

Quantification of the weight of fingerprint evidence using a ROC-based Approximate Bayesian Computation algorithm for model selection*

Jessie Hendricks and Cedric Neumann and Christopher P. Saunders

*Department of Mathematics and Statistics
South Dakota State University
Brookings, SD, USA
e-mail:*

jessiehendricks@gmail.com; cedric.neumann@me.com; christopher.saunders@sdstate.edu

Abstract: For more than a century, fingerprints have been used with considerable success to identify criminals or verify the identity of individuals. The categorical conclusion scheme used by fingerprint examiners, and more generally the inference process followed by forensic scientists, have been heavily criticised in the scientific and legal literature. Instead, scholars have proposed to characterise the weight of forensic evidence using the Bayes factor as the key element of the inference process. In forensic science, quantifying the magnitude of support is equally as important as determining which model is supported. Unfortunately, the complexity of fingerprint patterns render likelihood-based inference impossible. In this paper, we use an Approximate Bayesian Computation (ABC) model selection algorithm to quantify the weight of fingerprint evidence. We supplement the original ABC model selection algorithm using a Receiver Operating Characteristic curve to mitigate its known shortcomings with respect to the choice of a suitable threshold and the curse of dimensionality. In a sense, we offer an alternative to other methods that have tried to address the same issues. Our method is straightforward to implement, computationally efficient, and visually intuitive for lay individuals (i.e., jurors). We apply our method to quantify the weight of fingerprint evidence in forensic science, but we note that it can be applied to any other forensic pattern evidence.

Keywords and phrases: Approximate Bayesian Computation, Bayes factor, forensic science, weight of evidence, receiver operating characteristic.

Received July 2019.

Contents

1 Introduction	1229
--------------------------	------

*This project was supported in part by Award No. 2014-IJ-CX-K088 awarded by the National Institute of Justice (NIJ), Office of Justice Programs, U.S. Department of Justice, and in part by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute (SAMSI). The opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the U.S. Department of Justice or the National Science Foundation.

2 Approximate Bayesian Computation 1232

3 ROC-ABC algorithm for model selection 1235

 3.1 Empirical ROC 1237

 3.2 The non-central dual beta ROC model 1238

4 Weight of fingerprint evidence using ROC-ABC 1240

 4.1 Generation of pseudo-fingerprint data 1241

 4.1.1 Generation of pseudo-fingermarks under the prosecution
 model 1242

 4.1.2 Generation of pseudo-fingermarks under the defence model 1242

 4.2 Summary statistic 1243

 4.3 Kernel function 1244

 4.4 Number of simulations 1244

5 Datasets 1244

6 Results 1245

7 Discussion 1249

8 Conclusion 1252

A Kernel function development 1253

 A.1 Components of the kernel function 1254

 A.2 Optimisation of the kernel function 1254

B Results using Beaumont (2008)’s method 1255

C Proof of convergence result 1257

Acknowledgments 1259

References 1259

1. Introduction

For more than a century, fingerprints have been used with considerable success to identify criminals or verify the identity of individuals. In this paper, we define a *fingerprint* (left panel of Figure 1), or *mark*, as the impression resulting from the inadvertent contact between the finger of an unknown donor and a surface (e.g., at a crime scene). We define a *control print* (right panel of Figure 1), or *print*, as a finger impression collected under controlled conditions from a known donor (e.g., a suspect). The purpose of forensic fingerprint examination is to support the inference of the identity of the donor of a fingerprint (source level in Cook et al. (1998)). Currently, this inference process relies on the visual comparison between the fingerprint and control prints from one or more candidate donors who may have been selected through a police investigation or by searching the fingerprint in a database of prints from known individuals.

Following the examination of fingerprint evidence, it is currently customary for the examiner to report one of two possible conclusions: an opinion of “identification”, implying that the source of the fingerprint is the donor of a given control print; or an opinion of “exclusion”, implying that the source of the fingerprint is not a considered candidate. Alternatively, the examiner may find the examination “inconclusive”, indicating that the characteristics of the



FIG 1. *Left panel: fingermark. Right panel: fingerprint taken under controlled conditions. Ridges appear darker than background. Both impressions were made by the same finger. Their comparison shows that both ridge flows are affected by different distortion and degradation effects.*

impressions being compared are not sufficient to reach one of the two possible conclusions (e.g., when the impressions are too small or too degraded).

The categorical conclusion scheme used by fingerprint examiners, and more generally the inference process followed by forensic scientists, has been heavily criticised in the scientific and legal literature (Cole, 2004, 2005, 2009; Kaye, 2003; Saks and Koehler, 2005, 2008; Zabell, 2005). Instead, scholars have proposed to characterise the weight of forensic evidence using the Bayes factor as the key element of the inference process (see Aitken and Taroni (2004) and Aitken, Roberts and Jackson (2010) for a comprehensive discussion).

It is worth stressing that forensic scientists do not have a complete picture of all the evidence available in a given criminal case (e.g., eyewitness evidence; means, motive and opportunity of an individual of interest; other material and non-material evidence), which prevents them from assigning probabilities to the different propositions that the parties may have regarding a particular criminal activity. Furthermore, forensic scientists are not tasked with the fact-finding mission in the criminal justice system and are not in charge of the decision-making with respect to the propositions of the parties. Therefore, the role of the scientist is necessarily limited to reporting the amount of support of the forensic evidence for these propositions in the form of a Bayes factor. It is of primary importance to note that, in this setting, it is not only of interest to support the correct model, but also to support it with the appropriate magnitude. An appropriate magnitude of support is critical to ensure the coherent combination of the respective weight of the multiple pieces of evidence that may be considered in a given case. This imperative requirement is the main motivation behind the present work and is the main force that drives us away from more deterministic pattern recognition algorithms.



FIG 2. Details of the ridge pattern. Ridges appear darker than background. The ridge pattern can be disturbed by different events, such as: A - ridge ending; B - bifurcation; C - enclosure or lake; D - island (very short ridges are called dot); E - incipient ridges (short, narrow and non-continuous ridges that appear between two parallel ridges). White dots within the ridges are sweat pores and are sometimes used by fingerprint examiners (see F).

Assuming a fingermark recovered in connection with a crime and a suspect, Mr. X., we consider the following two alternative propositions:

H_1 : the fingermark originates from Mr. X.

H_2 : the fingermark does not originate from a finger of Mr. X., but from another person in a relevant population of potential donors.

To address the so-called prosecution and defence propositions, H_1 and H_2 , we consider two models, M_1 , which represents how Mr. X. generates fingermarks, and M_2 , which represents how fingermarks are generated by the donors in the population of alternative sources. These models enable us to sample fingermarks under the two alternative propositions and are necessary to generate the data used in the algorithm proposed in this paper. These models and their use are described in more details below. Let e_u denote the set of observations made on the fingermark which originates from an unknown donor u . We are interested in evaluating the following Bayes factor:

$$BF = \frac{\pi(M_1|e_u)}{\pi(M_2|e_u)} \cdot \frac{\pi(M_2)}{\pi(M_1)} = \frac{m(e_u|M_1)}{m(e_u|M_2)}, \tag{1}$$

where $\pi(\cdot)$ is a measure of belief about the model and its parameters (see Robert (2007), chapter 7, for a formal discussion on Bayesian model selection), and $m(\cdot)$ represents a (marginal) probability density function.

Fingerprints are usually characterised through certain features of the ridge pattern, such as the general pattern formed by the friction ridge flow and events disturbing the continuity of the ridges. These events, traditionally called minutiae, occur when a ridge ends (ridge ending) or bifurcates (bifurcation). Other types of events exist but are mainly combinations of these two basic types. Figure 2 illustrates details of the ridge pattern on a finger impression.

When comparing two fingerprints, an examiner first verifies the compatibility of their general patterns and then determines whether the spatial relationships,

types, and orientations of the minutiae on both impressions correspond. Mathematically characterising fingerprint patterns results in heterogeneous, high-dimension random vectors: minutia locations and spatial relationships are continuous measurements; minutia types are discrete observations; and minutia orientations are circular measurements. In addition, the dimensionality of the problem increases with the number of minutiae observed on a given impression. Therefore, modelling the joint likelihood of the characteristics of multiple features observed on a finger impression seems to be an unreasonable challenge.

The complexity of modelling the probability distribution of fingerprint patterns explains why, despite the many attempts that have been made to assign Bayes factors for fingerprint evidence (see Abraham et al. (2013) for a recent review), no algorithm has gained wide acceptance. In particular, while the results obtained with the algorithm proposed by Neumann, Evett and Skerrett (2012) were used to support the general admissibility of fingerprint evidence in U.S. courts (State v. Hull, 2008; State v. Dixon, 2011), commentators of the Neumann, Evett and Skerrett (2012) paper argued that the model had two main shortcomings. Several researchers discussed that the algorithm did not result in a formal Bayes factor as it does not formally incorporate user beliefs (Kadane, 2012; Lauritzen, Cowell and Graversen, 2012; Stern, 2012). Others noted that the algorithm relied on an ad-hoc weighting function used to palliate the inability of the authors to assign Bayes factors *at* 0, and that this function had no other justification than convenience (Balding, 2012; Jandhyala and Fotopoulos, 2012; Kadane, 2012).

In this paper, we take advantage of the similarities between the algorithm proposed by Neumann, Evett and Skerrett (2012) and the original Approximate Bayesian Computation (ABC) model selection framework (Grelaud et al., 2009; Toni and Stumpf, 2010) to provide a method to formally and rigorously assign Bayes factors to fingerprint evidence. Our method leverages the well-known Receiver Operating Characteristic (ROC) curve (Pepe, 2003) to address shortcomings in the original ABC model selection algorithm. Our application addresses the issues raised in relation to Neumann, Evett and Skerrett (2012) and provides a much needed general framework for the quantification of the weight of any type of forensic pattern evidence, as long as a similarity measure can be defined to compare two pieces of evidence.

2. Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) originated as a class of algorithms designed to sample from the approximate posterior density of a vector of parameters, θ , given an observed data set, \mathbf{D} , without direct evaluation of the likelihood function, $f(\mathbf{D}|\theta)$. This class of algorithms is especially useful in complex and high-dimensional settings where the likelihood function is not available in a usable form (see Sisson, Fan and Beaumont (2019) for a recent overview).

To sample from an approximate posterior density, vectors of parameter values are first sampled from a prior distribution over the parameter space, and

then used to generate pseudo-observations (each denoted \mathbf{D}^*) from a generative model. In the original approach, a vector of parameter values is retained if the distance measured by a kernel function, $\Delta\{\cdot, \cdot\}$, between values of a summary statistic, $\eta(\cdot)$, of \mathbf{D} and \mathbf{D}^* is less than some tolerance, $t > 0$. In other words, the i^{th} sampled parameter vector, $\boldsymbol{\theta}^{(i)}$, is retained if the corresponding distance satisfies $\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\} < t$.

In some application problems, such as forensic science, statisticians are not concerned with assigning posterior probability distributions in the parameter space, but are interested in performing model selection. However, performing likelihood-based inference using patterns of forensic interest, such as fingerprints, shoe sole impressions, or striations on bullets, is not feasible in the original feature space of the data. Starting with Pritchard et al. (1999), ABC has evolved into an algorithm that can be used for model selection by considering a model index parameter, \mathcal{M} , and its prior distribution over model indices. The model index determines the prior distribution over the parameter space and the likelihood structure used to generate pseudo-observations. In the original approach, the i^{th} sampled model index parameter, $\mathcal{M}^{(i)}$, is retained if the corresponding score satisfies $\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\} < t$.

