

Andrew W. Hartshorne,<sup>1</sup> B.Sc. and David K. Laing,<sup>1</sup> Ph.D.

## Color Matching Within a Fiber Data Collection

---

**REFERENCE:** Hartshorne, A. W. and Laing, D. K., "Color Matching Within a Fiber Data Collection," *Journal of Forensic Sciences*, JFSCA, Vol. 33, No. 6, Nov. 1988, pp. 1345-1354.

**ABSTRACT:** The fiber data collection in use in the Home Office Forensic Science Service uses complementary chromaticity coordinates as a method of color description. Color variation within samples is represented by an elliptical error region around the mean chromaticity point of the sample.

For searching this collection, a method of matching single colors to the collection ellipses has been assessed. However, the necessary calculations are complex, so a number of simpler alternatives has been investigated. A simple box search has been found to be convenient, albeit with some reduction in specificity.

**KEYWORDS:** forensic science, fibers, colors (materials), matching

The microspectrophotometer can be used to obtain visible absorbance spectra allowing visually indistinguishable fibers to be discriminated [1]. Such spectra may also be used to calculate complementary chromaticity coordinates,  $x'$  and  $y'$ , which describe fiber colors objectively [2]. By using such a system, variations within replicate measurements on a sample can be represented by an elliptical error region located around the mean chromaticity point [3].

This notation has been used to record fiber colors in a large data collection [4]. The aim of this archive is to provide information to forensic scientists on the rarity of specific fiber features to help assess the evidential value of fiber matches. Within the data base, color is probably the most important feature recorded, and so a prime need is to be able to find all entries matching the color of a given fiber.

One solution is to consider a match to occur when the complementary chromaticity point of the subject fiber falls inside the error ellipse of a collection entry. Estimation of ellipse size is thus critical; overestimation will give many false matches, while an underestimate will lead to some true matches being rejected.

One way to determine the optimal ellipse size is to abstract one replicate chromaticity point from each sample in the data collection. If the ellipse size is correct, each abstracted "daughter" point will fall in the ellipse calculated for its "parent" group. In practice, however, daughter points may match other parents because the two parents are of similar color. A number of nonparental matches will be produced depending on the color distribution within the collection. The ideal is to have error ellipses large enough to maximize parental matches but small enough to minimize nonparental matches.

A possible drawback is that the necessary calculations may become unwieldy within a large data base. A simplified model then becomes attractive. However, this may produce an

Received for publication 3 Dec. 1987; accepted for publication 20 Jan. 1988.

<sup>1</sup>Forensic chemists, Central Research Establishment, Home Office Forensic Science Service, Aldermaston, Reading, Berkshire, U.K.

unacceptable level of mismatches so the simplified forms need to be tested against the ellipse model for both convenience and accuracy.

This work examines several methods of color matching within the fiber data collection. Some preliminary information on the distribution of fiber colors is also presented.

## **Experimental Procedure**

### *Parent and Daughter Data Sets*

From the fiber data collection [4], a total of 499 wool, 614 cotton, and 403 polyester records were obtained. Each wool and cotton record included 10 replicate chromaticity points from different fibers within the sample. The polyester records contained only 5 points. From each record, a single chromaticity point (the daughter) was abstracted. In addition, 390 cotton samples from blue denim garments were identified within the collection. These were similarly separated into parent and daughter sets.

### *Matching Parent and Daughter Sets Using Elliptical Error Regions*

For each parent set, the position and inclination of the error ellipse axes were found and the standard deviations of the distances of chromaticity points from the axes calculated [2]. A range of differently sized ellipses for each parent was then produced by setting the ellipse axis lengths equal to multiples of these standard deviation values. Increments of 0.25 standard deviations were used in all cases.

Each daughter chromaticity point was compared with every parent error ellipse to identify matches, this being accomplished in two stages. If the distance between the point and the ellipse center was greater than the major axis length, the pairing was rejected as a match. If not, the ellipse radius in the direction of the point was calculated. If this radius was greater than the distance between the point and the center, the chromaticity point lies inside the ellipse and the colors match.

