Abstract

In numerous major crime cases undertaken at our laboratory the recovery of large numbers of fibres (> 1,000), particularly in cases with no known source, presented several challenges. These included the inability to effectively manage the data (i.e. combination of MSP spectra, microscopic characteristics, composition, etc.) and perform comparisons in an efficient manner. To address these challenges, and in response to a growing need for performing fibre comparisons, we developed a database of textile fibre microspectrophotometric (MSP) spectra. The database, designed to compare MSP spectra using a modified Pearson method of correlation, currently contains over 20,000 normalised and first derivative spectra of casework, validation and reference textile fibres. A comparison strategy for cases with a large number of questioned samples was devised, involving identification of critical fibres in the casework data set, development of preliminary fibre groups classified according to their corresponding/similar MSP spectra, and verification of the preliminary groups via brightfield and fluorescence comparison microscopy. The database has successfully been utilised for proficiency trials and casework with small questioned fibre sets. Furthermore, in a case involving a larger dataset (>4,000 "unknown" fibres) the database assisted in the efficient classification of 156 distinct groups of interest, highlighting its utility in providing investigative leads for the identification of potential sources of the recovered fibres.

Introduction

Casework submitted to trace evidence laboratories for textile fibres analysis typically requires comparison of fibres recovered from a victim or scene to constituent fibres from a known item of interest (such as a clothing item from a suspect). As the known fibres are of a defined colour and type, the search for corresponding recovered fibres is narrowed significantly as those of similar colour and type are targeted and collected. Known and recovered fibres are then analysed and compared using a range of long-established techniques, including comparison microscopy (CM), fluorescence microscopy (FM), microspectrophotometry (MSP), and Fourier-transform infrared (FTIR) spectroscopy [1].

In some instances, a known or comparison garment or fabric may not be submitted for analysis. This may include situations where investigation through other forensic avenues such as DNA analysis has not yielded a person of interest. In such scenarios, trace evidence laboratories may be required to analyse recovered fibre evidence and indicate a possible source material; establish links between victims and scenes; or in cases with multiple victims, to establish links between victims [2]. These cases are approached from an investigative perspective, and all recovered fibres may be of potential evidential value as the colour and type of fibres of interest is unknown. The quantity of textile fibres collected for these cases is typically much larger than cases where known garments are submitted,

and may be reflective of the severity of the crime. As the quantity of fibres increases, the number of fibre comparisons required is likely to render the standard fibres methodology of the trace evidence laboratory impractical, and a computerised approach to data evaluation becomes more valuable in terms of the required time and effort.

The value of textile fibre reference collections or databases to allow a stronger assessment of the frequency (and therefore significance) of a particular fibre colour and type has been long understood [1]. Most trace evidence laboratories have access to or maintain a fibre reference collection [3], typically used to identify unusual fibre polymer types. However, larger databases or collections of fibres designed to address complex cases or address questions pertaining to frequencies of fibre colour or type in the general population are rare in the literature. Home and Dudley reported on a collection of 10,034 fibres from casework materials collated by eight laboratories in the United Kingdom (UK) [4]. Brightfield and polarised light microscopy (PLM) were used to identify fibre type, and fibre colour was recorded by comparison to the Methuen Handbook of Colour. This project was later expanded to produce a data collection of 19,959 fibres from 7,367 casework garments processed in UK laboratories [1, 5]. Fibre colour was determined via MSP, but converted into complementary chromaticity coordinates as computer memory limitations prevented storage of the entire visible spectral range.

Biermann and Grieve described a novel approach to estimating the frequencies of fibre colours and types by recording data from mail-order clothing catalogues into a Catalogue Data Base (CDB) [6-8]. This allowed for rapid and 'real-time' collection of data; as of 2007, the CDB contained information on approximately 133,000 clothing textiles [9]. Fibre frequency estimates were generally in agreement with world textile production figures [8]. The main disadvantages to this approach are that garments are not 'on-hand' in the laboratory to compare against casework samples, and colour information is described in a subjective manner.

Biermann and Deck described development of a casework database of textile fibres in response to complex and high profile cases during the 1990s in Germany, including incidents of terrorism and assassination attempts [2]. These cases required analysis and comparison of thousands of fibres and resulted in a review of the strategy used and an improvement of the analysis protocol. The process included analysis of fibres by brightfield, polarised light and fluorescence microscopy, with visible and (if possible) fluorescence spectral analysis of fibres conducted by a diode array spectrometer coupled to the microscope. Data collection was comprehensive and involved storage of a wide range of fibre attributes such as cross-sectional shape, delusterant presence, fibre colour and type, and fluorescence behaviour among others, as well as associated case-related information. Spectral information was stored using Spectralys software, which allowed for creation of spectral libraries and comparison of fibre spectra using various metrics. The authors stated that complex cases required an "enormous amount of work and time" [2], and as such it is clear that due to the laborious and time-consuming process of data collection, large spectral fibre databases are likely to be created as required for a specific case rather than in a proactive fashion. In the absence of in-house databases, analysts have typically relied on various target [10, 11] and population [12, 13] studies, or on data from non-spectral databases such as the CDB to aid in assessment of the significance of a fibre match.