The original ABC approach for assigning a posterior probability to a model, given the observed data, is a function of the ratio of the number of times that model index has been retained over the total number of times any model was retained (Grelaud et al., 2009; Toni and Stumpf, 2010; Didelot et al., 2011; Robert et al., 2011). Mathematically, the ABC posterior odds for the comparison between two models is given by:

$$\frac{\pi_t(\mathcal{M} = 1|\mathbf{D})}{\pi_t(\mathcal{M} = 2|\mathbf{D})} = \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N \mathbb{I}[\mathcal{M}^{(i)} = 1] \cdot \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\} \leq t]}{\sum_{i=1}^N \mathbb{I}[\mathcal{M}^{(i)} = 2] \cdot \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\} \leq t]}, \quad (2)$$

where $\mathbb{I}(\cdot)$ is the indicator function and the subscript t in π_t indicates that this measure is a function of the choice of the tolerance level. The original ABC Bayes Factor, BF_{abc} , is assigned using the ABC approximation of the posterior odds, divided by the prior odds

$$BF_{abc} = \frac{\pi_t(\mathcal{M} = 1|\mathbf{D})}{\pi_t(\mathcal{M} = 2|\mathbf{D})} \cdot \frac{\pi(\mathcal{M} = 2)}{\pi(\mathcal{M} = 1)}. \quad (3)$$

The original ABC algorithm for comparing two models is summarised in Algorithm 1.

The benefits of Approximate Bayesian Computation methods are not without cost. Model selection using ABC is subject to two primary sources of error:

1. the quality of the approximation to the Bayes factor due to the use of the tolerance, t ,
2. and the loss of information engendered by using insufficient summary statistics.

The effect of the tolerance level, t , on the performances of ABC algorithms has been widely discussed since the inception of ABC methods. In short, if t is

Algorithm 1: Original ABC model selection algorithm

Data: Observed data, D .

Result: ABC Bayes factor, BF_{abc} .

for $i = 1$ to N **do**

Sample a model index $\mathcal{M}^{(i)}$ from the model prior, $\pi(\mathcal{M} = m)$, where $m = 1, 2$;
 Sample a vector of parameters, $\theta^{(i)}$, from the prior density, $\pi(\theta|\mathcal{M}^{(i)})$;
 Generate a pseudo-observation, $D^{*(i)}$, from the generative model, $f(D^*|\theta^{(i)}, \mathcal{M}^{(i)})$;
 Compute $\Delta\{\eta(D), \eta(D^{*(i)})\}$;

Calculate the ratio of counts as defined in Equation (2);

Assign BF_{abc} as in Equation (3);

too large, too many samples from the prior distribution are accepted and the approximation becomes invalid; and, if t is too small, the rate of acceptance is too small to produce a stable result.

The motivation for using summary statistics, rather than the full data, stems from the curse of dimensionality encountered with the high-dimensional data sets that are common to scenarios in which ABC is necessary. To enable convergence of BF_{abc} to the Bayes factor based on the full data, a summary statistic that is sufficient across models must be used (Robert et al., 2011; Didelot et al., 2011). However, in general, it seems difficult (if not impossible) to find sufficient summary statistics for the types of data that typically require ABC methods. The goal is then to minimise the loss of relevant information to the model selection task that may be encountered when using insufficient statistics. If the curse of dimensionality was less of an issue, a general solution addressing the issue of the potential lack of sufficiency of summary statistics could consist in using a large set of summary statistics with the hope that they will jointly tend towards sufficiency by decreasing the loss of information.

Several modifications of the original ABC algorithm for model selection (Algorithm 1) have been proposed with the goal of addressing some of the issues related to the determination of the tolerance and the selection of low-dimensional summary statistics that minimise the loss of information that is relevant to the model selection task. Most of these modified algorithms are aimed at alleviating issues posed by use of the tolerance, t . Most of these algorithms are rooted in the one proposed by Beaumont (2008), which uses a polychotomous weighted logistic regression model that has been trained on the summary statistics of the pseudo-observations and corresponding model indices to predict the posterior probability of a model. Modifications include constructing low-dimensional vectors of summary statistics that are informative about the model index parameter (Estoup et al., 2012; Prangle et al., 2014) and replacement of logistic regression by artificial neural networks (Blum and Francois, 2010) or random forest algorithm (Pudlo et al., 2016). While these methods may help reduce the loss of information due to the use of insufficient summary statistics, computational challenges are associated with using summary statistics of high dimension.

The literature shows that the issues of sufficiency, variable selection and tol-

erance level are usually entangled. The solutions to these issues proposed to date, and summarised above, have their own sets of complications:

1. None of these solutions enables formal monitoring of the convergence of BF_{abc} to the Bayes factor as the value of the tolerance, t , is reduced, or as the number of samples increases.
2. Most of these solutions are focusing on selecting the correct model, but are not necessarily designed to support the selected model with the appropriate magnitude, which is a critical requirement in forensic science.
3. Some of the model selection algorithms, such as the one proposed by Beaumont (2008), are trained using a limited subset of pseudo-observations. These pseudo-observations are selected or weighted based on the proximity of their summary statistics with those of the observed data. Unfortunately, this results in replacing user-defined prior probabilities on model indices with probabilities assigned in unpredictable ways by the algorithm based on the proportions of training data selected from each model.
4. Methods based on generalised linear models (Beaumont, 2008), artificial neural networks (Blum and Francois, 2010), random forests (Pudlo et al., 2016) or other classifiers can be very computationally intensive depending on the dimension of the summary statistics or the number of pseudo-observations generated from the considered models. They also heavily rely on appropriately estimating a very large number of parameters.

In this paper, we propose a novel extension to the original ABC model selection algorithm that utilises a relationship between the ABC Bayes factor and the Receiver Operating Characteristic (ROC) curve. Our ROC-based approach addresses issues with the original ABC Bayes factor posed by the use of a heuristic tolerance level. Our algorithm shifts the focus away from the tolerance level, and in doing so, can accommodate larger sets of summary statistics in a computationally efficient manner. Our method is not designed to extend upon the models based on logistic regression and other classifiers, but rather proposes an alternative development of the original ABC algorithm that addresses shortcomings related to the choice of a suitable threshold in the original algorithm. Our method preserves user-defined priors, is straightforward to implement, computationally efficient, and promotes inferences that are visually intuitive and can be communicated effectively. Critically, our solution enables us to rigorously control its convergence as the number of simulations increases.

3. ROC-ABC algorithm for model selection

ROC curves have traditionally been used to measure the performance of a binary classifier as the decision threshold, t , is varied across the domain of (dis)similarity scores, $\Delta\{\cdot, \cdot\}$, that can be returned by the classifier (Pepe, 2003). In this paper, the ROC curve will be considered as a tool to convey the relationship between two distributions of scores as a function of a threshold, t .

ROC curves are constructed by plotting the rate of correct decisions in favour of the first model against the rate of incorrect decisions (false positives) in favour of the first model (i.e. when the second model is true) for all possible values of the decision threshold. Defining $F(\cdot)$ and $G(\cdot)$ as the cumulative distribution functions of scores arising under models 1 and 2, respectively, the general form of the ROC is given by (Pepe, 2003) as $\text{ROC}(p) = F(G^{-1}(p))$, where p denotes the rate of false positives in favour of the first model and $G^{-1}(\cdot)$ denotes the quantile function (van der Vaart, 1998) for scores under the second model.

Here, we show that the Bayes factor depending only on the summary statistic of the data, BF_η , is a function of the ROC curve constructed on the set of $\Delta\{\eta(\mathbf{D}), \cdot\}$ arising under model 1 and the set of $\Delta\{\eta(\mathbf{D}), \cdot\}$ arising under model 2.

First note that, from Robert et al. (2011), we have

$$\begin{aligned} BF_\eta &= \lim_{t \rightarrow 0^+} BF_{abc} & (4) \\ &= \lim_{t \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N \mathbb{I}[\mathcal{M}^{(i)} = 1] \cdot \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\} \leq t]}{\sum_{i=1}^N \mathbb{I}[\mathcal{M}^{(i)} = 2] \cdot \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\} \leq t]} \cdot \frac{\pi(\mathcal{M} = 2)}{\pi(\mathcal{M} = 1)} & (5) \end{aligned}$$

Define K as the total number of times pseudo-data is sampled from model 1 out of N trials and let $L = N - K$. Note that, by construction of the sampling algorithm, K/L converges almost surely to the ratio $\pi(\mathcal{M} = 1)/\pi(\mathcal{M} = 2)$, as $N \rightarrow \infty$, for $\pi(\mathcal{M} = 1) \neq 0$ and $\pi(\mathcal{M} = 2) \neq 0$.

Now let

$$\hat{F}_K(t) = K^{-1} \sum_{k=1}^K \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(k)})\} \leq t | \mathcal{M} = 1] \quad (6)$$

denote the proportion of times that pseudo-data is accepted for threshold t when generated from model 1 and

$$\hat{G}_L(t) = L^{-1} \sum_{l=1}^L \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(l)})\} \leq t | \mathcal{M} = 2] \quad (7)$$

be defined in an analogous manner for pseudo-data from model 2. Noting that as $N \rightarrow \infty$ both $\sup_t |\hat{F}_K(t) - F(t)| \xrightarrow{p} 0$ and $\sup_t |\hat{G}_L(t) - F(t)| \xrightarrow{p} 0$ (please see Appendix C for a proof), this implies that

$$BF_\eta = \lim_{t \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{\sum_{k=1}^K \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(k)})\} \leq t | \mathcal{M} = 1]}{\sum_{l=1}^L \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(l)})\} \leq t | \mathcal{M} = 2]} \cdot \frac{\pi(\mathcal{M} = 2)}{\pi(\mathcal{M} = 1)} \quad (8)$$

$$= \lim_{t \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{K \cdot \hat{F}_K(t)}{L \cdot \hat{G}_L(t)} \cdot \frac{\pi(\mathcal{M} = 2)}{\pi(\mathcal{M} = 1)} \quad (9)$$

$$= \lim_{t \rightarrow 0^+} \frac{F(t)}{G(t)} + o_p(1) \quad (10)$$

$$= \lim_{t \rightarrow 0^+} \frac{F(G^{-1}(G(t)))}{G(t)} + o_p(1) \tag{11}$$

$$= \lim_{p \rightarrow 0^+} \frac{F(G^{-1}(p))}{p} + o_p(1) \tag{12}$$

$$= \lim_{p \rightarrow 0^+} \frac{\text{ROC}(p)}{p} + o_p(1) \tag{13}$$

where $\text{ROC}(p) = F(G^{-1}(p))$.

The relationship expressed between Equations (8) and (13) shows that the ABC Bayes factor for two alternative models of interest can be assigned using the ratio between $\text{ROC}(p)$ and p as the rate of false positives in favour of model 1 approaches 0. This notable result allows us to express the convergence of the ABC Bayes factor as a function of the rate of false positives in favour of model 1. Our result has significant practical implications when it comes to using ABC for model selection:

1. Our solution allows to monitor the convergence of the output of the algorithm to BF_η as a function of a single, well-defined, measure, p , that depends only on the data generated under one of the considered models, as opposed to t which depends on both models and is usually set arbitrarily.
2. Our solution addresses the curse of dimensionality affecting the original ABC model selection algorithm, since it does not require any of the $\Delta\{\eta(\mathbf{D}), \cdot\}$ to be close to 0. Indeed, our solution considers only the relative ranks of the $\Delta\{\eta(\mathbf{D}), \cdot\}$ calculated for the data generated under models 1 and 2.
3. Our approach provides a natural setting for selecting summary statistics that are informative for the model selection task. As with random forest based ABC model selection algorithms (Pudlo et al., 2016), our approach can accommodate large vectors of summary statistics. It requires only the careful design of a kernel function that ensures that the distributions of $\Delta\{\eta(\mathbf{D}), \cdot\}$ are well-separated under the competing models. In other words, we are replacing the variable selection that is part of classifier-based ABC model selection algorithms by choosing a kernel function that weights the components of $\eta(\cdot)$ in a particular and influential way.
4. Our solution uses the entire pseudo-data generated and does so in a computationally efficient manner.
5. Finally, our solution formally preserves the user’s priors on the model indices.

