The evaluation of evidence for microspectrophotometry data using functional data analysis

Journal Pre-proof

The evaluation of evidence for microspectrophotometry data using functional data analysis

Colin Aitken, Ya-Ting Chang, Patrick Buzzini, Grzegorz Zadora, Genevieve Massonnet

PII: S0379-0738(19)30419-0

DOI: https://doi.org/10.1016/j.forsciint.2019.110007

Reference: FSI 110007

To appear in: *Forensic Science International*

Received Date: 6 July 2019

Revised Date: 1 October 2019

Accepted Date: 21 October 2019

Please cite this article as: Colin Aitken, Ya-Ting Chang, Patrick Buzzini, Grzegorz Zadora, Genevieve Massonnet, The evaluation of evidence for microspectrophotometry data using functional data analysis, *<![CDATA[Forensic Science International]]>* (2019), doi: https://doi.org/10.1016/j.forsciint.2019.110007

The evaluation of evidence for microspectrophotometry data using functional data analysis

Abstract

Microspecrophotometry data arise in the study of many forensically applicable situations. The situations here are those of ink and fibres. In a criminal investigation, data associated with a crime scene are compared with data associated with a person of interest. Methods based on the likelihood ratio are often used to evaluate such evidence. A technique known as functional data analysis for determining likelihood ratios using the full spectrum is described. It provides support comparing a proposition of common source with a proposition of different sources for data from the crime scene and from the person of interest. Data are available from ink, woollen and cotton fibres. The effectiveness of the method is assessed using false positive and false negative rates and Tippett plots in the comparison of data from known sources.

Keywords: Microspectrophotometry, Evidence evaluation, Likelihood ratio, Functional data analysis

1. Introduction

1.1. Evidence evaluation

A novel approach to the evaluation of evidence in the form of microspectrophotometry (MSP) data using functional data analysis is described. The data take the form of functions rather than measurements or counts. There are three data sets to which the approach is applied, one from ink, one from red woollen fibres and one from red cotton fibres. There are training data for each set. The evidence to be evaluated is considered to be in two parts, a *control* part in which the source of the data is known and a *recovered* part in which the source of the data is not known. Often the control data will be a pen, source of ink, found in the possession of the person of interest, or fibres found on clothing of the person of interest. The recovered data will then be the ink on a document of interest or fibres found at a crime scene thought to have been transferred from clothing worn by the criminal.

The evidence is evaluated through use of the likelihood ratio (LR). In this context, the ratio is that of the probability density function of the control and recovered data[1] under each of two propositions, where the data are themselves functions.

- H_p: the control and recovered data come from the same source, and
- H_d: the control and recovered data come from different sources.

The microspectrophotometry (MSP) data used for the analysis, $\mathbf{Y}_c$ and $\mathbf{Y}_r$, for control and recovered data, respectively, are absorbance (A) in the case of ink or transmittance (T) in the case of woollen and cotton fibres versus wavelength w, where $A = -\log(T)$. The evidence of $\mathbf{Y}_c$ and $\mathbf{Y}_r$ in support of H_p or H_d is to be evaluated. Training data $\mathbf{W}$ are available from ink and woollen and cotton fibres. These training data are used to develop the functional

[1]The ratio is known as a likelihood ratio as this is a technical phrase in statistical theory in which a probability density function for data given parameter values may be thought of as a likelihood of the parameter values given the data. The phrase has been transferred in the forensic statistic literature to refer to the ratio of the probability density functions given propositions. Note that as the data are continuous it is not possible to refer to the probability of the data.

data model used for evaluation. In order to reduce the dimensionality of the problem, analysis is conducted with transformed data $\mathbf{Z} = g(\mathbf{Y})$ where the transformation g is described in the Model section

The model developed is a multivariate random effects model for $\mathbf{Z}$. The within-group distribution is assumed normal with a mean $\boldsymbol{\theta}$ and fixed covariance matrix U, so $\mathbf{Z} \sim N(\boldsymbol{\theta}, U)$. The mean $\boldsymbol{\theta}$ is also assumed to have a normal distribution with fixed mean $\boldsymbol{\eta}$ and a fixed covariance matrix C, so $\boldsymbol{\theta} \sim N(\boldsymbol{\eta}, C)$. The current exposition is an initial attempt to evaluate evidence in the form of functions. Future work could include consideration of a non-parametric distribution for the between-group variation.

The likelihood ratio LR associated with the propositions H_p and H_d is given by:

$$\frac{f(\mathbf{z} \mid H_p)}{f(\mathbf{z} \mid H_d)} = \frac{\int f(\mathbf{z}_c \mid \boldsymbol{\theta}, U) f(\mathbf{z}_r \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\eta}, C) d\boldsymbol{\theta}}{\int f(\mathbf{z}_c \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\eta}, C) d\boldsymbol{\theta} \int f(\mathbf{z}_r \mid \boldsymbol{\theta}, U) f(\boldsymbol{\theta} \mid \boldsymbol{\eta}, C) d\boldsymbol{\theta}}.$$

1.2. Functional data analysis

Functional data analysis is the analysis of data in the form of functions. The examples discussed here relate to MSP data on inks (Section 2.1) and red woollen and cotton fibres (Section 2.2). For ink the data are absorbance *vs* wavelength and for fibres the data are transmittance *vs* wavelength. The observed values of absorbance or transmittance are the sum of an underlying function $f(x)$ of wavelength x and an error term to account for random variation. See Burfield et al. [8] for a description of proofs of concept for comparison, classification and database search of forensic ink chromatograms. Ramsay and Silverman [10] provides the theoretical background for functional data analysis.

