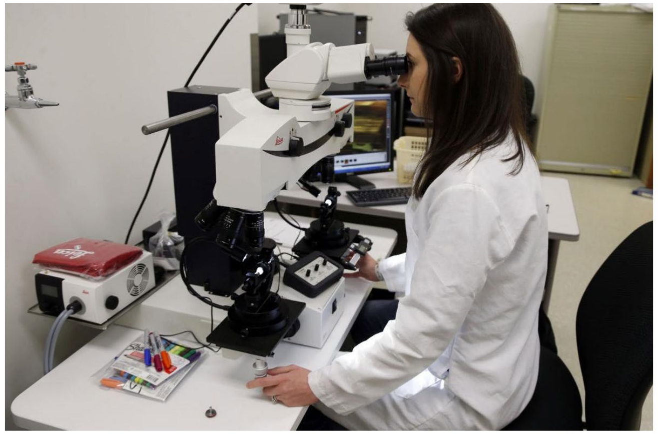
- Biological evidence: science to support forensic conclusions.
- Physical evidence: no *generative models* but we can obtain measurements.
- Pattern evidence:
  - No measurements.
  - Visual inspection.
  - Subjective conclusions.

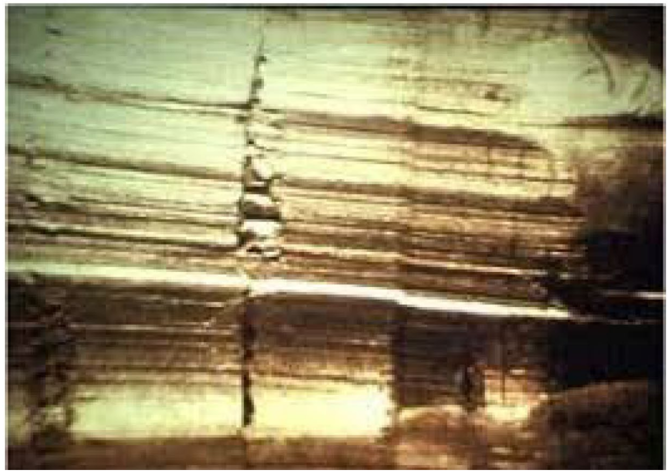


Firearms as illustration

The AFTE Theory of Identification:

1. …. *enables opinions of common origin to be made when the unique surface contours of two toolmarks are in **sufficient agreement**.*
2. *…. Agreement is significant when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool.*





Dubious conclusions

- "No other gun could have fired this round".
- "To a high degree of ballistic certainty".
- **In fact:**
  - **No well-designed studies.**
  - **"Inconclusive" instead of "exclusion".**

- Motivation
- Same or different source?
- Bayes' Rule and the likelihood ratio
- When data are images
  - Footwear analysis
- Parting thoughts

Formalizing the source problem

- Two competing hypotheses:
  - H_p: Defendant is source of evidence
  - H_d: Someone else is
- **Lindley, *Biometrika* 1977.**

Biometrika (1977), **64**, 2, *pp.* 207–13
Printed in Great Britain

A problem in forensic science

By D. V. LINDLEY
Department of Statistics and Computer Science, University College London

SUMMARY

The problem of deciding whether two sets of fragments have come from a common source frequently arises in forensic science. This paper provides a solution in the realistic case where the distribution is nonnormal. The normal case is also discussed because it is there easier to understand the nature of the solution and, in particular, its relationship to significance tests. The solution requires the distribution function of the product of standardized normal quantities which is tabulated in the appendix.

Some key words: Discrimination; Factor multiplying odds; Identification; Product of normal quantities; Significance test.

1. Solution in the normal case

The following problem arises in forensic science. Material is found at the scene of a crime and similar material is found on a suspect: do the two come from the same source, thereby suggesting the suspect was at the scene of the crime? An example, to which we will refer throughout the paper, occurs with window glass when breakage takes place on forcibly entering a building. Measurements are made of the refractive indices of the pieces of glass at the scene and of the fragments of window glass found on the suspect's clothing. Because of the small sizes of some of the fragments the measurements are subject to error. In this paper a solution to this identity problem is obtained and some interesting features of it discussed.

We suppose that the measurements are normally distributed about the true values with a known, constant variance σ^2. If m measurements are made at the scene a sufficient statistic is their mean X, normally distributed about the unknown true value θ_1 with variance σ^2/m. Let Y denote the mean of n similar measurements made on material found on the suspect: this is $N(\theta_2, \sigma^2/n)$. In the case of identity $\theta_1 = \theta_2$: otherwise it is supposed that $\theta_1 \neq \theta_2$. One remaining piece of information is the distribution of the true values. In the case of window glass there is considerable evidence about the distribution of refractive indices, some values being common, some rare. That such information is relevant is seen intuitively by considering the case where X and Y are close together, both being unusual indices. This gives greater evidence of identity than does the case where X and Y are equally close but are frequently occurring indices. In much of this paper we assume that the true values are normally distributed about μ with variance τ^2, both values being known. Typically τ will be larger, sometimes much larger, than σ. The normality assumption does not correspond to the practical situation where the distribution of refractive indices has a pronounced peak and a long tail to the right. Our justification for using the normal distribution is that we can get analytic results and consequently understand the situation more easily. We extend the argument to a general distribution where resort may have to be made to simple numerical integration though a possible approximation avoiding this is also provided.

Let I denote the event that the two sets of fragments come from the same source ($\theta_1 = \theta_2$)

Lindley's reasoning

- $X_1, X_2, \ldots, X_m$, $\bar{X} \sim N\left(\theta_1, \frac{\sigma^2}{m}\right)$ are crime scene measurements.
- $Y_1, Y_2, \ldots, Y_n$, $\bar{Y} \sim N\left(\theta_2, \frac{\sigma^2}{n}\right)$ are from suspect.
- H_p: $\theta_1 = \theta_2$, versus H_d: $\theta_1 \neq \theta_2$.
- Evidence in favor of same source (S):
  - $| \bar{X} - \bar{Y} |$ small.
  - Matching values are ***rare.***
- $\theta_1, \theta_2 \sim p(\mu, \tau^2)$

Therefore....

- Odds in favor of same source:

$$\frac{p(\bar{X}, \bar{Y} | S)}{p(\bar{X}, \bar{Y} | \bar{S})}$$

- Numerator: $\int p(\bar{X} \,|\theta) p(\bar{Y} \,|\, \theta)\, p(\theta\,) d\theta$
- Denominator: $\int p(\bar{X} \,|\, \theta_1)\, p(\theta_1) d\theta_1 \int p(\bar{Y} \,|\, \theta_2)\, p(\theta_2) d\theta_2$
- Intuition: ratio is function of product of two terms:

  $(\bar{X} - \bar{Y}\,)^2$: closeness, $(Z - \mu)^2$: rarity.
- Z is weighted average of $\bar{X}, \bar{Y}$

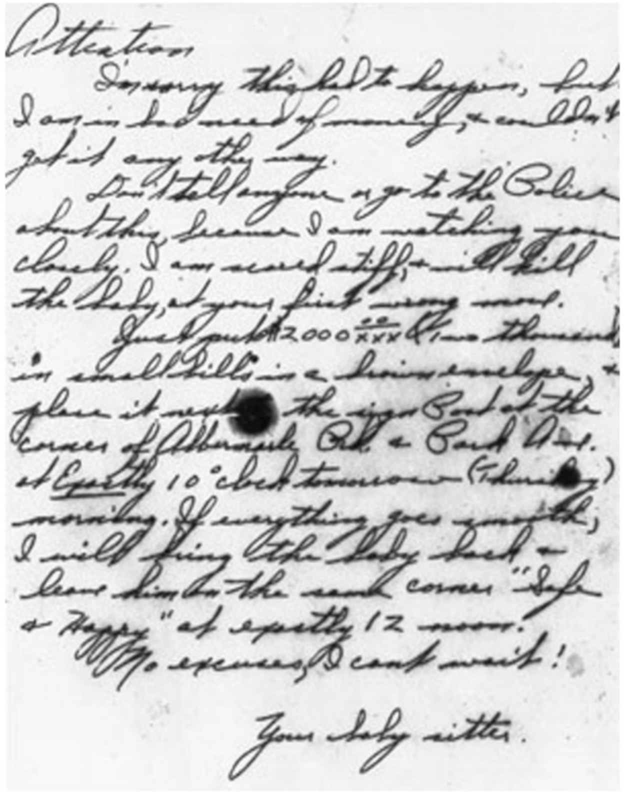
Bayes' Rule

- Imagine you are in the jury.
- Defendant is innocent until proven guilty.
- Evidence against defendant must be overwhelming.