Microspectrophotometry allows for the objective measurement of fibre colour in a rapid and non-destructive manner. Fibres may be coloured using a variety of dyes, with the type of dye used often

determined by the fibre polymer type (e.g. disperse dyes for polyesters, etc.) [1]. Since spectral collection times on modern instruments are measured in milliseconds, MSP is an appropriate choice of analytical technique for formation of a textile fibre database. Several target fibre studies have indicated that fibre analysis by MSP is highly discriminatory, particularly when used in conjunction with other techniques such as compound microscopy and thin layer chromatography [10, 11]. Wiggins and Drummond reported on the separation of a collection of 2,740 blue wool fibres into 300 unique MSP spectral groups [14]. MSP also allows for discrimination of metameric samples (samples with different colourant chemistries, but which appear to the observer to be the same colour) [15]. Analysis of first derivative MSP spectra has been observed to aid in making comparisons and provide further discrimination of textile fibre samples [16, 17].

In recent years, trace evidence laboratories have been challenged to improve efficiency of analytical processes, timeliness of reporting, and clarity as to the evidential value of results [18]. The European Network of Forensic Science Institutes (ENFSI) has recently released a Guideline for Evaluative Reporting in Forensic Science, which encourages adoption of a standardised 'likelihood ratio' approach to reporting the significance of case findings [19]. Examples of the use of likelihood ratios in interpretation of glass, gunshot residue (GSR), shoe impression, voice recording and CCTV evidence types are provided. Regardless of method of reporting, it is clear that there is an international movement amongst practitioners to better express the significance of findings. Population databases for fibres and various other trace evidence types are likely to become more common in future as they can help facilitate this process.

This contribution describes the development of a microspectrophotometric (MSP) database approach for textile fibres analysis, and its use in handling cases involving large bodies of textile fibre evidence. The database was initially developed in response to a growing need for performing fibre comparisons in Western Australian 'major/serious crime' investigations. It was developed to address the challenges posed by cases where large numbers of fibres are recovered, and/or the potential sources have not been recovered. The primary purpose of the database upon its creation was to provide investigative intelligence. To achieve this, a comparison strategy was developed allowing the analyst to focus on the most critical fibres of a casework set (i.e. those with the largest number of close correlations), and thereby more efficiently identify large groups of casework fibres with corresponding or similar MSP spectra. The database now contains over 20,000 fibres, including fibres recovered from reference fabrics (predominantly from motor vehicles) for source identification of casework fibre groups. Further, through fibre collection for an ongoing fibre population study, we envisage that the utility of this fibre database will assist in our progression towards interpretation of textile fibre casework in our laboratory via a Bayesian framework.

Materials and Methods

Data collection

Recovered casework, reference and validation textile fibres were mounted onto glass microscope slides in XAM (no longer used) or Entellan New mounting media under a protective glass coverslip. For major investigations with no known source fabrics, typically all foreign fibres were collected from exhibits, excluding colourless fibres. Other fibre colours and types considered to be of little

evidential significance were not exhaustively collected (e.g. blue denim / black / grey cotton). Fibres were initially examined under brightfield and polarised light using a Leitz Diaplan comparison microscope with 40x objective, and imaged using 'IM1000' software (Leica Microsystems). Image histograms were adjusted in a consistent manner to regulate brightness and ensure consistent lighting conditions for each brightfield image. Perceived fibre colour and fibre polymer type inferred from the polarised light birefringence pattern were recorded for subsequent entry into the database.

Fibre MSP spectra were collected in absorbance mode using a CRAIC Technologies (San Dimas, CA) QDI 2010 UV-visible microspectrophotometer with 36x reflecting objective. The instrument was calibrated for wavelength accuracy with National Institute of Standards and Technology (NIST) traceable holmium oxide and didymium standards and photometric accuracy with three neutral density standards at the beginning of each day of use. Absorbance MSP spectra were collected from 310 to 800 nm with the following parameters: scans to average = 50; resolution factor = 2; aperture size typically 8 μm x 8 μm (size 4), or 4 μm x 4 μm (size 5) where required for thinner fibres. Ten spectra were collected from synthetic fibres, and twenty spectra were collected from natural fibres to account for increased dye variation. The repeat spectra were collected from differing locations along the fibre length. The multiple measurements were averaged using CRAIC MSP software to form a representative spectrum for each fibre. This software process also generates spectra reflective of one- and two- standard deviations from the average spectrum. Representative spectra were exported from CRAIC MSP software into comma separated values (CSV) files for upload into the database.

Analysis of fibres by FTIR was conducted using a Thermo Nicolet 6700 micro-FTIR spectrometer (ThermoFisher Scientific, Madison, WI). Subsamples were flattened via microdiamond cell and analysed in transmission mode and the remaining fibre re-mounted.