Functions can be represented by Fourier series for cyclical data, such as those relating to seasonal data or by B-splines for non-cyclical data such as the MSP data of the examples under discussion. Spline functions are the most common choice of approximation system for non-periodic functional data, A spline function is a piecewise polynomial. The spline is determined by the order o (degree $o-1$) of the polynomial and a non-decreasing knot sequence τ where adjacent polynomial pieces meet. The number B of splines (or bases) is related linearly to the order o of the polynomial and the number N of internal knots such that $B = N + o$. B-splines and equidistant knots are assumed so a choice has to be made in the selection of B and of o.

Consider a set $\mathbf{x} = (x_1, \ldots, x_m)^T$ of m observed values of wavelength and corresponding function values $f(\mathbf{x})$. A system of basis functions is a set of B known functions $\phi_b, \quad b = 1, \ldots, B$, that are orthogonal and have the property that any function can be approximated arbitrarily well by a linear combination of spline functions for a sufficiently large value for B. Thus the function values $f(\mathbf{x})$ may be approximated by a finite linear combination of basis functions

$$\sum_{b=1}^{B} \theta_b \phi_b(\mathbf{x}) = \Phi(\mathbf{x})\Theta. \tag{1}$$

where $B < \infty, \Theta = (\theta_1, \ldots, \theta_B)^T$ and

$$\Phi(\mathbf{x}) = \begin{pmatrix} \phi_1(x_1) & \ldots & \phi_B(x_1) \\ \vdots & & \vdots \\ \phi_1(x_m) & \ldots & \phi_B(x_m) \end{pmatrix}$$

The data to be analysed here are hierarchical. There are several samples taken to be a random sample from some relevant population. Within each sample there are several MSP data which are bivariate with absorbance (for ink) or transmittance (for fibre) and wavelength recorded. Consider transmittance or absorbance $\mathbf{W}_{ki} = (w_{ki1}, \ldots, w_{kim})^T$ measured at m wavelengths $\{\ell_{kij}; \ k = 1, \ldots, K, i = 1, \ldots, n, j = 1, \ldots, m\}$ from K groups with n measurements from each group. The m wavelengths are taken to be equally spaced with $\ell_{ki1} < \ldots < \ell_{kim}$ and $\ell_{k_{p1}i_{q1}j} = \ell_{k_{p2}i_{q2}j}$ for all $p1 \neq p2$ and $q1 \neq q2, \ j = 1, \ldots, m$.

The data are expressed as

$$w_{kij} \quad = \quad f_k(j/m) + r_{kij}, \quad k = 1, \ldots, K, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m, \tag{2}$$

for some smooth functions f_k, dependent on the group k, and measurement errors $\mathbf{r}$ that are independent of f_k. The space covered by f_k is standardised to $[0,1]$ by division of j by m. The functions $f_k(\mathbf{x})$ are estimated by B-splines. In general, $f_k(\mathbf{x}) \simeq \sum_{b=1}^{B} \theta_b^k \phi_b(\mathbf{x}) = \Phi(\mathbf{x})\boldsymbol{\theta}_k$ as in (1). Note that for the analysis that follows the standardised x is only available at a sample of discrete points $(\frac{1}{m}, \ldots, \frac{m-1}{m}, 1)$.

2. Data

2.1. Ink

Forty blue inks (36 ballpoint and 4 gel) were analysed. They came either from the Polish market, or were gifts presented to the Institute of Forensic Research in Krakow. Lines were made by drawing on white printing paper (80 gm^2, A4). A fragment of the paper was then cut and fixed to a microscope base slide, and placed on the stage of the microscope with the MSP instrument. The data are reported in Mar [4]. See also Zad [6] for further details of the analytical procedures.

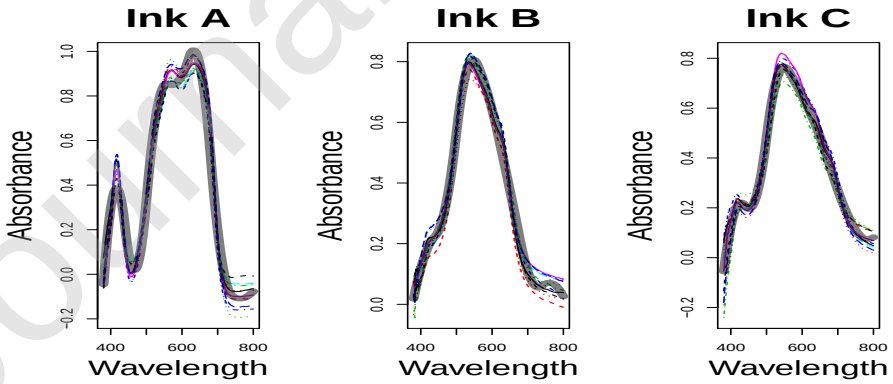
Measurements were made upon blue inks using a microspectrophotometer Zeiss Axioplan 2 with a J & M Tidas Diode Array Detector (DAD; MCS/16 1024/100-1, Germany), which was configured for the VIS range (380–800 nm) analyses with intervals of 1 nm. The inks were measured in reflection mode using an integration time of 2.5 seconds, the magnification being 400×. Sample plots of absorbance *versus* wavelength for three of the forty inks, together with fitted curves

$$\begin{aligned} \hat{y}_k &= \Phi\hat{\theta}_k \text{ where} \\ \hat{\theta}_k &= (\Phi^T\Phi)^{-1}\Phi^T \sum_{i=1}^{k} \mathbf{w}_{ki}/n_k \end{aligned} \tag{3}$$

with $\mathbf{w}_{ki} = \{w_{ki1}, \ldots w_{kim}\}$ and nine bases of order three are shown in Figure 1.