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \frac{\Pr(H_p)}{\Pr(H_d)}$$

- Motivation
- Same or different source?
- Bayes' Rule and the likelihood ratio
- When data are images
  - Bullets
  - Handwriting
- Parting thoughts

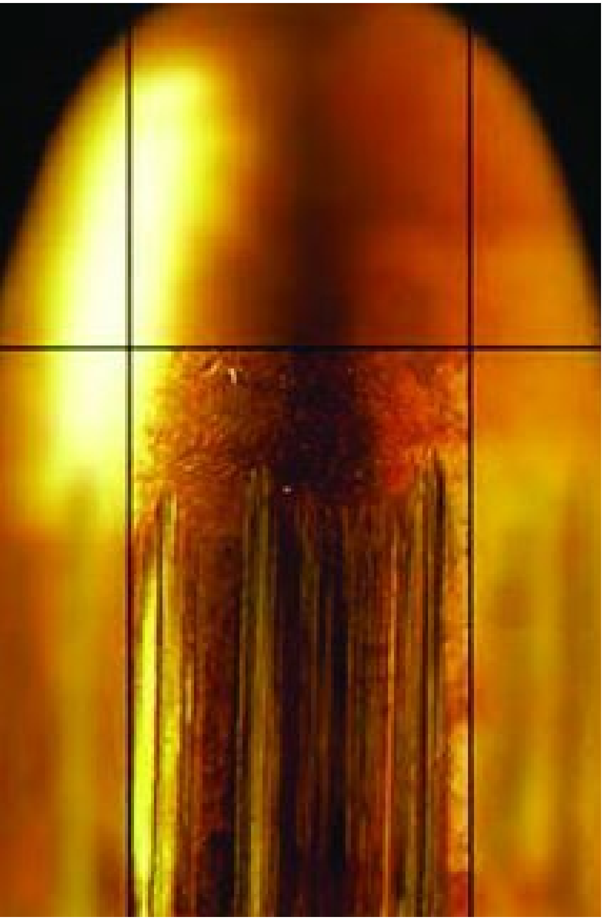
When data are images


Attention
I'm sorry this had to happen, but I am in bad need of money, & couldn't get it any other way.
Don't tell anyone or go to the Police about this, because I am watching you closely. I am scared stiff, & will kill the baby, at your first wrong move.
Just put $2,000 (two thousand) in small bills in a brown envelope, & place it [illegible] the sign post at the corner of Albemarle Rd. & Park Ave. at Exactly 10 o'clock tomorrow (Thursday) morning. If everything goes smooth, I will bring the baby back & leave him at the same corner "Safe & Happy" at exactly 12 noon.
No excuses, I can't wait!
Your baby sitter.

First Ransom Note in Weinberger Kidnapping

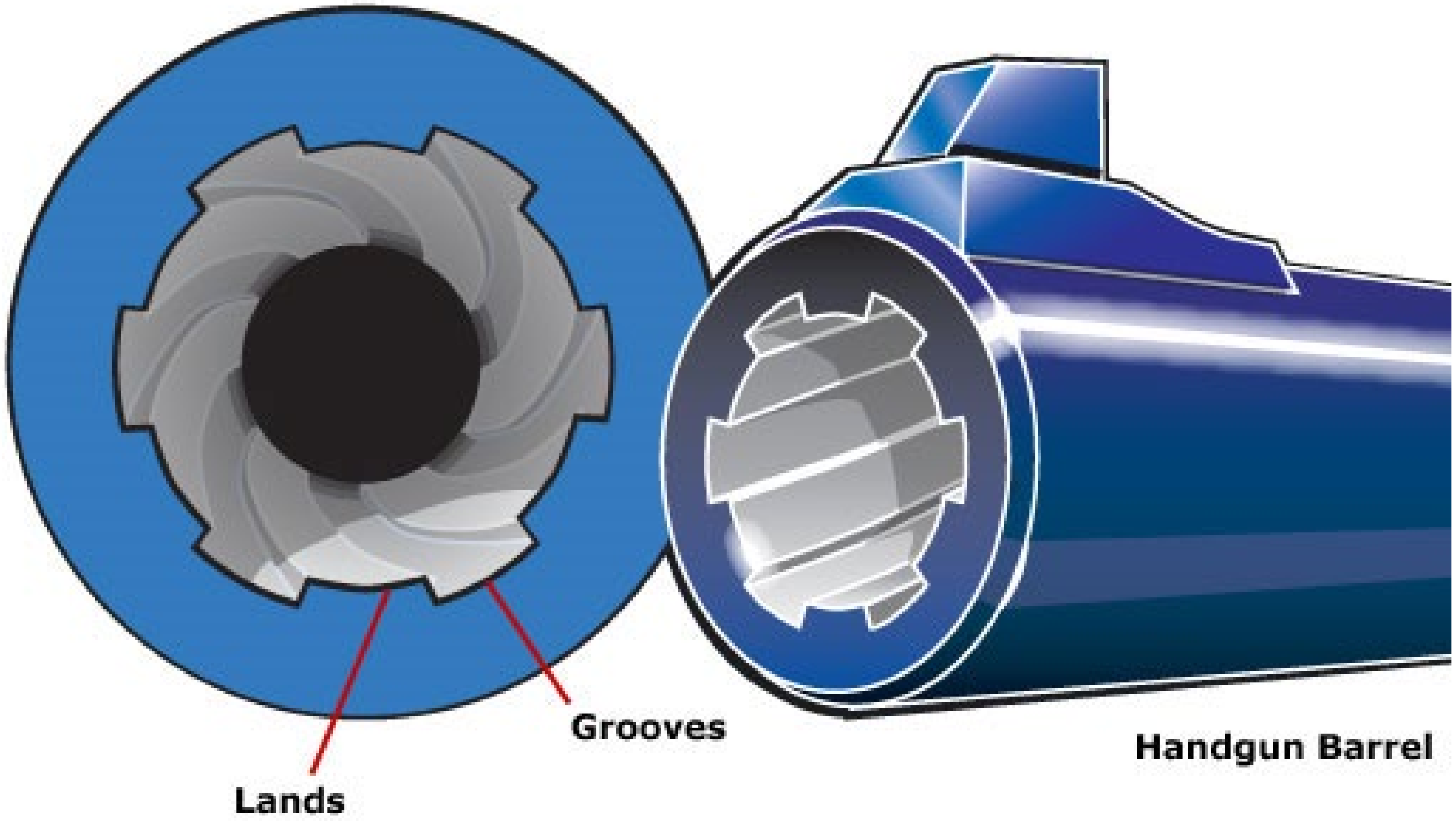
Bullets as example

Comparing striations on bullets



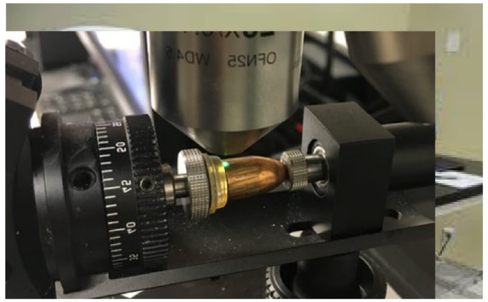
- Were two bullets fired from the same gun?
- Examiners compare striations with a comparison microscope.
- “Enough” matching striations: *identification* conclusion.
  - “Enough” is subjectively determined.
  - How many is enough?

