Journal of Chromatography A, 1218 (2011) 2793–2811

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Review

Forensic examination of ink by high-performance thin layer chromatography—The United States Secret Service Digital Ink Library

Cedric Neumann[a,*], Robert Ramotowski[b], Thibault Genessay[c]

[a] Forensic Science Program, Eberly College of Science, The Pennsylvania State University, 107 Whitmore Lab, University Park, PA 16802, USA
[b] Forensic Science Division,USA Secret Service, 950 H Street NW, Washington, DC 20223, USA
[c] Ecole des Sciences Criminelles, University of Lausanne, CH-1015 Dorigny, Switzerland

ARTICLE INFO

Article history:
Available online 22 December 2010

Keywords:
Forensic science
Ink
Secret service
Analysis
Comparison
Library
Standardization

ABSTRACT

Forensic examinations of ink have been performed since the beginning of the 20th century. Since the 1960s, the International Ink Library, maintained by the United States Secret Service, has supported those analyses. Until 2009, the search and identification of inks were essentially performed manually. This paper describes the results of a project designed to improve ink samples' analytical and search processes. The project focused on the development of improved standardization procedures to ensure the best possible reproducibility between analyses run on different HPTLC plates. The successful implementation of this new calibration method enabled the development of mathematical algorithms and of a software package to complement the existing ink library.

Contents

* Corresponding author.
E-mail address: cedric.neumann@me.com (C. Neumann).

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2010.12.070

1. Introduction

It is fair to say that the forensic examination of writing fluids is a specialty within the various fields of expertise of modern criminalistics. The impact of ink evidence on the detection and conviction of offenders is not as significant, immediate and visible as for other evidence types such as DNA, fingerprint or shoe impressions. However, the U.S. Secret Service has used its large collection of writing inks, known as the International Ink Library, to help investigate forged documents, financial fraud, threats against government officials, suspected terrorism, child pornography, medical fraud, war crimes, and obstruction of justice cases. Specifically, the library has been used in the last decade to examine evidence from several high profile cases, including the Martha Stewart ImClone stock trading case, the DC Sniper case, and war crimes cases (e.g. U.S. vs. John Kalymon [1]).

The forensic examination of writing inks is usually performed for identification, comparison or dating purposes [2–6]. The identification and comparison processes aim at inferring the identity of the source [7] of an ink specimen sampled from a questioned document. In other words, the purpose of the identification process is to investigate the brand, model, manufacturer and year of production of the instrument(s) that could have produced the ink specimen. This is achieved by comparing, and selectively associating, the questioned ink specimen with samples contained in a library of controls from known sources, such as the International Ink Library (see Section 3). The aim of the comparison process is to assess, by a direct side-by-side comparison, whether a particular writing instrument has been used to write the questioned ink specimen.

The dating exercise focuses on the determination of the time of a particular entry on a document, usually by quantifying one or several components of the ink. Dating an ink is not a straightforward examination and has been discussed by many authors [8–14]. The dating process is not the topic of this paper, which focuses on the first two uses of ink analysis in forensic science.

The forensic examination of ink is not recent. Carvalho relates that the first report in court of the results of the chemical examination of an ink on a questioned document occurred in 1889 in New York County [15]. This interest for the examination of ink has naturally evolved with the developments of ink and writing instruments: from India ink, and iron gall inks used with dip pens, through water-, oil- and glycol-based inks used in ballpoint and felt-tip pen, to toners and pigment-based inks used in modern printers. In parallel, the technological development of analytical methods has offered criminalists with a wide range of possibilities to analyze the various components of inks. Early examinations were performed by reacting a minute drop of chemical reagent on a portion of the ink stroke [16–19]. Chromatography of ink was introduced in the early 1950s [20–22], soon followed by planar thin-layer chromatography (TLC) [23]. More recently, gas chromatography/mass-spectroscopy (GC/MS), Fourier transform infrared (FTIR), Raman spectroscopy, high performance liquid chromatography (HPLC) and high-performance thin layer chromatography (HPTLC) have been used for the analysis of the vehicles, dyes and pigments present in ink samples. Reviews of these techniques can be found in [24–29]. Importantly, these reviews show that planar chromatography – more specifically TLC and HPTLC – has remained the technique of choice for the analysis of ink in criminalistics. It combines rapidity, efficiency in the screening of ink dye composition and is cost effective. Planar chromatography also suffers from some limitations that were not properly addressed in the field of forensic ink examination. The main limitations are presented in this paper, together with solutions that were developed during a research project aimed at creating the new Digital Ink Library for the U.S. Secret Service.

In the next section, a summary of the scientific and legal contexts defining the admissibility of ink evidence in U.S. courts is provided; Section 3 briefly presents the International Ink Library; Section 4 looks at the main limitations of the current analysis of ink by planar chromatography and lists a series of improvements needed to meet the criteria of the admissibility of scientific evidence in courts; Sections 5 and 6 present the research projects that were designed to investigate these improvements and their results; Section 7 describes the practical implementation of the results of the research projects in the Digital Ink Library (DIL) at the U.S. Secret Service; and finally, Sections 8 and 9 discuss the immediate implications of this work for the ink examiner community and present some scope for future work.

2. Scientific and legal contexts

The main purpose of forensic ink examination is to contribute to the determination of facts in civil and criminal investigations. Ultimately, every ink analysis is susceptible to being presented in court and used by the trier of facts. In 1984, Brunelle and Reed [3] presented a series of historical and more recent precedents of the admissibility of ink evidence in U.S. civil and criminal courts. At the time, Brunelle considered that ink identification satisfied the criteria of the general acceptance of the technique in the relevant scientific community [30]. And indeed, to this day, ink evidence has only been successfully challenged on one occasion: in *U.S. v. Bruno* in 1971 [31].

The Judge's opinion in the *Bruno* case was that the "state of the art in this field of ink identification is not yet sufficiently advanced to be reasonably scientifically certain that an ink of unknown composition is the same as a known ink". His justification was mostly based on:

(i) The "great number of variables" that "affected to an unknown extent" the analysis and their results, including analytical and environmental factors, and the lack of uniformity in the production of inks. The demonstration of this effect was performed by the expert himself by presenting to the court "a slide in which three chromatograms of the same ink showed entirely different results" and having to "concede that he could not tell [. . .] that they were all prepared from the same ink";
(ii) The "rough subjective judgments as to differences in hue and intensity of color" between ink samples made by the expert when comparing the chromatograms and inferring the commonality of source;
(iii) The serious lack of representativeness of the ink library used by the expert to identify the brand and date of first introduction of the ink in the case, to the exclusion of "all [other] ball point inks manufactured anywhere in the world for an indefinite period before December, 1965";
(iv) The poor documentation and laboratory notes maintained by the expert with respect to the analytical procedure and its

apparent lack of standardization "did make cross-examination difficult" and "made it impossible for a jury to evaluate" the findings.

The arguments raised in the *Bruno* case were considered in other contemporaneous cases (see [3] for a list of cases). The *Bruno* case ultimately had a very limited impact on the judicial system. Courts, since that time, have admitted the testimony of ink examiners without any limitations. Nevertheless, the *Bruno* ruling triggered some significant changes in the way ink evidence was examined. The ink library used by the expert was significantly expanded with foreign inks, leading to the creation of the International Ink Library (see Section 3). Improved documentation of the analytical procedure and chain of custody were required from ink examiners, and more importantly, the development of standardized analytical procedures was initiated [32,33]. These changes were deemed sufficient at the time, and to this day ink evidence is accepted in court without much challenges.

A recent, widely publicized, report has once again brought the general issue of the validity and reliability of various forensic examination techniques under the scrutiny of the Courts and the public. In 2005, following more than a decade of challenges to scientific evidence in U.S. court, the U.S. Congress mandated the National Academy of Sciences to conduct a study on forensic science. In 2009, an ad hoc committee from the National Academy issued a report containing several observations and recommendations [34]. Overall, the committee felt that there were significant shortcomings in the demonstration of the scientific bases for most of the evidence types that were considered (including fingerprint, trace evidence, questioned documents, toolmark impressions and firearms).

In the particular case of the examination of ink, the committee found that (p. 167) it is "based on well-understood chemistry" and that it "presumably rests on a firmer scientific foundation". However, the committee "did not receive input on these fairly specialized methods and cannot offer a definitive view regarding the soundness of these methods or their execution in practice." A recent paper by Neumann and Margot [35] continues this analysis. In their review of the state of ink analysis in forensic science, and more specifically of the two ASTM standards supporting it [32,33], the authors show clear potential for improvements. Based on the same observations and arguments developed in the *Bruno* case, the authors reached very similar conclusions to the ones already presented in 1971: a better implementation of quality assurance processes for the analysis and comparison of ink samples is needed. Naturally, the developments and improvements made in the 1970s were based on the technology and knowledge available at the time; however, the improvements that were made in 1971 do not take advantage of modern technology and answer more modern court requirements on the admissibility of scientific evidence [36]. In other words, falling short of these needed improvements, ink evidence may experience the same challenges that other types of evidence are currently encountering [34,37,38].

In addition to the scientific and legal challenges focusing on the analytical aspects of ink examination, modern criminalists are increasingly pressed to assess and report the meaning of the scientific evidence in an objective and transparent manner [34]. Neumann and Margot [39] have shown that evaluating the weight of ink evidence using a probabilistic framework, in a similar fashion as DNA evidence [40], enables the maximization of the benefits and contribution of ink evidence to the criminal and civil justice system. Such probabilistic framework demands the establishment of reference collections for the determination of the frequency of questioned specimens. The establishment and use of these collections naturally appears to face the same challenges of reproducibility of the analysis and objectivity of the comparison of ink samples as introduced above for the comparison and identification processes. The implementation of a probabilistic framework will therefore also benefit from the analytical research presented in this paper.

3. The International Ink Library of the United States Secret Service

The U.S. Secret Service maintains a reference collection of more than 10,800 domestic and international writing ink samples, some of which date back to the 1900s [8]. The library also includes a smaller collection of toners as well as printing, stamps pad, typewriter, and inkjet inks. The origin of the collection dates back to the 1960s when Werner Hoffman (of the Zurich Cantonal Police in Switzerland) amassed a modest collection of European blue and black ballpoint inks. This collection, along with samples from the Federal Bureau of Investigation, United States Postal Service, and several other private examiners, would form the nucleus of the International Writing Ink Library in the late 1960s [3].

In 1968, Richard Brunelle, a chemist with the Internal Revenue Service's (IRS) Alcohol and Tobacco Tax Division Laboratory, began to contact domestic ink producers and quickly amassed between 1500 and 2000 samples to add to the library. This collection was soon transferred to the new Bureau of Alcohol, Tobacco, and Firearms (BATF) Laboratory on July 1, 1972 (the old IRS Alcohol and Tobacco Tax Division laboratory was incorporated into this new entity). The IRS laboratory in Chicago ultimately contributed approximately 1500 open market samples (i.e., purchased in stores or collected from public sources) to the library by the late 1980s. The IRS laboratory also obtained portions of the remaining samples contained in the BATF library at that time. In 1988, the BATF transferred custody and responsibility for maintaining the library to the U.S. Secret Service, Forensic Services Division [41]. The two libraries are now maintained cooperatively and are constantly being updated with samples from major ink manufacturers (primarily in the United States, Europe, and Japan) as well as with open market purchases from numerous domestic and international sources.