### *Matching Parent and Daughter Sets Using Simplified Error Models*

Three simpler alternatives to the error ellipse model were investigated: a circle and a rectangle circumscribing the ellipse and a rectangle inscribed inside the ellipse, as shown in Fig. 1.

Circles were constructed so that the major axis of the ellipse became the diameter of the corresponding error circle. This produced a range of error circles for each parent group corresponding to the different standard deviation values used to construct the ellipses. A color match was said to occur when the distance between the daughter point and the center of the circle was less than the radius.

Both rectangles were constructed so that the sides were parallel to, and equidistant from, the ellipse axes. The sides of the rectangle circumscribing the ellipse were made equal to the axis lengths. For the rectangle inscribed by the ellipse, the sides were set equal to the ellipse axis lengths multiplied by 0.7071, which makes its area a maximum. Again, a range of rectangle sizes based on the ellipse standard deviation values was obtained. In both cases, a match occurred when the perpendicular distances of the daughter point from the ellipse axes were both less than the half lengths of the respective sides.

### *Matching Parent and Daughter Sets Using A Simple Box Search*

An even simpler alternative to the use of ellipses is the construction of square "boxes," centered on the mean chromaticity point of each parent set. The box size was initially set at a side length of 0.01 chromaticity units, subsequently increasing in steps of 0.01 units to pro-

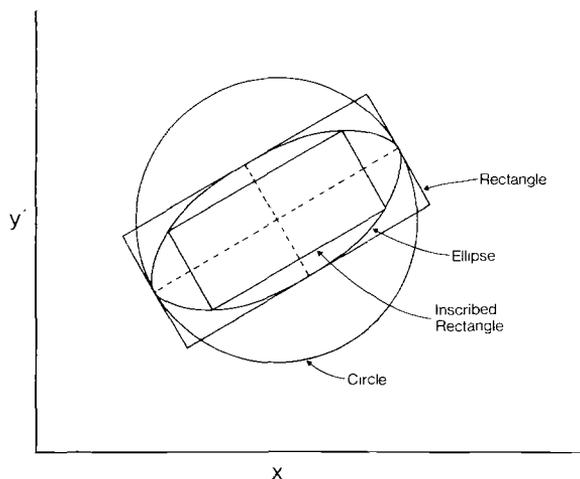


FIG. 1—Three simpler alternatives to the error ellipse model are a circle and a rectangle circumscribing the ellipse and a rectangle inscribed inside the ellipse.

duce a range of box sizes for each parent. Daughter chromaticity points were judged to match when the  $x'$  and  $y'$  differences between a daughter point and the box center were less than half of the box dimensions.

#### Color Distributions

To enable the color distributions within the wool, cotton, and polyester data sets to be estimated, the mean complementary chromaticity coordinates of each parent were converted into polar coordinates. This meant that hue, represented by  $\theta$ , and saturation, represented by 1, could be separated [5]. The hue and saturation frequencies were then obtained by sorting the values into intervals of  $10^\circ$  and 0.05 saturation units, respectively.

### Results and Discussion

#### Wools

The numbers of parental matches found using error ellipses, circles, and rectangles are shown in Table 1, the information being presented diagrammatically in Fig. 2. For the same number of standard deviations, ellipses produce more parental matches than inscribed rectangles, but less than the other two models. This is expected, as it is also the order of increasing physical area. The numbers of nonparental matches produced are also shown in Table 1 and plotted in Fig. 3. Again, the number of nonparental matches increases with increasing area.

Taking identification of 95% of parental matches as an acceptable level, elliptical error regions also classify 2.9% of the total possible number of nonparental comparisons as matches. However, some of these are almost certainly not false matches. A plot of hue angle frequency for wool (Fig. 4) shows two sizeable maxima, representing yellow and blue hues. It is thus extremely likely that some of the 499 wools are indistinguishable by colorimetric methods. Inscribed rectangles, circumscribed rectangles, and circles give 3.2, 3.3, and 7.2% of nonparental comparisons as matches. On this basis, circles are a poor substitute for ellipses in color matching, while both rectangles are reasonable alternative models.

