*Submitted to the Annals of Applied Statistics*

# A ROTATION-BASED FEATURE AND BAYESIAN HIERARCHICAL MODEL FOR THE FORENSIC EVALUATION OF HANDWRITING EVIDENCE IN A CLOSED SET

BY AMY M. CRAWFORD[1], DANICA M. OMMEN[2,*] AND ALICIA L. CARRIQUIRY[2,†]

[1]*Berry Consultants, LLC, amy.m.crawf@gmail.com*

[2]*CSAFE/Department of Statistics, Iowa State University, *dmommen@iastate.edu; †alicia@iastate.edu*

Forensic handwriting examiners are often tasked with identifying the writer of a particular document. Examples of handwriting evidence include ransom notes, forged documents and signatures, and threatening letters. At present, examiners rely on visual inspection of similarities and differences between the questioned document and reference writing samples. Here, we propose a principled modeling approach to compute the posterior predictive probability of writership when the author of the questioned document is part of a closed set of writers. Given a handwritten document, we extract document-level and character-level measurements which are the response variables in a multi-level model. We fit the model and test its posterior predictive performance using writing samples from the United States and from Europe. We find that as long as the questioned document is longer than a sentence or two, it is possible to correctly associate a writer with a document that he or she wrote with high probability. Earlier versions of this work have been well received by the community of forensic document examiners.

**1. Introduction.** Handwriting evidence has been used in a number of high profile cases, including the Lindbergh (Young, 2018a) and Weinberger (Young, 2018b) baby kidnappings in the 1930s and 1950s, respectively, and the BTK killer in the 2000s (Federal Bureau of Investigation, 2009). Forensic document examiners (FDEs) are the practitioners tasked with comparing handwriting on a questioned document and control samples written by the suspect to determine whether the suspect could have written the questioned document. At present, the comparison relies almost exclusively on visual inspection of the differences and similarities between the two samples of handwriting. Within the realm of forensic evidence interpretation, there are two subsets of problems, the closed-set and the open-set problem (Kwan, 1977). The following explanations of the closed-set and open-set problems in forensic document examination parallel the definitions of closed-set and open-set problems in machine learning (Choi et al., 2020). In the closed-set problem, investigators have some information that narrows the pool of suspects to a *watch list* of potential writers of the questioned document. In this case, the writer is assumed to be on the watch list. The goal of the FDE is to determine which person on the watch list wrote the questioned document. The closed-set problem can occur, for example, when a threatening note is found on school premises, and the writer can only be one of the students or staff members. From a statistical perspective, this problem consists of assigning a questioned document to its most likely writer. In contrast, the open-set problem occurs when one single person is suspected of committing a crime. Here, the FDEs job is to determine whether or not the specific suspect could have written the questioned document. In this article, we focus solely on the closed-set problem.

During an examination of a questioned document, the FDE relies on the subjective evaluation of features of handwriting and on the subjective assessment of their significance (Huber

*Keywords and phrases:* handwriter, rotation angle, London Letter, IAM data, K-means clustering, forensic identification.

and Headrick, 1999). There are many reasons for the subjective nature of the process, including the fact that “different examiners may detect or focus on different features, attach differing levels of importance to the same features, and have different criteria for reaching a conclusion" (Expert Working Group for Human Factors in Handwriting Examination, 2020). In 2009, the National Research Council (NRC) published a report on the state of forensic practice in the U.S. which stated that “the goal is to make scientific investigations as objective as possible so the results do not depend on the investigator" (National Research Council, 2009). The field of document examination had a head-start on developing algorithmic tools to objectively extract quantifiable features of handwriting using pattern recognition techniques even before the NRC report appeared. These algorithmic methods can assist trained FDEs as they perform their jobs by reducing the subjectivity related to feature detection (Leedham and Srihari, 2003).

Software packages that perform quantitative handwriting feature extraction include FLASH ID® (Sciometrics LLC, Chantilly, VA, USA) (Miller et al., 2017), WANDA (Franke et al., 2004), CEDAR-FOX (Srihari, Srinivasan and Desai, 2007), and `handwriter` (Crawford, Berry and Carriquiry, 2021) systems. WANDA and CEDAR-FOX provide options for interactive intervention by the FDE. For example, examiners can select different regions of interest in a document, like a particular word or letter, and then the system extracts the measurements from the selected component. These functions provide computer based assistance to the daily workflow of the examiner, but also the opportunity for cognitive biases to influence the resulting decisions. Neither of these systems has been updated in many years.

In contrast, the FLASH ID® system runs automatically without the need for frequent examiner intervention to extract handwriting features, with the exception of standard settings and input document selection. FLASH ID® extracts an extensive set of detailed measurements from each component of writing, leading to good discrimination between writers in a closed set. However, the FLASH ID® system can be expensive and the underlying algorithms are proprietary. While FLASH ID® is an undeniably useful and accurate tool, the use of proprietary software in the United States courts of law, has been questioned because the algorithms underlying those systems are not available for examination by both sides (see, e.g., United States v. Johnson (2016); State v. Pickett (2021)).

In the spirit of reproducible and open-source research, and to provide a non-proprietary alternative to existing commercial systems, the Center for Statistics and Applications in Forensic Evidence (CSAFE) developed an `R` package called `handwriter` (Berry, Taylor and Baez-Santiago, 2019). Details of the package utility can be found in Crawford, Berry and Carriquiry (2021). To extract features from writing, the primary function in the package takes a *png* image and binarizes the pixels so that each is either black or white to isolate handwriting from background noise. Then, the black pixels are reduced to a pixel-wide skeleton of the writing which is then decomposed into a series of smaller skeleton structures. These smaller graphical structures with nodes and edges characterize the writing. Furthermore, Crawford, Berry and Carriquiry (2021) describe a method of clustering these graphical structures using a modified $K$-means algorithm. The proportions of graphs assigned to each cluster in a document are the *features* that (Crawford, Berry and Carriquiry, 2021) propose be used to identify the writer of a document. The features showed promising performance in the closed-set writer identification case on a small example that used handwritten documents from the CVL database (Kleber et al., 2013). We use `handwriter` for feature extraction in this article.

After the features are extracted from the documents, Crawford, Berry and Carriquiry (2021) built a framework for closed-set writer identification tasks using statistically grounded methodology and publicly available datasets that is transparent, model-based, and provides probabilistic outcomes for closed-set writer identification. In this article, we improve on their

work in several ways. First, we extend the method of Crawford, Berry and Carriquiry (2021) by defining a graph-level feature that reflects writing slant (a feature that FDEs often consider in their subjective visual assessment). To do so, we compute the rotation-angle corresponding to the principal component decomposition of the graph embedded in the unit square. We propose a statistical model for these new features that takes into account their non-standard sample space. This model is incorporated into a probabilistic assessment of closed-set writership in a manner similar to Crawford, Berry and Carriquiry (2021), with the intention of improving predictive accuracy and robustness. Then, we apply these methods to two different datasets, one collected by CSAFE and another by the Research Group on Computer Vision and Artificial Intelligence at the University of Bern. Finally, we show that the updated probability model utilizing both document-level and rotation-based, graph-level features outperforms the method of Crawford, Berry and Carriquiry (2021) that relies solely on document-level features. On the downside, we find that closed-set writer identification methods which rely on clustering the graphs into a pre-specified number of groups exhibit worsening performance as the length of the questioned document decreases relative to the length of the training documents.

## 2. Data.

2.1. *CSAFE Data.* Some of the data used in this manuscript were contributed by 90 writers in the CSAFE database (Crawford, Ray and Carriquiry, 2019). Every participant in the study completed three sessions of writing, with a waiting period of at least three weeks between sessions. During each session, participants were asked to complete a short survey, and transcribe three passages (or prompts), three times each (nine samples per writer, per session). The same three prompts were used in each session. The longest, the London Letter (Osborn, 1929), is a common handwriting exemplar. It includes every letter, both in upper- and lower-case forms, and every number. The next longest is a short passage from *The Wonderful Wizard of Oz* by Baum, L. Frank, illustrated by W.W. Denslow (1900), chosen for its more natural, conversational tone than that of the London Letter. The third and shortest prompt is the phrase: "The early bird may get the worm, but the second mouse gets the cheese."

The feature extraction and model development which we describe in the remainder of this paper use three London Letters and one Wizard of Oz sample from the first data collection session of each participant in the CSAFE database. We use the three London letters obtained from each writer to construct the training set, on which we rely to estimate model parameters. A single Wizard of Oz sample from each writer is used as the testing set, which we use to examine the performance of the model. The goal is to correctly identify the writer of each Wizard of Oz sample using the posterior predictive probability of writership computed from a Bayesian hierarchical model developed from the training samples. For a small number of the writers (9), we use only two London Letter samples in the training set, the rest (81) contribute all three samples. Thus, we have 261 training samples of handwriting and 90 testing samples. The London Letter and the passage from the Wizard of Oz we used in this study are shown in Figure 1.

2.2. *IAM Data.* The IAM Handwriting Database, developed by the Research Group on Computer Vision and Artificial Intelligence at the University of Bern (Marti and Bunke, 2002), contains handwriting samples from 657 individuals. The forms that participants transcribed were built using content from the Lancaster-Oslo/Bergen (LOB) Corpus (Johansson, Leech and Goodluck, 1978). After scanning, the writing was segmented into lines, sentences, and words (Zimmermann and Bunke, 2002). We use the sentence-parsed data to investigate

Our London business is good, but Vienna and Berlin are quiet. Mr. D. Lloyd has gone to Switzerland and I hope for good news. He will be there for a week at 1496 Zermott Street and then goes to Turin and Rome and will join Colonel Parry and arrive at Athens, Greece, November 27 or December 2. Letters there should be addressed King James Blvd. 3580. We expect Charles E. Fuller Tuesday. Dr. L. McQuaid and Robert Unger, Esq., left on the 'Y. X.' Express tonight.