The software used for spectra collection was Spectralys version 1.82 from J & M Tidas. It was also used to provide the tristimulus values CIE-XYZ, chromaticity coordinates CIE-xyz, and CIE-Lab values for the 2 degree standard observer. Thus, the dimensionality of data for determination of the likelihood ratio by Mar [4] was reduced to one, two or three dimensions.

Figure 1: Plot of absorbance *versus* wavelength for three of the forty inks together with fitted mean curves (3) with nine bases of order 3



2.2. Fibres

Data are available from MSP spectra for samples of red woollen fibres which can be easily told apart with a discriminating power of 98% from visual inspection and samples of red cotton fibres which cannot be easily told apart

with a discriminating power of 58% using visual inspection. Both datasets consist of 20 samples. Nine replicate MSP spectra were collected for each sample, three spectra in three different regions of a single fibre and three different fibres known to be from the same textile. Buz [3] contain more detail on these data. Plots of transmittance *versus* wavelength for three of the twenty woollen fibres and three of the twenty cotton fibres together with fitted curves (3) with eight bases of order three are shown in Figures 2 and 3, respectively.

It is not important to be able to determine that a spectrum is from a particular fibre within a set of fibres from the same textile. Such a determination is only important in cases where fibres of similar type and colour (e.g., red wool) are dyed with different dyes which is not the case here. It is more important to gauge intra-sample variation based on all nine measurements.

The red wool dataset included data of spectra from 350 to 690 nm with intervals of five nanometers. The red cotton dataset included data from 240 to 690 nm also with intervals of five nanometers. The difference in range is because wool absorbs UV radiation and therefore the signal of the fibre prevents the detection of the signal of the dye in the UV range between 240 and 350 nm as already noted by Wiggins and Drummond [14] and Almer et al. [7]. The red cotton dataset does not exhibit much variation from visual inspection of the spectra while the red wool dataset exhibits large inter-sample variation.

Figure 2: Plot of transmittance *versus* wavelength for three of the twenty woollen fibres together with fitted mean curves (3) with eight bases and order 3.

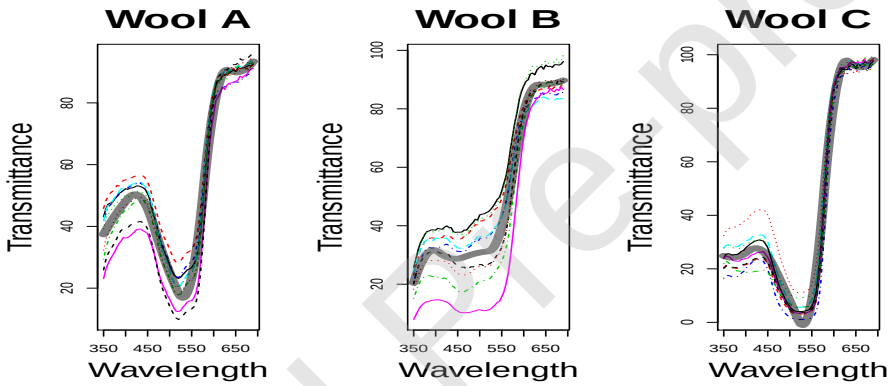
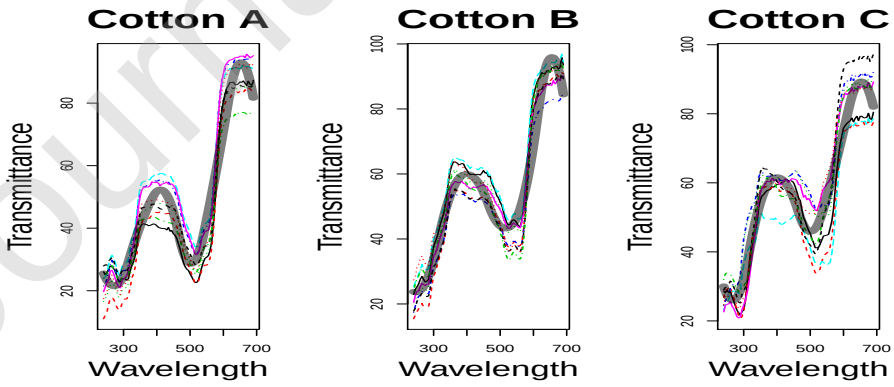


Figure 3: Plot of transmittance *versus* wavelength for three of the twenty cotton fibres together with fitted mean curves (3) with eight bases and order 3.



3. Likelihood ratio for functional data

The training data $\mathbf{W}$ have two levels. There are K groups and within each group there are n members. For ink, $K = 40$ and $n = 10$. For wool and cotton, $K = 20$ and $n = 9$. Measurements of absorbance or transmittance are taken at each of m wavelengths, with constant intervals between wavelengths. For ink the interval is 1 nm and for wool and cotton 5 nm. An individual measurement is thus represented by $w_{kij}; k = 1, \ldots, K, i = 1, \ldots n, j = 1, \ldots, m$ which is the absorbance or transmittance at wavelength j for member i of group k. Each member of each group $\mathbf{w}_{ki} = \{w_{ki1}, \ldots, w_{kim}\}$, is assumed to be a randomly selected member of a function f_k representing the relationship between wavelength and absorbance or transmittance for group k. The K groups are assumed to be a randomly selected number of groups from a relevant population. For model fitting, wavelengths are replaced by increments j/m for $j = 1, \ldots, m$.