Previous work [2] stated that an ellipse with axis lengths based on two standard deviations

TABLE 1—Parental and nonparental matches for wool samples using different models to describe the measurement errors in the parent set.<sup>a</sup>

Number of Standard Deviations <sup>b</sup>	Inscribed Rectangle		Ellipse		Rectangle		Circle	
	Matches		Matches		Matches		Matches	
	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent
0.25	3	43	6	68	9	84	42	382
0.50	20	173	32	278	44	347	111	1 419
0.75	54	387	84	589	102	739	189	3 028
1.00	91	653	140	991	162	1 239	272	4 916
1.25	142	985	197	1 455	237	1 830	340	7 144
1.50	187	1 361	268	2 077	313	2 604	401	9 650
1.75	235	1 794	323	2 806	361	3 531	432	12 234
2.00	286	2 316	373	3 633	407	4 554	460	14 989
2.25	334	2 952	412	4 549	435	5 671	478	17 798
2.50	365	3 601	442	5 545	462	6 908	489	20 748
2.75	400	4 316	460	6 601	479	8 155	497	23 821
3.00	419	5 107	478	7 689	485	9 492	499	26 727
3.25	440	5 917	488	8 867	491	10 963		
3.50	459	6 778	493	10 109	494	12 459		
3.75	470	7 674	495	11 411	495	14 018		
4.00	483	8 582	495	12 770	495	15 665		
4.25	485	9 527	495	14 250	496	17 374		
4.50	488	10 552	495	15 711	498	19 101		
4.75	493	11 588	497	17 143	498	20 818		
5.00	494	12 679	499	18 655	499	22 607		
5.25	495	13 799						
5.50	495	14 942						
5.75	495	16 087						
6.00	496	17 317						
6.25	496	18 553						
6.50	498	19 764						
6.75	498	20 966						
7.00	499	22 231						

<sup>a</sup>Maximum number of parental matches = 499.

Maximum number of nonparental matches = 248 502.

<sup>b</sup>Derived from the ellipse model—see Fig. 1.

should enclose about 90% of measured values. By analogy, 90% of possible parental matches should occur with ellipses of this size. Table 1 shows that this is not so, only 75% being identified at this level. However, only nine chromaticity points were used for the calculation of each ellipse, leading to a poor estimate of the population standard deviation. A better estimate uses a correction factor, or *t*-value, to account for the small data sets [6]. This correction multiplies the ellipse axis lengths by the magnitude of the *t*-value.

The numbers of parental and nonparental matches produced for ellipses modified by the *t*-value are shown in Table 2. Over 90% of parental matches are now identified at the two standard deviation level, with the relationship between the numbers of parental and nonparental matches being unchanged.

#### Cottons

Table 3 shows the numbers of parental matches produced by elliptical and circular error regions at different standard deviation levels. It can be seen that 95% of parental matches are identified at a level comparable to the wool samples. However, to identify the remaining

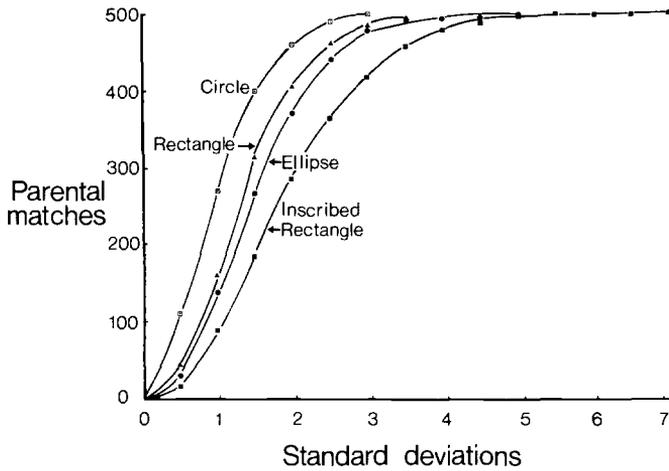


FIG. 2—Numbers of parental matches found using error ellipses, circles, and rectangles.