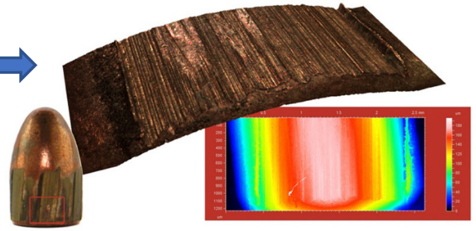
Inside of a barrel



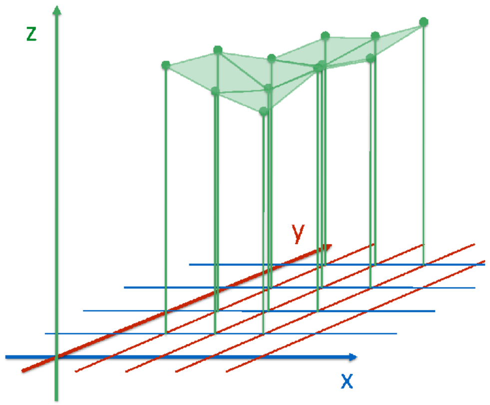
A quantitative approach

- 3D microscopes capture the surface of a bullet.



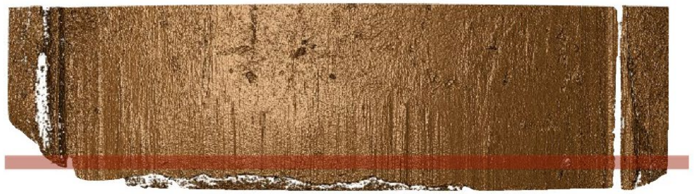


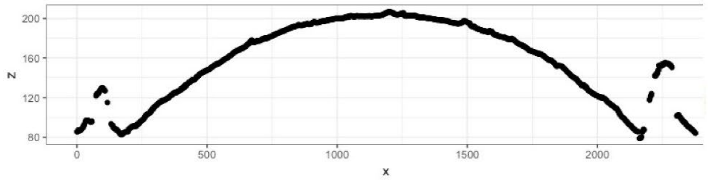
- Image represented as x-y-z coordinates:
    - x and y are coordinates on surface.
    - z is depth of surface at each x-y location.

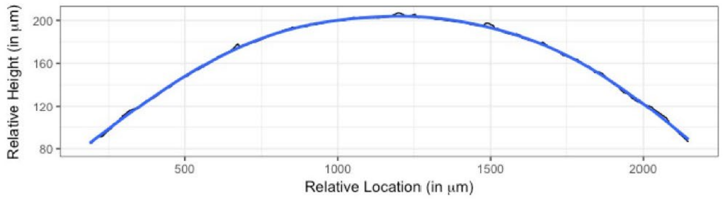


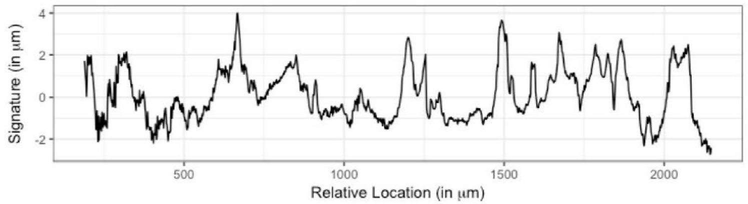
x	y	z
18.705	0.000	-25.221138
19.350	0.000	-25.253155
19.995	0.000	-25.335022
20.640	0.000	-25.418171
21.285	0.000	-25.477917
21.930	0.000	-25.541687
22.575	0.000	-25.673903
23.220	0.000	-25.966341
23.865	0.000	-40.070286
24.510	0.000	-40.407612
25.155	0.000	-40.587063
25.800	0.000	-33.437973
26.445	0.000	-33.691895
27.090	0.000	-39.690674
27.735	0.000	-40.317741