Database description and functionality

An in-house database used for profiling of illicit drugs was modified to accept fibre MSP spectral data. The software was limited to 50 variables; therefore, to preserve resolution each 649-point representative spectrum was reduced to a 50-point spectrum by performing a 5-point ensemble boxcar average at approximately 10 nm intervals. This process therefore only used 250 of the 649 data points of the raw spectrum, and the remaining 399 data points were not considered. The resulting 50-point spectra were zero adjusted to eliminate negative absorbance values, normalised by the sum of absorbance values, and the fourth root applied to each data point. The database utilised a modified Pearson correlation coefficient (MPCC) metric to compare MSP spectra, which has been previously used in a database approach for illicit drug profiling [20]. The modified Pearson metric removes negative values as the range of possible values is transformed from between -1 and 1 to between 0 and 100.

$$MPCC = (1 - r)/2 * 100$$

Grouping of casework fibre MSP spectra was initially attempted via principal components analysis (PCA) (Unscrambler®). Fibres were separated by polymer type prior to PCA. In all attempts, no clear grouping was achieved. It is our belief that this is due to MSP spectra generally not being highly

detailed, and the extremely wide variety of possible colours of the total collection of casework recovered fibres.

Subsequently, the database was reproduced in Microsoft Excel (Microsoft Corp., CA) to allow for the development of a fibre grouping methodology. All fibre data reside in a single spreadsheet, with information for each fibre contained in individual rows. This information includes the unique fibre identifier, case number, perceived colour by compound microscopy, fibre type, slide box location, and numeric normalised and first derivative MSP spectral data (all in individual columns). Microsoft Excel's native 'sort and filter' function allowed for simple separation of data types by column.

Database functions were created using Visual Basic (VBA) and embedded in the Microsoft Excel control ribbon. Basic functionality includes:

- calculation of a single fibre's match score against all other fibres;

- plotting a fibre's normalised and first derivative MSP spectra against one or several other fibres;

- displaying brightfield and polarised light microscopic images of fibres;

- displaying all fibres which correlate with a particular fibre below a user-defined match score threshold; and

- correlating all fibres against all other fibres, and displaying those correlations which fall below a user-defined match score threshold.

The database has been designed to allow for automated upload of numerical MSP fibre data from CSV files and accept spectra from MSP instruments of differing wavelength spacing to our instrument. Interpolation was performed between data points of raw spectra from foreign instruments to ensure spectra were all expressed with the same wavelength spacing. Complementary chromaticity coordinates were also calculated for each spectrum during importation to the database, allowing for approximation of the colour of the fibre that produced the spectrum.

Method Development

Initial validation of the fibres database created from our in-house illicit drugs database (50 data point spectra) involved analysis of a set of red, purple and pink natural and synthetic fibres by four analysts over an extended period of time. Analysis of the same sample by different operators provided good agreement of MSP spectra collected up to 189 days apart for samples exhibiting relatively high absorbance.

In order to compare data pretreatment methods and their effect on correlation, MSP spectra of a set of 226 'same-garment' fibres were collected from 46 garments of varied colours and fibre polymer compositions, including cotton, wool, polyester, rayon, nylon, acrylic and lyocell. These fibres also exhibited wide variation in physical characteristics (diameter, cross-sectional shape, etc). Spectra of 3,865 casework fibres (also of varied colours, compositions and physical characteristics) were added to the set, allowing for a total of 444 intra-garment comparisons and 898,471 'non-match' comparisons (between different garments, or between garment and casework sample). Correlation

threshold values were determined where 95% of intra-garment fibre comparisons were included. The correlation efficacy of each data pretreatment method was evaluated by expressing the amount of non-matches excluded at this correlation threshold as a percentage of total non-matches. Correlation efficacy values for data pretreatment methods for the 50 data point spectra are provided in Table 1.

Table 1. Correlation efficacy for pretreatment methods for 50-point spectral data.

Data treatment method	CT	%exc
Normalised only	1.80	95.67
Normalised, fourth root	1.30	92.94

CT = correlation threshold (expressed as MPCC), %exc = percentage of comparisons excluded

The success of this approach with relatively low-resolution spectra encouraged the reproduction of the database in Microsoft Excel. The reproduction removed the limitation posed by reducing the MSP spectra to 50 data points, which allowed for a revisit to data pretreatment experimentation. The raw 649-point spectral data were again normalised by the sum of absorbance values, as casework fibres may vary in sample history and may be subject to environmental dye fading. A three-point ensemble boxcar average was used to reduce the database size and improve calculation speed almost twofold. The three-point averaged data (216 data points) exhibited a wavelength spacing of approximately 2.3 nm. Spectra with narrow peaks and small shoulders were examined and found to maintain these features, i.e. spectral integrity was preserved.