Leveraging this result requires estimation of the ROC curve, which is the topic of the next two sections.

3.1. Empirical ROC

We can leverage the relationship between Equations (8) and (13) to assign an ABC Bayes factor in several ways. Our first method is purely data driven and

uses the following approximation of the ratio in Equation (13):

$$\frac{\widehat{\text{ROC}}(p)}{p} = \frac{\hat{F}_K \left(\hat{G}_L^{-1}(p) \right)}{p} = \frac{\hat{F}_K(t)}{\hat{G}_L(t)} \tag{14}$$

$$= \frac{K^{-1} \sum_{k=1}^K \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(k)})\} \leq t | \mathcal{M} = 1]}{L^{-1} \sum_{l=1}^L \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(l)})\} \leq t | \mathcal{M} = 2]} \tag{15}$$

where $\widehat{\text{ROC}}(\cdot)$ is the empirical estimate of the ROC curve constructed from the set of K $\Delta\{\eta(\mathbf{D}), \cdot\}$ generated under model 1 and the set of L $\Delta\{\eta(\mathbf{D}), \cdot\}$ generated under model 2. Expression (15) can be interpreted as the ratio of the proportion of $\Delta\{\eta(\mathbf{D}), \cdot\}$ arising under model 1 that are less than t and the proportion of $\Delta\{\eta(\mathbf{D}), \cdot\}$ arising under model 2 that are less than t . While this interpretation is still stated in terms of the tolerance, t , the convergence can be expressed in terms of p , by defining t such that $\sum_{l=1}^L \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(l)})\} \leq t | \mathcal{M} = 2]$ is fixed (e.g., define t such that $\sum_{l=1}^L \mathbb{I}[\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(l)})\} \leq t | \mathcal{M} = 2] = 10$), while allowing the total number of simulations, N , and necessarily, L , to increase. As a result, the expression in the denominator of Equation (15) will be driven to 0; hence, approximating the limit as $p \rightarrow 0^+$ in Equation (13). This approach has several major advantages as compared to the original approach for assigning ABC Bayes factors:

1. t is chosen as a function of only the distance scores between the value of the summary statistic of the observed data and the value of the summary statistic of the data generated under model 2 (versus all distance scores in other implementations of the ABC algorithm).
2. t is chosen such that the number of accepted distance scores under model 2 is fixed (versus a fixed value of t arbitrarily close to 0, or a varying value of t based on a fixed quantile of the empirical distribution of all the scores from models 1 and 2 combined).

For a given set of observed data, the original ABC model selection algorithm results in unpredictable variations of both the numerator and the denominator of the ratio in Equation (5) as N increases, which makes its convergence difficult to monitor. By fixing the rate of convergence of the denominator in Equation (5), our approach has the potential to better plan computing resources and monitor convergence in comparison to the original ABC model selection algorithm.

3.2. The non-central dual beta ROC model

Our second approach extends further the relationship between the ABC Bayes factor and the ROC curve by noting (Pepe, 2003):

$$\lim_{p \rightarrow 0^+} \frac{\text{ROC}(p)}{p} = \frac{d}{dp} \text{ROC}(p) \Big|_{p=0} \tag{16}$$

$$= \frac{d}{dp} F(G^{-1}(p)) \Big|_{p=0} \tag{17}$$

$$= \frac{f(G^{-1}(p))}{g(G^{-1}(p))} \Big|_{p=0} \tag{18}$$

$$= \frac{f(G^{-1}(0))}{g(G^{-1}(0))} \tag{19}$$

where $f(\cdot)$ and $g(\cdot)$ denote the probability density functions of distance scores under models 1 and 2, respectively. Assigning an ABC Bayes factor using equation (19) requires evaluation of the first derivative of $\text{ROC}(p)$, evaluated at $p = 0$, or estimation of the ratio of densities, $\frac{f(\cdot)}{g(\cdot)}$ evaluated at $G^{-1}(0)$.

Following Metz, Herman and Shen (1998), Mossman and Peng (2016), and Chen and Hu (2016), we fit a parametric model to the empirical ROC curve obtained from the finite number of $\Delta\{\eta(\mathbf{D}), \cdot\}$ generated by the algorithm. As noted by these authors, fitting a model to each of the score distributions separately requires a direct assumption about the distribution of each set of scores. Instead, fitting a model directly to the ROC curve relies on the weaker assumption that there exists a monotonic transformation of the observed scores that results in scores whose distributions can be described by a simple model. Since the ROC curve is invariant to monotonic increasing transformations of the scores (Pepe, 2003; Metz, Herman and Shen, 1998; Mossman and Peng, 2016; Chen and Hu, 2016), the simple model can be used to represent the ROC curve without knowledge of the underlying score distributions.

We considered the common binormal representation of the ROC (Pepe, 2003), however, it can be easily shown that the first derivative of the binormal model, at $p = 0$, takes a value of 0, 1 or ∞ , depending on the values of its parameters. Instead, we use a model based on two non-central beta distributions (Johnson, Kotz and Balakrishnan, 1995). Placing a restriction on the first shape parameter of each distribution ($\alpha_F = \alpha_G = 1$) results in the following parametric model for the ROC curve

$$\text{ROC}(p) = F(G^{-1}(p) \mid \alpha_G = 1, \beta_G, \lambda_G) \mid \alpha_F = 1, \beta_F, \lambda_F) \tag{20}$$

where $F(\cdot)$ and $G(\cdot)$ are non-central beta distribution functions with parameters $\alpha_F, \beta_F, \lambda_F$, and $\alpha_G, \beta_G, \lambda_G$ corresponding to F and G respectively. We will refer to this parametric model for the ROC curve as the *non-central dual beta ROC model* for the remainder of this paper (Mossman and Peng, 2016). Note that, contrary to current ABC methods for model selection based on pattern recognition algorithms, our method only requires the estimation of four parameters. The restriction on the first shape parameter of the densities guarantees that the first derivative of $\text{ROC}(p)$, at $p = 0$, produces a stable result for all parameter values within the support:

$$\frac{d}{dp}\text{ROC}(p) \Big|_{p=0} = \frac{f(G^{-1}(0))}{g(G^{-1}(0))} \tag{21}$$

$$= \frac{f(0)}{g(0)} \tag{22}$$

$$= \frac{\exp\{-\frac{1}{2}\lambda_F\}}{B(1, \beta_F)} \frac{B(1, \beta_G)}{\exp\{-\frac{1}{2}\lambda_G\}}. \tag{23}$$

Fitting the parametric model requires the use of numerical optimisation techniques to estimate the parameter values. We apply a two step fitting procedure:

1. Obtain an initial set of parameter estimates for $\beta_F, \lambda_F, \beta_G, \lambda_G$ using a Maximum Likelihood Estimation approach (Metz, Herman and Shen, 1998; Mossman and Peng, 2016; Chen and Hu, 2016) for fitting a non-central dual beta ROC model to continuous data.
2. Refine the initial estimates by minimising some distance between the parametric model and the empirical ROC. A distance that is designed to weight differences near $p = 0$ more heavily is preferable. In this case:

$$\sum_{i=1}^n \left(p_i^{-1} \left(F(G^{-1}(p_i | 1, \beta_G, \lambda_G) | 1, \beta_F, \lambda_F) - \widehat{\text{ROC}}(p_i) \right) \right)^2,$$

where $0 < p_1, \dots, p_n \leq 1$.

The first step of our procedure enables us to include information on which model produced each distance score $\Delta\{\eta(\mathbf{D}), \eta(\mathbf{D}^{*(i)})\}$. However, during our implementation we found that the fit of the parametric model to the empirical ROC curve could be improved from this method. Meanwhile, the second step alone did not allow us to fit an adequate model when the score distributions overlapped heavily. Combining both processes gave the best results.

Once estimates for $\beta_F, \lambda_F, \beta_G$, and λ_G are obtained, it is trivial to use Equation (23) to assign the ROC-ABC Bayes factor. In practice, when a limited number of distance scores near 0 are observed, the quality of the fit of the ROC model near 0 can produce Bayes factors of meaningless magnitude (e.g., larger than 10^{100} or smaller than 10^{-100}); these Bayes factors would vary wildly from one computation to another using the same observed data. We found that evaluating the ratio in the left side of Equation (16) using some low value for p , dependent on the sample size, produced more robust ABC Bayes factors in comparison to those assigned using Equation (23) directly. This is likely due to the fact that $p = 0$ is at the boundary of the support of the ROC function.

4. Weight of fingerprint evidence using ROC-ABC

ABC for model selection possesses some obvious similarities with the algorithm proposed by Neumann, Evett and Skerrett (2012). To assign a Bayes factor for an observed fingerprint, Neumann, Evett and Skerrett (2012) algorithm considers two sets of scores. The first set contains scores measuring the level of dissimilarity between the observed fingerprint and *pseudo-fingermarks* generated by Mr. X. The second set contains scores measuring the dissimilarity between the observed fingerprint and *pseudo-fingermarks* originating from a sample of individuals from a population of potential alternative sources. Pseudo-fingermarks from any given individual can be directly obtained by repeatedly sampling impressions from that individual's finger, or by generating artificial impressions (see section

4.1). The general idea behind Neumann, Evett and Skerrett (2012)’s model is that, if H_1 is true, Mr. X. would generate many more pseudo-fingermarks that are similar to the observed fingerprint than the individuals in the population of alternative sources.

If we draw a parallel between Algorithm 1 and the fingerprint problem, we can:

1. replace the observed data \mathbf{D} , by observations, e_u , made on the fingerprint that was recovered at the crime scene and whose donor is *unknown*;
2. replace models 1 and 2 in Equations (2) and (3) by methods to generate pseudo-fingermarks, e_u^* , from any individual considered under the prosecution or the defence model (please refer to section 4.1 which describes generation of pseudo-fingermarks and section 5 which describes the sample of the population of potential sources);
3. define a kernel function, $\Delta\{\cdot, \cdot\}$, to compare pairs of finger impressions;

to obtain an ROC-ABC algorithm to assign a Bayes factor for fingerprint evidence. The ROC-ABC algorithm for fingerprint evidence is summarised in Algorithm 2. The ROC-ABC Bayes factor will converge to the Bayes factor under the same conditions as discussed above (i.e. sufficient statistic across all models, optimal kernel function, infinite/large number of simulations, etc.).

Algorithm 2: ROC-ABC algorithm for fingerprint evidence

```

for  $i = 1$  to  $N$  do
    Sample a model index,  $\mathcal{M}^{(i)}$ , from  $\pi(\mathcal{M} = m)$ , where  $m = 1, 2$ ;
    if  $\mathcal{M}^{(i)} = 1$ : Select the control print;
    else: Sample a print from the alternative source dataset;
    Sample a set of distortion parameters;
    Generate a pseudo-fingerprint,  $e_u^{*(i)}$ , by distorting the selected print using the
    sampled distortion parameters;
    Compute  $\Delta\{\eta(e_u), \eta(e_u^{*(i)})\}$ ;

```

Assign the ROC-ABC Bayes factor using methods from Section 3.1 or 3.2.