4. Forensic examination of writing inks

4.1. Background information

The analysis and comparison of ink in criminalistics is more of a screening problem than a detection and quantification one. When inferring the source of an ink specimen, criminalists are not so much concerned with the specific identification and quantification of its components: their main interest is the verification that the components *are the same* in the questioned and control ink samples.

Unfortunately, criminalists usually work in adverse conditions. Firstly, ink examination is part of a more comprehensive forensic document examination process. It is important to minimize the impact of the ink sampling process in order to avoid important alteration of the questioned document. Usually, a few millimeters of ink stroke are sampled for the examination and only a small amount of ink material is available for the chemical analysis. Since the ink is usually not homogeneously deposited on the paper, it is not known precisely how much ink is extracted and analyzed.

Secondly, ink examination is often performed on documents that are several months/years old. The conservation conditions of the document are not necessarily optimal, or even known to the criminalist, and environmental factors may degrade/modify the chemistry of the ink. As an example, the degradation of the dyes in an ink due to its exposure to sunlight over the course of several

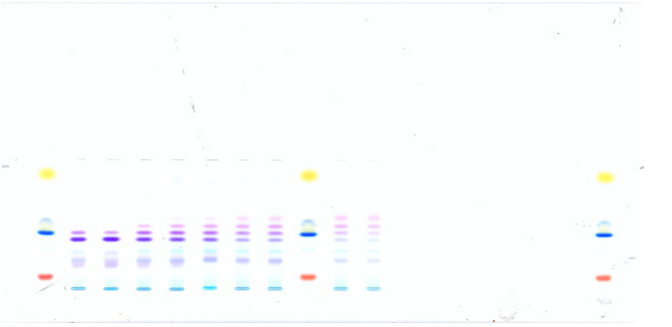

Fig. 1. Degradation of an ink containing Basic Violet 3 (Methyl Violet 10B – C.I. 42555 [42]) due to exposure to sunlight. Exposure went from 0 to 8 weeks (tracks 2–11). Dye ladders are located on tracks 1, 9 and 18; solvent and paper controls are on tracks 16 and 17.

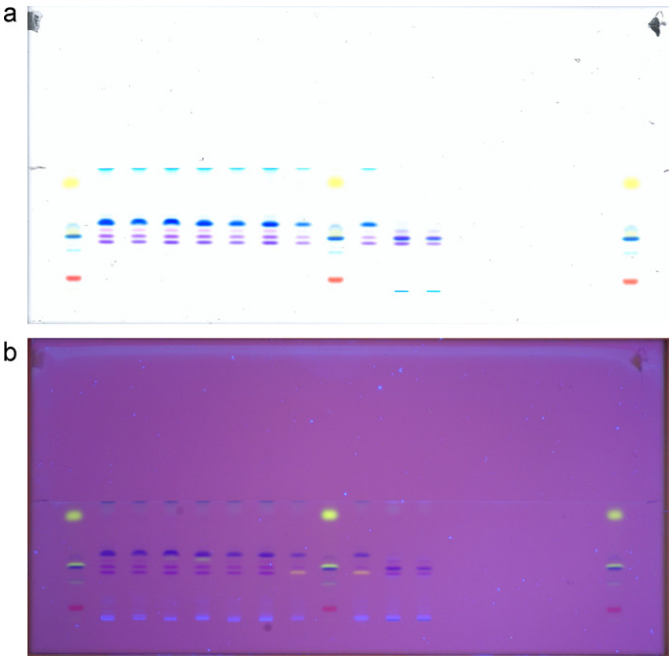

Fig. 2. (a) Comparison between 5 pairs of blue ballpoint pen ink from different batches used in Caran d'Ache Goliath refills (tracks 2 and 3, 4 and 5, 6 and 7, 8 and 10, 11 and 12). Dye ladders are located on tracks 1, 9 and 18; solvent and paper controls are on tracks 16 and 17. (b) Observation of the results presented in (a) under UV–visible luminescence (excitation at 254 nm—observation in the visible range). Dye ladders are located on tracks 1, 9 and 18; solvent and paper controls are on tracks 16 and 17.

weeks is presented in Fig. 1. The comparison of two samples to infer the source of a questioned ink needs to properly take into account these degradation factors. This is a difficult task. It is especially difficult to account for these factors when simply comparing samples side-by-side without supporting data.

Thirdly, the aim of ink manufacturers is to satisfy writing instrument manufacturers' requirements in terms of viscosity, opacity and drying time of the ink. Ink manufacturers are more interested in the physical properties of the ink than in the chemical ones. They also pay a lot of attention to economical considerations and do not hesitate to replace components by cheaper ones providing that the physical properties of the ink remain the same. Thus criminalists cannot rely on the characterization of every ink by the list of their components. The individual identification of ink components would require an exhaustive reference collection of all possible chemical compounds that were, are and will be used in the manufacture of inks. The creation and maintenance of such collection does not appear realistic in practice.

Finally, in today's globalized economy and mass-manufacturing world, instrument manufacturers may use inks from different suppliers for a given model of instrument and are selling their product worldwide. The variation in the composition of the ink of a given brand/model of instrument is illustrated in Fig. 2: tracks 2–10 are samples from different batches of ink produced for the Caran d'Ache Goliath refills by National Ink Inc. (USA), while tracks 11 and 12 are samples of ink produced for the same refills by Dokumental GmbH (Germany). Fig. 2 shows that only the first (tracks 2 and 3) and third batches (tracks 6 and 7) – both produced by National Ink Inc – cannot be differentiated by HPTLC, while all other batches have different dye compositions.

Reciprocally, ink manufacturers may sell their ink to multiple instrument manufacturers, thus different brand/models of pens may share the same ink. This absence of direct relationships between ink and instrument manufacturers impacts the process of the identification of the source of an ink sample using a reference library, and is a challenge for the interpretation of the weight of ink evidence.

Back in 1971, the judge in the *Bruno* case [31] already formed his opinion based on the elements presented in this section, and on the fact that the ink expert at that time did not properly address them. These background elements are still very much influencing ink analysis today. It is possible that under certain environmental conditions some components of questioned ink specimens will be absent, or only weakly showing on their chromatogram; if they do appear, it may be under some degraded form. Examiners need to have a deep understanding of the environmental and analytical factors influencing ink samples in order to objectively analyze and compare them. The scale of the problem is evidently magnified when considering the hundreds of comparisons that need to be performed when identifying the source of an ink sample through a library.

4.2. In practice

The current practice results from the improvements made following the *Bruno* case. Forensic ink examinations are still mostly performed by HPTLC since it remains the most versatile technique. The analysis of the samples is guided by two ASTM standards [32,33]. These standards lack any form of control designed to prevent, minimize, or at least measure, the variability occurring between analytical runs. Over the years, some accredited laboratories have instituted quality control measures to address some of these issues. However, further work need to be done on the development and implementation of internal and external standards aimed at warning the examiner against failing analyses and at reducing variability between different runs.

Following its analysis, the inference of the source of a questioned ink is performed by direct visual comparison of the chromatograms of the questioned and control ink samples. The respective composition of the samples is usually not taken into account for the reasons expressed above: criminalists only consider whether these chromatograms *look alike*, taking into account *explainable differences* [33,35]. Side-by-side comparison of the two chromatograms on the same plate is the norm in the comparison scenario. The search of the source of a questioned ink in a reference library requires the repetitive direct comparison of the questioned specimen with the control samples from the library; however, in this scenario, the questioned specimen is analyzed on one plate, and is compared to reference samples, which were analyzed on different plates, at a different time, potentially at different places and by different individuals. The comparison process is essentially subjective and experience based, and lacks the necessary structure to account for analytical and environmental influencing factors. In turn, the identification process

encapsulates the limitations of the comparison process and may magnify them. Verification procedures aim at minimizing these limitations by exposing the samples to the same environmental conditions on a single plate. In addition, challenges related to the efficiency and reliability of searching the library may be manifested into longer than expected search times.

Finally, given the perceived difficulties of interpreting the meaning of ink evidence within the context of the case, examiners usually revert to fairly non-informative statements, such as "ink A is of the same formulation of ink B". They do not take full advantage of the statistical information contained in the available ink libraries. Such conclusions do not provide the courts with any indication on how common the formula of the ink is, and thus the weight associated with the considered evidence [35].

Overall, this short summary of current practice shows that ink examination could face validity and reliability challenges based on the requirements detailed by the National Academy of Science [34] and the courts [36].

4.3. Requirements of modern forensic ink examination

We outlined in the previous sections that the examination of ink can only satisfy the most recent and stringent criteria for the admissibility of scientific evidence if it meets some fundamental requirements: (i) reproducibility of the analytical process and results, (ii) understanding of the potential differences that may exist between the chromatograms of two samples of the same ink, (iii) efficient and reliable searching and retrieval of ink samples in the library and (iv) documentation of the analytical conditions, the results and of the chain of custody. More specifically:

(i) The analysis of the ink sample needs to be reproducible enough to ensure that differences in analytical conditions will not negatively affect the comparison of samples analyzed on different plates, by different individuals, or at different times and locations. The aim is to reduce analytical variability to a point where it is negligible when compared to the effect of other environmental factors, and that it will not have to be used to justify differences observed between two ink samples.
(ii) Differences between multiple chromatograms of samples from the same ink that were exposed to different environmental conditions, prior to their analyses, need to be better studied and documented. This will assist in the transparent and objective assessment of *explainable differences* between these chromatograms.
(iii) Automated algorithms are needed for the comparison and retrieval of ink samples in large libraries (as opposed to the current manual, tedious and subjective practice). The algorithms would naturally benefit from digital storage of the samples and be robust to the environmental conditions mentioned above.
(iv) Modern quality assurance processes in forensic science also require the implementation of efficient laboratory information management and case file systems. These requirements are not exclusively aimed at improving the quality of the analysis of ink samples. However, they are necessary when considering the wider challenges of reporting the results in court, and the maintenance of the chain of custody and quality assurance in the forensic laboratory.

A collateral requirement focuses on the implementation of a probabilistic framework for the interpretation of the ink evidence. From the analytical perspective, we have already mentioned that the application of such framework will take advantage of the same technology that needs to be developed to fulfill the requirements outlined above. We will not consider the statistical aspect any further. More information about probabilistic framework and its benefits in the context of the examination of ink evidence can be found in [39].

5. The project

A research program was designed in order to investigate how modern technology could support the requirements presented in the previous section to strengthen the validity and reliability of ink evidence. This joint program was led by the University of Lausanne (Switzerland) and by CAMAG AG (Switzerland). It was supported by the Science and Technology Directorate of the U.S. Department of Homeland Security and by the U.S. Secret Service.

The project was delivered in 2 phases. The aim of the first phase (research phase) was essentially to build a proof of concept of a computerized library for the storage, searching and retrieval of ink samples analyzed by planar chromatography [43]. During this PhD research, different technologies were investigated and tested to improve the reproducibility of the ink analyses and the reliability of the search and comparison processes. Subsequently, the analytical method was optimized and a software package for the Ink Library was developed and delivered to the U.S. Secret Service [44] (implementation phase). The implementation project considered both the analytical aspects and the needs in terms of laboratory information management and chain of custody.