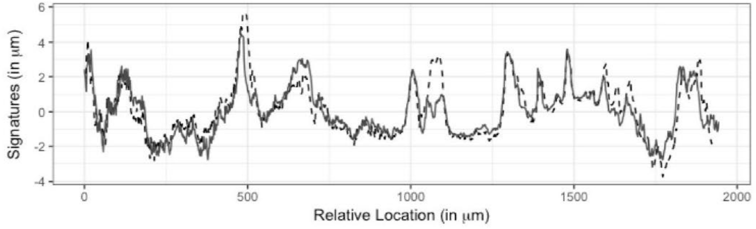
Algorithm - I





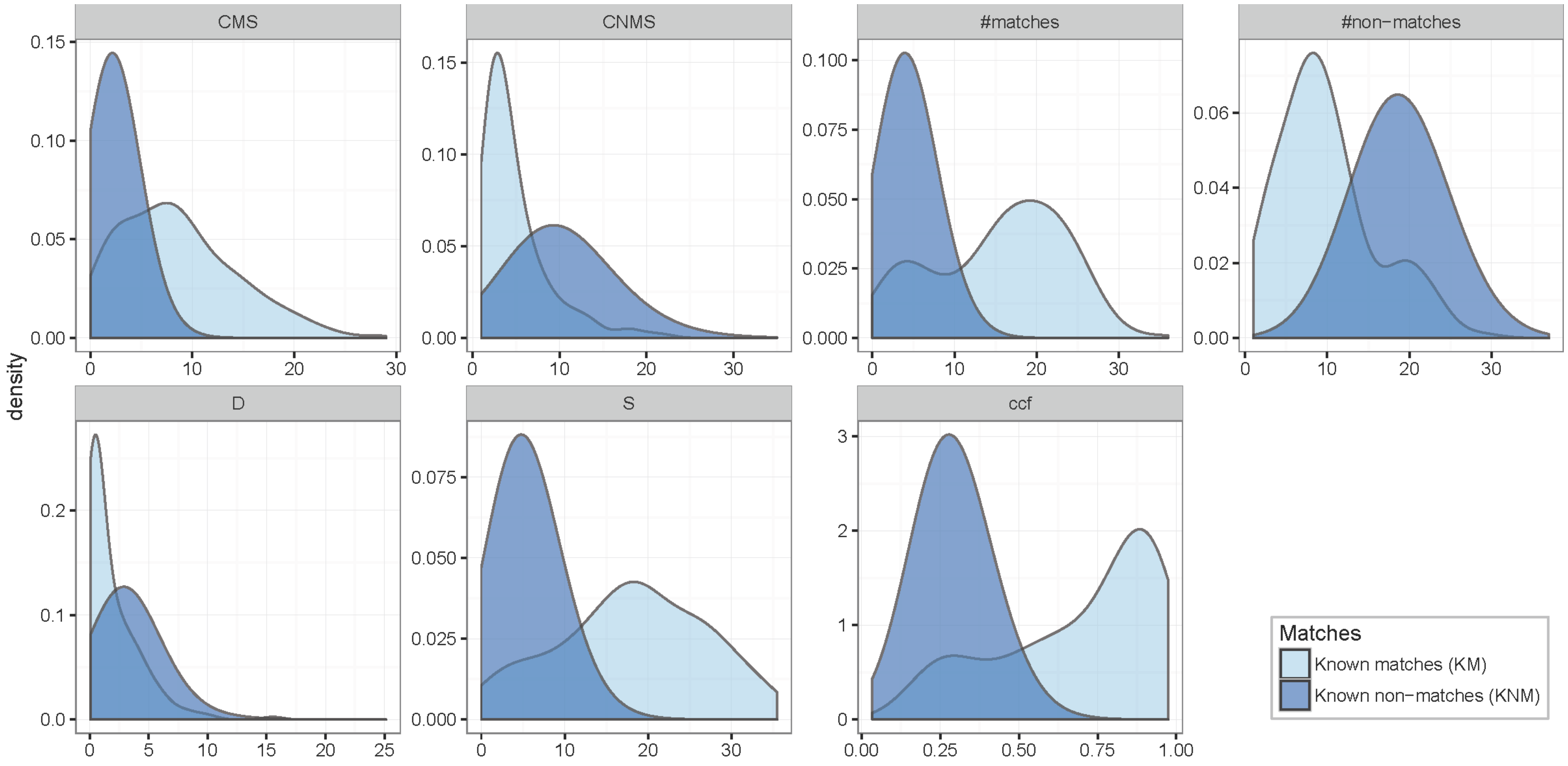






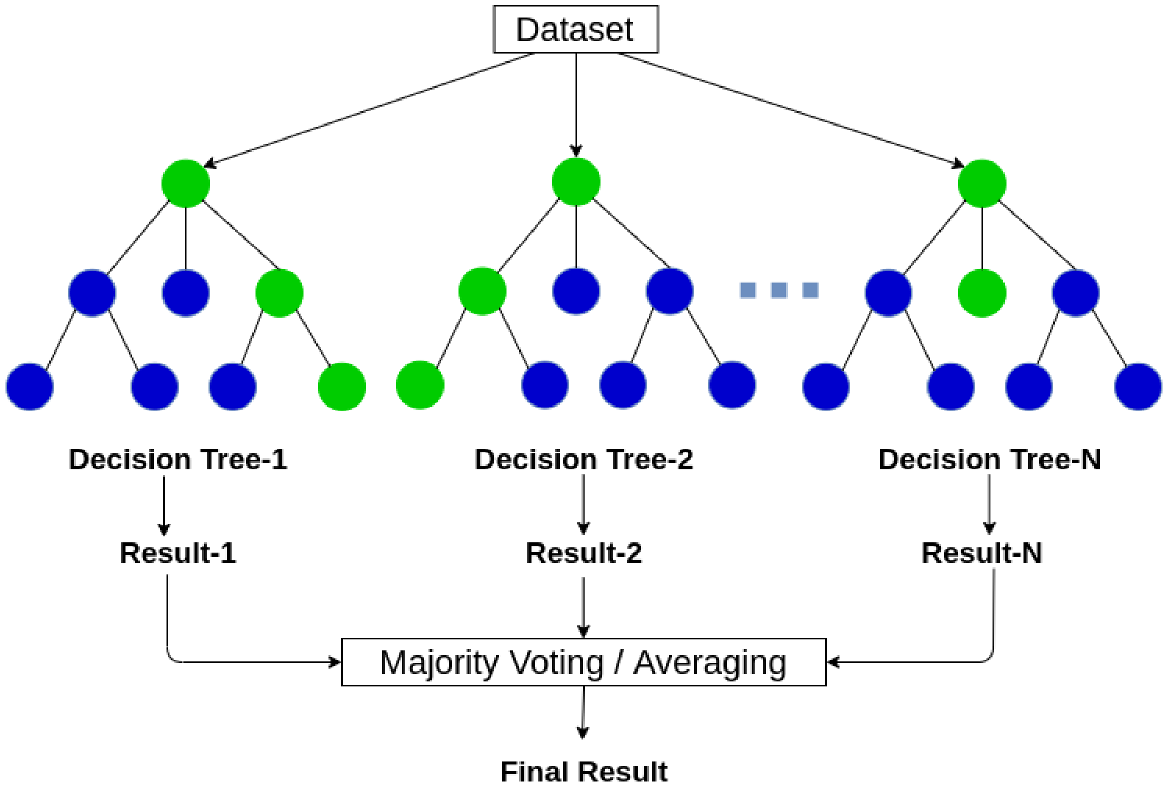
- Data are z values along cross-section.
- “Flatten” land: remove curvature.
  - Resulting series is the *signature.*
- Given signatures from two different bullets:
  - Overlay them.
  - Measure differences

Can we discriminate between SS and DS?



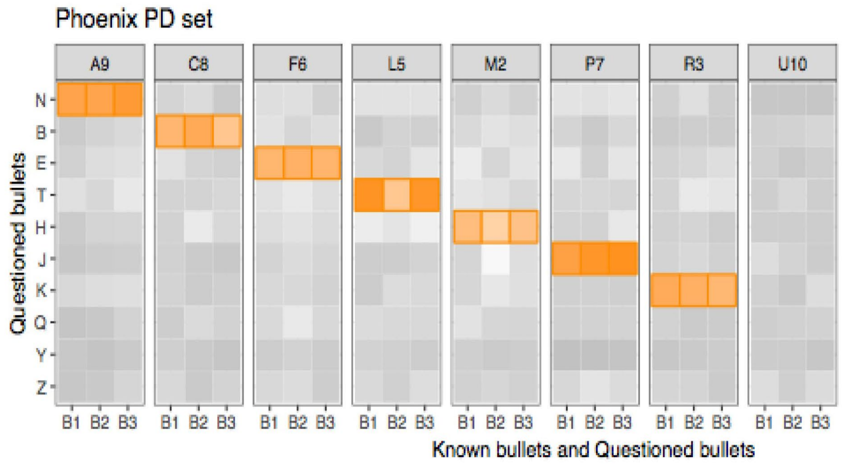
Algorithm - II

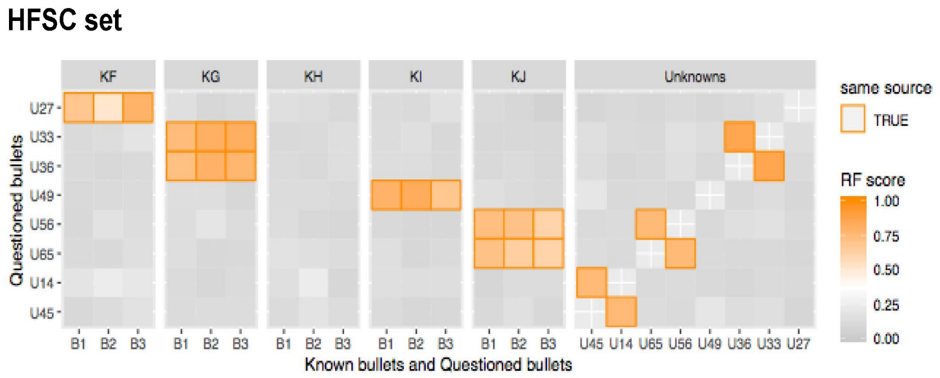
- We use 7 measurements.
- Combine 7 values into a **single similarity score** using a random forest.
- Random forest: collection of decision trees.
- Random forest must be trained.