4.1. Generation of pseudo-fingerprint data

Implementation of the ROC-ABC algorithm for fingerprint evidence requires a model from which pseudo-fingermarks can be generated. We utilise the same fingerprint distortion model as Neumann, Evett and Skerrett (2012) to generate pseudo-fingermarks from any given control print. This model mimics the way fingerprint features are displaced as the skin on the tip of a finger is distorted when pressed against a flat surface. This model relies on thin-plate splines (Bookstein, 1989) and involves the estimation of the parameters of a 2-dimension vector field (Allasonniere et al., 2013). The parameter space of the model represents a wide variety of distortion directions and pressures. These parameters were estimated

using 704 replicate impressions taken under 11 different controlled directions and 4 pressure conditions from 17 fingers. The model allows many different distortions to be produced from a single finger. Our distortion model assumes that fingers distort in the same way, independently of factors related to the donor (e.g., age, weight, profession) and to the finger number (e.g., right vs. left hand, thumb vs. index finger). The model obviously does not cover all possible distortion and pressure conditions, and donor related factors; however, it is currently the only option to obtain a large number of pseudo-fingermarks from any given individual.

4.1.1. Generation of pseudo-fingermarks under the prosecution model

When comparing fingerprints, an examiner first detects k features of interest on the fingermark. Second, the examiner compares it to the 10 control prints from the donor considered by the prosecution proposition and selects the finger appearing to be the most likely source of the mark. Finally, the examiner attempts to identify the most similar k corresponding features out of the n features present on the control print from the selected finger. Note that, once selected, the sets of features on the fingermark and the control print remain fixed for the duration of the experiment, and that the selection process results in a unique bijective pairing between the two sets of k features. Our algorithm assumes that this selection and pairing process has been done before generating pseudo-fingermarks under M_1 . When M_1 is selected by the algorithm, a pseudo-fingermark is generated from the k features selected on the control print using the distortion model. By construction, the features on this pseudo-fingermark have the same pairing with the features of the fingermark as the ones on the control print. By repeating this process each time M_1 is selected, a set of pseudo-fingermarks can be obtained from the k features selected on the control print of Mr. X.

4.1.2. Generation of pseudo-fingermarks under the defence model

The defence proposition considers that, if the fingermark was not left by Mr. X, it must have been left by another person in a relevant population of alternative sources. We can generate data under M_2 , first, by randomly selecting an individual from a representative sample of donors from that population, and secondly, by generating a pseudo-fingermark from this individual's k minutiae configuration that is most similar to the k minutiae observed on the fingermark. As in the previous section, the features of this pseudo-fingermark have a unique bijective pairing with the features on the fingermark by construction. This process is repeated each time M_2 is selected to obtain a random set of pseudo-fingermarks from the population of potential donors considered by M_2 . Since it would be unrealistic to repeat manually the selection of the most similar k minutiae for each individual in a large sample, we use a commercially available automated fingerprint matching system to perform this task.

TABLE 1
Variable types of various measurements from a friction ridge pattern.

	measurement	variable type
minutia locations	Cartesian coordinates in pixels	continuous
minutia orientations	angular measure	circular
minutia types	bifurcation vs. ridge ending	discrete

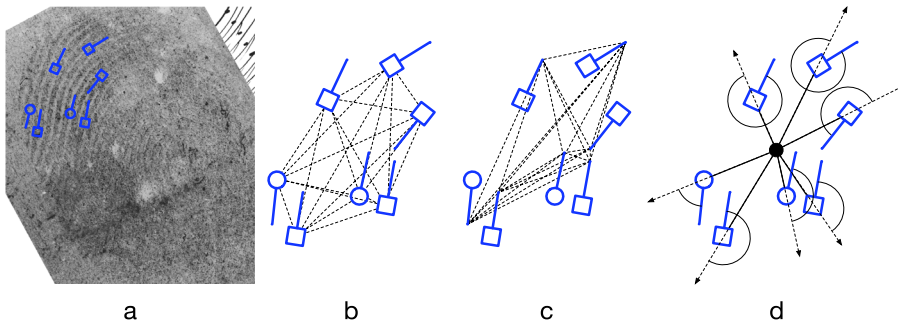


FIG 3. (a) An annotated configuration of features on a friction ridge pattern. Squares and circles denote the types of the features and serve as markers for the locations of the features, while the extending line indicates the direction of the feature. (b) Cross-distances between minutiae locations. (c) Cross-distances between the ends of the fixed-length segments originating from the centre of the minutiae and oriented in the same direction as the minutiae. (d) Angles between the axes going from the centroid through the locations of the minutiae and the segments representing the minutiae directions are indicated by the solid line.

4.2. Summary statistic

Configurations of minutiae can be described numerically in the form of heterogeneous multi-dimensional random vectors containing the measurements summarised in Table 1. Minutiae locations and orientations are taken with respect to a coordinate system based on the frame of the impression’s picture (Figure 3 (a)). Different framings of the same impression result in different measurements for the same set of features. For this reason, the original measurements need to be summarised in a way that is rotation and translation invariant.

Several invariant measurements capturing the spatial relationships between the minutiae in a configuration can be calculated, such as the distances between every pair of minutiae in the configuration (Figure 3 (b)). A similar approach can be used to define invariant summaries of the direction of each feature by using fixed-length segments to represent minutiae directions and by taking the distances between the ends of these segments for every pair of minutiae in the configuration (Figure 3 (c)). We also choose to represent minutiae directions as a function of the axes going from the centroid through the location of each minutia (angles are measured counterclockwise) (Figure 3 (d)). Feature types can be directly recorded since types are rotation and translation invariant.

All of these measurements can be brought together to create a vector of sum-

mary statistics of the original representation. For the example in Figure 3 with 7 features, a vector of summary statistics would include 21 cross-distances between pairs of minutiae, 21 cross-distances to capture the spatial representation of the minutiae directions, 7 angles and 7 types, for a total length of 56. For 10 features, the length of the vector of summary statistics would be 110; and for 15 features, it would be 240.

Given the nature of the problem, it is unlikely that a sufficient summary statistic exists for fingerprint data. Hence, we adopt an approach, similar to the one taken by Pudlo et al. (2016), which consists in pooling together as many individual summary statistics as possible in order to minimise the loss of information with respect to the original data. However, contrarily to previously proposed ABC model selection approaches, our method replaces the summary statistic selection process by a careful design of the kernel function that weights the different components of $\eta(\cdot)$ in order to maximise the separation between the distributions of $\Delta\{\eta(\mathbf{e}_u), \cdot\}$. Therefore, the dimensionality of the summary statistics and the computational aspects are not so problematic for our algorithm.

4.3. Kernel function

Our ABC algorithm depends on a kernel function, $\Delta\{\cdot, \cdot\}$, which compares pairs of summarised configurations of minutiae. As mentioned in Section 3, we wish to use a kernel function that best separates the distributions of $\Delta\{\eta(\mathbf{e}_u), \cdot\}$ obtained under the competing models considered in Section 1 and which satisfies the conditions for a metric to be used in an ABC algorithm (Prangle, 2017).

We carefully designed a kernel function to compare pairs of summarised configurations of minutiae, and optimised the weights of several components to best separate the two distributions of $\Delta\{\eta(\mathbf{e}_u), \cdot\}$. For completeness, we have included this development and optimisation process in Appendix A; however, we stress that other summary statistics, kernel functions and optimisation procedures could be considered without loss of generality of our proposed ROC-ABC method.

4.4. Number of simulations

For the purpose of this application, we limited the total number of pseudo-fingermarks generated for each test configuration to 500,000 (approximately 250,000 under each model, due to our choice of equal model priors). If it was of interest to assign the ROC-ABC Bayes factor for a specific fingermark in forensic casework, a much larger number of pseudo-fingermarks could be generated.

5. Datasets

The algorithm has been developed, optimised, and tested using two datasets. The first dataset (Population Dataset), containing prints taken under controlled conditions from more than 400,000 individuals (only identified through ran-

domly assigned ID numbers), was used as a sample of a population of potential sources. The size of the dataset was not driven by scientific considerations, but corresponds to the number of control prints that the authors managed to gather for research purposes. This dataset was used to generate pseudo-fingermarks under the defence model (see Section 4.1.2) and to generate special test cases under the prosecution model (see below).

The second dataset (Test Dataset) contains 207 quadruplets of finger impressions. Each quadruplet includes:

1. a fingermark obtained from casework archives. These fingermarks were developed on multiple surfaces using different physicochemical methods and represent a range of fingermarks that can be observed in casework;
2. a control print from an individual deemed to be the true source (TS) of the fingermark by an experienced fingerprint examiner;
3. the closest matching control print from another individual (CNM) than the true source, as selected by an automated fingerprint matching system among the individuals contained in Population Dataset;
4. a random control print from another individual (RS) than the true source selected among the individuals contained in Population Dataset;

Within each quadruplet, the features observed on the fingermark and the TS print were manually labelled, while the features observed on the CNM and RS prints were automatically labelled. Each feature observed on the fingermark was manually paired to a single feature on the TS print. Each feature observed on the fingermark was automatically paired with the single best corresponding feature on the CNM print and on the RS print by the automated fingerprint matching system used in this study.

Each dataset comprises fingermarks or control prints captured digitally at 1:1 magnification and a resolution of 500 pixels per inch. The following features were extracted from every fingerprint used in the study:

1. the finger of origin of the impression, from 1 – right thumb – to 10 – left little finger (for control prints only);
2. the Cartesian coordinates of each minutia in pixels, using the bottom left of the image as the origin;
3. the direction of each minutia in radians, using the bottom left of the image as the origin and measuring counterclockwise;
4. the type of each minutia: ridge ending, bifurcation or unknown;
5. the Cartesian coordinates of the centre of the impression (for control prints only).

6. Results

We sampled 4067 minutiae configurations ranging from 3 to 25 minutiae from the 207 fingermarks contained in the Test Dataset (one configuration with a given number of minutiae per fingermark). For each minutiae configuration, the corresponding minutiae on the TS, CNM and RS prints were also selected, thus

TABLE 2
Number of test configurations for each number of minutiae in the Test Dataset.

# of minutiae	3	4	5	6	7	8	9	10	11	12
# of configurations	207	207	207	205	203	203	199	197	190	187
# of minutiae	13	14	15	16	17	18	19	20	21	22
# of configurations	184	179	174	173	170	168	162	154	151	144
# of minutiae	23	24	25							total
# of configurations	142	133	128							4067

TABLE 3
Rates of misleading evidence for the experiments presented in Figures 4 and 5.

method	scenario	# of minutiae													
		3	4	5	6	7	8	9	10	11	12	13	14		
empirical ROC	TS	0.44	0.47	0.28	0.21	0.07	0.06	0.04	0.02	0.02	0.02	0.01	0.01		
	CNM	0.20	0.09	0.09	0.06	0.16	0.17	0.10	0.19	0.14	0.14	0.16	0.18		
	RS	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.01		
non-central dual beta ROC	TS	0.40	0.32	0.15	0.07	0.03	0.04	0.04	0.01	0.01	0.01	0.01	0.01		
	CNM	0.42	0.33	0.51	0.43	0.59	0.58	0.40	0.49	0.40	0.35	0.37	0.38		
	RS	0.12	0.04	0.05	0.06	0.03	0.03	0.01	0.03	0.02	0.01	0.01	0.04		
method	scenario	# of minutiae													
		15	16	17	18	19	20	21	22	23	24	25			
empirical ROC	TS	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.02		
	CNM	0.12	0.16	0.11	0.11	0.10	0.10	0.07	0.09	0.06	0.06	0.06			
	RS	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.01	0.01	0.00			
non-central dual beta ROC	TS	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.00	0.00		
	CNM	0.37	0.30	0.22	0.21	0.19	0.21	0.15	0.22	0.07	0.14	0.12			
	RS	0.01	0.03	0.02	0.02	0.01	0.03	0.04	0.03	0.00	0.02	0.02			

providing quadruplets of “matching” configurations. Table 2 summarises the number of configurations per number of minutiae.