The project was then oriented around 3 main axes:

(i) The development and optimization of methods for standardizing the analysis and measurement of ink samples by HPTLC. The aim was to ensure a level of reproducibility for the analysis of ink by planar chromatography that would be compatible with the establishment of a searchable ink library.
(ii) The development and optimization of algorithms for the comparison and searching of ink samples stored digitally. The aim was to improve the reliability of the search and comparison process, and to objectively take into account the environmental factors that can influence ink chromatograms.
(iii) The development of a computer-based laboratory management and electronic case-file system to assist with the monitoring of the proposed quality assurance process and chain of custody of the ink samples.

The project also ultimately included the complete reanalysis of nearly all of the inks in the International Ink Library using the newly proposed methods and processes. The Digital Ink Library of the United States Secret Service was delivered on January 30th 2009.

5.1. Sampling

Blue ballpoint pens are the most widespread type of writing instrument. Thus the concepts investigated during the research phase were tested and validated using a subset of these inks.

The sampling was designed to inform on the reproducibility of ink analysis when performed repeatedly on different plates using different batches of the solvent system. The sampling was also intended to provide data on the natural differences in the chromatograms of samples from the same ink exposed to different environmental conditions.

Thirteen inks were selected following their analysis by HPTLC. All 13 inks have a different dye profile from each other. Multiple samples from every ink (96 samples per ink) were repeatedly analyzed under a variety of analytical and conservation conditions known to influence their chromatograms [32,33,45,46]. The details of the sampling are reported in Table 1. For all conditions except the influence of paper, ink strokes were drawn on ISO 12757 certified write test paper produced by Baumgartner Papier Holding SA

Table 1
List of the samples and conditions studied during the development of the prototype.

#	Condition	Exposure	Same plate	Visual description of the samples after analysis
1	Ink concentration	4 samples of 5 mm of ink stroke in 10 μL of solvent system 5 samples of 10 mm of ink stroke in 10 μL of solvent system 4 samples of 20 mm of ink stroke in 10 μL of solvent system	Yes	The overall chemical profiles for samples of the same ink are similar. Some dyes are not present in the lower ink concentrations; some extra dyes are present in the upper ink concentration. On some samples with the upper ink concentration, major dye components seem to drag minor ones on the elution distance
2	Influence of HPTLC plate	10 samples analyzed on 10 different plates using the same preparation of the elution solvent system	No	No significant differences between samples from the same ink. Some weak dye components are not always present. Some differences in the respective concentration of some dyes are noticeable
3	Influence of the solvent system	10 samples analyzed on 10 different plates using 10 different batches of elution solvent system	No	Similar observation as in 2 above
4	Ink homogeneity in the cartridge—different plates	10 samples from 10 ink strokes generated at 5 min scribbling intervals analyzed on 10 different plates using the same batch of solvent system	No	Similar observation as in 2 above
5	Ink homogeneity in the cartridge—same plates	10 samples from 10 ink strokes generated at 5 min scribbling intervals analyzed on the same plate	Yes	Similar observation as in 2 above
6	Exposure to dry heat (temperature of 100 °C)	10 ink strokes exposed to 0, 20, 45, 90 min, and then 3 h, 6 h, 1 day, 2 days, 4 days and 8 days	Yes	Overall, there is a tendency for inks to produce faded or weakened dye profiles for 2 days onward
7	Exposure to humidity (RH 95%)	10 ink strokes exposed to 0, 20, 45, 90 min, and then 3 h, 6 h, 1 day, 2 days, 4 days and 8 days	Yes	No significant differences between samples from the same ink. Some weak dye components are not always present. Some differences in the respective concentration of some dyes are noticeable.
8	Ageing in darkness	9 ink strokes left in the dark at an ambient room temperature and %RH during 0, 6, 12, 18, 24, 30, 36, 42 and 48 weeks	Yes	Similar observation as in 7 above
9	Exposure to sun	9 ink strokes exposed to sun through a glass window during 0, 1, 2, 3, 4, 5, 6, 7 and 8 weeks	Yes	Clear degradation of the dye components, starting week 1. Few dyes remain visible after 8 weeks
10	Influence of paper	5 ink strokes drawn on the 5 different following papers: Antalis coloraction — Blue — 80 g/m^2 Antalis coloraction — Green — 160 g/m^2 Pelikan post-it — Yellow — 80 g/m^2 Xerox business — White — 0 g/m^2 Baumgartner ISO 12757 — White — 80 g/m^2	Yes	Paper influences the samples in two ways: (1) paper dyes are visible along the ink dyes; (2) non-visible paper components disturb ink dyes elution.

(Switzerland). In some experiments, all samples of the same ink were analyzed together on the same HPTLC plate, while in others, the samples were analyzed on different plates (see Table 1).

The development and optimization projects occurring in the implementation phase benefited from a larger sample of inks. A series of ballpoint inks, non-ballpoint inks (consisting of gel, roller ball, and felt types), and pure dyes were used to evaluate and optimize the analysis process. Sixteen black and 16 blue inks as well as one red ink were used in this study. Twenty-one pure dyes of various colors were also used in this evaluation. Ink samples were taken from Whatman grade 2 filter paper scribble sheets using a 1.0 mm diameter "Harris Micro-Punch" (Electron Microscopy Sciences, catalog #69034-10). Various factors like plate variation, humidity, saturation, extraction solvent, and mobile phase solvents were optimized using the aforementioned inks.

5.2. Analytical equipment

All analyses for the development of the prototype were performed on Merck® silica gel 60 plates (Merck KGaA, Germany). Prior to being used, the plates were washed (entirely eluted; the direction of elution is recorded) with methanol and dried in an oven for 20 min at 110 °C. The samples were deposited on the plates using a semi-automatic deposition apparatus LINOMAT IV from CAMAG AG (Switzerland). The plates were developed in horizontal chambers. The chromatograms on the plates were digitally acquired using the TLC Scanner III piloted by the WinCats software (CAMAG AG, Switzerland).

All analyses for the development and population of the Digital Ink Library were performed on Merck® silica gel 60 F_{254} (Merck KGaA, Germany). All the analytical equipment was provided by CAMAG AG (Switzerland). The samples were deposited on the plates using an automated system, the Automatic TLC Sampler (ATS) 4; developed in an environmental chamber, the Automatic Developing Chamber (ADC) 2; imaged (clean and developed plates) in the visible (transmitted and reflective modes) and in the short- and long-wave ultraviolet using a digital image capturing system, the DigiStore 2; the spectral information was digitally acquired using the TLC Scanner III run by the WinCats software.

5.3. Solvent systems and procedures

During the development of the prototype, the inks were extracted from the paper in a solution of ethanol/water: 1:1 (v/v). Ten millimeters of ink stroke was carefully scratched from the surface of the paper inserted into a micro-capillary tube together with

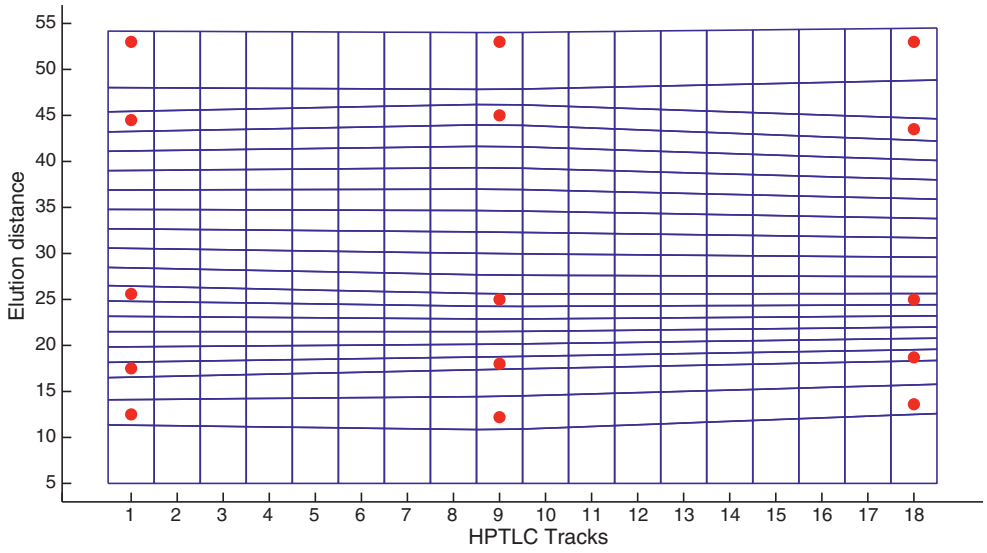

Fig. 3. Schematic of the dye ladder (dots) used as anchor points for the grid of the new coordinate system (lines).

10 μL of the extraction solvent system. The tubes were sealed on both sides and heated at 100 °C for 15 min. The full 10 μL was deposited on the HPTLC plates in bands 4 mm wide and 10 mm apart from each other. The samples were deposited 10 mm from the bottom of the plates, starting 15 mm from the left side of the plate. This configuration allows for the deposition of 18 tracks on a 20 cm wide plate.

Dye ladders (see Section 6.1) were deposited on each plate (tracks 1, 9 and 18). Blank samples from the extraction solvent system alone and from blank paper in the extraction solvent system, were also added to each plate (tracks 16 and 17).

The plates were eluted horizontally using an elution solvent system containing 1-butanol/ethanol/water/acetic acid: 60:10:20:0.5 (v/v) [47]. All solvents were HPLC grade. The plates were eluted for a distance of approximately 55 mm from the bottom of the plate (around 1 h of elution).

The extraction and elution solvent systems and the *dye ladders* were optimized during the implementation phase of the project. The optimization process, the solvent systems and the dye ladder that were used to re-analyze the samples of the International Ink Library are described in Section 6.2.

6. Results

6.1. Development and implementation of the quality assurance procedure

A significant improvement of this quality assurance procedure over the one described in the ASTM documents [32,33] can already be achieved by using the analytical equipment and method described in Sections 5.2 and 5.3. Nevertheless, it was deemed necessary to include the analysis of known quality standards alongside the ink samples. The purpose of these external standards is dual: on the one hand, the analysis of known standards on every plate acts as a warning mechanism against any failure in the analytical process; on the other hand, it enables the further processing of the analytical results in order to maximize their level of reproducibility.

Traditionally, elution distances in planar chromatography are characterized by the measurement of retention factors (Rf). Several solutions were proposed to improve the reproducibility of the Rf, such as the use of relative and corrected retention factors. In this study, we choose to expand on these concepts to emulate the use of allelic ladders in forensic DNA profiling [48,49] or the use of Kovat indices in gas chromatography [50].

A dye ladder was developed using a mixture of 3 dyes: Acid red 18 (Color Index (C.I.) [42] 16255)/Basic Blue 26 (C.I. 44045)/Solvent Orange 3 (C.I. 11270:1) in methanol—6:1.5:1:60 (w/w/w/v). These dyes were chosen on the basis of the limited dispersion of their spots at the end of the elution with the selected solvent system. The concentration of each dye was determined empirically. The dye ladder can be seen in the tracks 1, 9 and 18 in Fig. 2. The particular use of the dye ladder to calibrate samples analyzed on different plates has been described in [51]. In summary, the dyes of the ladders are used as anchor points for a calibration grid, which can be seen as a new coordinate system. The calibration grid is used to 'correct' the elution distance of every position on the plate by a two-dimensional interpolation between the metric system and the new coordinate system (Fig. 3).