Performance

- RF trained on “mated” and “non-mated” pairs of bullets.
- Testing on different sets –thousands of bullets.
- Accuracy high (for now)
  - Few false positives or negatives.



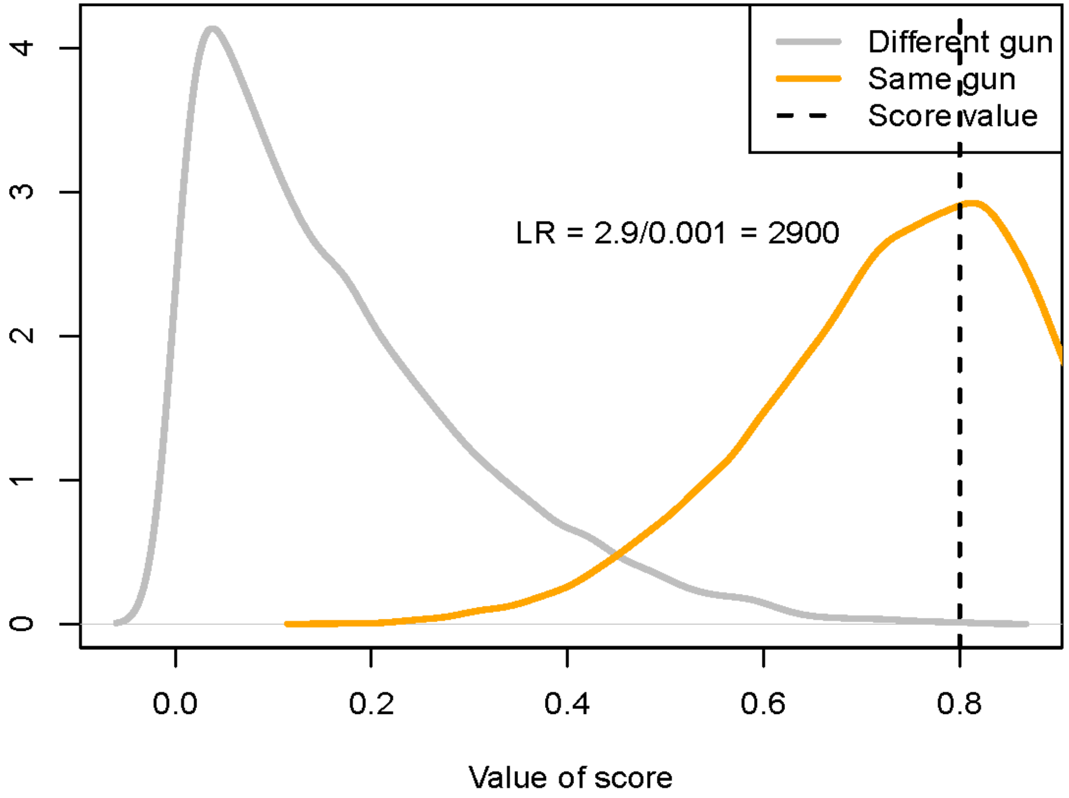


Score-based LR

- Empirical calculation of LR:
  - Quantify similarity of many pairs of mated and non-mated bullets
  - Get distribution of scores.
- Suppose that in a real case, similarity is equal to 0.8:

$$LR = \frac{height\ of\ same\ gun\ distribution\ at\ 0.8}{height\ of\ diff\ gun\ distribution\ at\ 0.8}$$

It is 2900 times more likely to observe 0.8 if bullets were fired from same gun.



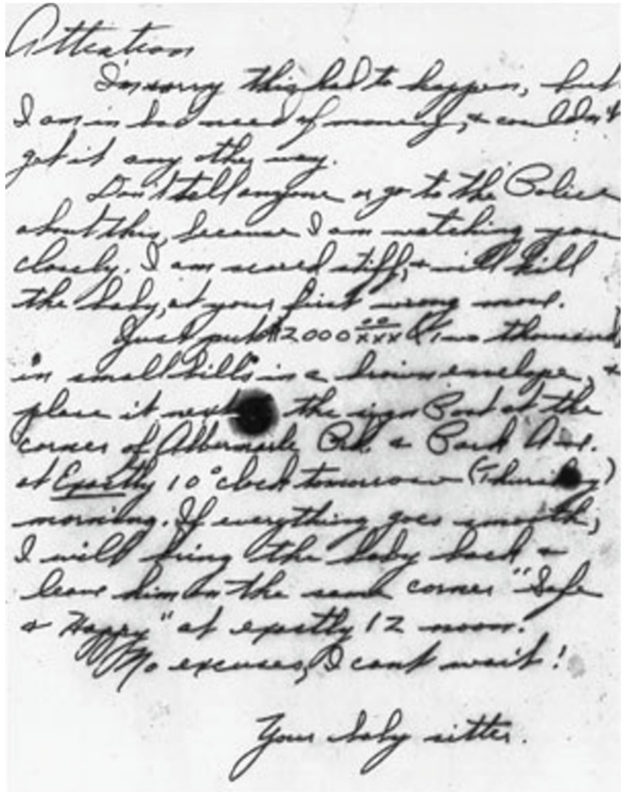
A few notes

- In study to be published soon, algorithm outperformed examiners:
  - No incorrect conclusions.
  - **No inconclusive conclusions** – about 30% inconclusives among examiners.
  - ***Vast majority of inconclusives corresponded to exclusions*.**
- Limitations:
  - Performance tests conducted on few gun/ammo combinations.
  - Expensive equipment.

Handwriting as data

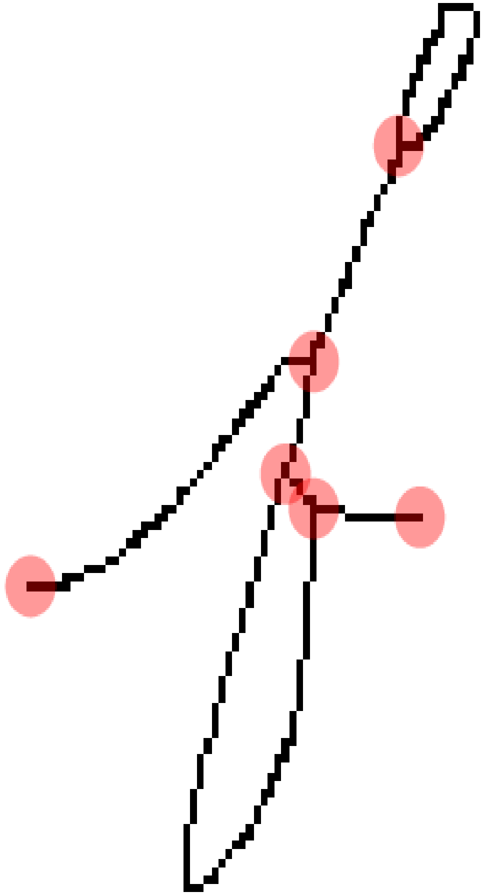
Handwriting as data

- Gray-scale scanned image of handwritten document.
- Focus is on *shape of writing*, not on content.
- A computer can identify:
  - Small graphical structures that <u>roughly</u> correspond to characters.
  - Individual words.
  - An entire line.
- We can extract "features" or data from any of them.