The configurations sampled from the fingerprint were used as e_u when generating the results presented below. The corresponding configurations on the TS, CNM and RS prints were used as control prints from Mr. X to test the ROC-ABC algorithm under three different prosecution propositions, H_1 :

- TS: The fingerprint originates from the donor of the TS print;
 CNM: The fingerprint originates from the donor of the CNM print;
 RS: The fingerprint originates from the donor of the RS print.

In these experiments, the configurations on the TS, CNM and RS prints were used to resample pseudo-fingermarks under M_1 , and the Population Dataset was used to resample pseudo-fingermarks under M_2 . Results for all three experiments can be found in Figures 4 and 5, and in Table 3. ROC-ABC Bayes factors presented in Figure 4 were assigned using the empirical ROC method discussed in Section 3.1. Figure 5 presents ROC-ABC Bayes factors that were assigned using the non-central dual beta ROC model described in Section 3.2. Note that the vertical axis in Figure 5 has been truncated to focus on the mass of the distributions and that some extreme outliers may not be represented.

In experiment TS, the control prints originate from the true sources of the marks, and so H_1 is true. Both figures show a similar behaviour where the magnitude of the ROC-ABC Bayes factor increases as the number of minutiae increases. In both cases, the ROC-ABC Bayes factors appear bounded. The bound for the empirical ROC-ABC Bayes factors stems from Equation (14), which requires us to set a value for p . In this case, we defined p as the false

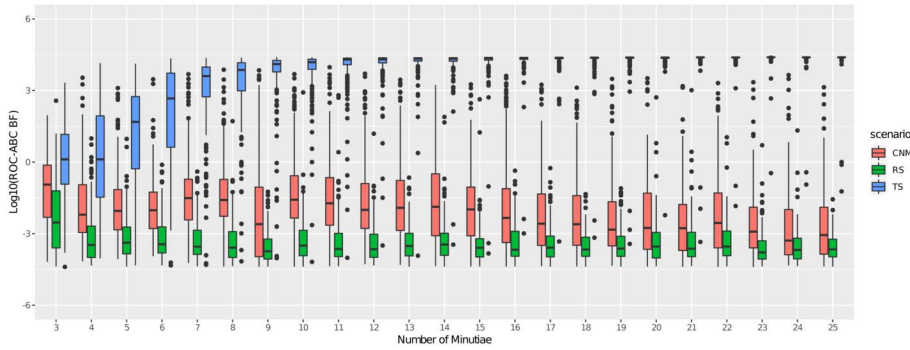


FIG 4. Results obtained using the empirical ROC method. Blue: results from the experiment where the control prints originate from the true sources (TS). Red: results when the control prints originate from sources with close non-matching prints (CNM). Green: results when the control prints originate from randomly selected sources (RS).

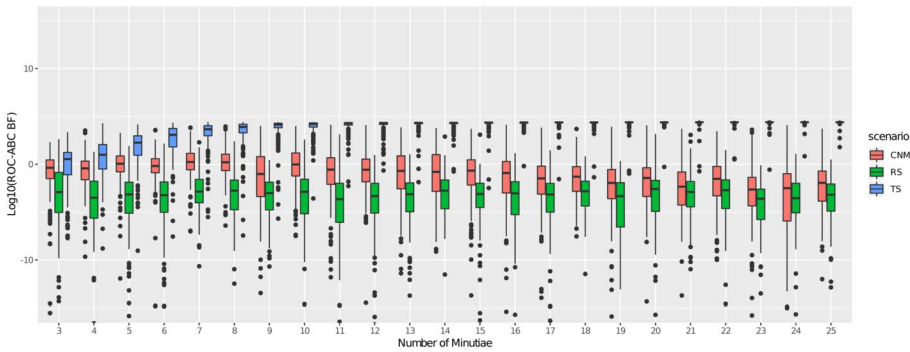


FIG 5. Results obtained using the non-central dual beta ROC model. Blue: results from the experiment where the control prints originate from the true sources (TS). Red: results when the control prints originate from sources with close non-matching prints (CNM). Green: results when the control prints originate from randomly selected sources (RS).

positive rate evaluated at the 10^{th} smallest $\Delta\{\eta(e_u), \cdot\}$ generated under H_2 . This corresponds to approximately 1 in 25,000. The bound for the non-central dual beta ROC model stems from using the same p to evaluate the ratio in the left side of Equation (16) as in the empirical model. Bayes factors erroneously supporting H_2 are noted in both series of results for smaller configurations of minutiae (from 40% of configurations with 3 minutiae down to 5% of configurations with 7 minutiae, as seen in Table 3). This result is not surprising as these configurations contain less discriminative information and are more likely to be observed by chance in fingers from different individuals. Nevertheless, we note that only around 1% of the ROC-ABC Bayes factors support the wrong proposition for configurations with 10 minutiae or more (Table 3). Upon further investigation, we found that these cases involve configurations displaying unusual distortion

that cannot be handled by the current generation of the distortion algorithm, or minutiae that are miss-paired by the fingerprint analyst between the fingermark and the control print. Improvement of the summary statistic, kernel function, distortion algorithm as well as procedures for documenting comparisons may be able to minimise further the number of cases where the Bayes factor erroneously supports H_2 . Finally, we observe that the range of values taken by the ROC-ABC Bayes factor for configurations with different numbers of minutiae overlap. This result supports the observations made by Neumann, Evett and Skerrett (2012) that there is no scientific justification for the use of a fixed number of minutiae as a decision point to distinguish between H_1 and H_2 , and that the evidential value of each configuration needs to be quantified based on its own characteristics.

In experiment CNM, the control prints originate from non-mated sources that are chosen due to their similarity to the minutiae configuration on the fingermarks, and so H_2 is true. In both Figures 4 and 5, we observe that a large proportion of the ROC-ABC Bayes factors erroneously support the hypothesis of common source, H_1 , although the algorithm using the empirical ROC appears to perform significantly better (Table 3). The high rate of misleading evidence is not a surprise for low numbers of minutiae (e.g., between 20% and 42% for configurations with 3 minutiae, as seen in Table 3) since it is not difficult to find multiple similar configurations on different fingers in large a dataset. The high rate of misleading evidence is more surprising for configurations with more minutiae (e.g., between 18% and 38% for configurations of 14 minutiae, as seen in Table 3). Larger values of the ROC-ABC Bayes factor when H_2 is true occur when the kernel function used by the algorithm considers the pseudo-fingermarks generated using M_1 more similar to the observed fingermark than they really are. As mentioned before, improvement of the summary statistic and the kernel function should significantly reduce the rate of misleading evidence in favour of H_1 . In practice, we believe that examiners comparing close non-matching finger impressions would be able to exclude that they originate from a common source by visual inspection using friction ridge characteristics that are not taken into account by our kernel function.

In experiment RS, the control prints are obtained from randomly selected sources, and so H_2 is also true. In both cases the majority of observations correctly support the hypothesis of different sources, H_2 . As in the second experiment, the algorithm using the empirical ROC curve significantly outperforms the one based on the non-central dual beta ROC model (Table 3).

Overall, we note that the parametric modelling of the ROC curve needs to be improved. In many situations, it appears that the non-central dual beta ROC model is not optimal. We are currently investigating more flexible models.

These experiments were repeated using the logistic regression method by Beaumont (2008) (see Appendix B). Results can be found in Figure 8 and Table 4 of Appendix B. It is important to note that the logistic method does not present an obvious upper bound for the results in experiment TS and assigns Bayes factors with notably large magnitudes. This property is not desirable since it may lead to unrealistic magnitude of support.

A comparison of the computational times for the empirical and parametric

ROC methods and the logistic method is presented in Figure 9 of Appendix B. The empirical ROC-ABC method was without rival in terms of computation time. The logistic method performed at a much slower rate, with the difference between the two methods increasing with the dimensionality of the data. In comparison to the widely used Beaumont (2008) ABC algorithm, we find that our method provides the high computational efficiency that is necessary to provide real-time calculation of the weight of forensic evidence in casework.

7. Discussion

The contribution of this paper is twofold. Firstly, we have proposed a method to rigorously quantify the weight of fingerprint evidence using the formal statistical framework provided by ABC. Our application addresses the issues raised in relation to Neumann, Evett and Skerrett (2012) and provides a much needed general framework for the quantification of the weight of any type of forensic pattern evidence, as long as a kernel function can be defined to compare two pieces of evidence. Secondly, our ROC-ABC algorithm addresses some issues associated with the original ABC model selection algorithm. Our intention is not to extend on the more modern approaches, but to propose an alternative approach to these strategies while serving to address shortcomings related to the choice of a suitable threshold in the original ABC model selection algorithm.

Overall, our results are consistent with the results presented by Neumann, Evett and Skerrett (2012) while capturing the user's belief about the parameters of the two competing models and providing an alternative to the weighting function that they proposed:

1. The probability of misleading evidence in favour of the defence proposition decreases dramatically as the number of minutiae increases.
2. The probability of misleading evidence in favour of the prosecution proposition is very low for configurations with more than 7 minutiae, when the donor has been randomly selected. As expected, this probability is higher when the donor has been selected based on the similarity of its fingerprints with the fingermark. Improvements in the summary statistic used to describe fingerprint patterns and in the kernel function used to compare them, together with the ability of fingerprint examiners to account for more discriminative information than our model, will certainly reduce the rate of misleading evidence in favour of the prosecution proposition in an operational implementation of the model.
3. The overlap between the ranges of values of the ROC-ABC Bayes factor across different numbers of minutiae confirms that the use of the number of corresponding minutiae is only one of the criteria for inferring the identity of the source, and that the contribution of additional information regarding fingerprint pattern needs to be taken into account when determining the donor of a fingermark.

In addition to the straightforward operational implementation of our ROC-ABC approach, Figures 6 and 7 illustrate the powerful visual representation

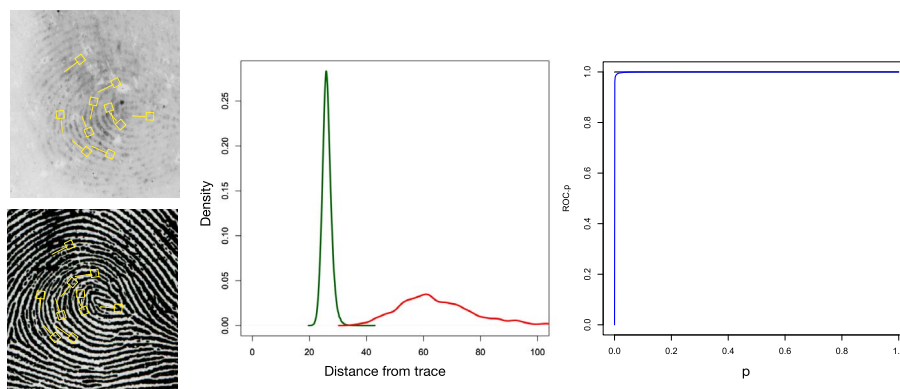


FIG 6. Left panel: Fingermark (top) and control print (bottom), both originating from the same source. Corresponding features between the two impressions have been annotated in yellow. Middle panel: Kernel density estimates of the densities of the scores generated under H_1 (green) and H_2 (red). Note that the distribution of scores generated under H_1 is situated much closer to 0, but neither of the curves cover 0. Right panel: Empirical ROC curve (black) which is nearly perfectly overlapped by the parametric ROC curve (blue) generated from the distributions of scores in the middle panel. Note the steep slope of the curve near $p = 0$.