Evidently the use of the proposed calibration process cannot be achieved without the digital acquisition of the chromatograms and some computerized signal processing. Traditionally, TLC scanners first detect the position of major components on a track at a given wavelength and, in a second stage, measure some other properties of the components at the detected positions (e.g. the absorption spectrum of the component). Since criminalists are more interested in the creation of a chemical profile of the ink sample rather than the identification of some/all of its components, our TLC scanner was modified to acquire the absorption intensity of each point of the elution track directly at multiple wavelengths. This approach provides a tridimensional representation of the sample elution track (Fig. 4), using the intensity of absorption at 31 wavelengths (between 200 and 700 nm) at each point of the elution distance (in this study, a measurement was taken every 50 μm over the elution distance). It does not require any manual verification of the selection of the relevant components since the whole elution distance is automatically scanned.

Preliminary experiments allowed for the correlation and calibration of the anchors of the grid with respect to the metric positions of the dyes in the ladder. This approach was evaluated using some of the data described in Table 1 (conditions #2–5). The improvements in reproducibility were tested for ink samples analyzed on different plates using the same batch of solvent system; on different plates using different batches of the solvent system; and when analyzed by different operators (different plates, same batch of solvent system).

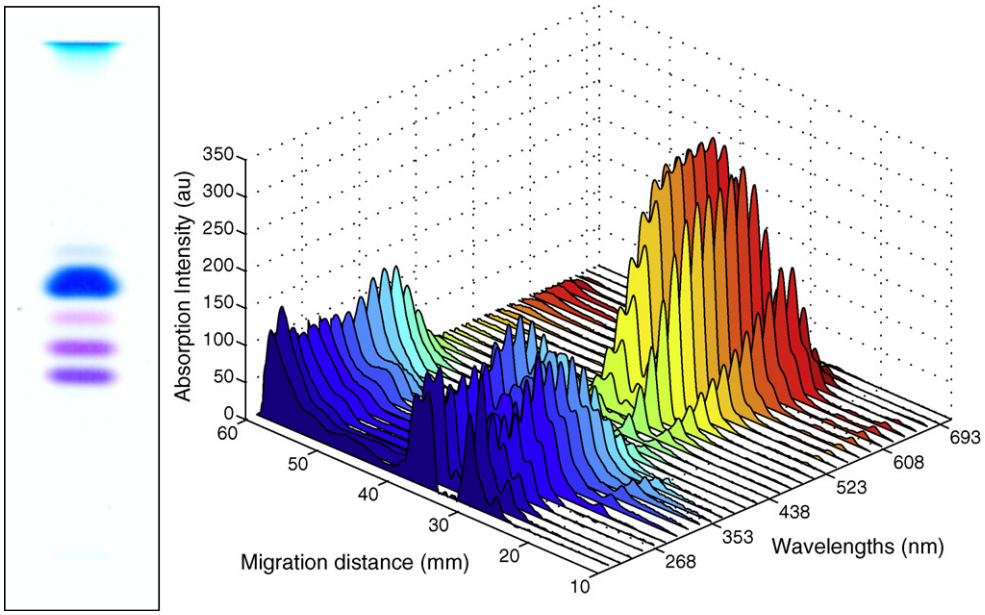

Fig. 4. Ink sample (on the left) with its tridimensional representation following its digital acquisition with the TLC Scanner III.

The results presented in Fig. 5a and b show the improvements in the reproducibility of the measurement of the position of the same ink components when analyzed on different plates. The improvements obtained by the use of the calibration grid over the use of the Rf, measured by the standard deviation of the positions of the components in the metric system and in the new coordinate system, are evident. When using the corrected Rf, Stead et al. [52] described an improvement of the mean standard deviation around the components' position from 0.06 Rf to 0.03 Rf. This can be compared to the reduction (in comparable scale) of the reduction of the variability of the positions of ink components from approximately 0.012 to 0.005 Rf achieved in the present research. Overall, we observe that analyzing samples of the same ink on different plates is the main source of influence on the variability of the position of the dyes, when compared to the examiner who performs the analyses or the batch of the solvent systems. While the proposed calibration process allows for the controlling of most of this variability, it is not able to reduce it to the low level of variability observed when analyzing several samples of the same ink on the same plate.

6.2. Optimization of solvent system and extraction solvent for the Digital Ink Library

Several different plates from multiple batches of Merck HPTLC plates as well as Durasil and Nanosil plates (Macherey and Nagel, Germany) were used to investigate plate variations. Slight differences were observed during the intra- and inter-plate comparisons, and a decision was made to run all tests (and subsequent ink library analyses) on plates from Merck (batch HX745979). Several different humidity levels were evaluated for plates developed in the ADC 2. Of the levels tested, $36 \pm 4\%$ worked best (maintained using $MgCl_2$). The chamber was left unsaturated.

A total of 21 different individual solvents and solvent mixtures were evaluated as potential extraction solvents. For all of the 60 inks and dyes evaluated, tetrahydrofuran/water 4:1 (v/v) was found to be the best for extracting both ballpoint and non-ballpoint inks. Eleven solvent systems were evaluated (Table 2). The system that achieved the best resolution of dye component bands was a mixture (by volume) of n-butanol/ethanol/water 50:10:15. All solvents were HPLC grade.

A total of 9200 samples were selected for re-analysis using this newly optimized analytical procedure. For each analysis, the ink punches were added to a vial containing 80 μL of a tetrahydrofuran mixture (4:1) and extracted for 15 min at room temperature.

A new dye ladder was developed to be compatible with the selected solvent system. The dye ladder consisted of four components (approximately 176 mg of each dye): crystal violet (Fluka, CI 61135 [42]); rhodamine 6G (Fluka, CI 83698); Metanil Yellow (Fluka, CI 64010); and acid red 52 (Transvase, CI 45100). A 10 mg portion of this mixture was dissolved in 10 mL of tetrahydrofuran/water 4:1 (v/v) and sonicated for 2 min.

Initial testing indicated that there could be considerable variation between plates even if they originate from the same batch. This variation would actually affect the relative elution of some dyes. In other words, dyes would switch places on the developed plate. After contacting and visiting the manufacturer (Merck), it was believed that the differences could originate in small changes in the pH of the silica gel on the plate during the manufacturing process. To detect these variations, a system suitability test (SST) was created by dissolving 4 mg of a particular blue ballpoint ink into 10 mL of tetrahydrofuran/water 4:1 and sonicated for 2 min. When analyzed on an unsuitable plate, the dyes in the SST would change position (Fig. 6) and inform the operator that an error occurred. The samples on the plate would then be reanalyzed on another plate.

Table 2
List of solvent systems evaluated during the implementation project and the approximate time required to process a plate.

Mobile phase	Ratio	Time (h)
1-Butanol:2-propanol:water:acetic acid	20:10:10:1	1:57
1-Butanol:2-propanol:water:acetic acid	20:10:10:1	1:25
1-Butanol:ethanol:water	11:3:3	1:45
Methyl ethyl ketone:2-propanol:25% ammonia	6:4:1	0:45
Ethyl acetate:ethanol:water	26:13:11	0:54
Ethyl acetate:ethanol:water	26:13:11	0:45
1-Butanol:ethanol:water:acetic acid	60:10:20:0.5	>2:00
1-Butanol:ethanol:water:acetic acid	40:10:20:10	>2:00
Ethyl acetate:methanol	2 cm:5 cm	n/a
1-Butanol:ethanol:water (lichosphere plate)	11:3:3	0:29
1-Butanol:ethanol:water	50:10:15	1:10

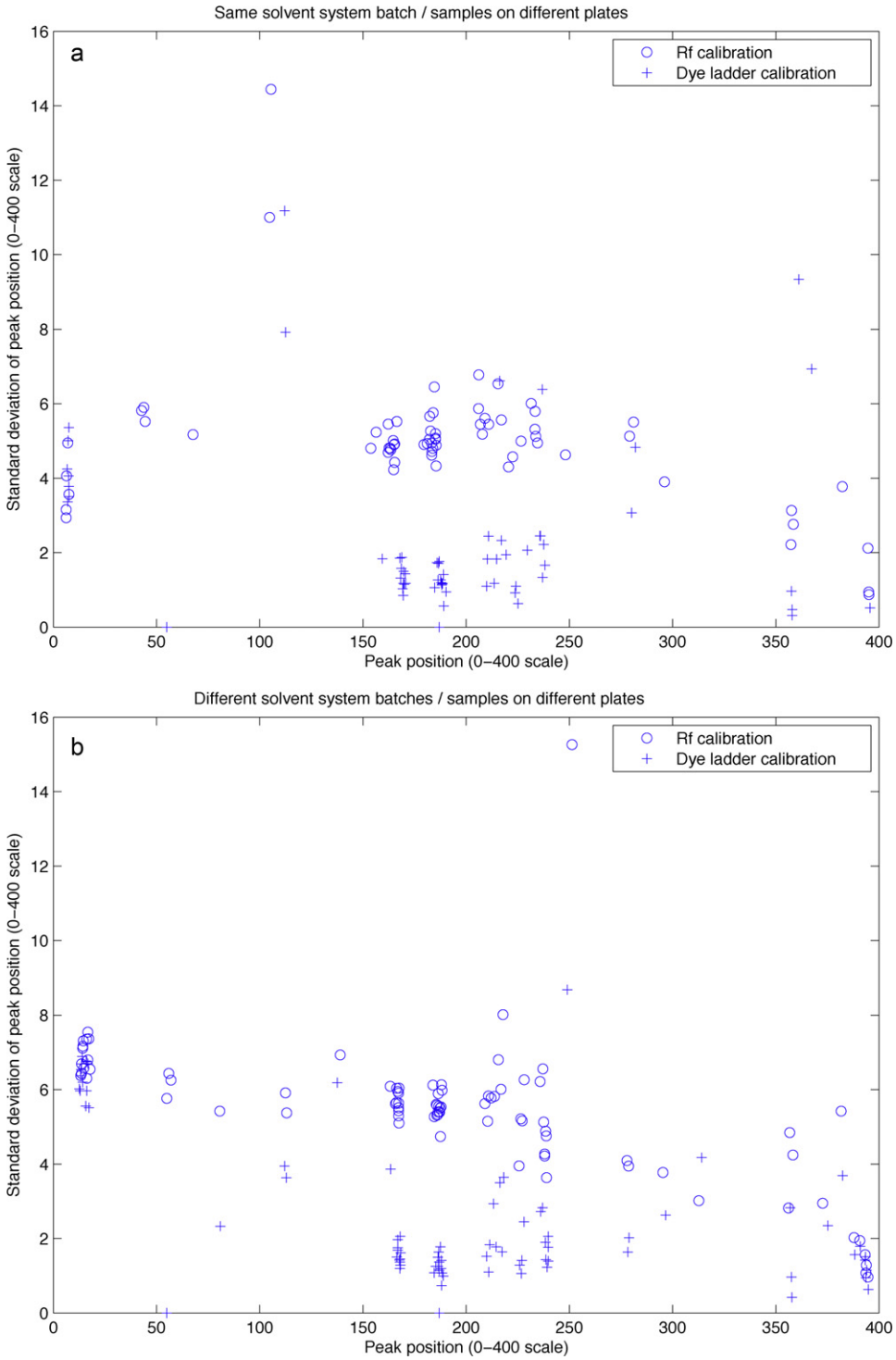

Fig. 5. (a and b) Comparison between the reproducibility of the peak positions obtained with the traditional Rf values versus the dye ladder calibration. The figure on the top presents the reproducibility between different plates when the samples were analyzed with the same batch of solvent systems over several days—the figure on the bottom presents the results when the solvent system was prepared freshly.