Various correlation processes on the 216-point data were compared in the same manner as for the 50-point data. A range of Savitzky-Golay first-derivative convolutions (3- to 17-point quadratics [21]) were performed on the 216-point data, and the correlation efficacy of each of the resulting first derivative datasets was evaluated. First derivative spectra were examined visually by selecting spectra with small peaks and shoulders, and a 9-point first derivative convolution was chosen as it offered the best performance without over-smoothing small spectral features. Data pre-treatment correlation comparisons are summarised in Table 2.

Table 2. Various means of treatment for 216-point data and their effect on correlation efficacy.

Data treatment method	CT	%exc
Normalised only	1.78	95.54
First derivative convolutions		
3-point	7.57	95.49
5-point	6.66	95.64
7-point	5.48	96.27
9-point	4.51	96.93
11-point	4.22	96.90
13-point	3.84	97.10
15-point	3.60	97.17
17-point	3.36	97.25
*Correlation method**		
Normalised only	1.78	95.54
First derivative only	4.51	96.93
(Norm x Deriv)	6.54	97.25

CT = correlation threshold (expressed as MPCC), %exc = percentage of comparisons excluded
*performed on 9-point Savitzky-Golay first-derivative convolution

In order to consider both the normalised and first-derivative spectra in the final correlation, the process above was followed with the normalised and 9-point first-derivative MPCC values calculated independently, and then the MPCC values multiplied together (NxD). The resulting correlation efficacy was compared to normalised-only and first-derivative-only MPCC correlations. The combined NxD correlation was found to outperform both the normalised-only and first-derivative-only correlations (Table 2). The weakest 5% of intra-garment correlations were investigated for each of the three correlation methods. These comparisons often involved noisy spectra of poor intensity, or large differences in dye intensity between the two samples (typically for natural fibres) leading to differences in peak intensities, or in extreme cases differences in peak shape. Four intra-garment comparisons of fibres with differing baseline slopes fell outside of the correlation cutoff for the normalised MPCC, but were included within the cutoff for both the first derivative and NxD methods. The normalised MPCC is more appropriate for noisy spectra of poor intensity, as the noise is increased in the first derivative spectrum (Fig 1, c and d; Fig 2, c and d). The database allows the analyst to select from these three methods of correlation as appropriate.

Inter-laboratory validation was performed by selection of a set of fifty fibres from proficiency trials, known reference samples, previous validation studies, and control and recovered casework samples. Fibres were selected with the aim of inclusion of a wide range of fibre colours, types and MSP spectra intensities. The fibre set was re-analysed by our laboratory, and distributed to three other Australian laboratories and analysed via MSP under the same or similar instrumental conditions. Two laboratories used a CRAIC QDI2010 instrument and one used a CRAIC QDI1000 instrument. One laboratory collected spectra in transmission mode, and the data were converted into absorbance

mode during importation into the database. Several slides in the set contained multiple fibres, and in two instances the incorrect fibre was analysed, resulting in six comparisons being discarded. The remaining 294 inter-laboratory comparisons were performed using the NxD correlation, and 272 (92.5%) fell below the correlation threshold of 6.54. Indicative MSP overlays of inter-laboratory comparisons are provided in Fig. 1. The comparisons which did not fall below the expected correlation threshold were typically of fibres of poor MSP spectral intensity (Fig. 1, c and d), or where a significant difference in intensity was observed between laboratories in the ultraviolet region of the spectrum (approximately 310-340 nm) (Fig.1, e and f).

(insert fig 1 here – 1.5x column width)

Fig. 1. Normalised (left) and first derivative (right) MSP overlays of inter-laboratory spectral comparisons: a) and b) Red cotton fibre, NxD = 0.002; c) and d) Yellow acrylic fibre, NxD = 41.9; e) and f) Grey polyester fibre, NxD = 12.5. The dotted lines in the first derivative spectra represent the zero-intercept.

The re-analysed spectra were also compared to our previous analyses of the same fibres using the NxD correlation. The time difference in analyses ranged from 148 days to 1,983 days. 49 of the 50 comparisons (98%) fell below the correlation threshold of 6.54. Indicative MSP overlays are provided in Fig. 2. The comparison which did not fall below the expected correlation threshold was of low MSP spectral intensity. This lead to poor reproducibility of the normalised spectra in the 550 to 800 nm region (varying noise levels) and different intensities below 550 nm (Fig. 2, c and d).

(insert fig 2 here – 1.5x column width)

Fig. 2. Normalised (left) and first derivative (right) MSP overlays of intra-laboratory spectral comparisons: a) and b) Blue wool fibre (1,754 days between analyses), NxD = 0.003; c) and d) Yellow nylon fibre (1,622 days between analyses), NxD = 126. The dotted lines in the first derivative spectra represent the zero-intercept.