First Ransom Note in Weinberger Kidnapping

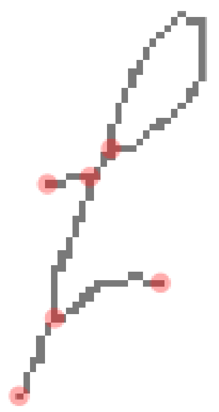
Characters as graphs

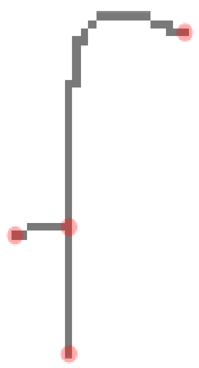


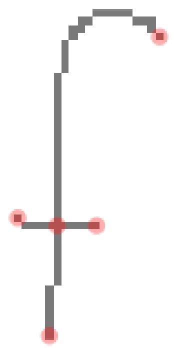
- Existing systems decompose writing into small graphical structures.
  - *Graphemes* in FLASH ID®.
- Graphs can be characterized by:
  - Number and geometric arrangement of nodes.
  - Attributes of edges connecting nodes.
  - Other attributes: slant, compactness,….
- `handwriter.R` uses a sequence of rules to section writing into graphs as in the left.

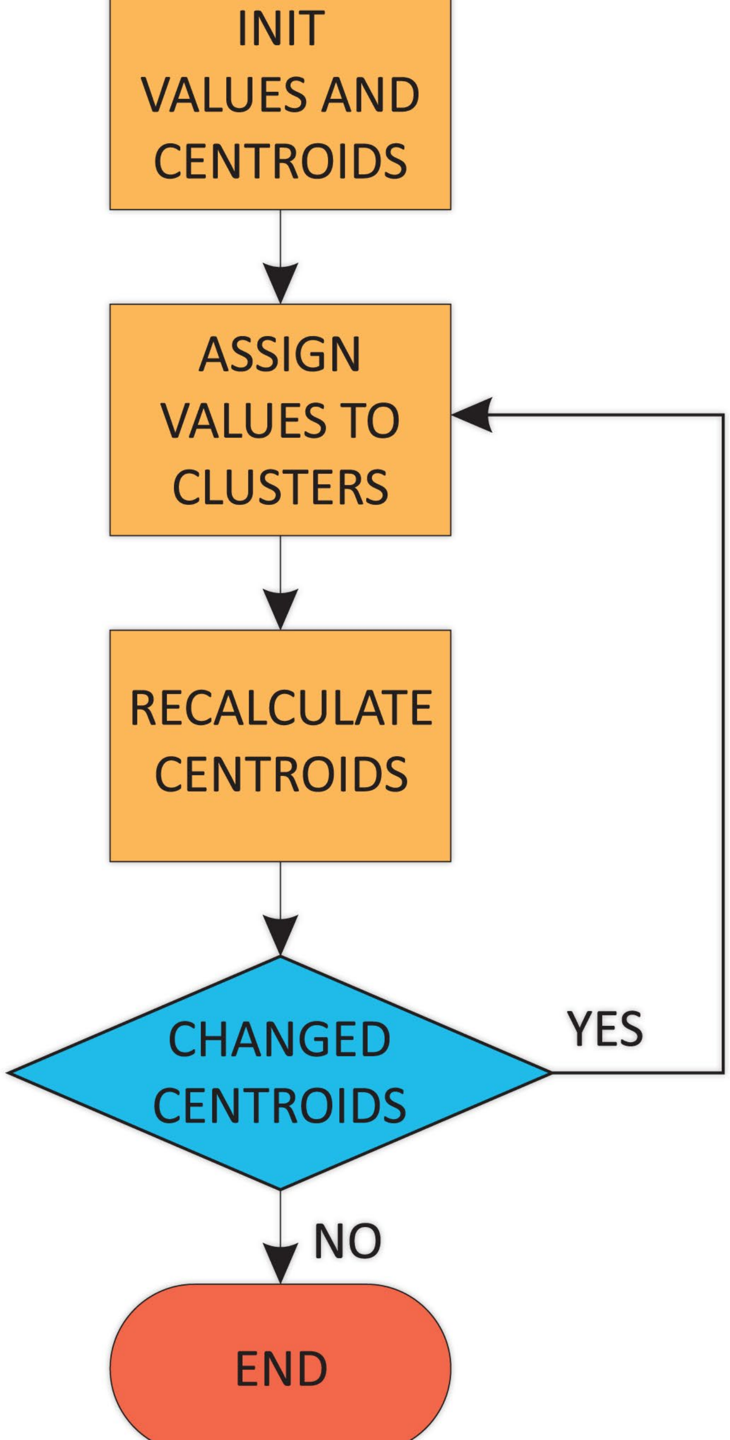
Grouping graphs

- A one-page document may contain hundreds of graphs.
- Can we group them into clusters of *similar* graphs?
- A strict definition of "similar" leads to thousands of groups:



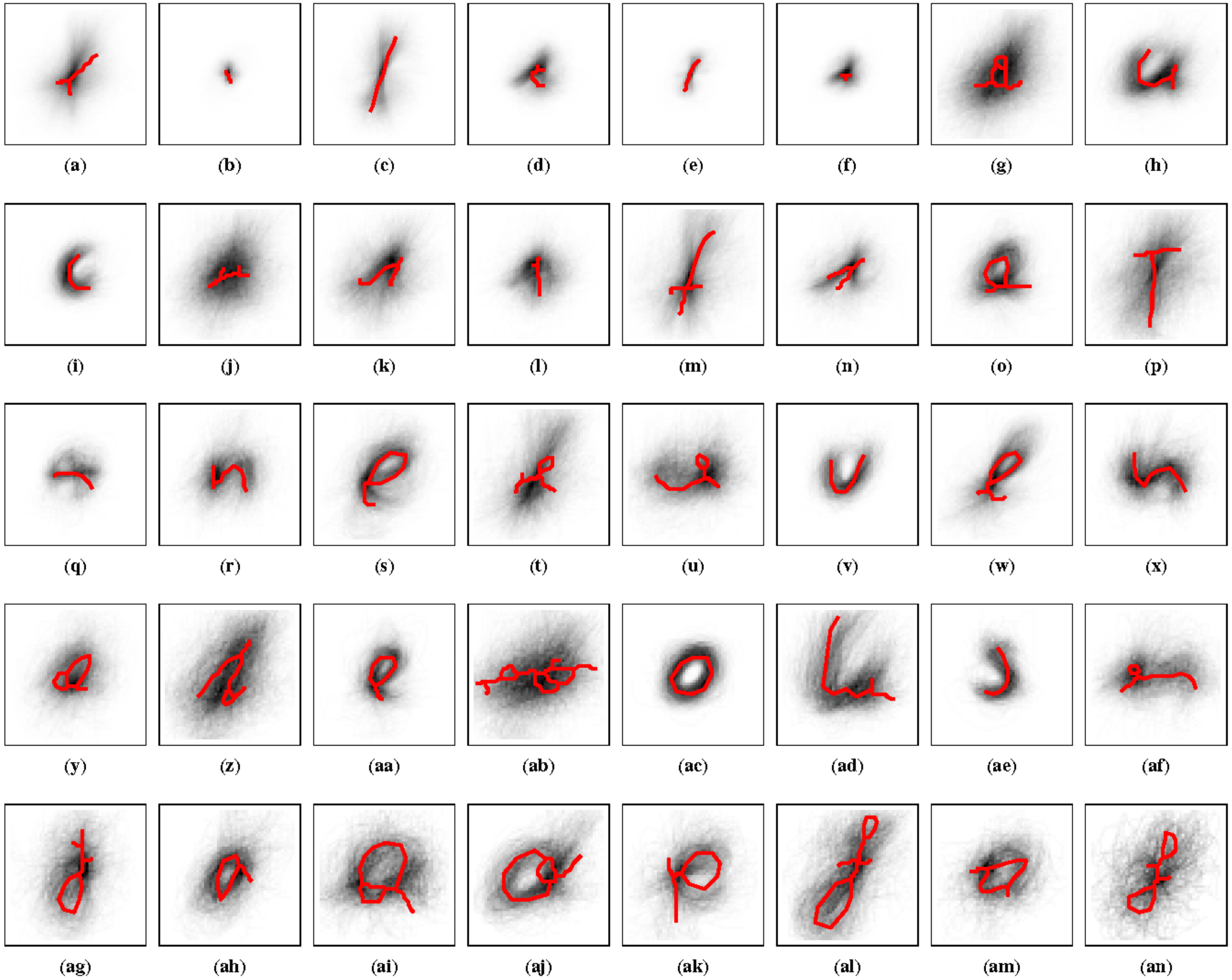






K-means clustering

- Choose a value *K*, say *K*=10.
- Pick 10 random graphs from the dataset to “initialize” each cluster.
- Now select another graph from the dataset and allocate it to the closest cluster.
- Iterate thousands of times, moving graphs between clusters.
- **Result: 10 clusters with “similar” graphs.**