A non-parametric regression model (2) is assumed for $\mathbf{W}$ where $f_k(\mathbf{x})$ is estimated by a B-spline $\sum_{b=1}^{B} \theta_b^{(k)} \phi_b(\mathbf{x}) = \mathbf{\Phi}(\mathbf{x})\Theta_k$and $\mathbf{\Phi}(\mathbf{x})$ is an m by B matrix of B-spline bases of order o; $x = \{1/m, \ldots, m/m\}$ and r_{kij} is an error term.

The data $\mathbf{w}_{ki}$ are m-dimensional where m can be very large, of the order of hundreds. The dimension is reduced with the transformation

$$\mathbf{Z}_{ki} = \mathbf{\Phi}^T \mathbf{W}_{ki}$$

so $\mathbf{Z}_{ki}$ is a $B \times 1$ vector. The problem of the evaluation of evidence is then transformed to one in which the data are a vector of B terms.

Analysis is as follows

- $\mathbf{Z}_{ki} \mid \boldsymbol{\theta}_k \sim N_B(\boldsymbol{\theta}_k, U), \quad k = 1, \ldots, K, \quad i = 1, \ldots, n$
- $\boldsymbol{\theta}_k \sim N_B(\boldsymbol{\eta}, C)$.

There are control data and recovered data forming the evidence to evaluated. Denote the control data by $y_{cij}, i = 1, \ldots, n_c; j = 1, \ldots, m$ with transformed data $\mathbf{z}_{ci} = \mathbf{\Phi}^T \mathbf{y}_{ci}$ with $\mathbf{Z}_{ci} \sim N(\boldsymbol{\theta}_c, U)$. Denote the recovered data by $y_{rij}, i = 1, \ldots, n_r; j = 1, \ldots, m$ with transformed data $\mathbf{z}_{ri} = \mathbf{\Phi}^T \mathbf{y}_{ri}$ with $\mathbf{Z}_{ri} \sim N(\boldsymbol{\theta}_r, U)$. The means $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}_r$ are distributed $N(\boldsymbol{\eta}, C)$.

3.1. Likelihood Ratio Calculation - Numerator

The numerator of the likelihood ratio is evaluated under the proposition that the data for the recovered curve and the control curve come from the same origin, or $\boldsymbol{\theta}_r = \boldsymbol{\theta}_c$.

$$\int \prod_{i=1}^{n_c} f(\mathbf{z}_{ci}|\boldsymbol{\theta}, U) \prod_{i=1}^{n_r} f(\mathbf{z}_{ri}|\boldsymbol{\theta}, U) f(\boldsymbol{\theta}|\boldsymbol{\eta}, C) d\boldsymbol{\theta}$$

which can be shown [1] to simplify to

$$|2\pi U|^{-(n_c+n_r)/2} |2\pi C|^{-1/2} |2\pi((n_c + n_r)U^{-1} + C^{-1})^{-1}|^{1/2} \exp\{-\frac{1}{2}(H_1 + H_2 + H_3)\}$$

where

$$
\begin{aligned}
H_1 &= \sum_{i=1}^{n_c}(\mathbf{z}_{ci}-\bar{\mathbf{z}}_c)^T U^{-1}(\mathbf{z}_{ci}-\bar{\mathbf{z}}_c) + \sum_{i=1}^{n_r}(\mathbf{z}_{ri}-\bar{\mathbf{z}}_r)^T U^{-1}(\mathbf{z}_{ri}-\bar{\mathbf{z}}_r) \\
&= \text{tr}(\sum_{i=1}^{n_c}(\mathbf{z}_{ci}-\bar{\mathbf{z}}_c)^T U^{-1}(\mathbf{z}_{ci}-\bar{\mathbf{z}}_c)) + \text{tr}(\sum_{i=1}^{n_r}(\mathbf{z}_{ri}-\bar{\mathbf{z}}_r)^T U^{-1}(\mathbf{z}_{ri}-\bar{\mathbf{z}}_r)) \\
&= \text{tr}(\sum_{i=1}^{n_c}(\mathbf{z}_{ci}-\bar{\mathbf{z}}_c)(\mathbf{z}_{ci}-\bar{\mathbf{z}}_c)^T U^{-1}) + \text{tr}(\sum_{i=1}^{n_r}(\mathbf{z}_{ri}-\bar{\mathbf{z}}_r)(\mathbf{z}_{ri}-\bar{\mathbf{z}}_r)^T U^{-1}) \\
&= \text{tr}(S_c U^{-1}) + \text{tr}(S_r U^{-1}) \\
H_2 &= (\mathbf{z}^*-\boldsymbol{\eta})^T(\frac{U}{n_c+n_r}+C)^{-1}(\mathbf{z}^*-\boldsymbol{\eta}) \\
H_3 &= (\bar{\mathbf{z}}_c-\bar{\mathbf{z}}_r)^T(\frac{U}{n_c}+\frac{U}{n_r})^{-1}(\bar{\mathbf{z}}_c-\bar{\mathbf{z}}_r)
\end{aligned}
$$

with

$$\mathbf{z}^* = \frac{n_c\bar{\mathbf{z}}_c + n_r\bar{\mathbf{z}}_r}{n_c+n_r}$$

3.2. Likelihood Ratio Calculation - Denominator

The denominator of the likelihood ratio is evaluated under the proposition that the data for the recovered curve and the control curve come from different origins, independently.

$$\int\prod_{i=1}^{n_c} f(\mathbf{z}_{ci}|\boldsymbol{\theta}_c,U)f(\boldsymbol{\theta}_c|\boldsymbol{\eta},C)d\boldsymbol{\theta}_c \int\prod_{i=1}^{n_r} f(\mathbf{z}_{ri}|\boldsymbol{\theta}_r,U)f(\boldsymbol{\theta}_r|\boldsymbol{\eta},C)d\boldsymbol{\theta}_r$$