The ladder was applied to tracks 1 and 18 and the SST was applied to track 8. Ink samples were applied in three different concentrations (3 μL, 6 μL, and 12 μL) in tracks 2–7 (two samples) and 9–17 (three samples). All 18 samples were applied to the HPTLC plates as 5 mm wide bands. The developing distance was 55 mm, as measured from the bottom of the plate. The plates used for the library were Merck HPTLC silica gel 60 F_{254} (batch HX745979).

6.3. Development, test and implementation of search algorithms in the prototype of the ink library

The digital capture and the calibration process described in Section 6.1 enables the representation of each ink sample by a two-dimensional matrix containing 400 rows (migration distances in the new coordinate system) and 31 columns (wavelengths). The matrix is populated with the absorption intensities (in absorp-

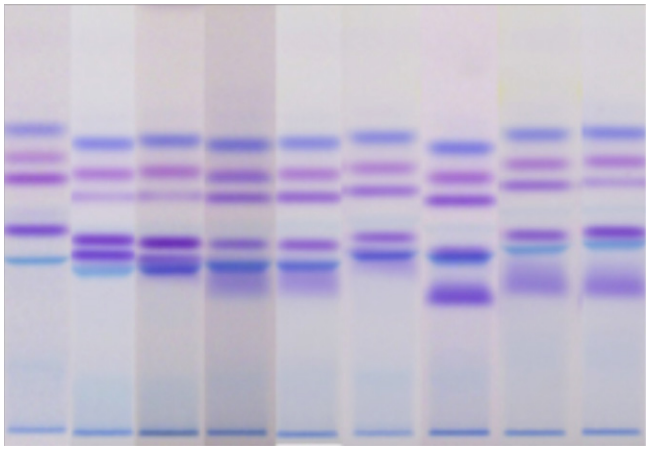

Fig. 6. Changes in the SST dye separations on different plates: the SST is used to detect plate-to-plate variation in dye separation profile.

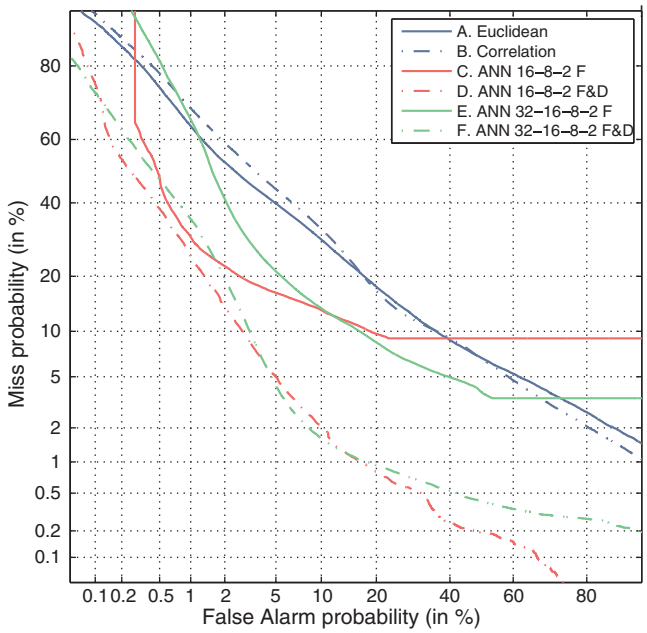

Fig. 7. DET curve presenting the performances of the six algorithms used during the development of the prototype when processing all samples from Table 1. The algorithms are referenced by their letter in Table 3.

tion units (au)) of the sample track at each migration distance and for each wavelength. The matrix captures digitally the analog information perceived by the human eye when observing ink chromatograms (i.e., the position and color of each dye).

The digital capture of chromatograms enables their post-processing (e.g. normalization of the intensity, calibration of the elution distances and removal of background spectra) and their objective comparison. It also allows for the systematic study of the influence of analytical and environmental conditions. Formatting every sample into a standard matrix provides the immense advantage of outputting standard digital files for all samples, independently from the type of ink, the number of ink components detected in the sample or their spectral characteristics. This enables the creation of digital libraries for the identification of the brand/model of questioned ink specimens.

As previously mentioned, the difference between the comparison and the identification processes does not reside in the way the comparison is performed, but in the number of comparisons that are carried out. Hence, the mathematical comparison algorithm designed in this study can serve both purposes.

A comparison algorithm could have been designed based on the facilities provided by the software packages implemented in most current analytical instruments (i.e., detection and identification of each component). The difficulties linked to this approach have already been explained (see Section 4.1). Mimicking the side-by-side comparison of the samples as done by forensic ink examiners could have been chosen (i.e., detection of the presence/absence of each component of the first sample in the composition of the second one and reciprocally). However, when considered in the context of the automation process, it appears clearly that the mathematical comparison of an unknown, potentially different number of – degraded, weakly showing, etc. – components between two samples or specimens requires a complex implementation. Instead, it was decided to take advantage of the standardized format under which the ink samples are recorded after the analytical, acquisition and calibration processes: we designed algorithms, which compare pairs of samples by computing the mathematical differences between their matrices, thus resulting in the computation of a *similarity score* for each pair of samples. This strategy avoids the reference to a collection of known components, and remains free from the difficulties of comparing the characteristics of an unknown and different number of components.

The production of similarity scores is well known to scientists working in the field of biometry (see for example fingerprint matching [53]). A significant body of theory and tools already exists to support the development and the evaluation of the performance of scoring algorithms. During the development of the prototype, we created and tested six scoring algorithms (Table 3): two algorithms based on linear distances and two pairs of non-linear algorithms using artificial neural networks (ANNs). The structure of the ANNs in each pair is the same; however one of the ANNs is trained using ink samples exposed to analytical conditions exclusively and the other one is trained using ink samples exposed to analytical and environmental conditions (see Table 1). While ANNs are commonly used in classification problems, we chose, following Srihari et al. [54], to train our ANNs to recognize between samples that are from the same ink versus samples that are from different inks (bear in mind that the inks in Table 1 were all chosen because

Table 3
List of the algorithms designed and tested during the development of the prototype.

#	Description	Input size (rows × columns)	ANN design (neurones in layer)	Training datasets	Size of training dataset
A	Euclidean distance	400 × 31			
B	Pearson correlation	400 × 31			
C	ANN—distance preANN compiled along migration distances	400 × 1	16–8–2	Fresh ink samples (datasets 1–5)	10′140 vectors
D	ANN—distance preANN compiled along migration distances	400 × 1	16–8–2	Fresh and degraded ink samples (datasets 1–10)	20′000 vectors
E	ANN—distance preANN compiled along wavelengths	31 × 1	32–16–8–2	Fresh ink samples (datasets 1–5)	10′140 vectors
F	ANN—distance preANN compiled along wavelengths	31 × 1	32–16–8–2	Fresh and degraded ink samples (datasets 1–10)	59′186 vectors

they had a different dye profile from each other). Our algorithms produced a score of 0 for two perfectly identical samples (equal matrices), and produced scores greater than 0 for samples that differ to some extent: the greater difference between the samples, the greater the number. The details of the algorithms can be found in [55].

The overall performances of all six algorithms were measured and compared by means of Detection Error Trade-off curves or DET curves [56] (Fig. 7). DET curves are commonly used in biometry in order to assess the performance of matching algorithms [57,58]. DET curves report the relationship between miss probabilities (or false negatives) and false alarm probabilities (or false positives). DET curves are constructed for any given algorithm by measuring the respective miss probability rate for each possible false alarm rate of the algorithm [56]. In other words, in reference to this particular study, a DET curve indicates the risk that the algorithm will miss an association between two ink samples of common origin, for an accepted risk of wrongfully associating ink samples of different origin. To construct the DET curves, multiple pairs of samples of common/different origin were repeatedly compared using each of the proposed algorithms. For each pair of ink samples, the score was computed and related to the expected theoretical outcome (i.e., common/different ink). Due to the limited sample size, the datasets used to train the 4 ANN-based algorithms were included in the datasets used to measure their performances. The datasets used to train the ANNs represent up to 10% of the data from the test datasets (the test datasets are the sets listed in Table 1).

The tests performed on the algorithms and reported in [55] were focusing on the following aspects:

(i) Influence of the calibration process: does the variability between HPTLC plates influence the performance of the algorithms? Can the proposed calibration process based on the dye ladders minimize this influence?
(ii) Comparison of the algorithm's performance: does an algorithm (or a class of algorithms) perform consistently better than others?
(iii) Effects of the different analytical and environmental factors on the algorithms: what is the impact of the various conditions on the performance of the algorithms? Does a particular algorithm (or class of algorithms) show a greater level of robustness to the various conditions?

The results of the tests show that all algorithms benefit from the dye ladder calibration process. Nevertheless, the calibration process does entirely compensate for the variability existing between samples of the same ink analyzed on multiple HPTLC plates: all algorithms show a decrease in performance when comparing these samples versus when comparing samples of the same ink analyzed side-by-side on the same plate. Neural network-based algorithms perform logically better than linear algorithms since they have been trained to take into account between-plate variability.

As a general rule, neural network-based algorithms perform better than the two other ones. The ANN algorithms trained using samples exposed to analytical and environmental factors show lower rates of false positives and false negatives. These algorithms also consistently perform better and show more robustness even when comparing heavily degraded or modified samples, such as the ones exposed to sunlight or extracted from colored paper.