Within a short time she was walking briskly toward the Emerald City, her silver shoes tinkling merrily on the hard, yellow roadbed. The sun shone bright and the birds sang sweet and Dorothy did not feel nearly as bad as you might think a little girl would who had been suddenly whisked away from her own country and set down in the midst of a strange land.

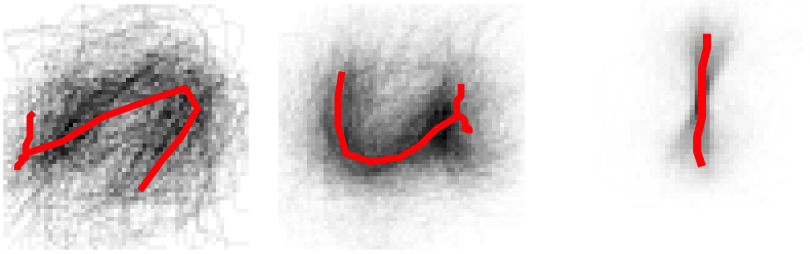
FIG 1. *The "London Letter" (top) and the "Wizard of Oz" (bottom) prompts.*

the performance of the hierarchical model that is fitted using 16 sentences for each writer in the closed-set, and tested on a decreasing number of sentences between four and one.

To build the analysis dataset, the full sentence repository is filtered based on two criteria. First, conscious of the fact that sentences are parsed based on form content, we exclude extremely short sentences (such as "No." or "They were not.") from consideration. Eligible sentences are those comprised of 30 or more graphs after decomposition by the `handwriter` R package (see Section 2.3 for further details). Second, we include writers who provided a sufficient number of eligible sentences. Requiring at least 20 eligible sentences per writer results in 45 viable writers for analysis. If a writer has an abundance of eligible sentences, we select 20 at random. The dataset used for analysis is therefore comprised of 900 sentences from 45 different writers. We divide the sentences into training and testing sets for model fitting and evaluation. For each writer, the sentences were ordered by length, with respect to the number of graphs extracted from each. Then, the 16 longest sentences for each writer were assigned to the training set, and the four shortest were held out for testing.

2.3. *Document Processing.* To begin, documents in both the training and testing sets were scanned, stored as *png* images at 72dpi, and processed by the `handwriter` package (Berry, Taylor and Baez-Santiago, 2019) to extract the graphs. Recall that eligible sentences in the IAM dataset consisted of 30 or more graphs. Grossly speaking, samples in the IAM testing set had fewer than 164 graphs, whereas samples in the IAM training set had more than 164 graphs. On average, a London Letter in the CSAFE training dataset is comprised of 376 graphs. The smallest number of graphs extracted from a CSAFE training document was 298, and the highest was 459. In the CSAFE testing set, the average number of graphs extracted from the Wizard of Oz passage was 302 (minimum of 228 graphs, maximum of 376 graphs). Documents in either of the test sets were, of course, left out of template creation and will not be used for model fitting. Other terminology used for the testing documents may include "questioned documents", or "holdout documents".

2.4. *Document-Level Features.* The primary feature extracted from each graph in a document is a cluster assignment based on a template similar to the one that was created and explained in Crawford, Berry and Carriquiry (2021). The template was generated via a $K$-means (Forgy, 1965; Lloyd, 1982) type algorithm, using novel definitions of center and distance to create 40 clusters. Each graph in a document is given a numerical cluster assignment (between 1 and 40) corresponding to the most similar cluster exemplar in the template, where similarity in the overall graph structures is quantified with the novel distance measure (Crawford, Berry and Carriquiry, 2021). Fig. 2 shows three of the clusters. The red graph is the



(a) Cluster #16 (b) Cluster #29 (c) Cluster #31

FIG 2. *The cluster exemplars (red) and cluster members (grey) from the final iteration of clustering template creation.*

cluster exemplar (the graph that is closest to every other graph in the cluster), whereas the shaded grey graphs in the background show the within-cluster variability in terms of graphs. Then, the document-level features are the numbers of graphs from that document in each of the 40 clusters. These features are often assembled into a vector of counts, similar to the method described in Saunders et al. (2011).

2.5. *Graph-Level Features.* Next, we introduce a graph-level feature, which we call *rotation angle*. This feature is extracted from every graph in a sample, and is used to quantify the inclination, lean, or slant of handwriting. Through a principal component decomposition, we find the direction of a graph with the greatest variability, and capture the angle of rotation corresponding to that direction, relative to the horizontal axis.

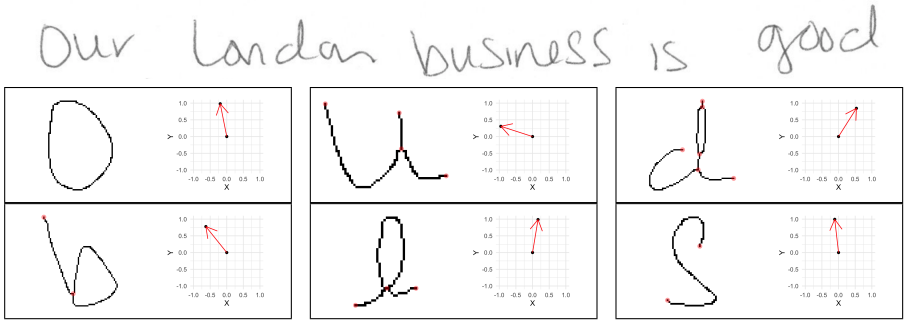
Suppose that we embed a graph in a unit square centered at $(0, 0)$ in Euclidean space while maintaining the original aspect ratio of the graph. For a graph with $n_p$ pixels, consider each pixel's coordinates $(x, y)$ in the unit square. Then the collection of all pixels in the graph can be expressed as $(X, Y) = (x, y)_1, \ldots, (x, y)_{n_p}$. The principal components are calculated through an eigen-decomposition of the covariance matrix,

$$\Sigma = \begin{bmatrix} var(X) & cov(X, Y) \\ cov(Y, X) & var(Y) \end{bmatrix},$$

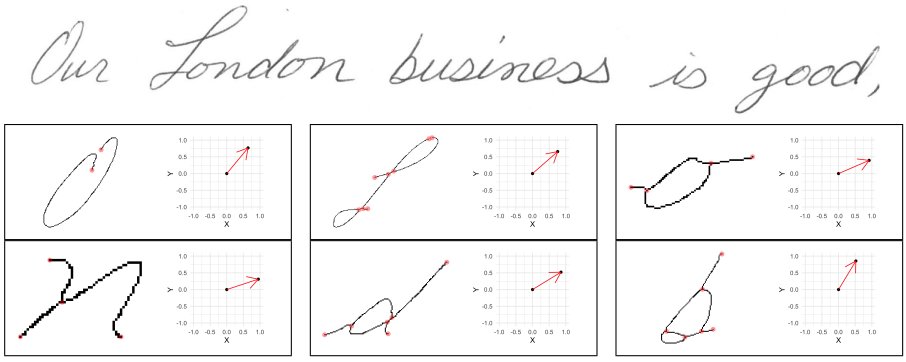
computed from the pixel coordinates. Since the first principal component can be identically defined in two opposite directions we simplify by enforcing that its direction lies in the first or second quadrant of the unit square. The rotation angle of the graph is defined as the angle between the unit vector (0,1) and the directional vector of the first principal component. Call the angle $\measuredangle \in (0, \pi)$.

Figure 3 shows a collection of graphs from both CSAFE Writer #1 and CSAFE Writer #95's training data along with the unit principal component vector on the upper half plane for each. CSAFE Writer #1 tends to use a connected print style of writing with characters that do not have a prominent slant one way or another. CSAFE Writer #95 uses a more formal cursive writing style, and has a consistent slant to the right.

We hypothesize that rotation angle comparisons are more meaningful when considered within cluster. Cluster assignment captures the overall shape of a graph, so two graphs with the same assignment will naturally have rotation angles that are similar. For example, clusters that tend to contain graphs that are taller than they are wide will have members with rotation angles near $\pi/2$, simply because of their structure. What is meaningful, are rotation angle comparisons within that cluster across the writers.



(a) Writer #1



(b) Writer #95.

FIG 3. *Writing samples from the CSAFE training documents of two writers, six graphs from each are included. Graphs are accompanied by a unit vector pointing in the direction of the rotation angle, ∠, for the corresponding graph, given by the principal component analysis. Graphs are plotted to fill the space in the figure, so pixels may be shown in different sizes.*

For each writer, it is the distribution of rotation angles within each cluster that we aim to capture in a Bayesian model. Rotation angles in all 40 clusters are used in the analysis, but we demonstrate data and results using three interesting clusters: #16, #29, and #31. These clusters are shown in Figure 2. Clusters #16 and #29 both include members of the training data that tend to be wider than they are tall. The opposite is true for Cluster #31.

The distribution of rotation angles for three different writer/cluster combinations are shown in Figure 4. Data for this figure are taken from all three training documents for each writer in the CSAFE dataset. On the left we show the distribution of rotation angles with traditional histograms. On the right, the angles are represented by a Nightingale rose diagram, or a radial histogram, in the polar coordinate system. See Brasseur (Brasseur, 2005) for a dialogue regarding the history of the rose diagram.