These two parts are similar, so the calculation is showed for the general case.

$$
\begin{aligned}
&\int\prod_{i=1}^{n_q} f(\mathbf{z}_{qi}|\boldsymbol{\theta}_q,U)f(\boldsymbol{\theta}_q|\boldsymbol{\eta},C)d\boldsymbol{\theta}_q = \\
&\quad |2\pi U|^{-n_q/2}|2\pi C|^{-1/2}|2\pi\Sigma_q^*|^{1/2}\exp\{-\frac{1}{2}\text{tr}(S_q U^{-1}) - \frac{1}{2}(\bar{\mathbf{z}}_q-\boldsymbol{\eta})^T(\frac{U}{n_q}+C)^{-1}(\bar{\mathbf{z}}_q-\boldsymbol{\eta})\}
\end{aligned}
$$

where

$$S_q = \sum_{i=1}^{n_q}(\mathbf{z}_{qi}-\bar{\mathbf{z}}_q)(\mathbf{z}_{qi}-\bar{\mathbf{z}}_q)^T$$

3.3. Estimate η, U and C from background

Given the transformed training data $\mathbf{Z}_{ki} \sim N_B(\boldsymbol{\theta}_k, U), \quad k = 1, \ldots, K, \quad i = 1, \ldots, n$
$\boldsymbol{\theta}_k \sim N_B(\boldsymbol{\eta}, C), \quad k = 1, \ldots, K$, the parameters η, U and C for the prior distributions have the following estimates.

$$\hat{\boldsymbol{\eta}} = \frac{1}{K}\sum_{k}\hat{\boldsymbol{\theta}}_k$$

$$\hat{\boldsymbol{\theta}}_k = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_{ki}$$

$$\hat{U} = \sum_{i=1}^{n}\sum_{k=1}^{K}(\mathbf{z}_{ki} - \overline{\mathbf{z}}_k)(\mathbf{z}_{ki} - \overline{\mathbf{z}}_k)^T/(Kn - K)$$

$$\hat{C} = \frac{1}{K-1}\sum_{k=1}^{K}\widehat{Var}(\boldsymbol{\theta}_k) = \frac{1}{K-1}\sum_{k=1}^{K}(\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\eta}})(\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\eta}})^T - \hat{U}/n.$$

Parameters were estimated from the full training data. Data for control and recovered evidence were chosen from the training data. For inks, the sizes (n) of control and recovered data were chosen to be 1, 3 and 5. For fibres, the sizes (n) of control and recovered data were chosen to be 1, 2 and 3.

4. Methods

The performances of various models were assessed by considering the rates of false positives and false negatives and Tippett plots. A false positive is a result from a different-source comparison which has a positive log likelihood ratio. A false negative is a result from a same-source comparison which has a negative log likelihood ratio.

Models for fitting were chosen by assessing their goodness of fit to the data using Akaike's Information Criterion (AIC) and R^2E, the sum of ratios of the squared residuals to the modulus of the fitted values. From this assessment, the best-fitting spline functions were of order 3. Thus, the spline functions chosen for assessment were all of order 3 or quadratic functions. Models of differing numbers of bases, varying from four to nine for the ink data and from four to eight for the woollen and cotton data, were chosen for the assessment of performance.

For ink, pairs of sizes 1, 3 and 5, were compared. For same-source comparisons the pairs were from the same group. For different-source comparisons the pairs were from different groups.

Ink Size 1: an item was compared with itself and all other members of its group. For ink, there were 10 members in a group and 40 groups. This led to 55 $\times$ 40 = 2200 same-source comparisons and (10 $\times$ 10) $\times$ (40 $\times$ 39)/2=78000 between-source comparisons.

Ink Size 3. Three sets of triplets were chosen. This meant one item was not chosen. The triplets chosen were $\{1, 2, 3\}$, $\{4, 5, 6\}$ and $\{7, 8, 9\}$ were formed. The arbitrary choice was made to ignore item 10. Each triplet was compared with itself and with the two other members of its group. This led to 6 $\times$ 40 = 240 same-source comparisons and (3 $\times$ 3) $\times$ (40 $\times$ 39)/2=7020 between-source comparisons of triplets.

Ink, Size 5: Sets of size 5 $\{1, 2, 3, 4, 5\}$ and $\{6, 7, 8, 9, 10\}$ were formed. Each set was compared with itself and with the other set in its group. This led to 3 $\times$ 40 = 120 same-source comparisons and (2 $\times$ 2) $\times$ (40 $\times$ 39)/2=3120 between-source comparisons of sets.

Wool and Cotton Size 1: an item was compared with itself and all other members of its group. For woollen and cotton fibre data, there were 9 members in a group and 20 groups. This led to 45 $\times$ 20 = 900 same-source comparisons and (9 $\times$ 9) $\times$ (20 $\times$ 19)/2= 15390 between-source comparisons.

Wool and Cotton Size 2. Pairs of items $\{1, 2\}$, $\{3, 4\}$, $\{5, 6\}$ and $\{7, 8\}$ were formed. The arbitrary choice was made to ignore item 9. Each pair was compared with itself and with the two other members of its group. This led to 10 $\times$ 20 = 200 same-source comparisons and (4 $\times$ 4) $\times$ (20 $\times$ 19)/2=3040 between-source comparisons of pairs.

Wool and Cotton Size 3: Sets of size 3 $\{1,2,3\}$, $\{4,5,6\}$ and $\{7,8,9\}$ were formed. Each triplet was compared with itself and with the other set in its group. This led to 6 $\times$ 20 = 120 same-source comparisons and (3 $\times$ 3) $\times$ (20 $\times$ 19)/2=1710 between-source comparisons of triplets.