Thirty-four of the 50 fibres (the non-casework fibres) were also sent to an international laboratory. The fibres were analysed using a Tidas 800 microspectrophotometer (J&M) coupled to a Zeiss microscope (Axioplan 2) and a photodiode detector (type MCS 1024, range 190 to 1020 nm). The spectra were collected from 380 to 800 nm with an integration time of 350 ms and 5 accumulations. After adjustment of the foreign spectra to our wavelength spacing, the 380 to 800 nm range represented 560 of the total 649 data points for our usual 310 to 800 nm range. Hence the normalised and first derivative spectra were scaled by a factor of 0.8629 (or 560 / 649). The spectra were compared to our spectra of the same samples via both NxD and first-derivative MPCC, due to the differing wavelength ranges. Thirty-three of the 34 correlations (97.1%) fell below the

correlation threshold of 6.54 for the NxD correlation, and thirty-two of the 34 correlations (94.1%) fell below the correlation threshold of 4.51 for the first derivative correlation. The comparisons which did not fall below the expected correlation thresholds were of poor MSP spectral intensity.

Several previous studies have highlighted errors in accuracy or precision of intermediate to advanced statistical processes in Microsoft Excel products [22]. Aside from simple multiplication and division processes, the only native MS Excel function used in the database algorithm was the 'PEARSON' function. This function was compared to output from R (R Foundation for Statistical Computing, Vienna, Austria) and MATLAB (MathWorks, Natick, MA) software and found to be accurate to 14 decimal places. Further, numeric data for 40 fibres (including 20 fibres that closely correlate to each other) were exported from the database, and the complete algorithm was replicated in MATLAB. Forty correlations were checked and MS Excel results were accurate in all instances to at least 12 significant figures.

Routine casework

Use of the database in routine fibres casework has typically involved studying the variation in match score for correlations of multiple fibres collected from known garments, and determining if a questioned fibre falls within that range of scores when correlated against a known fibre. This provides a higher level of objectivity for MSP analysis of fibres than previously possible in our laboratory. The database provides a convenient framework for management of routine casework. Data from previous routine casework and proficiency trials dating back to 2009 involving fibre MSP analyses with the same instrumentation have also been uploaded to the database. Retrospective review of this casework via the database confirmed previous findings. The database was also successfully used in assessment of two recent textile fibres proficiency trials.

Comparison strategy for large textile fibres cases

Fibres collected for a major crime investigation were processed via the earlier methodology prior to later development of the database. The comparison strategy was undertaken on 50-point spectral data with normalisation and fourth root pretreatments, before the introduction of first derivative spectra to the database.

The total collection of casework textile fibres for the investigation was initially segmented by fibre type. Predominant fibre types included polyester, cotton and wool. Unusual fibre types, which could not be readily identified by their brightfield image or birefringence pattern, were considered as a single type at this stage (e.g. 'other natural', 'other synthetic'). Fibres were not separated by perceived colour as this was considered analyst-dependent and hence subjective.

All fibres of a particular type were correlated against each other, and a list created of those correlations below a normalised MPCC threshold of 0.25. The list was then interrogated to identify critical fibres likely to form large groups ("target fibres"). This was performed by:

- sorting the list by descending match count, such that the fibre with most potential matches is listed first;

- proceeding down the list, and if fibre B (e.g. with seven potential matches) is already listed as a potential match for fibre A (e.g. with nine potential matches), then fibre B is deleted. Fibre A is considered to represent fibre B;

- if fibre C is found to not match any others preceding it in the list, fibre C is kept.

The resulting list of target fibres typically contained approximately 10-20% of the fibres from the initial pool. Each target fibre was considered critical as it was likely to be unique from other target fibres (normalised MPCC > 0.25), and also likely to be representative of a larger fibre group.

Target fibres were correlated against all fibres of that type for that case, and a list created of those correlations below a higher (more conservative) normalised MPCC (2.00 or greater). The produced list is highly likely to contain all actual matches to the target fibre, as well as several obvious non-matches. The database allows the user to proceed through the list of potential matches, providing a window with which to simultaneously view the brightfield and cross-polarised images of the target and potential match fibre collected via compound microscopy, and view overlays of the normalised (and after later development, first derivative) MSP spectra of both fibres. The user then chooses to retain or delete the potential match. Through viewing microscopic images of the fibres, the analyst considers the diameter, perceived cross-sectional shape, delusterant distribution (if present) and other physical characteristics in their decision. This process was conducted in a rapid, yet conservative fashion, with the intention of removing obvious non-matches. The conclusion of this process resulted in formation of a preliminary fibre group.

Target fibres were sequentially processed in this manner to form multiple preliminary groups. Each preliminary group was stored in a separate spreadsheet and removed from the initial fibre pool prior to interrogation with the next target fibre, in order to avoid classification of a single fibre into multiple groups. Once all target fibres had been interrogated, all fibres remaining in the ungrouped pool were correlated against each other at a higher normalised MPCC threshold of 2.00, and these correlations were investigated as a final search for possible groups that were not represented by target fibres. Preliminary grouping was considered finalised for the fibre type at this stage, and the grouping process was then repeated for the next fibre type.