## 3. Methods.

3.1. *Model for Document-Level Features.* Let $k$ denote a cluster assignment $k = 1, \ldots, K$; here, we fixed $K = 40$, (see Crawford, Berry and Carriquiry (2021) for a discussion regarding the choice of $K$). Let $\boldsymbol{Y}_{w(d)}$ be a $K$-dimensional vector with elements equal to the number of graphs assigned to each cluster for document $d$, within known writer $w = 1, \ldots, W$. As noted earlier, the handwritten samples were provided by $W = 90$ writers in the closed-set of CSAFE writers or by $W = 45$ writers from the IAM dataset. Then, the

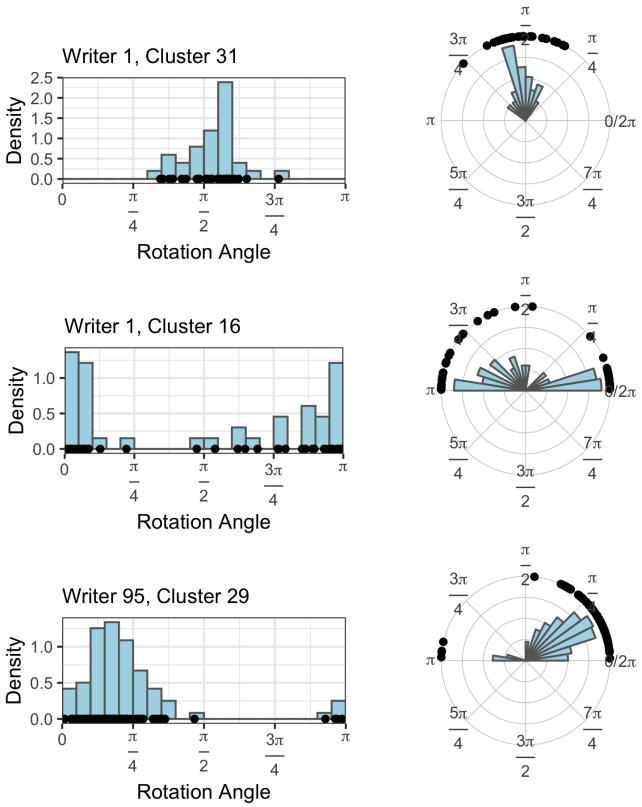



FIG 4. *For selected CSAFE Writer/Cluster combinations: Left, a traditional histogram of the rotation angles with 20 bins, a point for each observation is plotted on the x-axis. Right, a Nightingale Rose diagram displaying the rotation angles in 20 'petals'. A point for each observed rotation angle is plotted on the outermost ring of the diagram.*

model for the data is defined as

$$\boldsymbol{Y}_{w(d)}|\boldsymbol{\pi}_w \overset{ind}{\sim} Multinomial(\boldsymbol{\pi}_w), \tag{1}$$

where $\pi_w$ is the 40-simplex that captures the cluster-fill probabilities for writer $w$. We choose the prior for $\pi_w$ to be the Dirichlet distribution

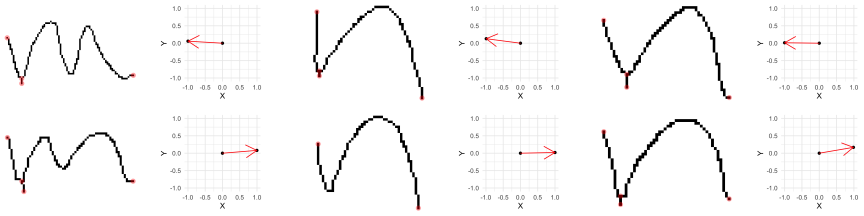
$$\boldsymbol{\pi}_w|\boldsymbol{\gamma} \overset{iid}{\sim} Dirichlet(\boldsymbol{\gamma}), \tag{2}$$

where $\boldsymbol{\gamma}$ is a vector of length 40 that can be thought of as pseudo cluster counts, across all writers. We specify independent priors for elements of the $\boldsymbol{\gamma}$ vector as

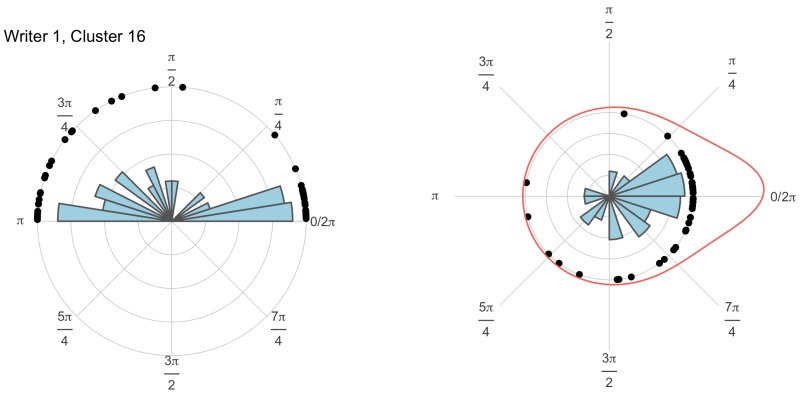
$$\gamma_k \overset{iid}{\sim} Gamma(a,\, b), \quad \forall k = 1, \ldots, K \tag{3}$$

and we use the $Gamma(shape, rate)$ parameterization so that the expectation of $\gamma_k$ is defined as $E(\gamma_k) = \frac{shape}{rate}$. A sensitivity analysis to changes in the final level prior parameters did not impact predictive outcomes, probably because we have a reasonably large set of data for each writer. A Bayesian probabilistic assessment of the questioned documents utilizing this model is conducted for a closed-set of CSAFE writers in Section 4.1.

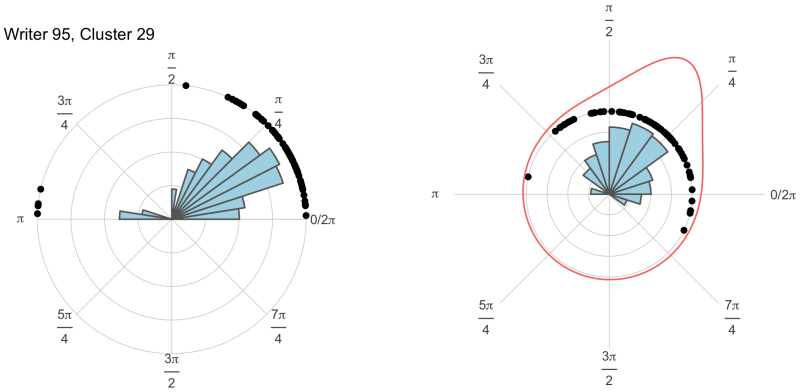
3.2. *Model for Graph-Level Features.* Rotation angles are circular random variables on the polar coordinate system, where the values 0 and $\pi$ are equivalent. Ignoring the circular nature of angles would not affect graphs assigned to clusters that are taller than they are wide,



(a.1) A collection of graphs from CSAFE Writer #1's training documents assigned to Cluster #16. To the right of each character graph is a depiction of the rotation angle measurement, $\measuredangle$.




(a.2) Distributions for CSAFE Writer #1, Cluster #16, corresponding to the elements in (a.1) above, on the half-circle (left, $\measuredangle$) and transformed to the full circle (right, $\theta$).




(b) Distributions for CSAFE Writer #95, Cluster #29, on the half-circle (left, $\measuredangle$) and transformed to the full circle (right, $\theta$).

FIG 5. *A demonstration of rotation angle behavior (a.1) and fully wrapped angles on the entire polar coordinate system (a.2) and (b). Left (a.2 and b): rose diagram with 20 petals for rotation angles on the half circle ($\measuredangle$), as originally measured. Right (a.2 and b): a 20 petal rose diagram for wrapped rotation angles on the full circle ($\theta$). A wrapped Cauchy density is shown in red with parameter values set to the MLEs for the corresponding data.*

such as Cluster #31, because most of the angles clump in the middle of the support. However, when the cluster includes graphs that are short and wide, like those in Cluster #29 or Cluster #16, modeling them as random variables on a circle cannot be avoided.

Figure 5(a.1) illustrates why it is important to model angles on a circle for clusters like #16 and #29. A few pixels added to the "corner" of a graph of this nature, can flip the rotation angle from near 0, to near $\pi$. Although the invariance of the principal component direction keeps the angles securely between 0 and $\pi$ during original feature extraction, the measure-

ments are meaningful in a full circle. Therefore, we take two times each measurement to achieve support on the full polar coordinate space. Figures 5(a.2) and 5(b) demonstrate the transition to the fully wrapped coordinate system, and a probability density function that is appropriate for circular data.

Let $\theta_{w,k,j} \in [0, 2\pi)$ denote the wrapped rotation angle of the $j^{th}$ graph that is assigned to cluster $k$ across all training documents from writer $w$. As before, $j$ is nested within the combination of $w$ and $k$, and $j = 0, 1, \ldots, J_{w,k}$. The relationship to the original measurement is

$$\theta_{w,k,j} = 2 * \measuredangle_{w,k,j},$$

where $\measuredangle$ denotes rotation angle and $\measuredangle \in [0, \pi)$.

Common probability models defined on the polar coordinate space include the wrapped normal distribution, the von Mises distribution, and the wrapped Cauchy distribution. The red circular densities shown in Figure 5 are estimated wrapped Cauchy densities with parameters set to the maximum likelihood estimators computed from the corresponding datasets. The wrapped Cauchy distribution is unimodal and symmetric, defined on the unit circle, and like many probability density functions, can be specified with a variety of parameterizations. We use the specification of Kent and Tyler (Kent and Tyler, 1988), who derived maximum likelihood estimates for the model parameters. The probability density function of the wrapped Cauchy distribution is

$$f(\theta; \mu, \tau) = \frac{1}{2\pi} \frac{1 - \tau^2}{1 + \tau^2 - 2\tau \cos(\theta - \mu)} \quad \theta \in [0, 2\pi),$$

where $\mu \in [0, 2\pi)$ is a location parameter, and $\tau \in [0, 1)$ controls the concentration of the observations around the location parameter. Setting $\tau = 0$ yields the circular uniform density.