Pairwise log-likelihood ratios (lLR_{ij}) (logs to base 10) are calculated for all pairs of curves according to the sampling schemes described above.

$$lLR_{i_1 i_2} = \log \frac{f(\text{data} \mid \text{ curves } i_1, i_2 \text{ are from the same source })}{f(\text{data} \mid \text{ curves } i_1, i_2 \text{ are from different sources })}.$$

Let S indicate the average $lLR_{i_1 i_2}$ for all pairs of curves $\{\mathbf{Y}_{ki_1}, \mathbf{Y}_{ki_2}\}$ where $k = u$ and let D indicate the average $lLR_{i_1 i_2}$ for all pairs of curves $\{\mathbf{Y}_{ki_1}, \mathbf{Y}_{ui_2}\}$ where $k \neq u$. FP and FN indicate the percentage of false positive (FP) and false negative (FN) results, respectively.

5. Results

Table 1: Ink: The percentage of false positives (FP) and false negatives (FN) for every possible pairwise comparison for the $CIE - Lab$, $CIE - xyz$ and $CIE - XYZ$ systems [4]. Reproduced with permission of the Royal Society of Chemistry.

$CIE - Lab$			$CIE - xyz$			$CIE - XYZ$		
	FP(%)	FN(%)		FP(%)	FN(%)		FP(%)	FN(%)
a	16.3	7.5	x	12.9	2.5	X	15.5	12.5
b	19.2	2.5	y	20.8	5.0	Y	17.1	15.0
L	16.2	10.0	z	18.7	2.5	Z	19.4	12.5
ab	3.5	2.5	xy	4.6	2.5	XY	6.8	10.0
aL	6.7	7.5	xz	4.6	2.5	XZ	5.4	7.5
bL	6.5	10.0	yz	4.6	2.5	YZ	6.3	7.5

Classification results for the tristimulus values $CIE{-}XYZ$, chromaticity coordinates $CIE{-}xyz$ and $CIE{-}Lab$ are given in Table 1. All possible within-source comparisons (780 under the different-source proposition using all ten replications and 40 under the same-source proposition using the first five replicates in comparison with the second five replicates) were made for each possible CIE element and each of the nine bivariate combinations of possible pairs of CIE elements. The pairs were ab, aL and bL from the $CIE - Lab$ system, zy, xz and yz from the $CIE - xyz$ system and XY, XZ and YZ from the $CIE - XYZ$ system. The numbers of false positives and false negatives are given in Table 1.

These results are to be compared with the results obtained using functional data analysis for inks. The CIE systems reduce the information in the spectra to only three numbers. The functional data analysis approach reduces the information to B numbers, the coefficients of the B-splines.

Table 2: Ink: mean log to base ten likelihood ratios for same-source (S) and different-source (D) pairs with false positive (FP) and false negative (FN) rates for various sizes (n) of the recovered and control sample and different numbers of bases (B) and order (o) for the fitted model

Bases B	Order o	S	D	FP(%)	FN(%)	n
4	3	1.98	−20.95	10.9	3.8	1
5	3	2.63	−28.33	6.2	2.7	1
6	3	3.21	−42.39	5.2	1.7	1
7	3	3.86	−49.76	4.1	1.0	1
8	3	4.46	−56.25	3.5	0.9	1
9	3	4.96	−63.87	3.2	0.7	1
4	3	2.91	−66.82	4.6	3.8	3
5	3	3.83	−89.89	2.3	3.3	3
6	3	4.68	−133.31	2.0	2.1	3
7	3	5.63	−156.68	1.4	1.7	3
8	3	6.56	−177.52	1.1	1.7	3
9	3	7.38	−200.96	1.1	1.2	3
4	3	3.31	−113.28	2.5	4.2	5
5	3	4.39	−152.95	1.5	5.0	5
6	3	5.32	−225.79	1.1	4.2	5
7	3	6.33	−265.58	0.9	3.3	5
8	3	7.37	−300.68	0.9	3.3	5
9	3	8.31	−341.00	0.6	2.5	5

Table 2 illustrates, for ink, the false positive (FP) and false negative (FN) rates for splines with bases 4 to 9 and order 3 with control and recovered samples of sizes 1, 3 and 5. Tables 3 and 4 illustrate for red woollen and red cotton fibres, respectively, the FP and FN rates with splines with bases 4 to 8 and order 3 with control and recovered samples of sizes 1, 2 and 3. These numbers of bases and the order were chosen as the ones that provided the best fit to the data, bearing parsimony in mind. In general, it can be seen that the false positive and false negative rates decrease as the number of bases increase. The false positive results are the ones that it is desired to keep as low as possible in order to minimise the risk of convicting an innocent person. These rates compare favourably with those in Table 1. They are reinforced visually with the Tippett plots in Figure 4 for ink with nine bases of order 3 and samples of size one for the control and recovered data and for wool and cotton in Figures 5 and 6, respectively, with eight bases of order 3 and samples of size one for the control and recovered data.

In Tippett plots, the lower of the two curves in each pair represent different source comparisons, the upper represents same-source comparisons. A curve represents the empirical probability of obtaining a log-likelihood ratio greater than the value x in the x-axis. Thus it is desirable that the upper (same-source) plot is no less than 100% for $x < 0$ and the lower (different-source) plot is no greater than zero for $x > 0$.

The right-hand plots are subgraphs on an expanded scale of the left-hand plots. The Tippett plots for all of ink, wool and cotton show a good discrimination between same-source and different-source comparisons.