Each preliminary group was confirmed via comparison microscopy (Leitz Diaplan). The target fibre was compared to each fibre in the group under brightfield conditions as well as by fluorescence microscopy with two filter cubes for incident ultraviolet (270 – 300 nm) and incident green light (530 – 560 nm). Due to the presence of large preliminary groups, a complete comparison of each fibre to every other fibre within a group was deemed impractical. Analyst observations were recorded for each comparison made within a preliminary group. Fibres were classified as either having 'corresponding' properties (i.e. where no significant differences were observed in the comparison microscopy images and MSP spectral overlays), or 'similar' properties (i.e. where differences in the observed properties were considered to be minor and within the range that could reasonably be expected from fibres from the same source). Allowances were made for potential differences in sample histories such as exposure to ultraviolet light or incorporation of biological material, which could reasonably result in some variation in the observed properties. Many groups contained fibres deemed as corresponding and also fibres deemed as similar. In some instances, brightfield and fluorescence comparison microscopy identified two or more subgroups initially classified in the same preliminary group.

Fibre groups where fibre type was not clearly identified by birefringence pattern were confirmed by demounting one fibre and analysing via FTIR spectroscopy.

Peer review

The database allows for peer review of comparison microscopy analyses in a digital format. The reviewer is able to view brightfield and fluorescence comparison microscope images, overlays of normalised (and later first derivative) MSP spectra, and the analyst's observations and comments for each fibre comparison. The reviewer may then either agree with the analyst's decision and proceed to the next comparison, or disagree and provide comment of their own. Groups were deemed as finalised when analyst and reviewer agreed on all findings.

A summary flowchart of the comparison strategy for large textile fibres cases is provided in Fig. 3.

(insert fig 3 here – 1.5x column width)

Fig. 3. Summary of comparison strategy for large textile fibres cases

Further analysis of groups of interest

In the event that groups of evidential significance are found, further analysis can be performed. This involves re-examination of the original spectra repeats, including the one- and two- standard deviation spectra generated during the averaging process. The fibres may also be re-analysed by MSP utilising the entire spectral range (250 to 900 nm), and may be further analysed by FTIR and/or TLC as required.

Results and Discussion

The database has been successfully utilised in interpretation of routine textile fibres casework and proficiency trials, as well as in evaluation of textile fibre evidence collected for major crime investigations.

Exemplar major-crime investigation

A total of 4,432 fibres were recovered. In the absence of control items (e.g. garments / fabrics), the purpose of interrogating data via the database was to streamline the approach of identifying relevant groups of fibres for intelligence purposes. The comparison strategy yielded 156 distinct groups of textile fibres, of which 129 were reported. Groups were predominantly of polyester, cotton or wool composition.

Efficiency of the comparison strategy

Prior to the aforementioned investigation, our approach to comparison of routine casework fibres was to store all fibre information on a Microsoft Excel spreadsheet. Data recorded for questioned fibres typically included perceived fibre colour and fibre type, presence of delusterant (e.g. none / low / high), birefringence observed (e.g. low / high), apparent cross-sectional shape, thickness, and perceived colour of fluorescence generated via incident UV and green light. Fibres with similar results were then compared via CM, FM, MSP and FTIR. This approach was deemed impractical for this investigation. Data collection for the investigation was performed by multiple analysts, which would introduce subjectivity into some data fields such as perceived colour. As an example, the investigation yielded 476 polyester fibres described as blue, blue-green, blue-grey, blue-purple, dark blue or light blue. A total of 113,050 comparisons could be made between fibres of this set alone. The number of comparisons would be significantly reduced using the spreadsheet approach, but comparisons would be eliminated based on potentially subjective data entered by multiple analysts.

An initial attempt was made to find potential links for the investigation using the MSP data and the modified in-house illicit drugs database. This process involved simply correlating each fibre sequentially against every other fibre in the collection, and recording comparisons of interest. The number of comparisons was effectively doubled, as while fibre A was correlated and found to match fibre B, fibre B would later be correlated and found to match fibre A. The analyst was unable to view images of the fibres, and as a consequence many comparisons were flagged where the MSP overlay appeared to correspond, only to be immediately discarded upon comparison microscopy due to obvious differences such as fibre thickness or presence of delusterant.

The generation of target fibres greatly improved the efficiency of the comparison strategy. The list of target fibres typically contained approximately 10 to 20% of the total number of fibres in the pool; therefore, the majority of fibres did not require a manual correlation to be performed. Large groups of corresponding textile fibres are most important for investigative purposes, so it is useful to be able to highlight these groups easily and with highest priority. The early removal of the largest groups from the fibre-type pool resulted in fewer comparisons required for subsequent target fibres against the pool. As an example, the largest reported group for the investigation contained 150 fibres. A total of 11,175 comparisons are possible within this group. The target fibre for this group was correlated against all other fibres collected for the investigation of the same fibre type. Therefore only 149 of these comparisons were considered, and the remaining 11,026 comparisons were effectively avoided.