The data model for the fully wrapped rotation angles is

$$\theta_{w,k,j} | \mu_{w,k}, \tau_{w,k} \overset{ind}{\sim} Wrapped\,Cauchy\,(\mu_{w,k}, \tau_{w,k})\,.$$

3.2.1. *Cluster Selection.* For this particular model, there is a large number of writer and cluster combinations for analyzing the rotation angles. To determine whether there was a subset of the clusters which provided more rotation angle information for distinguishing writers than other clusters, we examined each cluster separately to see whether the rotation angles were similar across all writers or variable enough to provide discriminating information. To this end, we choose a non-informative uniform prior for $\mu_{w,k}$. To choose a prior for $\tau_{w,k}$, we recall that a $\tau$ value of 0 in the wrapped Cauchy density function yields a circular uniform distribution. If the rotation angles in cluster $k$ computed from writer $w$ are widely distributed around the support $(0, 2\pi)$, we expect the estimated concentration parameter of the wrapped Cauchy distribution $\tau_{w,k}$ to be near 0 to reflect the uniform distribution of the angles in $(0, 2\pi)$. If all writers exhibit this behavior in cluster $k$, this would indicate that the rotation angles of graphs assigned to the cluster likely hold little or no information to aid in identifying the writer of a questioned document.

We can then select a mixture prior that forces each $\tau_{w,k}$ parameter to choose between a point mass at 0, or a larger value in the interval $(0, 1)$. The prior for $\tau_{w,k}$ is then

$$\tau_{w,k} \sim \phi_{w,k} Beta(4, 1) + (1 - \phi_{w,k}) Dirac(0),$$

with final level priors

$$\begin{aligned} \mu_{w,k} &\sim Uniform(0, 2\pi) \\ \phi_{w,k} &\sim Bernoulli(\rho_k) \\ \rho_k &\sim Beta(1, 1) \end{aligned} \tag{4}$$

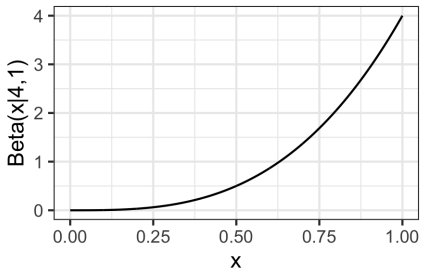



FIG 6. *The Beta(4,1) distribution.*

The $Beta(4,1)$ has mass concentrated on the upper end of the support, as shown in Figure 6. By constraining the shape of the Beta component we avoid identifiability issues in the mixture. Sensitivity analysis reveals that the estimated probabilities of writership that are obtained by fitting this model are only weakly sensitive to the choice of parameter values for the Beta distribution component in the mixture prior.

The model was fitted using a single chain, that was run for 3000 iterations after burn-in. At each MCMC iteration, the Bernoulli distributed mixing parameter $\phi_{w,k} \in \{0,1\}$ acts as a density selector for an entire set of rotation angles in a cluster for a writer. Figure 7 shows the distributions of angles for six writer and cluster combinations evaluated at the posterior means of $\mu_{w,k}, \tau_{w,k}$, along with the posterior average of the mixing parameter $\phi_{w,k}$, say $\hat{\bar{\phi}}_{w,k} = \frac{1}{3000}\sum_{m=1}^{3000} \phi_{w,k}^{(m)}$, which corresponds to the proportion of times that $\tau_{w,k}$ was sampled from the $Beta(4,1)$ density rather than the point mass at 0 in the MCMC routine.

In the top left of Figure 7, the data from CSAFE Writer #2 and Cluster #26 is abundant, but not concentrated in any particular direction. These data result in the lowest estimated $\hat{\bar{\phi}}_{w,k}$ across all writer and cluster combinations. In the 3000 MCMC draws, the parameter $\phi_{2,36}$ was equal to 1 only once. At the other extreme, the mixing parameter for CSAFE Writer #92, Cluster #38 data was sampled as 1 at every iteration. This was not a unique scenario, and a variety of writer, cluster combinations exhibit the same behavior. Figure 8 gives a summary of the distribution of estimated $\hat{\bar{\phi}}_{w,k}$ in each cluster $k$. It also includes a summary for the value of $\rho_k$ for each cluster (designated by a red $\times$), which lends insight into the overall rate with which rotation angle sets in the cluster tend to select the wrapped Cauchy distribution over the uniform.

There are clusters that exhibit rotation angles with a clear orientation for a great majority of writers. These clusters are on the right side of Figure 8, where the $\hat{\bar{\phi}}_{w,k}$ values are high for most writers, indicating that a wrapped Cauchy distribution is desirable to capture the shape of the wrapped rotation angle distributions.

The main goal of this exercise was to investigate whether it was possible to reduce the number of model parameters by identifying clusters that could be ignored without losing information about writership. For such clusters, the mixing parameters $\phi_{\cdot,k}$ would be close to zero for all writers, indicating that the cluster contain graphs with such large variability in rotation angles that a flat uniform distribution is preferred to a concentrated beta distribution, and result in a boxplot with all elements near zero. Rotation angles in such a cluster would be best modeled with a circular uniform density across all writers, effectively contributing a constant value to the likelihood evaluation in the predictive analyses. We would consider omitting such a cluster when modeling rotation angles, if one was observed. However, the boxplots in Figure 8 do not provide any evidence of such a cluster.

This approach is effective in capturing whether there is well-defined slant of the graphs in the cluster for each writer, but it does not indicate whether the slants are the same across all

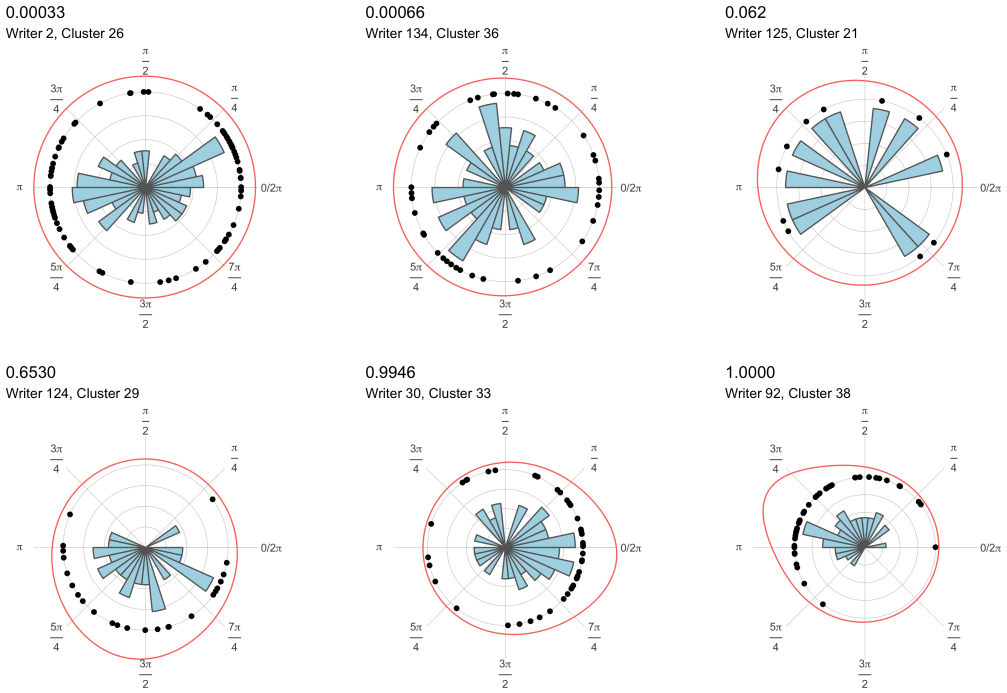



FIG 7. *Rose diagrams depicting wrapped rotation angles from six different writer/cluster combinations. Each includes a corresponding wrapped Cauchy distribution shown in red. The wrapped Cauchy distributions are evaluated at the posterior means of the location and concentration parameters. The posterior mean of* $\phi_{w,k}$ $\left(\bar{\phi}_{w,k} = \frac{1}{3000}\sum_{m=1}^{3000}\phi_{w,k}^{(m)}\right)$ *is shown on the top left of each panel, along with the writer and cluster to which the data were assigned.*

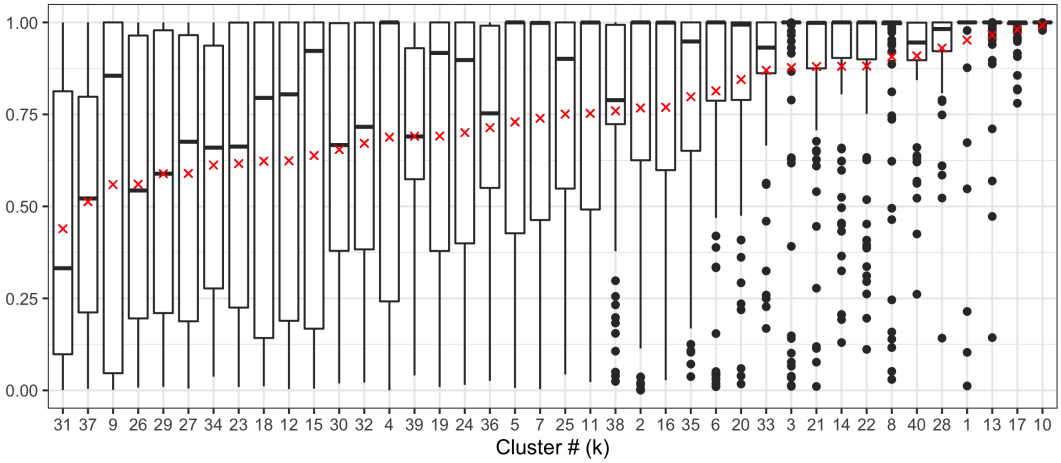



FIG 8. *For each cluster* $k$ *shown on the x-axis, a boxplot is constructed using* 90 *values of* $\hat{\bar{\phi}}_{w,k}$*, one for each writer. The red* $\times$ *shows the Monte Carlo estimate for the median of each* $\rho_k$*, from Equation 4.*

writers in the cluster. If so, then the angles in the particular cluster do not contribute to the estimation of probability of writership. If we wish to reduce the number of model parameters, it may be worthwhile determining the clusters for which the angle distributions have a mode

at the same location for all writers. These would also provide very little information to the final analyses.