The performances of the algorithms in the specific context of the search of questioned specimens through a reference library were also tested. The performances were measured using precision/recall curves (Fig. 8), which are commonly used to monitor the performances of information retrieval systems [59]. These plots compare the capacity of the system to retrieve only relevant information (precision) to its capacity of retrieving all the relevant information (recall). The precision is then defined as:

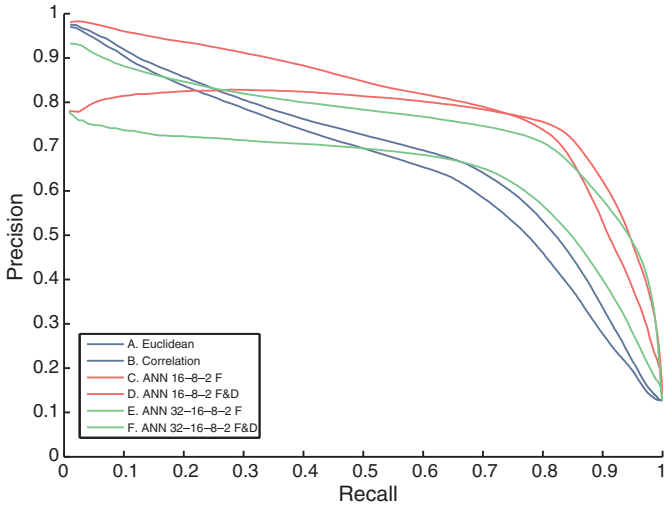

Fig. 8. Precision/recall curve presenting the performances of the six algorithms used during the development of the prototype when processing all samples from Table 1. Their letter in Table 3 references the algorithms.

$$\text{precision} = Pr(\text{relevant}|\text{retrieved}) = \frac{\#(\text{relevant items retrieved})}{\#(\text{retrieved items})}$$

And the recall capability as

$$\text{recall} = Pr(\text{retrieved}|\text{relevant}) = \frac{\#(\text{relevant items retrieved})}{\#(\text{relevant items})}$$

The performances of the algorithms were tested for the same three aspects described earlier in this section. The results also show that all algorithms benefited from the dye ladder calibration and that, on average, neural network-based algorithms perform better than the two other ones. However, linear algorithms were observed to have better precision than the ANN-based algorithms. Similarly, ANN algorithms trained on samples exposed to analytical factors only had a slightly better precision than the ANN algorithms trained on samples exposed to both analytical and environmental factors. We deduced that since ANN-based algorithms are trained to perform better *on average* in a multitude of conditions, they tend to lose their ability to strongly associate ink samples that are nearly identical (e.g. analyzed side-by-side on the same plate). On the contrary, linear algorithms do not perform as well on degraded samples for obvious reasons; however, they are very efficient at recognizing very similar dye profiles.

6.4. Implementation of the scoring algorithms in the Digital Ink Library

Between the development of the scoring algorithms for the prototype and the development of those for the Digital Ink Library, new requirements and challenges were raised. Firstly, each ink sample of the new library was analyzed at three different concentrations: therefore each control ink was represented by at least three samples. Furthermore, the ink library was designed in order to allow for the consolidation of each sample (i.e., the addition of new samples of the same ink to the samples already representing that ink), thus leading to a multiple, but potentially different, number of samples representing each ink.

Secondly, the implementation and training of artificial neural networks proved to be complex: the size (more than 10,000 inks) and structure (multitude of inks with a similar dye profile, three samples per ink) of the ink library rendered impractical the

Table 4
Description of the algorithms and weights used for the implementation phase.

k	Algorithm	Dimension of feature vector	Weight β_k
1	Tri-dimensional representation of ink chromatogram	512×31	10
2	Lightness L^* (in Lab color space)	512	1
3	Red	512	1
4	Green	512	1
5	Blue	512	1

development of a dataset suitable to train an ANN. Furthermore, the solvent system and the number of divisions in the coordinate systems used for the calibration of the samples were changed between the development of the prototype and the final version. This implied that the ANNs developed and trained for the prototype could not be reused.

Given the results presented in Section 6.3, we decided to use a combination of different linear algorithms to compute aggregated scores. In contrast to the algorithms presented above, where similarities were only measured between chromatograms, we also elected to have our algorithms measure the similarities between the images of the tracks.

Five algorithms were developed (Table 4). The first algorithm was designed to compare the tridimensional chromatograms presented in Fig. 4 and corresponds to algorithm A in Table 3. It was chosen based on its performance in the identification process as presented above. The remaining four algorithms were designed to compare the images of the elution tracks and are based on the bitmap information of these images. Note that all algorithms are in fact simple Euclidean distances computed on different feature vectors. The different algorithms were weighed and aggregated to produce a single resulting score.

More precisely, for two ink entries in the library, denoted A and B, there can be any number of samples, denoted $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_m$ (where n and m are typically equal to 3). A comparison between inks A and B with any of the algorithms in Table 4 would then result in nm scores. If k denotes the chosen algorithm in Table 4, the score for that algorithm is then:

$$score_k = \sum_{i=1}^{n}\sum_{j=1}^{m}|A_i - B_j|$$

And the aggregated score is:

$$score_{A,B} = \sum_{k-1}^{5}\beta_k score_k$$

where the coefficients β_k are the weights reported in Table 4 for each k algorithm. These weights were determined empirically and remain to be optimized. A significantly larger weight was assigned to the algorithm comparing the chemical profile of the ink measured with the TLC scanner since the reliability of the absorption spectra (the TLC scanner has internal calibration algorithms) is much higher than the expected reproducibility of the colors of the imaging device used to take the pictures of the ink tracks.

7. Implementation at the United State Secret Service—The Digital Ink Library

In order to manage the ink samples, and to exploit the technologies selected during the first phase of the project to their full efficiency, the Digital Ink Library (DIL) was created. It was implemented following the specifications outlined by the U.S. Secret Service. In short, the DIL is the digital version of the International Ink Library. It includes a content management system, a laboratory management system and a user interface. The content management system stores information on several thousands of inks from various origins and on ink suppliers, ink formulas, contacts, and case files. The laboratory management system records all actions performed on the ink samples, such as the introduction of new questioned and control samples, the creation of case files, and the identification of questioned specimens. The user interface supports all aspects of casework operations: it allows for monitoring the ink examination workflow, piloting the comparison process and managing the questioned and reference samples, the cases and all other relevant information pertaining to the ink examination.

The user interface contains a navigation bar on the left-hand side, which allows users to browse the entire file system of the library, and a main screen where they can perform actions and observe results (Fig. 9). It is designed to allow users to perform the 7 main actions derived from the user requirements for the DIL:

(i) Importation of questioned specimens and reference/control samples.
(ii) Browsing of the library and management of samples.
(iii) Management of cases' and suppliers' information.
(iv) Direct comparison between any samples contained in the DIL.
(v) Searching, identification and confirmation of the source of questioned ink specimens, using the reference samples contained in the DIL.
(vi) Audit of the creation and modification of DIL elements.
(vii) Production of reports for the cases examined using the DIL.

7.1. Importation of ink samples

Since its creation in 1968, the International Ink Library grew from a few hundred reference inks to nearly eleven thousands. Therefore, there is an obvious need to import samples into the DIL and the technologies included in the DIL allow for its exponential expansion. The importation procedure is a fundamental one. Indeed, it involves various calibration and quality assurance procedures, which are crucial to ensure the optimal performance of the DIL, and the tagging of the samples, which allows for their subsequent handling.

The DIL accepts the importation of *reference samples* (for library population and maintenance), *questioned specimens* (to investigate their provenance using the DIL), and of a mixture of questioned specimens and reference samples analyzed on the same plate (to *confirm* the source of the questioned specimens by comparing them directly to reference samples). For each ink sample (questioned or reference), the DIL stores its chemical profile, multiple images of the sample's HPTLC elution tracks and meta-data, such as brand, manufacturer, and years of production. The chemical profiles were measured in absorbance from 200 to 700 nm as described in Section 6.1. The images were taken under different light conditions (UV fluorescence, visible, IR luminescence). The DIL allows for important new data for samples that are already contained in the library; this process enables the consolidation of the samples by expanding the knowledge on their individual variability.

Because of the requirements of the calibration process, all samples and ladders analyzed on a given plate are imported together. The importation process is summarized in Fig. 10.

The calibration process follows the procedure summarized in Section 6.1 and described in [51]. In addition to the calibration of the chromatograms, the images of the plates are also calibrated to ensure the correspondence between the images of the dyes and their position in the chromatograms (Fig. 11).

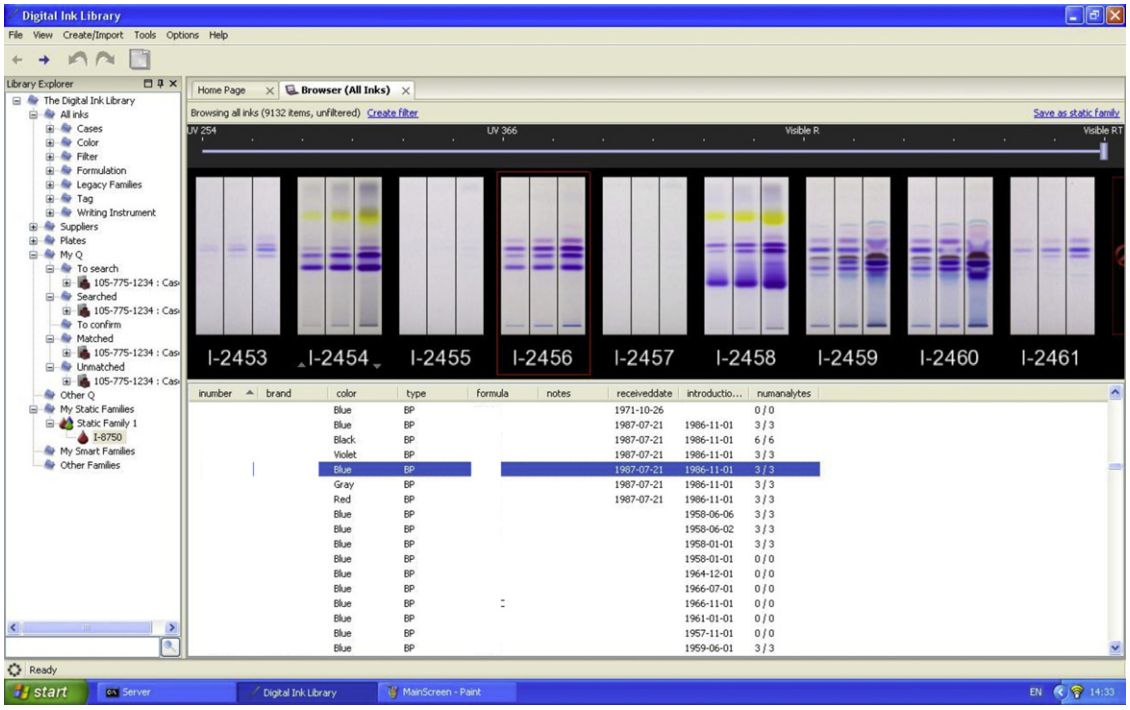

Fig. 9. Layout of the DIL's user interface. The navigation bar can be seen on the left hand side, while the list of the ink samples in the library can be browsed in the main screen (ink manufacturers information have been edited out).

7.2. Browsing of the DIL and sample management

Users can browse the file system of the DIL using the navigation bar on the side of the screen and obtain/modify information on questioned specimens and control samples. The navigation bar provides access to the entire file system of the library, including ink cases. Ink samples are referenced under different categories. A single sample can belong to one or more of these categories. Reference samples are categorized by color, type of writing instrument, chemical tag [3], formulation and manufacturer.

Questioned specimens are referenced by case. A distinction is made between questioned specimens that need to be searched and specimens that have already been searched. A further distinction is also made between questioned specimens, which were imported for investigation and those, which were imported for confirmation purposes (i.e., analyzed side-by-side with control samples).

The interface allows users to search and select particular samples and view their physical and chemical properties. It also permits to observe and compare samples visually. The observation and comparison can be done at various wavelengths providing that images of the samples at those wavelengths were imported into the DIL.

More importantly, the information available for the samples can be viewed and modified. The sample chromatogram and its spectra can be accessed, together with the plate on which it was analyzed, and a sheet presenting the meta-data gathered for that particular ink (Fig. 12).

A search engine, based on keywords, is available in the DIL. Users can search ink samples using any keyword in the suppliers, attached documents, reference inks and families contained in the DIL.