The process of identifying preliminary groups for the investigation required approximately two weeks to complete for all fibre types. Comparison by brightfield and fluorescence microscopy to confirm these groups required approximately four months. This requirement pales in contrast to the time spent recovering fibres and collecting data. It is unknown how much more efficient the comparison strategy was as opposed to the process of comparison of each fibre against every other fibre, but it is believed that many months of comparison microscopy was avoided.

The strategy outlined in this study is not an effective method for comparison and grouping of colourless fibres or carbon black pigmented fibres, as many have relatively featureless MSP spectra.

Colourless fibres were not specifically targeted and subsequently were not grouped for the investigation; however, we do continue to collect MSP spectra for colourless fibres and populate the database with these as occasionally such fibres do exhibit featured spectra. The validation process indicated that lightly coloured fibres with poor MSP spectral intensities may require a more conservative correlation threshold, or if of very poor intensity may not be suitable for the grouping process. Fifteen blue, light blue, or dark blue cotton groups were reported for the investigation; however, an indigo blue spectral group typical of denim cotton was not confirmed via comparison microscopy (and therefore not reported) due to being deemed of little evidential significance. Additionally, grey and black cotton fibres were not exhaustively recovered from items and not subjected to the full comparison procedure as these fibre types are so common as to have little evidential value.

Process improvement

Implementation of the database has resulted in discovery of ways to improve efficiency in other areas of our fibres analysis methodology. For example, Windows desktop automation software (AutoHotKey) was utilised to automate certain repetitive aspects of processes such as fibre image collection using the IM1000 software, and calculation of the average of repeat fibre MSP spectra using the CRAIC MSP software. These automations saved countless hours during data collection, and reduced the chance of human error in data entry. The ability to electronically peer-review casework findings using the database greatly increased the efficiency in reporting on the investigation, and will prove useful in future routine casework.

The recent development in use of the first derivative MSP spectrum in the algorithm and in viewing spectral overlays has proved worthwhile. Wiggins et al. reported in 2006 on their investigation into the use of the first derivative to aid in discrimination of fibre absorbance spectra, and were generally positive in their critique but stated that caution should be applied to avoid potential false exclusions, and that the first derivative spectrum should never be considered in isolation to the absorbance spectrum [16]. In this approach, normalised and first derivative spectra are considered in tandem by the analyst before inclusion into a preliminary group, and are also considered in tandem by the reviewer during the checking process.

Future directions

The authors are in the process of conducting a fibre population study by collecting foreign fibres via tapelift from public seating in the Perth metropolitan area. Fibres of all types are first counted via stereomicroscope. Fibres of colours and types considered to be of greater evidential significance are mounted, subjected to analysis via compound microscopy and MSP, and uploaded into the database. Ultimately the goal of the study is to in future be able to use the database to compare casework fibres against population data, and therefore estimate fibre frequencies with consideration of physical characteristics and MSP spectra. The ability to generate accurate frequency estimates for questioned fibres will aid in our progression towards interpretation of textile fibre casework in our laboratory via a Bayesian framework.

While the use of Microsoft Excel to develop and house the database has proven successful in this laboratory, it is anticipated that the continued growth of the database will necessitate migration to dedicated database software. It is envisioned that any future development will require the same flexibility that the current Excel version affords with regard to accessibility by other laboratories and ability to upload MSP spectra from other instrumentation. It is considered that a 'closed source' version could potentially be created to allow access to all parties of the judicial system, while protecting intellectual property.

Conclusion

A comparison strategy for large textile fibres cases was devised and facilitated using a textile fibres database. The strategy greatly reduced the total number of comparisons required by prioritising critical 'target fibres' which were considered to represent large groups of corresponding fibres. The approach was applied to a collection of over 4,400 fibres recovered in connection with a major crime investigation. Using spectra with just 50 MSP data points and normalisation and fourth root pretreatments, the strategy resulted in identification of 156 distinct groups of textile fibres, of which 129 were reported. Later database development allowed for more detailed spectra (216 data points), and inclusion of first derivative spectra, resulting in improved efficiency in the removal of non-matches.

Future endeavours of the project are focused on increasing the total number of collected fibres by addition of further casework and exemplar samples. The authors are also pursuing collaborations to potentially form large collections of exemplar materials (primarily motor vehicle fabrics) and progress towards providing accurate frequency data for specific fibre colours and types.