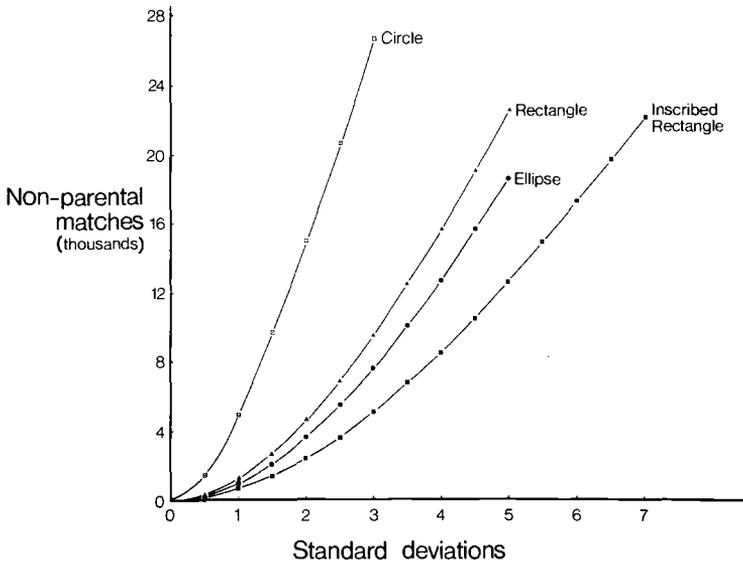


FIG. 3—Numbers of nonparental matches found using error ellipses, circles, and rectangles.

parental matches, large increases in size are required. This may be due to atypical values being selected as the daughter points in a few cotton samples. This reflects some of the difficulties endemic in color measurement of cotton fibers.

For a given proportion of parental matches, relatively more nonparental matches are identified as compared to the wools. However, Figs. 5 and 6 show the color distribution of the cottons is very restricted, reflecting the large number of blue denim garments in the general population. As a result, many parent sets are indistinguishable.

A *t*-value modification is required to give 90% of parental matches at the two standard deviation level. As was found for wools, the proportion of nonparental matches is relatively unaffected by this correction.

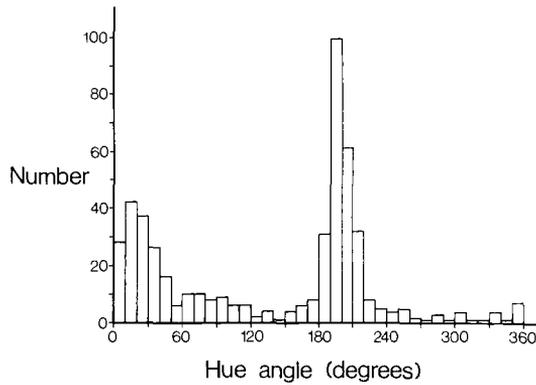


FIG. 4—Plot of hue angle frequency for wool.

TABLE 2—Parental and nonparental matches for wool samples correcting for the small replicate sets.<sup>a</sup>

Number of Standard Deviations <sup>b</sup>	Ellipse		Circle	
	Matches		Matches	
	Parent	Nonparent	Parent	Nonparent
0.25	48	349	143	1 861
0.50	172	1 278	306	6 273
0.75	316	2 742	431	12 036
1.00	419	4 781	480	18 509
1.25	472	7 208	498	25 441
1.50	493	9 914	499	32 397
1.75	495	13 013		
2.00	495	16 393		
2.25	499	19 947		

<sup>a</sup>Maximum number of parental matches = 499.