3.2.2. *Final Model Specification.* Based on the analysis of rotation angles in each cluster performed in Section 3.2.1, the final model specification incorporates rotation angle measurements from all $K = 40$ clusters. Because there were no clusters that consistently had uniformly dispersed rotation angles for all writers, the mixture model for the concentration parameter, $\tau$, was abandoned in favor of a simpler Beta model. Also, we implemented a more informative prior for the location parameter of the wrapped Cauchy rotation angles than the previous non-informative uniform prior. Therefore, the final model for the graph-level features is defined as follows:

$$\theta_{w,k,j}|\mu_{w,k},\tau_{w,k} \overset{ind}{\sim} Wrapped\,Cauchy\,(\mu_{w,k},\tau_{w,k}) \tag{5}$$

$$\tau_{w,k} \sim Beta(c,d) \tag{6}$$

$$\mu_{w,k} \sim WrappedCauchy(\eta_k,e) \tag{7}$$

$$\eta_k \sim Uniform(0,2\pi). \tag{8}$$

A Bayesian probabilistic assessment of the questioned documents utilizing this model is conducted for a closed-set of CSAFE writers in Section 4.2.

## 4. Results.

4.1. *Document-Level Features Only.* We conduct a posterior evaluation of a single test document with data $\boldsymbol{Y}_{w^*}$, by unknown writer, $w^*$. Consider a particular writer, $w'$, from the closed-set of CSAFE writers. The goal is to estimate the probability that this particular writer was the author of the test document. Then, $\boldsymbol{\pi}_{w'}^{(m)}$ is the $m^{\text{th}}$ MCMC sample of the $w'$ multinomial parameter vector for $m = 1,\ldots,M$. Evaluate the multinomial likelihood under writer $w'$ at MCMC iteration $m$ as

$$q_{w'}^{(m)} = Mult(\boldsymbol{Y}_{w^*}^{*};\boldsymbol{\pi}_{w'}^{(m)}), \tag{9}$$

where the model for $\boldsymbol{\pi}_w$ is given by Equations 2-3 with $a = 2$ and $b = 0.25$:

$$\boldsymbol{\pi}_w|\boldsymbol{\gamma} \overset{iid}{\sim} Dirichlet(\boldsymbol{\gamma})$$

$$\gamma_k \overset{iid}{\sim} Gamma(a,\,b)$$

The evaluation is conducted for all known writers in the closed-set and stated together as

$$\boldsymbol{q}^{(m)} = \left[\, q_{w_1}^{(m)},\, q_{w_2}^{(m)},\, q_{w_3}^{(m)},\, \ldots,\, q_{w_W}^{(m)} \,\right], \tag{10}$$

and at MCMC iteration $m$, we recognize the writer with the largest likelihood evaluation with a "vote",

$$v_{w'}^{(m)} = \begin{cases} 1 & \underset{w_1,\ldots,w_W}{\text{argmax}}\; q^{(m)} = w' \\ 0 & o.w. \end{cases} \tag{11}$$

Then, aggregating across all MCMC iterations, evaluate the proportion of MCMC samples in which each writer is most likely to be the true writer of the questioned document by

$$\bar{\boldsymbol{p}} = \frac{1}{M}\left[\sum_{m=1}^{M} v_{w_1}^{(m)}, \sum_{m=1}^{M} v_{w_2}^{(m)}, \sum_{m=1}^{M} v_{w_3}^{(m)}, \ldots, \sum_{m=1}^{M} v_{w_W}^{(m)}\right], \tag{12}$$

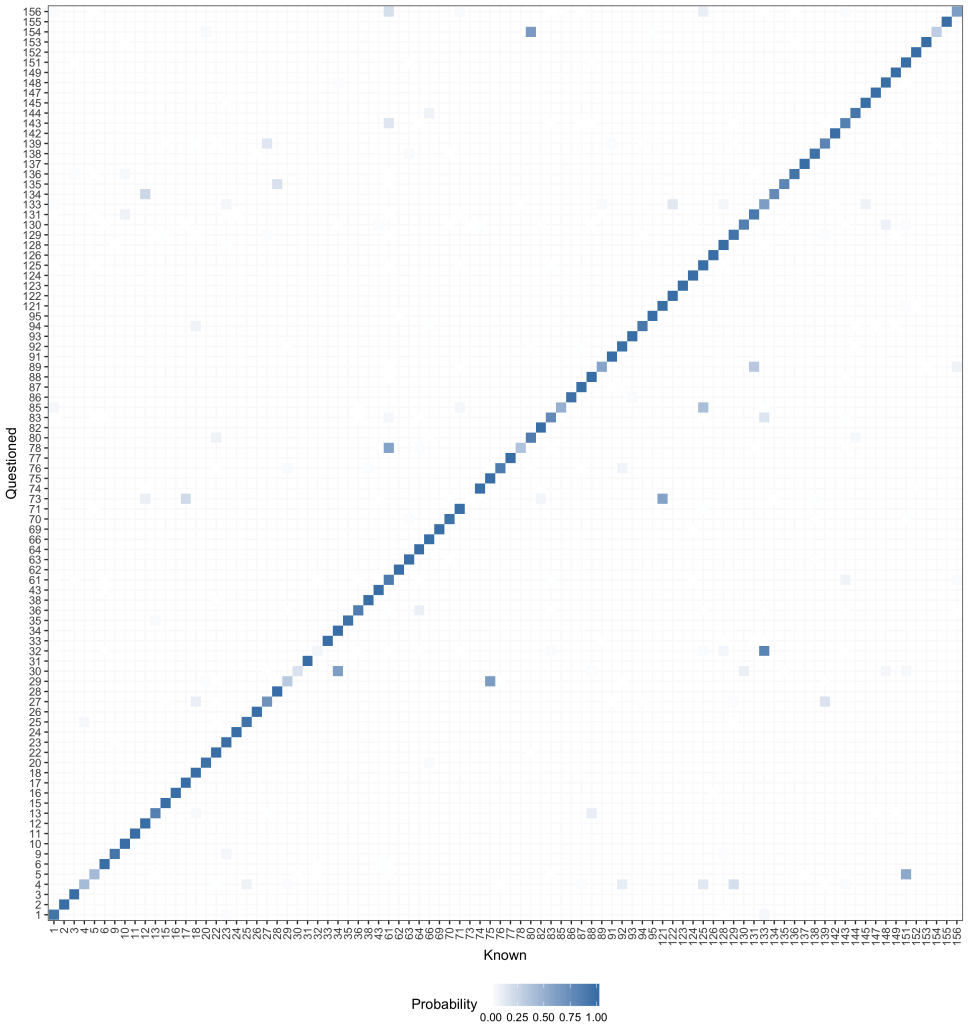



FIG 9. *Posterior probability of writership as defined in Equations 9 - 12 for each of the 90 CSAFE holdout documents, one row for each. True writers are labeled on the left-hand side. Columns are labeled by known writers. Cells are colored by elements of the $\bar{\boldsymbol{p}}$ vectors for each holdout document. Thus, each row sums to one and stands alone.*

where $\sum_{i=1}^{W} \bar{p}_{w_i} = 1$. Given restriction to the closed-set, the $\bar{\boldsymbol{p}}$ vector indicates the posterior probability of writership for each writer after accounting for variability in the parameters $\boldsymbol{\pi}_w$ through the MCMC samples. The general process shown in Equations 9-12 will be used to evaluate the holdout documents under each of the models that we will propose. With each new model, the definition of $q_{w'}^{(m)}$ will change to reflect the predictive distribution, the rest remains the same.

The set of 90 CSAFE testing documents are each evaluated in this way, and we obtain an estimated $\bar{\boldsymbol{p}}$ vector for each. The estimated probability assignments are depicted in Figure 9, where each row of the grid represents the $\bar{\boldsymbol{p}}$ vector for a test document. Under this reference analysis, $88.46\%$ (83.33, 93.33) of all probability assignment goes to the true writer. The remaining $11.54\%$ is assigned to an incorrect writer, and corresponds to off-diagonal elements in Figure 9.

TABLE 1

*Results of the prior sensitivity analysis conducted for the final model specification. The first three lines correspond to varying levels of prior information. The last two lines include parameter settings that affect the concentration of the wrapped Cauchy densities in Equations 5 and 7. An analysis of the holdout documents was conducted for each set of parameter settings, the percentage of correct assignment is included in the final column with a corresponding 95% credible interval.*

| Situation/Goal | a | b | c | d | e | True Writer $\bar{\boldsymbol{p}}$ (%) (95% credible interval) |
|---|---|---|---|---|---|---|
| Non informative | 1.1 | 0.1 | 1 | 1 | 0 | 97.05% (94.44, 98.89) |
| Moderately informative | 2 | 0.25 | 2 | 2 | 0.3 | 96.97% (94.44, 98.89) |
| Informative | 2 | 0.8 | 5 | 5 | 0.6 | 96.74% (94.44, 98.89) |
| Diffuse Wrap. Cauchy | 2 | 0.25 | 2 | 9 | 0.2 | 95.78% (93.33, 97.78) |
| Tight Wrap. Cauchy | 2 | 0.25 | 9 | 2 | 0.8 | 96.71% (94.44, 98.89) |

The models that we introduce in the following sections include graph-based measurements, in addition to cluster assignments. The measurements are assigned their own data models and build upon the reference model specified in Equations 1-3.