Table 3: Wool: mean log to base ten likelihood ratios for same-source (S) and different-source (D) pairs with false positive (FP) and false negative (FN) rates for various sizes (n) of the recovered sample and different numbers of bases and order for the fitted model

Bases	Order	S	D	FP(%)	FN(%)	n
4	3	1.39	−6.97	10.7	7.7	1
5	3	1.63	−7.96	8.4	9.2	1
6	3	1.48	−10.21	6.2	9.3	1
7	3	1.88	−13.47	5.0	8.8	1
8	3	2.11	−16.24	3.7	9.3	1
4	3	1.82	−15.40	6.6	7.5	2
5	3	2.18	−17.55	4.7	9.0	2
6	3	1.74	−22.44	3.0	9.5	2
7	3	2.36	−29.58	2.7	7.5	2
8	3	2.65	−35.64	1.6	8.5	2
4	3	1.88	−24.26	4.6	10.8	3
5	3	2.27	−27.86	3.3	14.2	3
6	3	1.28	−35.78	1.9	13.3	3
7	3	1.78	−47.03	2.0	14.2	3
8	3	1.92	−56.41	1.0	14.2	3

6. Discussion

Proof of concept for ink chromatograms is described in Burfield et al. [8]. They represent the data with principal component scores rather than the original coefficients of the basis functions and did not determine likelihood ratios. They then carried out classification on ink samples reporting very high discrimination results. In contrast, the model developed here uses a technique based on B-splines for the reduction of dimensions with fixed covariance matrices estimated from training data. This model enables the determination of likelihood ratios with few dimensions with the associated advantage over discrimination of the provision of a value for the evidence and the ability to update the odds in favour of the prosecution proposition in a coherent manner with other pieces of evidence.

The mean log-likelihood ratios for different-source comparisons can be very large, up to 341 in magnitude. These values are unrealistic for reporting purposes. It is proposed that there be a cap of 9 on these values. This implies an upper limit of 10^9, or one billion, for the likelihood ratios.

An interesting property of the likelihood ratio is that the likelihood ratio of the likelihood ratio is itself the likelihood ratio [13]. Van Es et al. [12] and Gupta et al. [9] use this result to do what they call calibration of the likelihood ratio. They view the likelihood ratio as a score, obtain the distribution of the scores from a training set and then report the likelihood ratio of the scores. However, the result that the likelihood ratio of the likelihood ratio is the likelihood ratio is a general result. In practice, interest is concentrated on one particular value of the likelihood ratio. In that context it makes no sense to talk of the probability of the likelihood ratio. In addition, calibration destroys the updating process of Bayes' theorem whereby the posterior odds of one piece of evidence becomes the prior odds for the next piece of evidence.

Table 4: Cotton: mean log to base ten likelihood ratios for same-source (S) and different-source (D) pairs with false positive (FP) and false negative (FN) rates for various sizes (n) of the control and recovered sample and different numbers of bases (B) and order (o) for the fitted model

Bases B	Order o	S	D	FP(%)	FN(%)	n
4	3	0.48	−0.90	34.9	12.3	1
5	3	0.62	−1.19	29.5	12.7	1
6	3	0.71	−2.38	20.8	10.8	1
7	3	1.16	−3.82	18.4	9.0	1
8	3	1.41	−5.87	13.2	10.3	1
4	3	0.83	−2.22	23.6	10.0	2
5	3	1.09	−2.94	19.3	12.5	2
6	3	1.17	−5.41	12.2	10.0	2
7	3	1.87	−8.76	10.7	10.0	2
8	3	2.14	−13.23	6.8	12.5	2
4	3	1.02	−3.71	17.1	11.7	3
5	3	1.39	−4.89	13.9	13.3	3
6	3	1.55	−8.77	8.5	9.2	3
7	3	2.52	−14.06	7.1	10.8	3
8	3	2.83	−20.84	4.5	12.5	3

Error rates are low for ink. For ink, from Table 2, it can be seen that with 9 bases of order 3 and a recovered sample with five members, the false positive rate is 0.6% and the false negative rate is 2.5%. This compares favourably with the results from Table 1.

For woollen fibres, the false positive rate from Table 3 with 8 bases of order 3 and a recovered sample with three members, the false positive rate is 1.0% and the false negative rate is 14.2%. Application to cotton fibres shows a higher false positive rate than in the application to woollen fibres. For cotton fibres, the false positive rate from Table 4, with 8 bases of order 3 and a recovered sample with three members, is 4.5% and the false negative rate is 12.5%. This is a consequence of the higher inter-sample variation for woollen fibres than for cotton fibres and the high intra-sample variation compared with the inter-sample variation for cotton fibres. Finally, the false positive and false negative rates arising from the use of B-splines and MSP data for inks and fibres are similar to those using similarity measures and FTIR data for the analysis of spray paint [5],

Visual comparisons of the woollen and cotton fibres gave discriminating powers [11] of 98% for red wool and 58% for red cotton fibres. The visual comparison considered (a) the general shape of the spectrum, (b) the location of maxima and minima values in terms of wavelength (nm) values and (c) the presence of shoulders and other flexion points. Natural fibres exhibit variation along the same fibre and also within a set of fibres from the same textile. Because such variation is part of the intra-source variation characteristic of the textile normalisation of the spectra was done as this would reduce this variation. Low differentiation for red cotton using MSP is not surprising as the primary dye for red cotton is known as *C.I. Reactive Red 180* (C.I.181055) [2], a very common dye.

Figure 4: Tippett plot for ink data for nine bases of order 3 with one recovered sample and one control sample. The right-hand plot is an expansion of the right-hand portion of the left-hand plot.