Maximum number of nonparental matches = 248 502.

<sup>b</sup>Derived from the ellipse model—see Fig. 1.

### Polyesters

Table 4 shows the number of parental matches produced using elliptical and circular error models. The number of standard deviations required for 95% of parental matches is much larger than those found for either wool or cotton fibers. However, as each ellipse is calculated from only four chromaticity points, the standard deviations are probably grossly underestimated. Applying the *t*-factor for this size of data set gives 90% of parental matches at a level more appropriate to this confidence level.

Also shown in Table 4 are the numbers of nonparental matches. For ellipses, recovery of 95% of parental matches produces a proportion of nonparental matches intermediate between that for wools and cottons. Again, this is probably related to the color distribution within the set.

### Denims

The numbers of parental matches produced using different ellipse sizes are shown in Table 5. As might be expected, the standard deviation needed for 95% of parental matches is

TABLE 3—Parental and nonparental matches for cotton samples using different models to describe the measurement errors in the parent set.<sup>a</sup>

Number of Standard Deviations <sup>b</sup>	Ellipse		Circle		Corrected Ellipse <sup>c</sup>		Corrected Circle <sup>c</sup>	
	Matches		Matches		Matches		Matches	
	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent
0.25	13	505	39	1 724	58	2 573	163	8 153
0.50	46	1 950	130	6 357	199	9 254	396	25 157
0.75	86	4 287	245	12 797	379	18 258	523	41 306
1.00	155	7 149	344	20 549	485	27 229	581	54 720
1.25	237	10 677	424	27 983	563	35 977	607	67 218
1.50	309	14 539	479	35 192	592	44 092	612	79 105
1.75	380	18 542	524	41 778	604	51 760	613	90 418
2.00	428	22 526	564	47 679	609	58 866	613	101 169
2.25	481	26 336	578	53 403	610	65 664	613	111 423
2.50	515	30 158	590	58 946	612	72 399	614	121 943
2.75	553	33 908	606	64 398	612	79 383		
3.00	570	37 577	608	69 492	613	86 060		
3.25	581	41 150	611	74 747	614	92 475		
3.50	592	44 566	612	79 779				
3.75	597	47 894	613	84 825				
4.00	603	51 158	613	89 584				
4.25	606	54 334	613	94 338				
4.50	609	57 360	613	98 992				
4.75	609	60 435	613	103 560				
5.00	610	63 377	613	107 930				
5.25	610	66 249	613	112 389				
5.50	612	69 197	614	116 926				
5.75	612	72 124						
6.00	612	75 104						
6.25	612	78 135						
6.50	612	81 087						
6.75	613	83 963						
7.00	613	86 811						
7.25	614	89 604						

<sup>a</sup>Maximum number of parental matches = 614.

Maximum number of nonparental matches = 376 382.

<sup>b</sup>Derived from the ellipse model—see Fig. 1.

<sup>c</sup>Corrected for the small replicate sets.

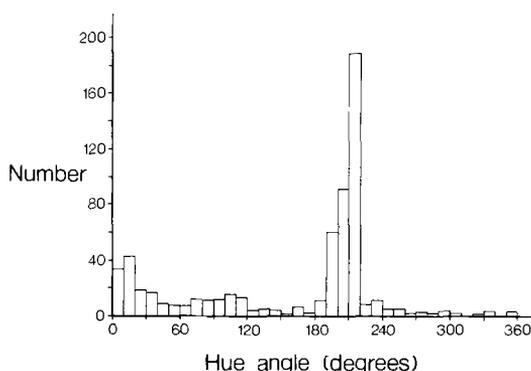


FIG. 5—Plot of hue angle frequency for cotton.