4.2. *Full Model including Graph-Level Features.* To evaluate a holdout document by unknown writer $w^*$, under known writer $w'$, with this model, we define $q_{w'}^{(m)}$

$$q_{w'}^{(m)} = Mult\left(\boldsymbol{Y}_{w^*}^*; \boldsymbol{\pi}_{w'}^{(m)}\right)\left[\prod_{k=1}^{K}\prod_{j=1}^{J_{w^*,k}} Wr.Cauchy\left(\theta_{w^*,k,j}; \mu_{w',k}, \tau_{w',k}\right)\right], \tag{13}$$

and the posterior probability of writership is summarized by $\bar{\boldsymbol{p}}$, which is found by using $q_{w'}^{(m)}$ in Equations 10-12, where the model for $\boldsymbol{\pi}_w$ is given by Equations 2-3 and the models for $\mu_{w',k}$ and $\tau_{w',k}$ are given in Equations 6-8.

$$\begin{aligned}
\boldsymbol{\pi}_w|\boldsymbol{\gamma} &\overset{iid}{\sim} Dirichlet(\boldsymbol{\gamma}) \\
\gamma_k &\overset{iid}{\sim} Gamma(a,\, b) \\
\tau_{w,k} &\sim Beta(c, d) \\
\mu_{w,k} &\sim WrappedCauchy(\eta_k, e) \\
\eta_k &\sim Uniform(0, 2\pi).
\end{aligned}$$

4.2.1. *Prior Sensitivity Analysis.* To demonstrate robustness of the model to final level prior parameter selection, a sensitivity analysis was conducted with two goals in mind. The first goal is to investigate the effect on the final posterior predictive analysis from varying the amount of information proffered by $a - e$. Parameters $a$ and $b$ express no information through a large variance. Parameters $c$ and $d$ control the beta distribution in Equation 6. To investigate the level of information introduced by these parameters it is reasonable to take $c = d$ in this first prior sensitivity stage to ensure that the center of the prior density remains at $0.5$, and increase the values to communicate higher levels of information. The second objective of the prior sensitivity analysis is to determine if the location specified by the beta parameters $c$ and $d$ impacts the posterior predictive accuracy. We first let $c = 2$ and $d = 9$ to concentrate the mass of the beta density near the lower end of the support. This encourages small concentration parameters, and thus a diffuse wrapped Cauchy distribution in Equation 5. Parameter $e$ operates in similar fashion for the wrapped Cauchy of Equation 7, where small values are associated with non-informative specification and a diffuse wrapped Cauchy distribution.

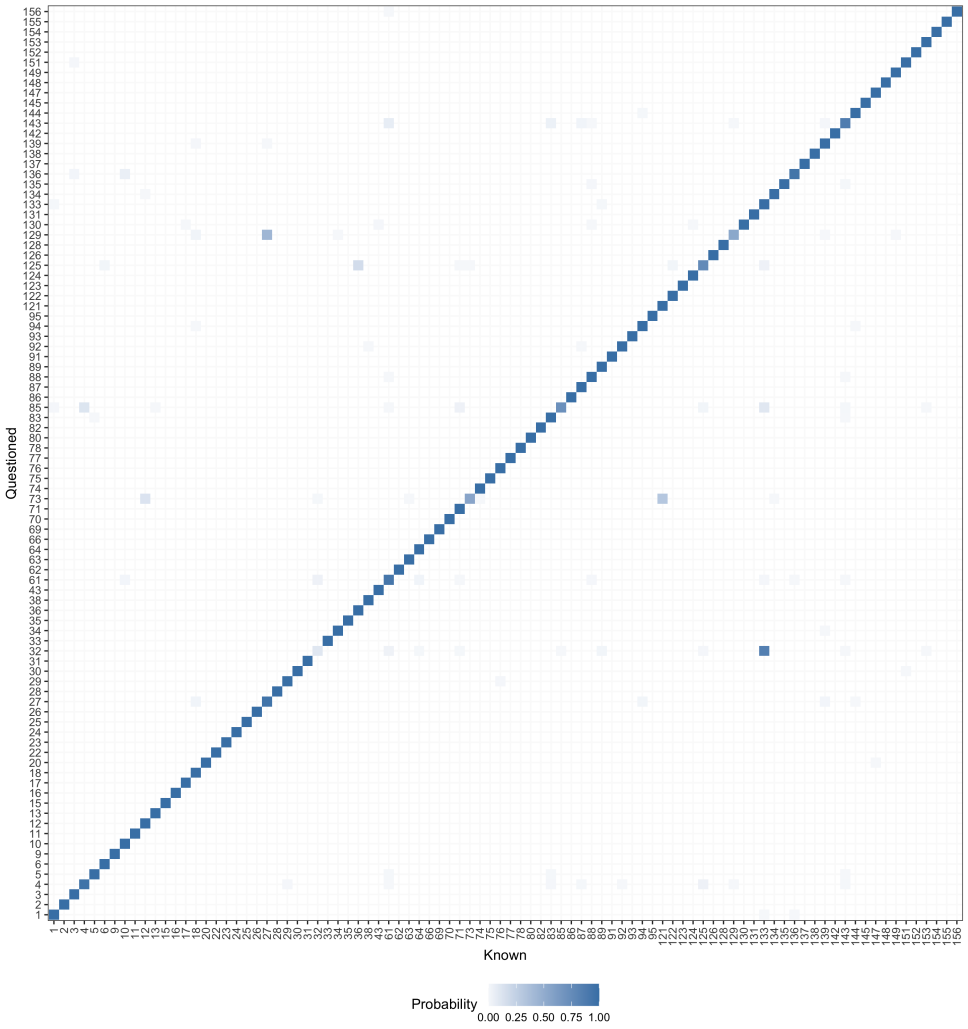



FIG 10. *Posterior probability of writership as defined in Equations 10 - 13, using the moderately informative prior parameter values, for each of the 90 CSAFE holdout documents, one row for each. True writers are labeled on the left-hand side. Columns are labeled by known writers. Cells are colored by elements of the* $\bar{\boldsymbol{p}}$ *vectors for each holdout document. Thus, each row sums to one and stands alone.*

Results of the analysis are included in Table 1, and provide evidence that results are relatively robust to prior parameter specification. This is particularly true for the information-based scenarios at the top of the table, which include reasonable values that could be chosen by another researcher who is familiar with the work. The two scenarios in the bottom of the table were included as more of an exercise, as these are not necessarily logical values to choose for parameters $c, d$ or $e$ apriori.

4.2.2. *Application of Full Model to CSAFE Data.* We will use the moderately informative parameter values as the final version of the hierarchical model. Posterior predictive evaluations are done as specified in Equation 13, and results of evaluation on the CSAFE closed-set are $96.97\%$ accurate assignment, as is stated in the previous table and shown in

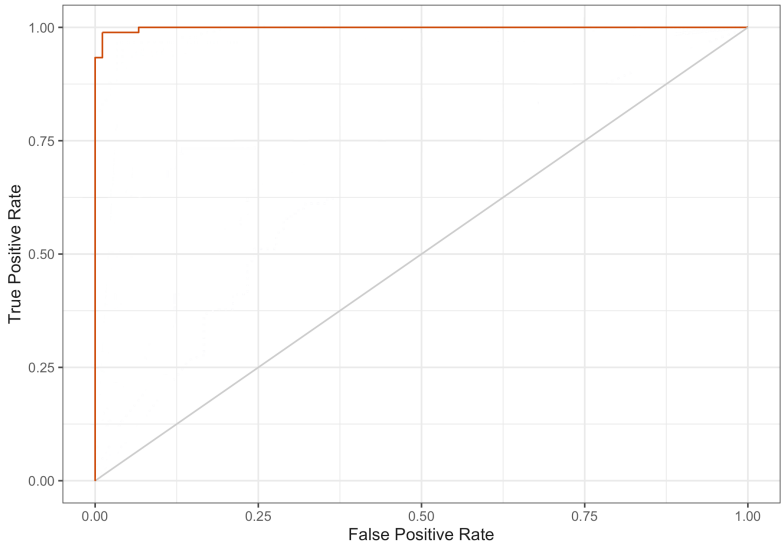



FIG 11. *ROC curve for the CSAFE test sets evaluated under the final model estimated for the CSAFE session #1 London Letter prompts. For a grid of cut-off values in (0,1), a true positive is declared if the value in the index of* $\bar{\boldsymbol{p}}$ *corresponding to the ground truth writer is greater than the cut-off. A false positive is declared if the value in that index is less than the cut-off.*

Figure 10. This analysis utilizing the graph-level features improves upon the previous analysis using only the document-level features (88.46% accuracy).

The proposed method was developed to produce an indication of which closed-set writers are most similar to the writer of the questioned document. If we wish to make a categorical decision regarding writership for each of the questioned documents, as is common in practice, a threshold would need to be chosen. The threshold would be used to dichotomize the space of $\bar{\boldsymbol{p}}$, where an identification would be declared for a writer if the corresponding entry of the vector exceeded the threshold. Rather than set a single hard threshold, we present a receiver operating characteristic (ROC) curve from which the accuracy of identifications can be observed for a variety of thresholds. The false positive and true positive rates of such identification tasks are calculated across all possible cutoff values. Identification results for this test set from the CSAFE database are shown via an ROC curve in Figure 11.

4.3. *Test on Decreasing Number of IAM Sentences.* In the previous analyses using the CSAFE data, the length of the test documents (302 graphs on average) is quite similar to the length of the training documents (376 graphs on average). Here, we explore whether the model is able to discriminate writers equally effectively when the length of the questioned document decreases. This is a typical situation in forensic document examination where the questioned document is something like a bank robbery note, i.e. "I have a gun, give me all the money," or an address on an envelope, but the samples collected from the closed-set writers are much longer. We explore the performance of our Bayesian hierarchical model on the IAM data where the training documents are similar in length to the London Letter, but the testing documents are increasingly shorter.