Finally, the interface permits the grouping of the samples in user-defined families.

7.3. Management of cases and other information

One of the primary functions of the DIL is to enable users to create and manage cases. Classic information pertaining to forensic cases is recorded in each case file: submission/analysis dates, submitting office, examiner name, notes on the case, ink samples analyzed in relation to the case, the audit trail of the activities performed in the DIL in relation to the case, and documents or emails that have been collected. The attribution of unique case identifiers to questioned specimens renders possible the archiving of search results and scores for confirmation matches, as well as the maintenance of the chain of custody of all actions performed on the specimens.

The DIL also stores other categories of data, such as contacts, manufacturer's information, ink formulation and tags. In addition to the preformatted fields storing information for each of these categories, the DIL allows for linking them together with any documents (e.g. Adobe PDF, Microsoft Word, emails, images, etc.).

7.4. Direct comparison of any samples contained in the DIL

The DIL undertakes three types of matching. Their main characteristics are presented in Fig. 13.

The comparison of samples is the first type of matching. It is a generic comparison process, which enables the direct algorithmic comparison of any two samples contained in the DIL. The two other types of matching can only be performed within an ink case. These two matching processes are part of the ink examination casework workflow and are described more extensively in the next section.

The comparison process allows for comparing samples outside of the usual casework 'identification' workflow, and can be used for research purposes or to group control samples together. It may also be used to compare two or more questioned specimens with a view to link them together. The comparison of questioned specimens to other questioned specimens can link these specimens within a case, or between cases.

Since this process performs comparisons outside of the regular casework workflow, the matching of two samples is not registered within a particular case. However, the association of two samples is recorded within their individual sheets.

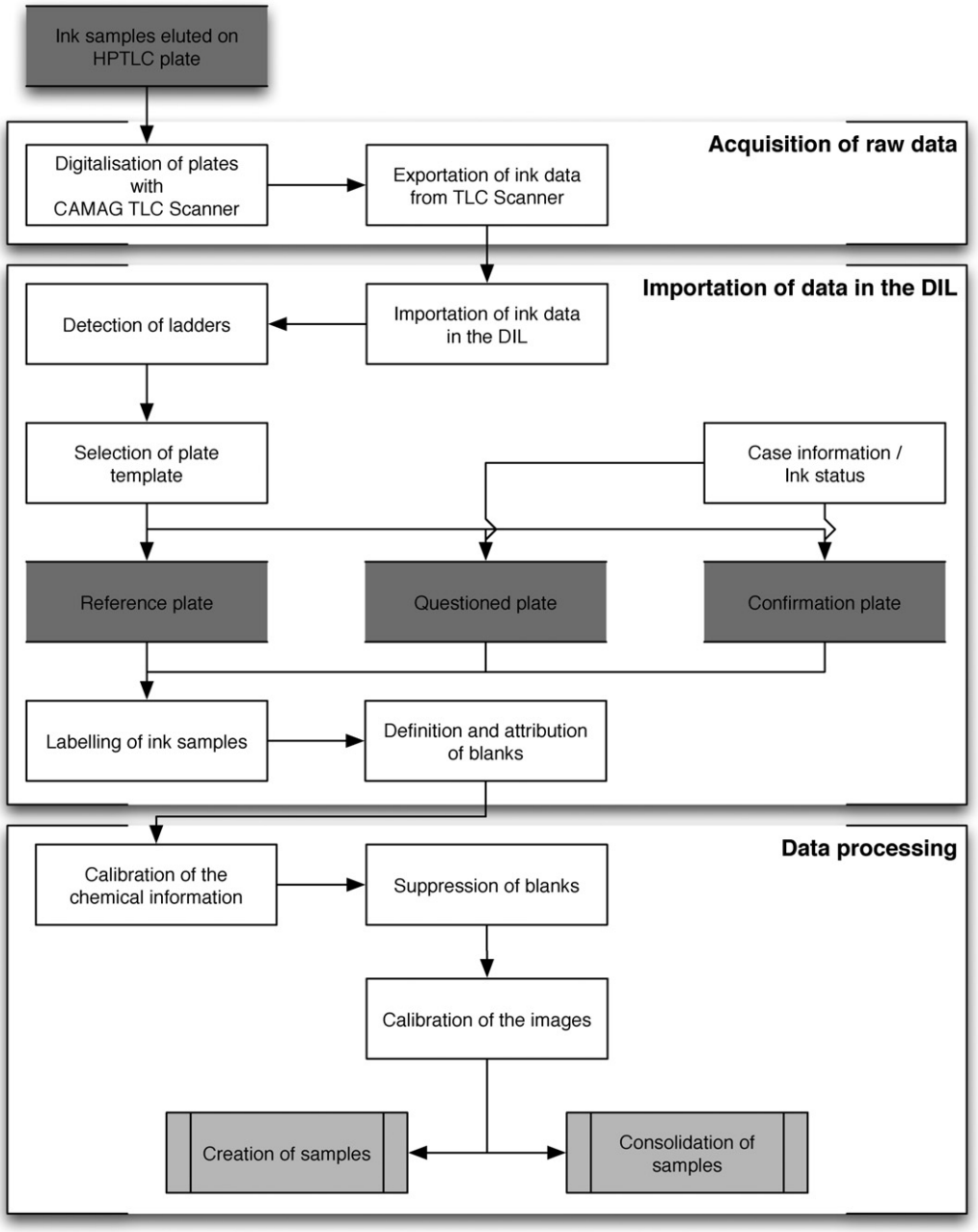

Fig. 10. Flow chart of the importation process.

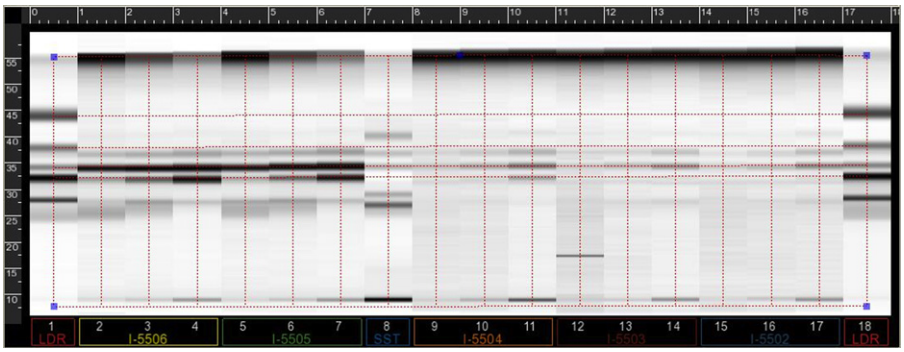

Fig. 11. Plate calibration: the positions of the dyes in the ladder are automatically detected and the frame of the new coordinate system is constructed. The handles in the corners of the grid allow for manually supplementing the automated calibration process.

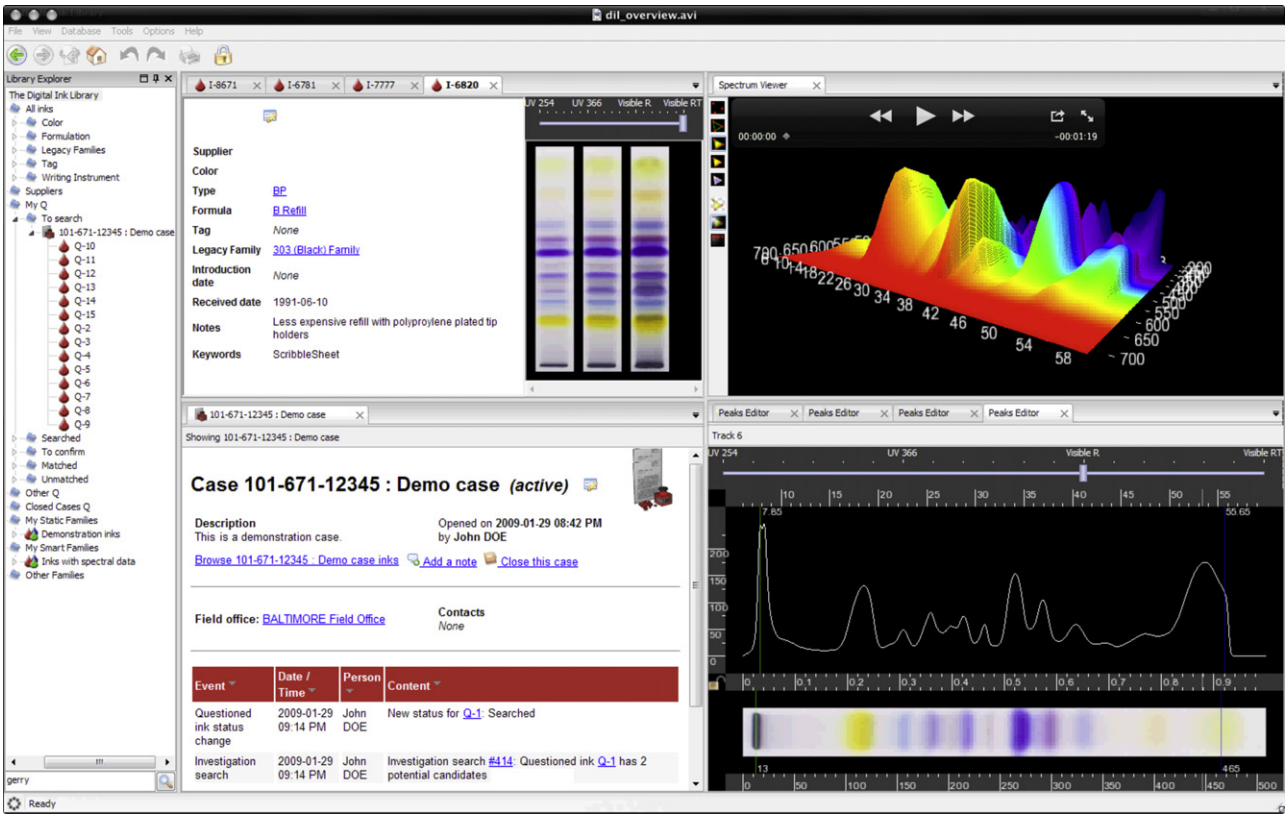

Fig. 12. Information available for an ink sample. Clockwise from the top left: sample sheet, 3D view of chromatogram, 2D view of chromatogram, audit trail.

7.5. *Identification and confirmation of the source of questioned specimens*

The determination of the source of questioned ink specimens is the main purpose and functionality of the DIL. It is carried out in two main stages [32]: the investigation and confirmation of the source of the specimen. The ink examination workflow commonly used by ink examiners in casework was implemented in the DIL to support them through the different steps of the forensic examination of an ink sample. The DIL guides the users at every step by indicating the required actions and reference material needed for the next steps.

The workflow is represented in Fig. 14 and its main steps are presented below:

(i) At first, a case is created or selected.
(ii) Questioned specimens are analyzed on a HPTLC plate and imported in the DIL.