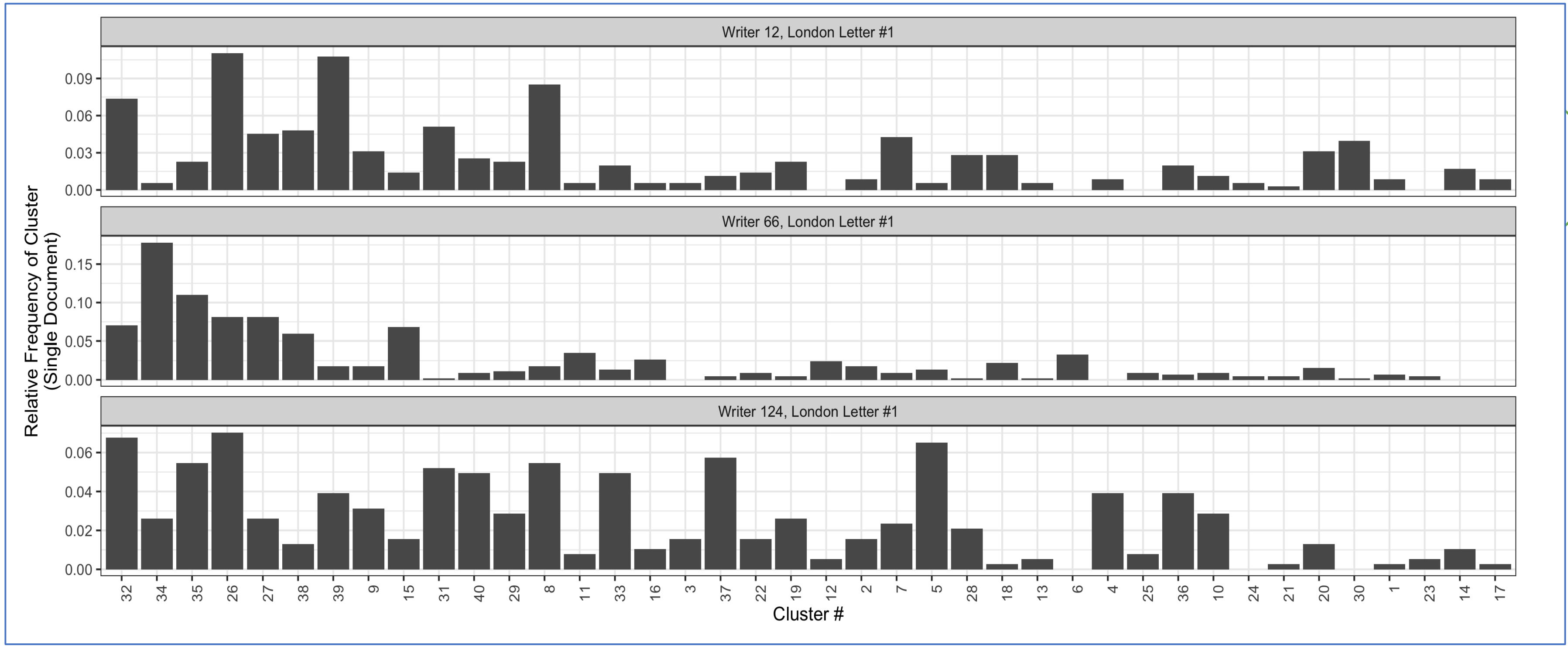

(a) (b) (c) (d) (e) (f) (g) (h)
(i) (j) (k) (l) (m) (n) (o) (p)
(q) (r) (s) (t) (u) (v) (w) (x)
(y) (z) (aa) (ab) (ac) (ad) (ae) (af)
(ag) (ah) (ai) (aj) (ak) (al) (am) (an)

Observed frequencies

$Y_{doc,writer}$	Cluster_1	Cluster_2	Cluster_3	Cluster_4	$\ldots$	Cluster_{39}	Cluster_{40}
$Y_{1,1}$	42	21	9	5	$\ldots$	1	1
$Y_{1,38}$	39	91	23	6	$\ldots$	0	1
$Y_{1,95}$	38	81	16	14	$\ldots$	0	0
$\vdots$							

Is cluster abundance informative?

- Can we use the frequency with which writers contribute graphs to clusters to identify a writer in a group?



Cluster sizes as data

- Extract and cluster graphs in each document, both Q and references.
- Y_{wj} = 40-dimensional vector of counts for the *j*th document written by the *w*th writer.
- Response vector is multinomial, so that:

$$Y_{wj} \sim \text{Multinomial}(\pi_w),$$

- The vector π_w is estimated for each writer: probability associated with each cluster.

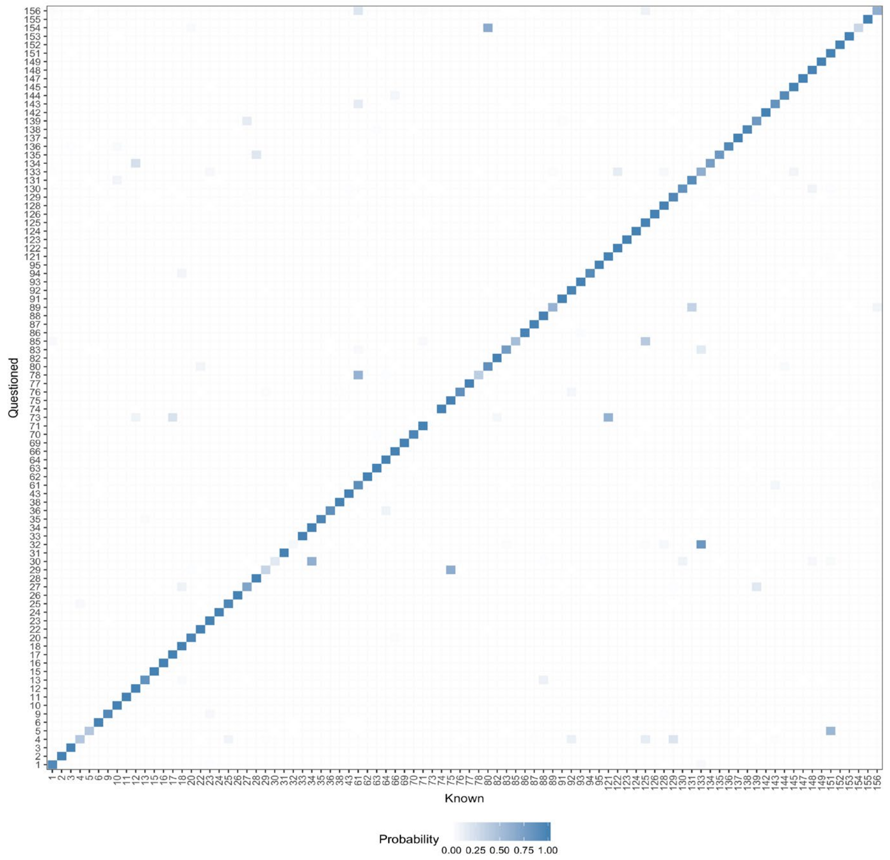
Implementation

- Create a cluster template using large, diverse collection of samples.
- Imagine creating a set of *K* “buckets”.
- Now, get a Q document and extract the graphs.
  - Each graph is put into the most similar bucket.
  - The proportion of graphs in each bucket for Q is the compared to the proportions observed for every other writer in the closed set.