References

[1] J. Robertson, M. Grieve, Forensic Examination of Fibres, 2nd ed., Taylor and Francis, Boca Raton, 1999. https://doi.org/10.1201/9780203484517

[2] T. Biermann, S. Deck, Computer-aided data handling in casework, Problems of Forensic Science 46 (2001) 183-197. http://www.forensicscience.pl/pfs/46_biermann.pdf

[3] K.G. Wiggins, Forensic textile fiber examination across the USA and Europe, J. Forensic Sci. 46 (6) (2001) 1303-1308. https://doi.org/10.1520/JFS15150J

[4] J.M. Home, R.J. Dudley, A summary of data obtained from a collection of fibres from casework materials, J. Forensic Sci. Soc. 20 (4) (1980) 253-261. https://dx.doi.org/10.1016/S0015-7368(80)71352-X

[5] D. Laing, A. Hartshorne, R. Cook, G. Robinson, A Fiber Data Collection for Forensic Scientists - Collection and Examination Methods, J. Forensic Sci. 32 (2) (1987) 364-369. https://doi.org/10.1520/jfs11139j

[6] T.W. Biermann, M.C. Grieve, A computerized data base of mail order garments - A contribution toward estimating the frequency of fibre types found in clothing. Part 1: The system and its operation, Forensic Sci. Int. 77 (1-2) (1996) 65-73. https://dx.doi.org/10.1016/0379-0738(95)01856-5

[7] T.W. Biermann, M.C. Grieve, A computerized data base of mail order garments - A contribution toward estimating the frequency of fibre types found in clothing. Part 2: The content of the data bank and its statistical evaluation, Forensic Sci. Int. 77 (1-2) (1996) 75-91. https://dx.doi.org/10.1016/0379-0738(95)01857-3

[8] T.W. Biermann, M.C. Grieve, A computerized data base of mail order garments - A contribution toward estimating the frequency of fibre types found in clothing. Part 3: The content of the data bank - is it representative?, Forensic Sci. Int. 95 (2) (1998) 117-131. https://dx.doi.org/10.1016/S0379-0738(98)00090-5

[9] T.W. Biermann, Blocks of colour IV: The evidential value of blue and red cotton fibres, Sci. Justice 47 (2) (2007) 68-87. https://dx.doi.org/10.1016/j.scijus.2006.08.001

[10] K. Wiggins, P. Drummond, T.H. Champod, A study in relation to the random distribution of four fibre types on clothing (incorporating a review of previous target fibre studies), Sci. Justice 44 (3) (2004) 141-148. https://dx.doi.org/10.1016/S1355-0306(04)71706-2

[11] M.M. Houck, Inter-comparison of unrelated fiber evidence, Forensic Sci. Int. 135 (2) (2003) 146-149. https://dx.doi.org/10.1016/S0379-0738(03)00195-6

[12] S. Cantrell, C. Roux, P. Maynard, J. Robertson, A textile fibre survey as an aid to the interpretation of fibre evidence in the Sydney region, Forensic Sci. Int. 123 (1) (2001) 48-53. https://dx.doi.org/10.1016/S0379-0738(01)00520-5

[13] M.C. Grieve, T. Biermann, The population of coloured textile fibres on outdoor surfaces, Sci. Justice 37 (4) (1997) 231-239. https://dx.doi.org/10.1016/S1355-0306(97)72196-8

[14] K. Wiggins, P. Drummond, The analysis and comparison of blue wool fibre populations found at random on clothing, Sci. Justice 45 (3) (2005) 157-162. https://dx.doi.org/10.1016/S1355-0306(05)71650-6

[15] M. Eng, P. Martin, C. Bhagwandin, The analysis of metameric blue fibers and their forensic significance, J. Forensic Sci. 54 (4) (2009) 841-845. https://dx.doi.org/10.1111/j.1556-4029.2009.01079.x

[16] K. Wiggins, R. Palmer, W. Hutchinson, P. Drummond, An investigation into the use of calculating the first derivative of absorbance spectra as a tool for forensic fibre analysis, Sci. Justice 47 (1) (2007) 9-18. https://dx.doi.org/10.1016/j.scijus.2006.11.001

[17] J. Jones, T. Coyle, Automotive flock and its significance in forensic fibre examinations, Sci. Justice 50 (2) (2010) 77-85. https://dx.doi.org/10.1016/j.scijus.2009.05.003

[18] M.C. Grieve, K.G. Wiggins, Fibers under fire: suggestions for improving their use to provide forensic evidence, J. Forensic Sci. 46 (4) (2001) 835-843. https://doi.org/10.1520/jfs15055j

[19] European Network of Forensic Science Institutes (ENFSI), Guideline for Evaluative Reporting in Forensic Science, 2015, http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

[20] C. Weyermann, R. Marquis, C. Delaporte, P. Esseiva, E. Lock, L. Aalberg, et al., Drug intelligence based on MDMA tablets data: I. Organic impurities profiling, Forensic Sci. Int. 177 (1) (2008) 11-16. https://dx.doi.org/10.1016/j.forsciint.2007.10.001

[21] A. Savitzky, M. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry 36 (8) (1964) 1627-1639. https://doi.org/10.1021/ac60214a047

[22] D.A. Heiser, B.D. McCullough, On the accuracy of statistical procedures in Microsoft Excel 2007, Comput. Stat. Data An. 52 (10) (2008) 4570-4578. https://dx.doi.org/10.1016/j.csda.2008.03.004