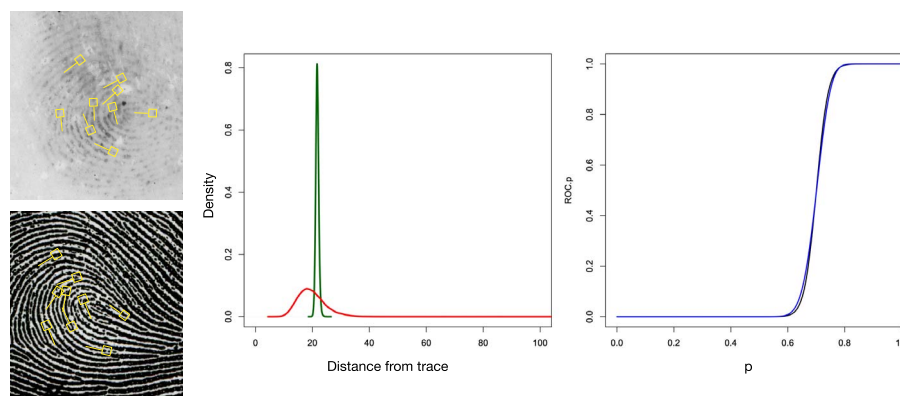


FIG 7. Left panel: Fingermark (top) and control print (bottom) originating from the different sources. Potentially corresponding features between the two impressions have been annotated in yellow. Middle panel: Kernel density estimates of the densities of the scores generated under H_1 (green) and H_2 (red). Note that the algorithm has detected many more random impressions that are more similar to the fingermark than the impression from the suspect. Right panel: Empirical ROC curve (black) and parametric ROC curve (blue) generated from the distributions of scores in the middle panel. Note the flat slope of the curve near $p = 0$ indicating that the algorithm supports H_2 .

of the probative value of fingerprint comparisons that is offered by the ROC-ABC algorithm. In Figure 6, the fingermark originates from same source as the control print and $BF_{abc} = 2.51 \times 10^7$ ($\log_{10} BF_{abc} = 7.4$). The score distributions

and the ROC curve in the middle and right panels intuitively show that the data supports the hypothesis that the fingerprint and control prints originate from the same source. In contrast, Figure 7 shows a situation where the same fingerprint as in Figure 6 is compared to a randomly selected control print. In this scenario, accounting for computational approximation, $BF_{abc} = 3.56 \times 10^{-153}$ is virtually 0. The clear support of the algorithm for the hypothesis that the fingerprint and control impressions do not originate from the same source can be seen using the score distributions and ROC curve in the middle and right panels. We believe that this intuition can easily be conveyed to jurors and other factfinders.

We do not claim that the kernel function that was used in this paper is optimal. It is worth exploring adaptive kernels that maximise the separation between the pseudo-data generated by both models in any specific case. Nevertheless, while the summary statistic and the kernel function used to generate the results presented in this paper can be improved upon, they are adequate to show the potential of the concept of the ROC-ABC algorithm. Operational implementation of the method would require further studies of the repeatability of the values obtained by the algorithm as a function of different samples (and different sample sizes) of the population of potential donors considered by H_2 .

Our proposed ROC-ABC algorithm results in several improvements to the original ABC model selection algorithm. Our approach, based on properties of the ROC curve, transforms the convergence of the algorithm into a function of the rate of false positives in favour of the model considered in the numerator of the Bayes factor. The original ABC model selection algorithm results in unpredictable variations of both the numerator and the denominator of the ABC Bayes factor as the number of simulations, N , increases, which makes the convergence of the algorithm more difficult to monitor. In our approach, the tolerance level, t , for a given data set is chosen such that the number of accepted samples under the model considered in the denominator of the Bayes factor is fixed for all N . Hence, as N increases, the approximation of the limit as the rate of false positives approaches 0 improves. Our approach has the potential to better plan computing resources. Critically, our method allows for rigorously monitoring convergence.

Additionally, the shift of focus from tolerance level on $\Delta\{\eta(\mathbf{D}), \cdot\}$ to the rate of false positives in favour of model 1 does not require any of the $\Delta\{\eta(\mathbf{D}), \cdot\}$ to be close to 0. Instead, only the relative ranks of the $\Delta\{\eta(\mathbf{D}), \cdot\}$ calculated for the data generated under models 1 and 2 are considered. This implies that the kernel function used to assess level of similarity can accommodate large vectors of summary statistics because there is no need for any of the scores to be close to 0. Our algorithm only requires a carefully designed kernel function that maximises the separation between the distributions of $\Delta\{\eta(\mathbf{e}_u), \cdot\}$. In addition to having the ability to accommodate large vectors of summary statistics, our method is also able to process the entire amount of pseudo-data generated in a computationally efficient manner and does not require filtering the pseudo-data: as the dimension of the summary statistic vector increases, the time required to assign Bayes factors using other methods (such as the logistic regression approach) increases

exponentially (Figure 9). Finally, our solution allows to formally preserve the user's priors on the model indices. Nevertheless, our approach may produce inconsistent Bayes factors as their inverses may no longer be the Bayes factors for the other model. We are currently investigating this and other properties of our approach.

In our approach, we have not formally addressed the issue of sufficiency of the summary statistic across models that is required for the convergence of the ABC Bayes factor to the Bayes factor. This convergence is extremely important in some contexts, such as forensic science, where fact-finders are as equally interested in the proposition supported by the Bayes factor as in the magnitude of this support. However, since the method that we propose is able to accommodate large vectors of summary statistics, it permits including as much information in the summary statistic vectors with the aim that they will jointly tend toward sufficiency by decreasing the loss of information.

To implement our approach in practice, we have proposed two methods to assign ROC-ABC Bayes factors. Our results show significant differences in performance between the empirical ROC and the non-central dual beta ROC method. Currently, the empirical model appears to produce more stable and meaningful results (i.e., ABC Bayes factors with reasonable magnitude). As we increase the number of simulations, the empirical model naturally approximates the limit as $p \rightarrow 0^+$ in Equation (13). A parametric ROC model has the potential to explore the limit as $p \rightarrow 0^+$ in Equation (13) with a smaller sample size. In our application using the non-central dual beta ROC model, we observed that the values obtained using Equation (23) for multiple runs of the ROC-ABC algorithm for a given set of observed data differ greatly from one another (many orders of magnitude on the \log_{10} scale). It appears that Equation (23) is very sensitive to small changes in the values of the estimates of the model's parameters. Instead, we evaluated the ratio in the left side of Equation (16) at approximately $\frac{1}{25,000}$ (corresponding to the false positive rate evaluated at the 10th smallest $\Delta\{\eta(e_u), \cdot\}$ generated under H_2) to obtain more robust values and generate the data in Figure 5. Our results show that in several cases our algorithm does not support the correct model; this may be due to our modelling of the ROC curve in the neighbourhood of 0 not being an accurate representation of the data. Once again, this shows that the non-central dual beta ROC model is very sensitive to small changes in the estimates of its parameters. Improvements could be made to the fitting procedure for the non-central dual beta ROC model, such as an explicit monotonic increasing transformation of the distance scores to initiate the numerical optimisation procedure with distributions that are closer to the assumed model; in addition, we are currently investigating other parametric approaches for estimating the first derivative of the ROC curve for $p = 0$.

8. Conclusion

In this paper, we propose an algorithm to formally and rigorously assign Bayes factors to forensic fingerprint evidence. Our ROC-ABC model selection algo-

rithm was used to address several criticisms of the model proposed by Neumann, Evett and Skerrett (2012) by framing the problem into a formal Bayesian framework. The results presented here show that our method is promising, with low rates of misleading evidence, and has the potential to be applied to many other complex, high-dimension evidence forms such as shoe prints, questioned documents, firearms, and paint fragments, glass fragments, and fibers characterised by analytical chemistry. Ultimately, the widespread use of statistical approaches to quantify the weight of forensic evidence to replace the existing inference paradigm can only be enabled by technology providers offering commercial products to the forensic community. Our method leverages currently available technology that was designed to search forensic traces recovered at crime scenes into large databases and retrieve the most likely candidates. For mainstream evidence types such as fingerprints, firearms, and shoe impressions, our algorithm can readily be implemented, validated, and integrated in current commercial offerings, such as Automatic Fingerprint Identification Systems. Furthermore, we note that the use of ROC curves in the algorithm will be naturally familiar to engineers and scientists designing these systems, which may facilitate the implementation of our method in commercial systems. In addition to its straightforward implementation, which leverages currently available technology and its computational efficiency, our method provides a visually intuitive presentation for lay individuals (i.e., jurors).

As an added benefit, our algorithm addresses several shortcomings of the original ABC model selection algorithm. We use properties of the Receiver Operating Characteristic curve to address the issue of choosing a suitable tolerance level when assigning ABC Bayes factors using the original approach. Our modification allows for a natural convergence of the algorithm as the number of simulations increases, and for monitoring this convergence as a function of the sole rate of false positives in favour of the model considered in the numerator of the Bayes factor. Focusing on the rate of false positives (rather than the tolerance level) allows our method to rely on the ordering of scores, rather than the magnitudes of scores, and thus, is less affected by the curse of dimensionality in comparison to the original ABC model selection algorithm. In addition, our method considers the entire amount of pseudo-data generated under the considered models in a computationally efficient manner, preserving the prior beliefs of the user.

Appendix A: Kernel function development

As mentioned in Section 4.3, we wish to use a kernel function that best separates the distributions of $\Delta\{\eta(\mathbf{e}_u), \cdot\}$ obtained under the competing models considered in Section 1. Our kernel function, $\Delta\{\cdot, \cdot\}$, is a linear combination of several metrics, denoted by $\Delta_1\{\cdot, \cdot\}$, $\Delta_2\{\cdot, \cdot\}$, and $\Delta_3\{\cdot, \cdot\}$, corresponding to the different summary statistics described above (see Section 4.2) and aimed at capturing differences in spatial relationships and directions of the features:

$$\Delta\{\cdot, \cdot\} = c_1\Delta_1\{\cdot, \cdot\} + c_2\Delta_2\{\cdot, \cdot\} + c_3\Delta_3\{\cdot, \cdot\} \quad (24)$$

where c_i , for $i \in \{1, 2, 3\}$, are real-valued constants. The components of the kernel function are described below. We remind the reader that by construction (see Section 4.1) the i^{th} measurement in \mathbf{e}_u is uniquely paired with the i^{th} measurement in any given pseudo-fingermark, \mathbf{e}_u^* .