Add more data

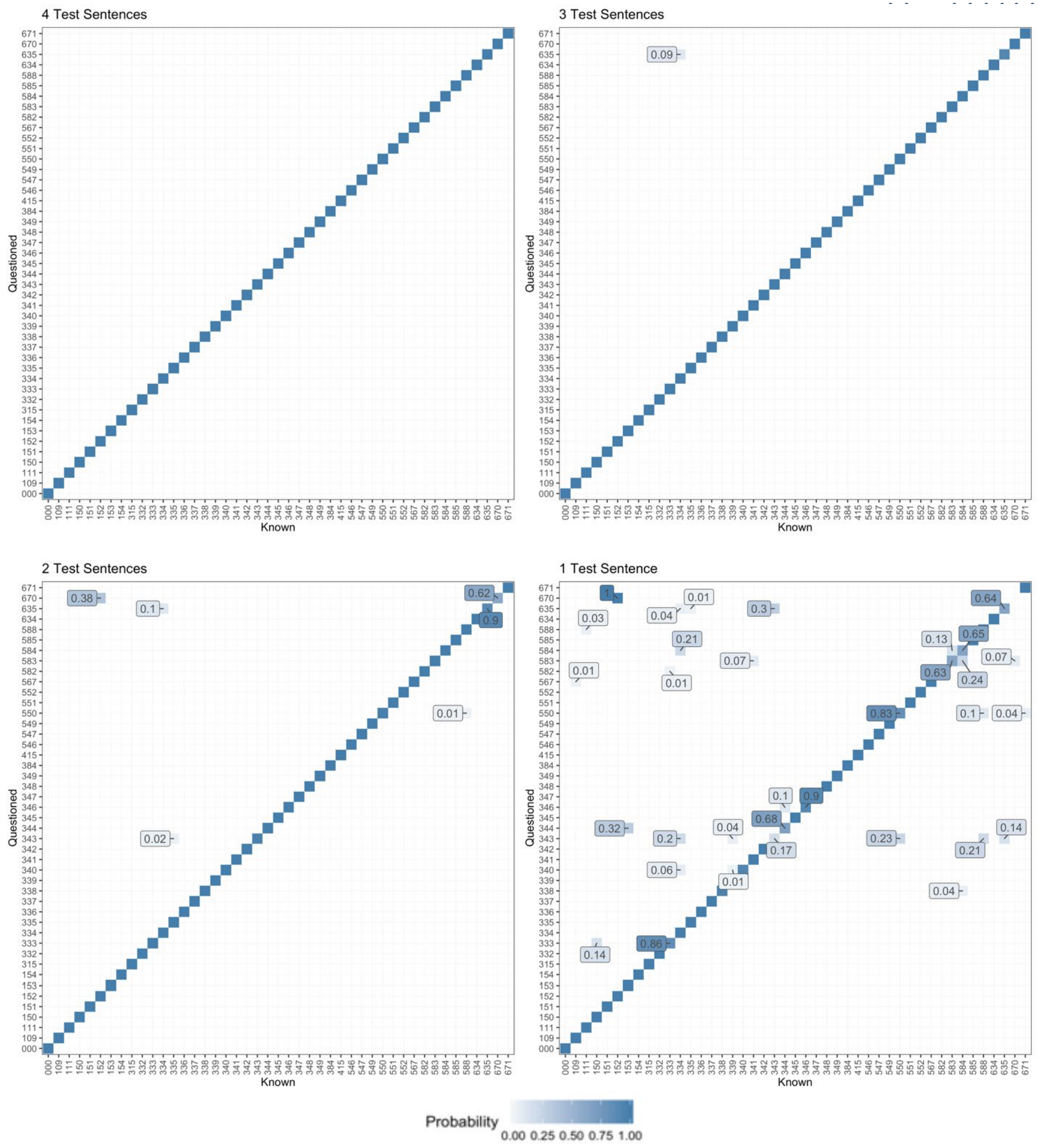
High accuracy

- Writers: mix of 90 writers from CSAFE, IAM, CVL databases.
- For 95% of Q documents, correct writer had > 0.9 probability of being identified.

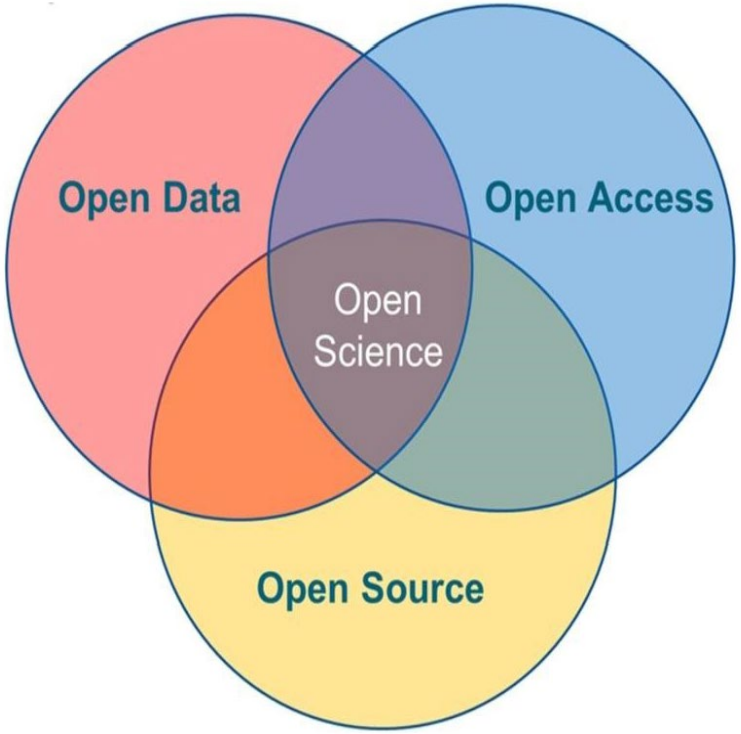


But…

- Accuracy decreases as evidence amount decreases.
- From the IAM database, we constructed test docs with 1, 2, 3 and 4 sentences.
- Issues:
  - Too few graphs
  - Too many clusters



Transparent, reproducible methods



- Trend is to expect ample access to data and code used by forensic practitioners.
- Proprietary software is difficult to test, must be trusted on faith.
- We produce open-source software to implement
  - Methods developed by CSAFE
  - Methods developed elsewhere (e.g., CMC method for cartridge case comparisons).

- Motivation
- Same or different source?
- Bayes' Rule and the likelihood ratio
- When data are images
  - Footwear analysis
- Parting thoughts

Much work to be done

- O

“All I had was plenty of time to die. Now, I don’t have enough time to live!”

THANKS

www.forensicstats.org

Or contact me:

Alicia@iastate.edu

csafe

Center for Statistics and Applications in Forensic Evidence