In order to properly learn the multinomial parameters during model fitting, the 16 training sentences for each IAM writer are taken in four groups, to make four "document" length observations on par with the writing samples from the CSAFE database. This effectively constructs four cluster-frequency data vectors, $\mathbf{Y}_{w(d)}$, for each writer, and the posterior distributions for the model parameters are estimated via MCMC using a single chain of length 4000.

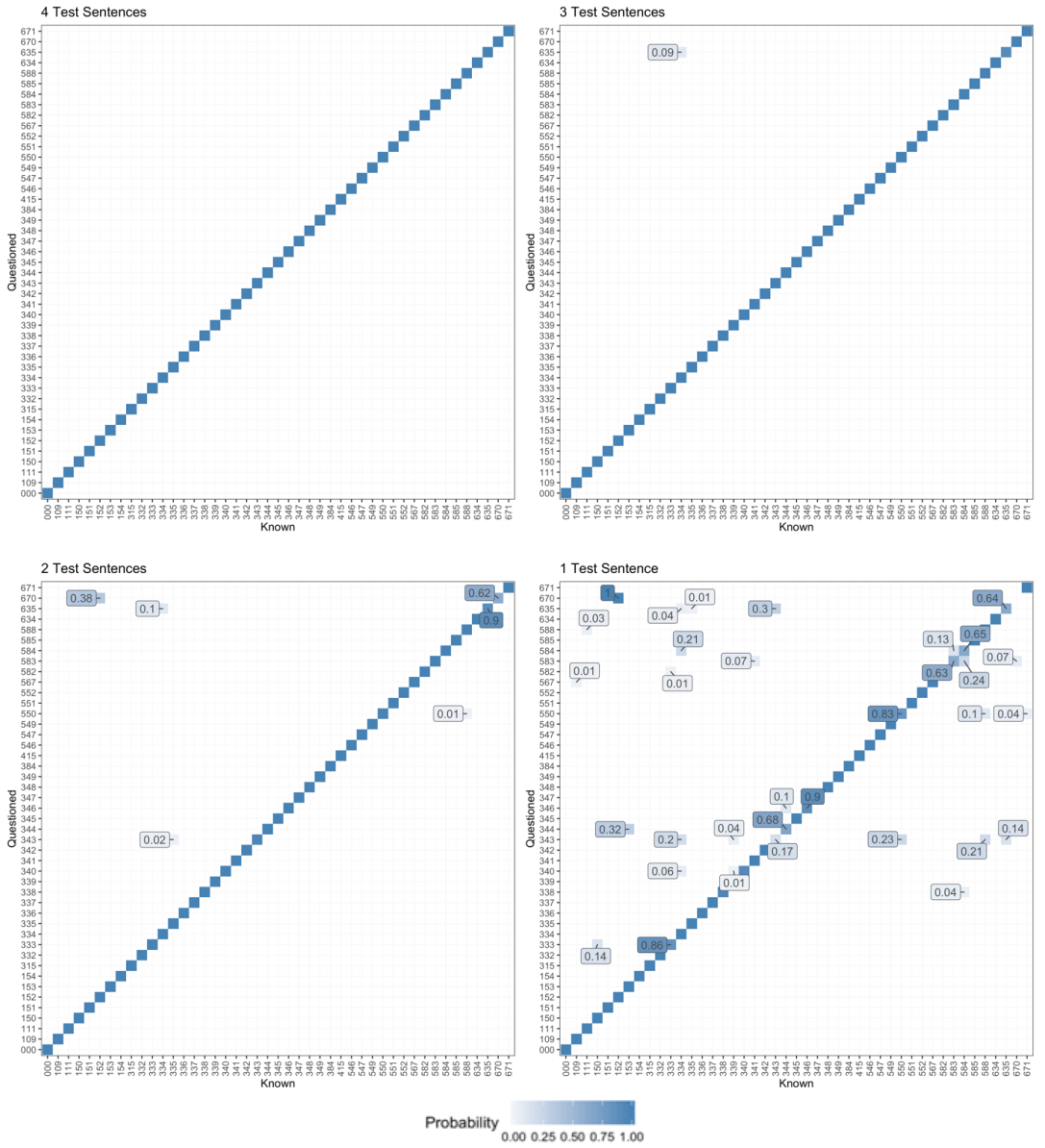



FIG 12. *Results of the posterior analysis under the final model, estimated using* 16 *sentences for each of the* 45 *writers in the closed-set taken from the IAM database. The four plots are shown on decreasing amounts of test data. True writers of each holdout document are labeled on the rows. Columns are labeled by known writers, and cells are colored by corresponding elements of the* $\bar{\boldsymbol{p}}$ *vectors. Vector elements* $< 0.01$ *are removed from the plot, showing the gridlines behind. Elements of the* $\bar{\boldsymbol{p}}$ *vector corresponding to the ground truth writer are labeled only if they are less than* 0.9*, and all other elements* $\geq 0.1$ *are labeled.*

We begin the test set evaluation by taking all four of the holdout sentences together as a single test sample for each writer. Each test sample is evaluated against the writers in the closed-set according to Equation 13, and elements of the resulting $\bar{\boldsymbol{p}}$ vectors are provided in the top left plot of Figure 12. Then, we remove the longest of the four holdout sentences and repeat the evaluation process using the shortest three together as a test sample for each writer. The same is done for the two shortest sentences. Finally, the single shortest sentence from the analysis dataset is evaluated for each writer.

TABLE 2

*Four amounts of testing data (in number of sentences) are evaluated. The median, minimum, and maximum number of graphs extracted from each test set are provided. The percentage of true writer assignment according to $\bar{\boldsymbol{p}}$ is included in the rightmost column with a 95% credible set.*

| # Test Sentences | Median # of Graphs (Min, Max) | % True Writer $\bar{\boldsymbol{p}}$ Assignment (95% credible interval) |
|---|---|---|
| 4 | 167 (131, 297) | $> 99.99$ [1] |
| 3 | 118 (95, 216) | 99.79 (97.78, 100.00) |
| 2 | 73 (61, 139) | 98.82 (95.56, 100.00) |
| 1 | 34 (30, 67) | 91.50 (84.44, 95.55) |

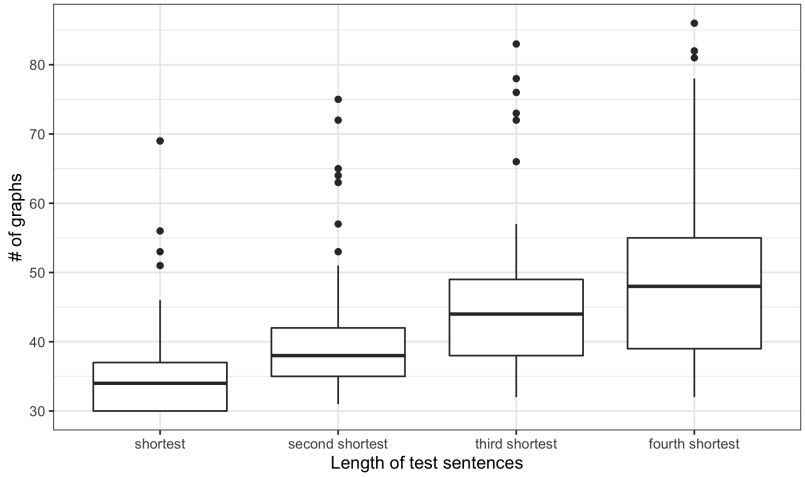


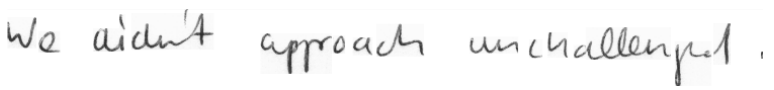

FIG 13. *Distributions of the number of graphs in the holdout sentences across all* 45 *writers in the IAM closed-set. To be eligible for the analysis dataset, sentences were required to contain a minimum of* 30 *graphs after decomposition by the* `handwriter` *R package.*

Results from all four evaluations are depicted in Figure 12. The percentage of $\bar{\boldsymbol{p}}$ probability assignment to the true writers for each test set is provided in Table 2 along with variability estimates. As the amount of testing data declines so does the accuracy of model predictions. An especially noticeable degradation occurs when we drop from two sentences down to the single shortest test sentence. Aside from the obvious decrease in available testing information, with only one test sentence we move into the space where representation of the 40-dimensional Multinomial data vector is likely poor.

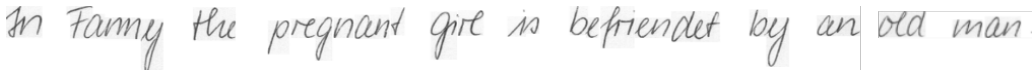
Figure 13 provides insight into the size of the test documents with respect to the number of graphs in each. We notice that, for most of the writers, the shortest test sentence is comprised of fewer than 40 graphs. Using these as test sentences provides little hope of a robust evaluation based on the cluster count data. Figure 14 provides the shortest test sentences for three writers. This gives the reader some insight into how much writing is required for a robust evaluation under the models we estimate.

**5. Discussion.** Huber and Headrick (1999) consider "slant, slope, incline, or lateral expansion" to be a "discriminating element of writing." We first attempted a centroid-based approach to capturing this feature of a graph. Centroid locations are calculated as the average position of all pixels in the graph. The centroid-based slope was found by locating the centroid of a graph, and splitting it into two sub-graphs consisting of the pixels to the right and

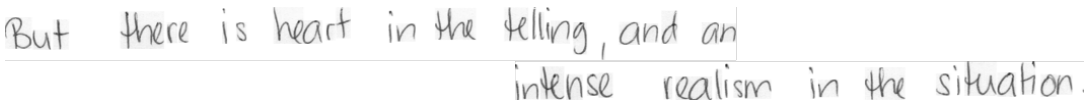
[1] One incorrect vote occurred in one iteration of the MCMC routine.