A.1. Components of the kernel function

The first component, $\Delta_1\{\cdot, \cdot\}$, captures the differences in cross-distances between the locations of the minutiae in a pair of configurations (as illustrated in Figure 3 (b)). Denoting the i^{th} cross-distance from \mathbf{e}_u by d_i , and the i^{th} cross-distance from \mathbf{e}_u^* by d_i^* , the first component of the kernel function is given by $\Delta_1\{\mathbf{e}_u, \mathbf{e}_u^*\} = \left(\sum_{i=1}^k \frac{(d_i - d_i^*)^2}{d_i} \right)^{1/2}$. The second component, $\Delta_2\{\cdot, \cdot\}$, takes the same form as $\Delta_1\{\cdot, \cdot\}$, but instead uses d_i and d_i^* as the i^{th} cross-distance between location markers for feature directions (as illustrated in Figure 3 (c)) on \mathbf{e}_u and \mathbf{e}_u^* respectively. This component captures differences in directions of the features. The third component, $\Delta_3\{\cdot, \cdot\}$, captures the difference in direction between the paired features for two configurations. Denoting the angle (measured in degrees) depicted in Figure 3 (d) for the i^{th} minutiae of \mathbf{e}_u by ρ_i , and the same from \mathbf{e}_u^* by ρ_i^* , the third component is given by

$$\Delta_3\{\mathbf{e}_u, \mathbf{e}_u^*\} = \sum_{i=1}^k \begin{cases} \frac{|\rho_i - \rho_i^*|}{\rho_i} & \text{if } |\rho_i - \rho_i^*| \leq 180 \\ \frac{\rho_i - |\rho_i - \rho_i^*| \pmod{180}}{\rho_i} & \text{if } |\rho_i - \rho_i^*| > 180 \end{cases}.$$

We considered including minutiae type in the kernel function, however, we discovered that this characteristic was not reliable enough to be included.

A.2. Optimisation of the kernel function

As noted in Section 4.2, our method replaces the summary statistic selection process by carefully designing a kernel function that weights the different components of $\eta(\cdot)$ in order to maximise the separation between the distributions of $\Delta\{\eta(\mathbf{e}_u), \cdot\}$. Values for all c_i in Equation (24) can be obtained by maximising the separation between the distribution of $\Delta\{\cdot, \cdot\}$'s from minutiae configurations generated by the same donor, and the distribution of $\Delta\{\cdot, \cdot\}$'s from minutiae configurations generated by different donors. To obtain the results presented later in this paper, we used numerical optimisation to maximise the average area under 450 ROC curves obtained from configurations with $k = \{5, 8, 12, 17, 23\}$ minutiae. Each ROC curve was based on 50,000 distance scores as calculated in Equation 24 and obtained by comparing k minutiae on a fingermark to pseudo-fingermarks resulting from the distortion of the true source of the fingermark (Section 4.1.1) and from the distortion of other fingers (Section 4.1.2). Our results indicated that component 2 was the most useful to maximise the average

area under the ROC curve ($c_2 = 6.5$), followed by component 1 ($c_1 = 1$) and by component 3 ($c_3 = 0.1$).

We stress that other summary statistics, kernel functions and optimisation procedures could be considered without loss of generality of our proposed ROC-ABC method.

Appendix B: Results using Beaumont (2008)'s method

Results for all three experiments were also generated using the logistic regression method proposed by Beaumont (2008) since this method is the basis for most current ABC model selection algorithms and has been widely used and tested. We use the logistic regression method with the same kernel function as the one used in our ROC-based ABC algorithm in order to have directly comparable results. We want to re-emphasise that it is critical for the forensic application to have a fair idea of the magnitude of support for a given model, and not only to be able to select the correct one. Thus, we did not compare our method to machine learning based ABC methods since they seem to be only focusing on the posterior distributions for the models (Pudlo et al., 2016).

These results are presented in Figure 8 and Table 4. A comparison of the computation times for each of the three methods (empirical ROC, parametric ROC, and logistic regression) is presented in Figure 9.

The general trend of the results of the logistic regression method are similar to those of the ROC-based methods. For the experiment TS, the magnitude of the ABC Bayes factor increases as the number of minutiae increases, while for experiments CNM and RS, the ABC Bayes factors tend to generally support H_2 .

However, the logistic method does not present an obvious upper bound for the results in experiment TS and assigns Bayes factors with notably large magnitudes. Based on the convergence results shown in Equations (8) to (13), we do not believe that the larger magnitude of the Bayes factors in Figure 8 can be justified by the number of simulations performed in this experiment. We can only conclude that these Bayes factors severely overstate the weight of the evidence observed and generated in these cases. In addition, we note the very large variance of the Bayes factor assigned using the logistic regression method during experiment TS. This large variance results in a high rate of misleading evidence in favour of H_2 . This rate is noticeably greater than that from the empirical method (Table 4).

Interestingly, results from experiments CNM and RS show that the logistic regression method produces less misleading evidence in favour of H_1 when H_2 is true, even when very similar prints are used (Table 4). The logistic regression method maximises the separation between the two models by leveraging all of the content of the vectors of summary statistics, while the kernel function described in Section 4.3 has been optimised for the average case. The ROC-ABC method may be improved in this aspect by using an adaptable kernel function that would also maximise the separation in each case.

TABLE 4
Rates of misleading evidence for the experiments presented in Figures 4, 5, and 8.

method	scenario	# of minutiae													
empirical ROC	TS	0.44	0.47	0.28	0.21	0.07	0.06	0.04	0.02	0.02	0.02	0.01	0.01	0.01	0.01
	CNM	0.20	0.09	0.09	0.06	0.16	0.17	0.10	0.19	0.14	0.14	0.16	0.18	0.18	0.18
	RS	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.01	0.00	0.01
non-central dual beta ROC	TS	0.40	0.32	0.15	0.07	0.03	0.04	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	CNM	0.42	0.33	0.51	0.43	0.59	0.58	0.40	0.49	0.40	0.35	0.37	0.38	0.38	0.38
	RS	0.12	0.04	0.05	0.06	0.03	0.03	0.01	0.03	0.02	0.01	0.01	0.01	0.01	0.04
Beaumont (2008)	TS	0.57	0.49	0.39	0.37	0.27	0.27	0.22	0.20	0.13	0.08	0.07	0.06	0.06	0.06
	CNM	0.13	0.06	0.05	0.02	0.02	0.02	0.02	0.00	0.02	0.00	0.01	0.00	0.01	0.00
	RS	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

method	scenario	# of minutiae											
empirical ROC	TS	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.02
	CNM	0.12	0.16	0.11	0.11	0.10	0.10	0.07	0.09	0.06	0.06	0.06	0.06
	RS	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.01	0.01	0.01	0.00
non-central dual beta ROC	TS	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.00	0.00	0.00
	CNM	0.37	0.30	0.22	0.21	0.19	0.21	0.15	0.22	0.07	0.14	0.12	0.12
	RS	0.01	0.03	0.02	0.02	0.01	0.03	0.04	0.03	0.00	0.02	0.02	0.02
Beaumont (2008)	TS	0.03	0.02	0.02	0.04	0.04	0.02	0.03	0.02	0.09	0.05	0.07	0.07
	CNM	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	RS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

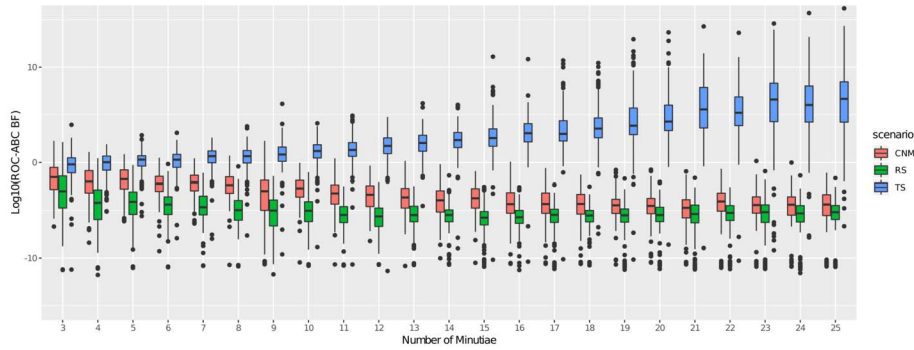


FIG 8. Results obtained using the logistic regression method. Blue: results from the experiment where the control prints originate from the true sources (TS). Red: results when the control prints originate from sources with close non-matching prints (CNM). Green: results when the control prints originate from randomly selected sources (RS).

Overall, while using the logistic regression method as proposed by Beaumont (2008), we noted that the weighting of the pseudo-data prior to fitting the logistic regression model resulted in the removal of large portions, if not all, of the data generated under either M_1 or M_2 . As discussed previously, this results in altering the user-defined priors on the model index and replacing them by unpredictable data-driven priors. Furthermore, this led to instability when fitting the logistic regression model.

A comparison of the computation time among the three methods (empirical ROC, parametric ROC, and logistic regression) is presented in Figure 9. This computation time represents the time required to assign ABC Bayes factors using the three different methods once the pseudo-data has been generated. The empirical ROC-ABC method was without rival in terms of computation time. Even as data complexity / dimensionality increased, computation time was rel-

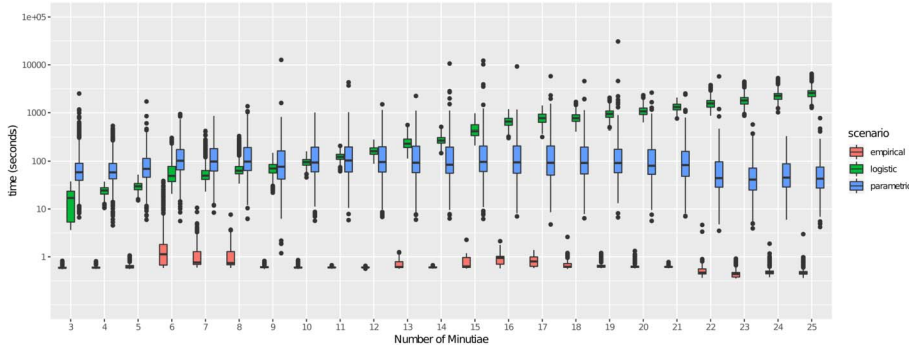


FIG 9. A comparison of the computation time for the empirical ROC method (red), the parametric ROC model (blue), and logistic regression method (green). The computation time represents the time required to assign Bayes factors once the pseudo-data has been generated.

atively constant. The logistic method outperformed the parametric ROC-ABC method up until 9 minutiae. At this point, the total computation time for the logistic method continued to increase at an exponential rate while the computation time for the parametric ROC-ABC method remained fairly uniform. This is unsurprising since the computational complexity of the parametric ROC-ABC is driven by the number of univariate scores in the ROC curve, and not by the dimension of the vectors of summary statistics. In addition to an increase in computing time, an increase in computing resources was also required by the logistic regression method such that fingerprint data with more than 22 features could not be processed on a standard desktop computer, while the ROC-based methods have a very small computing footprint (once the initial pseudo-data has been generated).

Appendix C: Proof of convergence result

Theorem. Main result. Let $Z_i \stackrel{iid}{\sim} \text{Bernoulli}(\pi)$, where $\pi \in (0, 1)$, and $X_i \stackrel{iid}{\sim} F$ for $i = 1, 2, \dots$. Define

$$\hat{F}_n(t) = \frac{\sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t]}{\sum_{i=1}^n Z_i} \tag{25}$$

where $t \in \mathbb{R}$. Then,

$$\sup_t |\hat{F}_n(t) - F(t)| \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

A proof of this theorem will follow after proofs of two preliminary results.