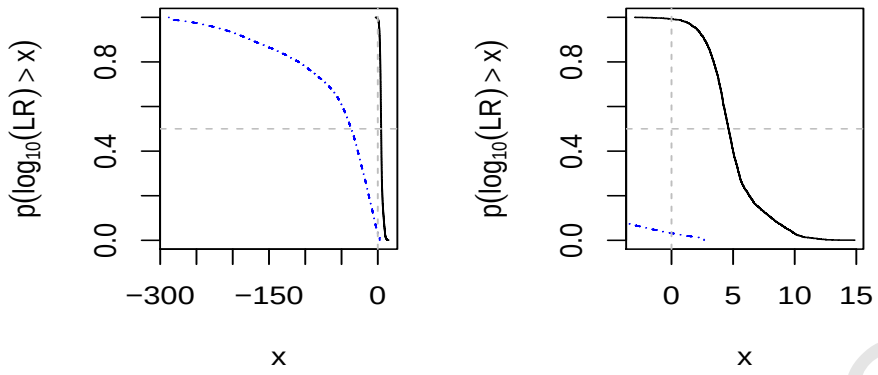
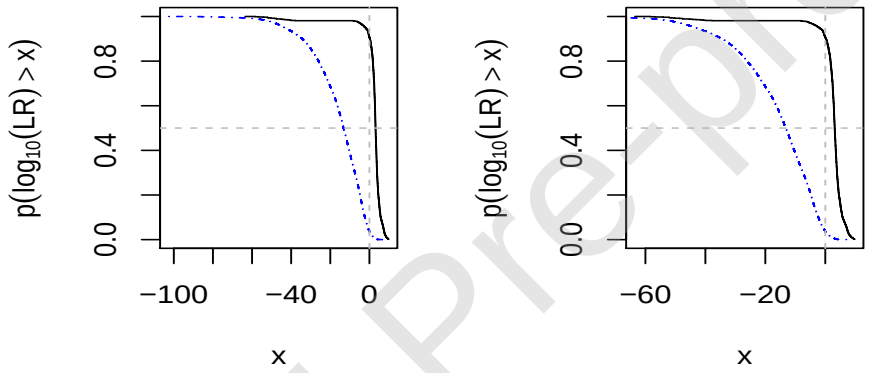


Figure 5: Tippett plot for wool data with eight bases of order 3 with one recovered sample and one control sample. The right-hand plot is an expansion of the right-hand portion of the left-hand plot.

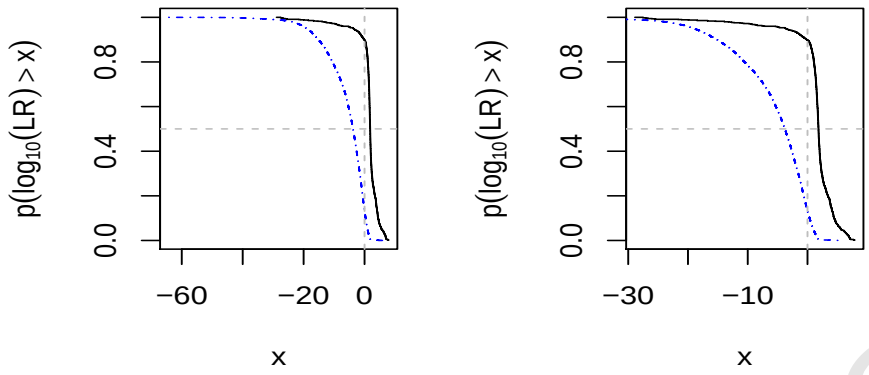


7. Conclusion

A novel approach for the evaluation of evidence in the form of MSP data is introduced. Ideas from functional data analysis are used to represent the spectra with linear functions comprised of piecewise quadratic characteristics with a limited number of coefficients. This approach is in contrast to the CIE approach for ink or the visual comparisons for woollen and cotton fibres. Performance is assessed by misclassification rates and Tippett plots and is found to be comparable to those provided by the other methods.

[1]
[2] .
[3] .
[4]
[5]
[6]
[7] J. Almer, E. Mcansh, and B. Doupe. Analysis by uv-visible microspectrophotometry. *Canadian Society of Forensic Science Journal*, 43: 16–30, 2010.
[8] R. Burfield, C. Neumann, and C.P. Saunders. Review and application of functional data analysis to chemical data - the example of the comparison, classification and database search of forensic ink chromatograms. *Chemometrics and Intelligent Laboratory Systems*, 149: 97–106, 2015. doi: 10.106/j.chemolab.2015.07.006.
[9] A. Gupta, C. Martinez-Lopez, J.M. Curran, and J.R. Almirall. Multi-element comparisons of tapes evidence using dimensionality reduction for calculating likelihood ratios. *Forensic Science International*, 301:426–434, 2019. https://doi.org/10.1016/j.forsciint.2019.06.002.
[10] J.O. Ramsay and B.W. Silverman. *Functional data analysis*. Springer, 2005.

Figure 6: Tippett plot for cotton data with eight bases of order 3 with one recovered sample and one control sample. The right-hand plot is an expansion of the right-hand portion of the left-hand plot.



[11] K. Smalldon and A. Moffat. The calculation of discriminating power for a series of correlated attributes. *Journal of the Forensic Science Society*, 13:291–295, 1973.

[12] A. Van Es, W. Wiarda, M. Hordijk, I. Alberink, and P. Vergeer. Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. *Science and Justice*, 57:181–192, 2017.

[13] D.A. Van Leeuwen and N. Brümmer. The distribution of calibrated likelihood ratios in speaker recognition. https://arxiv.org/pdf/1304.1199.pdf, 2013.

[14] K. Wiggins and P. Drummond. The analysis and comparison of blue wool fibre populations found at random on clothing. *Science and Justice*, 45:157–162, 2005.