(a) 30 graphs



(b) 46 graphs



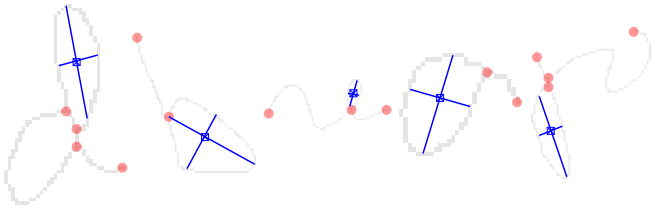
(c) 69 graphs

FIG 14. *The shortest sentence for three of the* 45 *writers in the closed-set. The number of graphs extracted from each sentence is listed.*

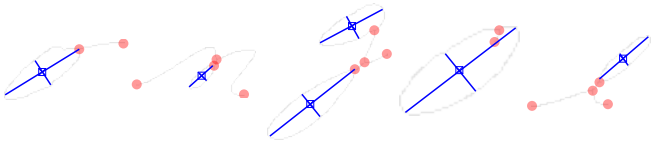
left of the centroid, respectively. We then found the centroid of each sub-graph, and calculated the slope of the line connecting them. This method to quantify the slope was undesirable for several reasons. For example, it was not robust to outliers, much like traditional arithmetic averages. In order to model these measurements, we used the average slant within each cluster for each document and assumed a Normal distribution with a hierarchical structure for the location parameter for each writer and cluster combination. Combining these graph-level features with the model for the document-level features resulted in $91.48\%$ (87.78, 94.44) accuracy on the test documents, which was only a minor improvement over the model using document-level features only.

Another approach we considered for capturing the "slant" of a graph was to naively treat the rotation angle values as if they are linear on the interval $(0,\pi)$ and then scale each value by $\pi$ to transform the support to $(0, 1)$. This approach corresponds to the traditional histograms in Figure 4, ignoring the fundamental coordinate system in which the angular measurements should be considered. Then, the scaled rotation angles were modeled with a beta distribution and weakly informative priors. This approach appropriately captured the shape of rotation angle distributions for a number of writer/cluster combinations, but was inadequate for a number of others. Furthermore, when we sampled simulated data from this model, the resulting simulated data did not match well to the observed data. This model for the graph-level features combined with the standard model for the document-level features resulted in $92.08\%$ (88.89, 95.56) accuracy. While this was a slight improvement over the centroid-based slope analysis, we found it much better to model the rotation angles in the polar-coordinate system.

Other features have been considered for inclusion in the model. A promising measurement was obtained from each loop in a graph. Loops are defined as a sequence of pixels with beginning and terminal nodes near each other, with no other structure interfering. We locate the centroid of each loop, and calculate the line that travels through the centroid having the longest distance. Then we find the perpendicular line that also passes through the centroid. Examples of these can be found in Figure 15. The log ratio of length to width of each loop was incorporated into the hierarchical model using a gamma distribution for the data. These response variables provided only marginal improvement over both the model with only document-level features, and the final model with the addition of the fully wrapped graph-level rotation angles.



(a) Writer #1



(b) Writer #95

FIG 15. *Loop measurement examples are shown in blue for graphs from the training documents of CSAFE Writers #1 and #95.*

**6. Conclusion.** In general, we observed acceptable results using both the document-level and graph-level measurements, with one exception; we saw relatively poor results when evaluating test documents with fewer than 70 graphs. This is likely due to the fact that for several of the 40 pre-determined clusters, we observe a frequency of zero graphs. Consequently, we cannot obtain reliable estimates of the elements of the vector $\boldsymbol{\pi}$ of probabilities. It will be left to future work to explore model fitting and performance for data with cluster assignments from a smaller template, with 20 or fewer clusters, for example.

Results presented here indicate that the analytical framework we propose may prove useful in actual case work. By evaluating a test document using our modeling framework, forensic document examiners can obtain a preliminary, objective and quantitative assessment of the probability of writership of a questioned document when the true writer is included in the pool of potential writers. These initial assessments may be accurate in those cases where the questioned writing has sufficient length. In future work, it would be worthwhile to develop a method that could evaluate the weight of handwriting evidence beyond a closed-set of writers.

**Acknowledgments.** The authors would like to thank Dr. Stephanie Reinders for her help with editing figures for the final manuscript.

**Funding.** This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreements 70NANB15H176 and 70NANB20H019 between the National Institute of Standards and Technology (NIST) and Iowa State University, which includes activities carried out at Carnegie Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln.

## REFERENCES

BAUM, L. FRANK, ILLUSTRATED BY W. W. DENSLOW (1900). *The Wonderful Wizard of Oz*. Chicago and New York: G.M. Hill Co.

BERRY, N., TAYLOR, J. and BAEZ-SANTIAGO, F. (2019). `handwriter`: Handwriting Analysis in `R`. https://CRAN.R-project.org/package=handwriter. `R` package version 1.0.1.

BRASSEUR, L. (2005). Florence Nightingale's Visual Rhetoric in the Rose Diagrams. *Technical Communication Quarterly* **14** 161–182.

CHOI, R. Y., COYNER, A. S., KALPATHY-CRAMER, J., CHIANG, M. F. and CAMPBELL, J. P. (2020). Introduction to machine learning, neural networks and deep learning. *Translational Vision, Science, and Technology*. doi: 10.1167/tvst.9.2.14.

NATIONAL RESEARCH COUNCIL (2009). *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press, Washington, D.C.

CRAWFORD, A., BERRY, N. and CARRIQUIRY, A. (2021). A Clustering Method for Graphical Handwriting Components and Statistical Writership Analysis. *Statistical Analysis and Data Mining*. doi: 10.1002/sam.11488.

CRAWFORD, A., RAY, A. and CARRIQUIRY, A. (2019). A Database of Handwriting Samples for Applications in Forensic Statistics. *Data in Brief*. doi: 10.1016/j.dib.2019.105059.

EXPERT WORKING GROUP FOR HUMAN FACTORS IN HANDWRITING EXAMINATION (2020). Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach. U.S. Department of Commerce, National Institute of Standards and Technology. NISTIR 8282.

FORGY, E. (1965). Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classifications. *Biometrics* **21** 768-780.

FRANKE, K., SCHOMAKER, L., VEENHUIS, C., VUURPIJL, L., VAN ERP, M. and GUYON, I. (2004). WANDA: A common ground for forensic handwriting examination and writer identification. *Enfhex News* **1** 23–47.

HUBER, R. A. and HEADRICK, A. M. (1999). *Handwriting Identification: Facts and Fundamentals*. CRC Press-Taylor & Francis, Boca Raton, FL.

JOHANSSON, S., LEECH, G. and GOODLUCK, H. (1978). *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computer*. Department of English, University of Oslo.

KENT, J. T. and TYLER, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics* **15** 247–254.

KLEBER, F., FIEL, S., DIEM, M. and SABLATNIG, R. (2013). CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting. In *2013 12th International Conference on Document Analysis and Recognition* 560-564.

KWAN, Q. Y. (1977). Inference of Identify of Source, Ph.D. Dissertation in Criminology, University of California, Berkeley.

LEEDHAM, G. and SRIHARI, S. (2003). A survey of computer methods in forensic handwritten document examination. In *Proceedings of the Eleventh International Graphonomics Society Conference* 278–281. Scottsdale, AZ.

LLOYD, S. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **28** 129-137.

MARTI, U. and BUNKE, H. (2002). The IAM-Database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition* **5** 39–46.

MILLER, J. J., PATTERSON, R. B., GANTZ, D. T., SAUNDERS, C. P., WALCH, M. A. and BUSCAGLIA, J. (2017). A set of handwriting features for use in automated writer identification. *Journal of Forensic Sciences* **62** 722–734.

FEDERAL BUREAU OF INVESTIGATION (2009). Getting Physical: Our Hardcopy Forensic Experts. https://archives.fbi.gov/archives/news/stories/2009/april/qdu_040909. Accessed November 12, 2021.

OSBORN, A. S. (1929). *Questioned Documents, 2nd edn.* Boyd Printing Company, New York, NY.

SAUNDERS, C. P., DAVIS, L. J., LAMAS, A. C., MILLER, J. J. and GANTZ, D. T. (2011). Construction and Evaluation of Classifiers for Forensic Document Analysis. *The Annals of Applied Statistics* **5** 381-399.

SRIHARI, S. N., SRINIVASAN, H. and DESAI, K. (2007). Questioned Document Examination Using. *Journal of Forensic Document Examination* **18** 1–20.

UNITED STATES V. JOHNSON (2016). No. 1:15-cr-00565 VEC, (S.D.N.Y. Jun. 7, 2016).

STATE V. PICKETT (2021). 246 A.3d 279, 466 N.J. Super. 270 (Super. Ct. App. Div. 2021).

YOUNG, E. (2018a). Lindbergh Kidnapping, FBI: Famous Cases and Criminals: Volume 2. Independently Published. Also available online: https://www.fbi.gov/history/famous-cases/lindbergh-kidnapping. Accessed November 12, 2021.

YOUNG, E. (2018b). Weinberger Kidnapping, FBI: Famous Cases and Criminals: Volume 3. Independently Published. Also available online: https://www.fbi.gov/history/famous-cases/weinberger-kidnapping. Accessed November 12, 2021.

ZIMMERMANN, M. and BUNKE, H. (2002). Automatic segmentation of the IAM off-line database for handwritten English text. In *Object recognition supported by user interaction for service robots* **4** 35–39. IEEE.