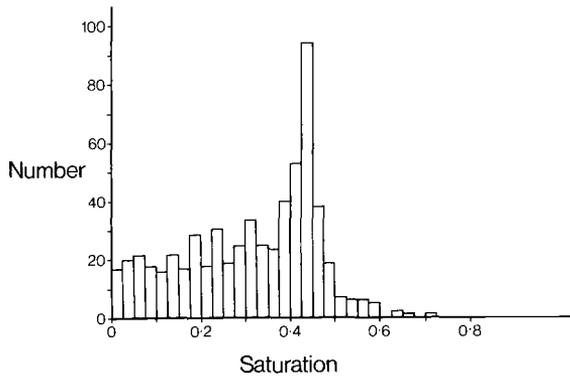


FIG. 6—Plot of saturation frequency for cotton.

TABLE 4—Parental and nonparental matches for polyester samples using different models to describe the measurement errors in the parent set.<sup>a</sup>

Number of Standard Deviations <sup>b</sup>	Ellipse		Circle		Corrected Ellipse <sup>c</sup>		Corrected Circle <sup>c</sup>	
	Matches		Matches		Matches		Matches	
	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent
0.25	8	23	43	230	57	226	167	1 815
0.50	22	88	99	775	166	870	302	5 753
0.75	53	205	157	1 637	261	1 901	360	10 377
1.00	83	353	208	2 711	310	3 253	389	15 412
1.25	114	540	262	4 000	338	4 754	398	20 484
1.50	148	780	294	5 261	361	6 549	403	25 778
1.75	191	1 047	319	6 648	370	8 492		
2.00	217	1 343	340	8 051	377	10 597		
2.25	242	1 708	355	9 513	379	12 718		
2.50	271	2 051	370	11 084	388	14 975		
2.75	287	2 472	381	12 639				
3.00	300	2 891	387	14 258				
3.25	312	3 377	390	15 839				
3.50	325	3 851	393	17 423				
3.75	332	4 308	396	19 062				
4.00	340	4 808	398	20 647				
4.25	347	5 330	402	22 322				
4.50	354	5 907	403	24 014				
4.75	360	6 507						
5.00	364	7 106						
5.25	367	7 725						
5.50	370	8 343						
5.75	371	8 996						
6.00	373	9 676						
6.25	377	10 316						
6.50	377	10 960						
6.75	378	11 642						
7.00	379	12 317						
7.25	379	12 968						
7.50	382	13 673						
7.75	387	14 385						

<sup>a</sup>Maximum number of parental matches = 403.

Maximum number of nonparental matches = 162 006.

<sup>b</sup>Derived from the ellipse model—see Fig. 1.

<sup>c</sup>Corrected for the very small replicate sets.

TABLE 5—Parental and nonparental matches for denim samples using ellipses to describe the measurement errors in the parent set.<sup>a</sup>

Number of Standard Deviations	Ellipse	
	Matches	
	Parent	Nonparent
0.25	4	1 325
0.50	36	5 318
0.75	75	11 633
1.00	125	19 636
1.25	169	29 088
1.50	224	39 190
1.75	259	49 352
2.00	290	59 238
2.25	322	68 380
2.50	339	76 564
2.75	356	83 876
3.00	363	90 324
3.25	372	96 119

<sup>a</sup>Maximum number of parental matches = 390.  
Maximum number of nonparental matches = 151 710.

identical to that for the cotton samples. Similarly, the proportion produced at a level of two standard deviations is comparable with the general cotton data set.

The numbers of nonparental matches are also shown in Table 5. At the 95% parental match level, 65.5% of all possible nonparental comparisons are recovered. As the blue denim samples were specially chosen to be of the same color, this result is not unexpected. However, it does show that even within a relatively small area of the chromaticity diagram, ellipses still provide a degree of color discrimination.

### Box Matching

Table 6 shows the numbers of parental and nonparental matches obtained for each of the fiber types examined. It can be seen that a box of  $\pm 0.025$  is necessary to identify 95% of parental matches in the three natural fiber groups. For polyesters, a box size of less than  $\pm 0.015$ , suffices, reflecting the greater precision of color measurements on synthetic fibers [2]. As box size is independent of statistical considerations, matching is relatively unaffected by the small number of data points in the polyester parent sets.