Theorem. Preliminary result (a). Define

$$\tilde{F}_n(t) = \frac{\sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t]}{\pi n}. \tag{26}$$

Then

$$\sup_t \left| \tilde{F}_n(t) - \hat{F}_n(t) \right| \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Proof. Consider

$$\sup_t \left| \tilde{F}_n(t) - \hat{F}_n(t) \right| = \sup_t \left| \frac{\sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t]}{\pi n} - \frac{\sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t]}{\sum_{i=1}^n Z_i} \right| \tag{27}$$

$$= \sup_t \left| \left(\frac{1}{\pi n} - \frac{1}{\sum_{i=1}^n Z_i} \right) \left(\sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t] \right) \right| \tag{28}$$

$$= \sup_t \left| \left(\frac{1}{\pi} - \frac{n}{\sum_{i=1}^n Z_i} \right) \left(n^{-1} \sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t] \right) \right| \tag{29}$$

$$= \sup_t \left| n^{-1} \sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t] \right| \left| \frac{1}{\pi} - \frac{n}{\sum_{i=1}^n Z_i} \right| \tag{30}$$

$$= O_p(1) o_p(1) \tag{31}$$

$$= o_p(1). \tag{32}$$

Equality (27) is by definition; equalities (28) and (29) are through algebraic manipulation; in equality (30), the second term has been removed from $\sup_t |\cdot|$ since it does not involve t . Because $Z_i \mathbb{I}[X_i \leq t]$ is bounded, then $\sup_t |n^{-1} \sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t]| = O_p(1)$. Also, $\frac{n}{\sum_{i=1}^n Z_i} \xrightarrow{a.s.} \frac{1}{\pi}$ as $n \rightarrow \infty$, so $\left| \frac{1}{\pi} - \frac{n}{\sum_{i=1}^n Z_i} \right| = o_p(1)$. \square

Theorem. *Preliminary result (b). Define $F_n(t) = n^{-1} \sum_{i=1}^n \mathbb{I}[X_i \leq t]$. Then*

$$\sup_t \left| \tilde{F}_n(t) - F_n(t) \right| \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Proof. Consider

$$\sup_t \left| \tilde{F}_n(t) - F_n(t) \right| = \sup_t \left| (\pi n)^{-1} \sum_{i=1}^n Z_i \mathbb{I}[X_i \leq t] - n^{-1} \sum_{i=1}^n \mathbb{I}[X_i \leq t] \right| \tag{33}$$

$$= \sup_t \left| (\pi n)^{-1} \sum_{i=1}^n (Z_i - \pi) \mathbb{I}[X_i \leq t] \right| \tag{34}$$

$$\xrightarrow{p} 0 \text{ as } n \rightarrow \infty. \tag{35}$$

Equality (33) is by definition. Equality (34) is a result of algebraic manipulation. Noting that $E[(Z_i - \pi)\mathbb{I}[X_i \leq t]] = E(Z_i - \pi) E(\mathbb{I}[X_i \leq t]) = 0$, the convergence on line (35) is implied by the Weak Law of Large Numbers. \square

A proof of the main result follows.

Proof. Consider

$$\sup_t |\hat{F}_n(t) - F(t)| = \sup_t |\hat{F}_n(t) - \tilde{F}_n(t) + \tilde{F}_n(t) - F(t)| \tag{36}$$

$$\leq \sup_t |\tilde{F}_n(t) - \hat{F}_n(t)| + \sup_t |\tilde{F}_n(t) - F(t)| \quad (37)$$

$$= \sup_t |\tilde{F}_n(t) - \hat{F}_n(t)| + \sup_t |\tilde{F}_n(t) - F_n(t) + F_n(t) - F(t)| \quad (38)$$

$$\leq \sup_t |\tilde{F}_n(t) - \hat{F}_n(t)| + \sup_t |\tilde{F}_n(t) - F_n(t)| + \sup_t |F_n(t) - F(t)| \quad (39)$$

$$= o_p(1) + o_p(1) + o_p(1) \quad (40)$$

$$= o_p(1) \quad (41)$$

Equation (36) is obtained by adding an identity for 0; equation (37) is by the triangle inequality; equation (38) is obtained by adding an identity for 0; equation (39) is by the triangle inequality; the first term in equation (40) is by *preliminary result (a)*, the second term is by *preliminary result (b)*, and the third term is a result of the Glivenko-Cantelli Theorem. Equality (32) is a result of the Continuous Mapping Theorem.

Therefore, $\sup_t |\hat{F}_n(t) - F(t)| \xrightarrow{p} 0$ as $n \rightarrow \infty$. \square

Acknowledgments

We would like to thank Dr. Danica Ommen for her help in the initial implementation of ABC for forensic evidence. We would also like to thank the anonymous reviewers whose comments and suggestions helped to improve and clarify our manuscript.

References

- ABRAHAM, J., CHAMPOD, C., LENNARD, C. and ROUX, C. (2013). Modern statistical models for forensic fingerprint examinations: a critical review. *Forensic Science International* **232** 131-150.
- AITKEN, C., ROBERTS, P. and JACKSON, G. (2010). 1. Fundamentals of Probability and Statistical Evidence in Criminal Proceedings. In *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice, Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses* Royal Statistical Society.
- AITKEN, C. G. G. and TARONI, F. (2004). *Evaluation of Evidence for Forensic Scientists*, 2nd ed. Wiley and Sons Ltd, Chichester.
- ALLASSONNIERE, S., BIGOT, J., GLAUNES, J. A., MAIRE, F., and RICHARD, F. J. P. (2013). Statistical models for deformable templates in image and shape analysis. *Annales Mathematiques Blaise Pascal* **20** 1-35. [MR3112238](#)
- BALDING, D. J. (2012). Comments on Neumann, C., Evett, I. W., Skerrett, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm in *J. R. Statist. Soc. A* **175** 317-396. *J. R. Statist. Soc. A* **175** 397-398. [MR2905045](#)

- BEAUMONT, M. A. (2008). Joint determination of topology, divergence time, and immigration in population trees. In *Simulation, Genetics and Human Pre-history*, (S. Matsumura, P. Forster and C. Renfrew, eds.). *McDonald Institute Monographs* 14, 135-154. Cambridge McDonald Institute for Archeological Research, UK.
- BLUM, M. and FRANCOIS, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing* **20** 63-73. [MR2578077](#)
- BOOKSTEIN, F. L. (1989). Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11** 567-585.
- CHEN, W. and HU, N. (2016). Proper bibeta ROC model: algorithm, software, and performance evaluation. In *SPIE Medical Imaging* 97870E-97870E. International Society for Optics and Photonics.
- COLE, S. A. (2004). Grandfathering evidence: fingerprint admissibility rulings from Jennings to Llera Plaza and back again. *American Criminal Law Review* **41** 1189-1276.
- COLE, S. A. (2005). More than zero, accounting for error in latent fingerprint identification. *Journal of Criminal Law and Criminology* **95** 985-1078.
- COLE, S. A. (2009). Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification. *Law, Probability and Risk* **In press**.
- COOK, R., EVETT, I. W., JACKSON, G., JONES, P. J. and LAMBERT, J. A. (1998). A hierarchy of propositions: deciding which level to address in case-work. *Science and Justice* **38** 213-239.
- DIDELOT, X., EVERITT, R. G., JOHANSEN, A. M. and LAWSON, D. J. (2011). Likelihood-free estimation of model evidence. *Bayesian Analysis* **6** 49-76. [MR2781808](#)
- ESTOUP, A., LOMBAERT, E., MARIN, J. M., GUILLEMAUD, T., PUDLO, P., ROBERT, C. P. and CORNUET, J. M. (2012). Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources* **12** 846-855.
- GRELAUD, A., ROBERT, C. P., MARIN, J. M., RODOLPHE, F. and TALY, J. F. (2009). ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis* **4** 317-335. [MR2507366](#)
- JANDHYALA, V. K. and FOTOPOULOS, S. B. (2012). Comments on Neumann, C., Evett, I. W., Skerrett, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm in *J. R. Statist. Soc. A* **175** 317-396. *J. R. Statist. Soc. A* **175** 404. [MR2905045](#)
- JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, N. (1995). *Continuous univariate distributions* **2**. Wiley, New York. [MR1326603](#)
- KADANE, J. B. (2012). Comments on Neumann, C., Evett, I. W., Skerrett, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm in *J. R. Statist. Soc. A* **175** 317-396. *J. R. Statist. Soc. A* **175** 404. [MR2905045](#)
- KAYE, D. H. (2003). Questioning a courtroom proof of the uniqueness of fin-

- gerprints. *International Statistical Review* **71** 521-533.
- LAURITZEN, S., COWELL, R. and GRAVERSEN, T. (2012). Comments on Neumann, C., Evett, I. W., Skerrett, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm in *J. R. Statist. Soc. A* **175** 371-396. *J. R. Statist. Soc. A* **175** 405-406. [MR2905045](#)
- MARIN, J. M. and ROBERT, C. P. (2014). *Bayesian Essentials with R*, Second ed. Springer Science + Business Media, New York. [MR3136532](#)
- METZ, C. E., HERMAN, B. A. and SHEN, J. H. (1998). Maximum Likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine* **17** 1033-1053.
- MOSSMAN, D. and PENG, H. (2016). Using Dual Beta Distributions to Create “Proper” ROC Curves Based on Rating Category Data. *Medical Decision Making* **36** 349-365.
- NEUMANN, C., EVETT, I. W. and SKERRETT, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J. R. Statist. Soc. A* **175** 371-396. [MR2905045](#)
- PEPE, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press Inc., New York. [MR2260483](#)
- PRANGLE, D. (2017). Adapting the ABC distance function. *Bayesian Analysis* **12** 289-309. [MR3620131](#)
- PRANGLE, D., FEARNHEAD, P., COX, M. P., BIGGS, P. J. and FRENCH, N. P. (2014). Semi-automatic selection of summary statistics for ABC model choice. *Statistical Applications in Genetics and Molecular Biology* **13** 67-82. [MR3159118](#)
- PRITCHARD, J. K., SEIELSTAD, M. T., PEREZ-LEZAUN, A. and FELDMAN, M. W. (1999). Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites. *Molecular Biology and Evolution* **16** 1791-1798.
- PUDLO, P., MARIN, J. M., ESTOUP, A., CORNUET, J. M., GAUTIER, M. and ROBERT, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics* **32** 859-866.
- ROBERT, C. P. (2007). *The Bayesian Choice*. Springer. [MR2723361](#)
- ROBERT, C. P., CORNUET, J. M., M., M. J. and PILLAI, N. S. (2011). Lack of confidence in ABC model choice. *Proceedings of the National Academy of Sciences of the United States of America* **108** 5112-5117.
- SAKS, M. J. and KOEHLER, J. J. (2005). The coming paradigm shift in forensic identification science. *Science* **309** 893-894.
- SAKS, M. J. and KOEHLER, J. J. (2008). The Individualization Fallacy in Forensic Science Evidence. *Vanderbilt Law Review* **61** 199-219.
- SHAO, J. (2003). *Mathematical Statistics*. Springer, New York. [MR2002723](#)
- SISSON, S. A., FAN, Y. and BEAUMONT, M. A. (2019). Overview of ABC. In *Handbook of Approximate Bayesian Computation* (S. A. Sisson, Y. Fan and M. A. Beaumont, eds.) 1, 3-54. CRC Press. [MR3889278](#)
- STERN, H. (2012). Comments on Neumann, C., Evett, I. W., Skerrett, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm in *J. R. Statist. Soc. A* **175** 317-396. *J. R. Statist. Soc. A* **175**

- 408-409. [MR2905045](#)
- TONI, T. and STUMPF, M. P. H. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* **26** 104-110.
- STATE V. DIXON (2011). No. 27-CR-10-3378. (*D. Ct. Cty. Hennepin, Minn.*).
- STATE V. HULL (2008). No. 48-CR-07-2336. (*Minn. D. Ct. Cty. of Mille Lacs*).
- VAN DER VAART, A. W. (1998). *Cambridge Series in Statistical and Probabilistic Mathematics: Asymptotic Statistics*. Cambridge University Press, Cambridge, UK. [MR1652247](#)
- ZABELL, S. L. (2005). Fingerprint evidence. *Journal of Law and Policy* **13** 143-170.