Comparing the numbers of nonparental matches, the box method is inferior. For example, when 95% of wool parental matches are identified, 9.5% of nonparental comparisons are matches. This compares with 2.9% using an ellipse model and 7.2% for circular error regions.

### Conclusions

The specificity of color matching depends on the description of uncertainty in measurement. The ellipse model produces the best discrimination, but match calculations between chromaticity points and ellipses are time-consuming and wasteful of computer resources if a large data base is to be searched. Of the approximations considered, circular error regions are a simpler method of matching, but specificity is reduced. Both rectangular approximations provide more specificity than circles, but are more difficult to use. A disadvantage of all approximations to ellipses is that the ellipse dimensions still have to be stored for calculation of the alternative models.

TABLE 6—Parental and nonparental matches using a simple box model.

Box Size	499 Wools		614 Cottons		403 Polyesters		390 Denims	
	Matches		Matches		Matches		Matches	
	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent	Parent	Nonparent
0.010	280	2 433	191	8 062	258	2 344	130	21 005
0.020	360	7 682	385	24 145	357	7 757	269	60 607
0.030	430	13 966	515	39 703	392	14 441	348	93 183
0.040	463	20 408	573	52 447	396	21 632	376	113 724
0.050	482	26 920	591	63 373	399	28 886	384	125 686
0.060	484	33 574	599	73 697	402	36 564	387	132 846
0.070	486	40 404	603	83 949	403	44 213	390	137 070
0.080	490	47 792	606	94 577				
0.090	491	55 738	608	105 297				
0.100	495	64 184	608	115 501				
0.110	496	72 823	611	125 481				
0.120	497	81 744	611	135 325				
0.130	498	90 478	611	145 096				
0.140	499	99 333	611	155 234				
0.150			612	165 754				
0.160			612	176 778				
0.170			613	188 376				
0.180			613	199 784				
0.190			613	210 284				
0.200			614	219 497				

The box search, while less specific than many of the previous models, has a number of advantages. As box size is independent of the data points, only the mean values need to be stored, making best use of computer space. The search system is extremely fast, and hence response times with a large data base are not a limitation. Taking all factors into consideration, the box search method is preferred. It has been shown that the optimum box size depends on fiber type. Adoption of the box size found for natural fibers will overestimate matches between synthetic fibers. However, if this method is used in a data base, overestimation will always tend to undervalue evidence, and thus err on the side of caution.

## References

- [1] Laing, D. K. and Isaacs, M. D. J., "The Examination of Paints and Fibres by Microspectrophotometry," *UV Group Bulletin*, Vol. 9, No. 1, June 1981, pp. 18-27.
- [2] Laing, D. K., Hartshorne, A. W., and Harwood, R. J., "Colour Measurements on Single Textile Fibres," *Forensic Science International*, Vol. 30, No. 1, Jan. 1986, pp. 65-77.
- [3] Hartshorne, A. W. and Laing, D. K., "The Definition of Colour for Single Textile Fibres by Microspectrophotometry," *Forensic Science International*, Vol. 34, No. 1/2, May/June 1987, pp. 107-129.
- [4] Laing, D. K., Hartshorne, A. W., Cook, R., and Robinson, G., "A Fiber Data Collection for Forensic Scientists—Collection and Examination Methods," *Journal of Forensic Sciences*, Vol. 32, No. 2, March 1987, pp. 364-369.
- [5] Venkataraman, K., Ed., *The Analytical Chemistry of Synthetic Dyes*, Wiley-Interscience, New York, 1977.
- [6] Miller, J. C. and Miller, J. N., *Statistics for Analytical Chemistry*, Ellis Horwood, Chichester, England, 1984.

Address requests for reprints or additional information to  
 Andrew W. Hartshorne  
 Central Research Establishment  
 Home Office Forensic Science Service  
 Aldermaston, Reading, Berkshire, RG7 4PN U.K.