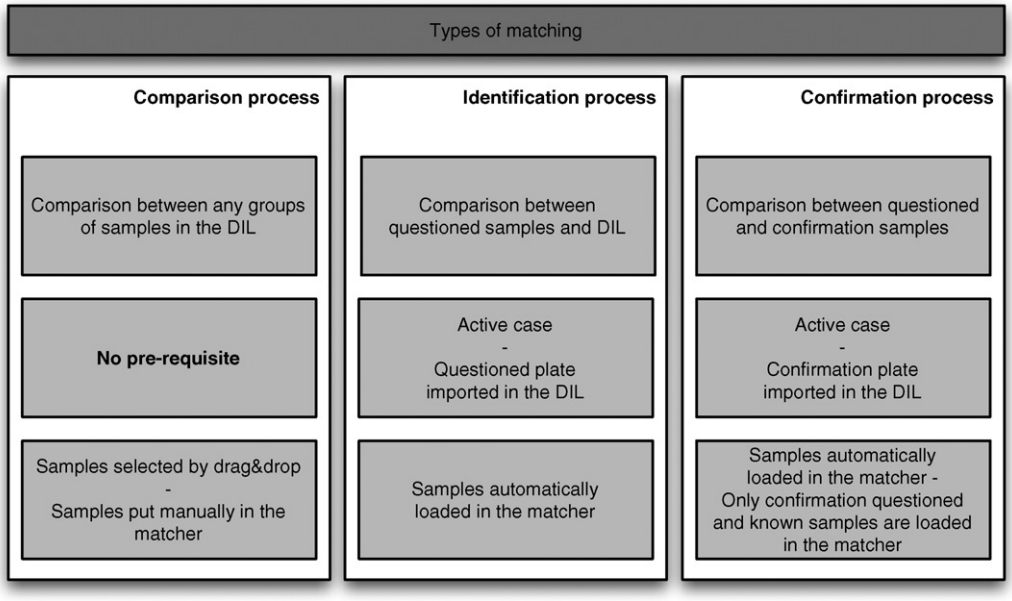

Fig. 13. Different types of matching allowed in the DIL.

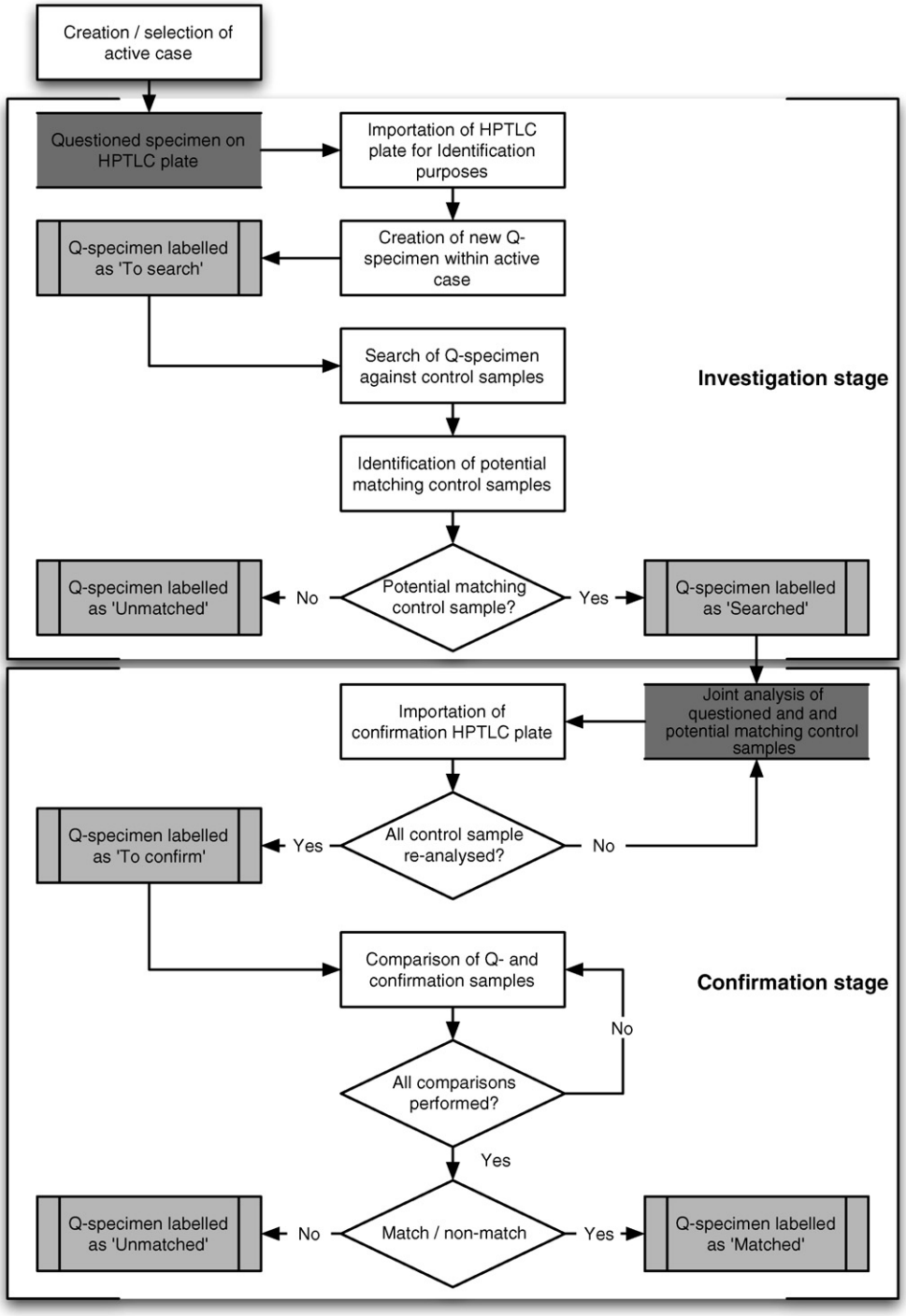

Fig. 14. Flow chart for the identification and verification of the source of a questioned ink specimen.

(iii) The source of each of the questioned specimens is investigated by searching similar control samples in the DIL. This stage is called the *Investigation stage* in the DIL. It allows for searching unknown specimens in the database in order to determine their potential model/brand/manufacturer/year of production.

(iv) Once the search is completed, the list of reference samples is presented to the user, ranked from the most similar to the most dissimilar to the questioned specimen. Users browse through the list of results and visually compare questioned specimens and control samples at multiple wavelengths. Users then select the most similar control samples.

(v) Each questioned specimen and its potentially 'matching' control samples are analyzed side-by-side on a new HPTLC confirmation plate.

(vi) The confirmation plate is imported into the DIL.

(vii) The confirmation samples from the questioned and the control inks are compared to each other using the DIL matching algorithm. This stage is called the *confirmation stage*. Questioned specimens are compared only with reference samples selected during the investigation search, in order to measure the quality of the association between the questioned and reference samples without any interference from the variability observed between plates.

Event	Date / Time	Person	Content
Note	2008-12-27 06:13 PM	John DOE	Example of a note
Searched questioned ink	2008-12-27 06:12 PM	John DOE	Investigation search #14688: Questioned ink Q-2 has 1 potential candidate : I-2459
Note	2008-12-26 10:53 AM	John DOE	twest note
Searched questioned ink	2008-12-26 09:01 AM	John DOE	Investigation search #14579: Questioned ink Q-3 has 1 potential candidate : I-2770
Imported confirmation plate	2008-12-26 08:37 AM	John DOE	Plate 070606-Rack5-Row1 with inks Q-3
Searched questioned ink	2008-12-25 02:41 PM	John DOE	Investigation search #14560: Questioned ink Q-7 has 1 potential candidate : I-8750
Searched questioned ink	2008-12-25 02:38 PM	John DOE	Investigation search #14559: Questioned ink Q-4 was searched but no potential candidates were found.
Added questioned inks	2008-12-24 08:44 AM	John DOE	Plate 070618-Rack13-Row8 with inks Q-1, Q-7, Q-8, Q-9, Q-10
Searched questioned ink	2008-12-20 01:45 PM	John DOE	Investigation search #14449: Questioned ink Q-1 has 1 potential candidate : I-2769
Searched questioned ink	2008-12-20 01:41 PM	John DOE	Investigation search #14448: Questioned ink Q-1 has 1 potential candidate : I-2773
Imported confirmation plate	2008-12-20 01:40 PM	John DOE	Plate 070606-Rack5-Row1 with inks Q-1
Searched questioned ink	2008-12-20 01:37 PM	John DOE	Investigation search #14426: Questioned ink Q-5 has 2 potential candidates : I-2773, I-2772
Searched questioned ink	2008-12-20 01:37 PM	John DOE	Investigation search #14425: Questioned ink Q-3 has 2 potential candidates : I-2771, I-2770
Searched questioned ink	2008-12-20 01:37 PM	John DOE	Investigation search #14424: Questioned ink Q-1 has 2 potential candidates : I-2773, I-2771
Added questioned inks	2008-12-20 01:33 PM	John DOE	Plate 070710-Rack22-Row3 with inks Q-1, Q-2, Q-3, Q-4, Q-5
Opened case	2008-12-20 01:32 PM	John DOE	Case 1

Fig. 15. View of the audit trail of an ink case as automatically recorded by the DIL.

(viii) Once the questioned and confirmation samples have been compared, the questioned specimen is declared 'Matched' or 'Non-matched'.

7.6. Auditing the DIL

The DIL includes an auditing mechanism. The creation and modification of any object in the DIL, such as ink samples, manufacturer, or a case, is logged into an audit file. The log registers the date and time of an activity performed on an object, the user performing the activity and the type of activity. The audit log is accessible to the users through the interface.

A special type of auditing is proposed for ink cases (Fig. 15). In addition to the standard log recording the creation and modification of case information, a complete history of the case is also available to the users. This case history lists:

(i) The plates imported for that case and more specifically the questioned specimens loaded into the DIL.
(ii) The investigation/confirmation searches performed on the questioned specimens.
(iii) The results of these searches and the control samples found to be similar.
(iv) The notes taken by the examiner.

The case history is available from the case sheet and questioned specimens' sample sheets.

7.7. Production of reports

Finally, the DIL is equipped with a report generator, which exports in a printable format the results of forensic investigations performed on questioned ink specimens. The following reports can be produced from the DIL: casework reports, reports on the actions performed on questioned specimens, and reports on the information gathered on control samples.

8. Discussion and scope for future work

The use of the dye ladder and the optimization of the solvent systems provide a mechanism to measure and reduce the analytical variability of the analyses of ink samples by HPTLC. Nevertheless, the variability between plates remains the main source of analytical uncertainty when inferring the commonality of the source of ink samples. Differences in the composition of the plates themselves (even within a same batch of plates) have also been found to influence the analytical results. It is not possible to have control over these elements. Using the described calibration technique, it is possible for different ink examiners to reliably compare ink samples analyzed on different HPTLC plates, and at different locations. This improvement in the reproducibility and reliability of the analyses enables the creation of ink reference libraries for the identification of the source of ink specimens or for the assessment of the weight of ink evidence. More generally, the use of the dye ladders permits the monitoring of the elution of given plates and the assessment of the quality of the analysis. Indeed, the use of adequate controls allows detecting lower quality analyses/plates and redoing them. A further increase in the reproducibility may be achieved by optimizing further the composition of the dye ladders and the number of dyes that are used. The optimal resolution of the new coordinate system used for the calibration process can also be investigated.

The digital acquisition of the samples with the TLC scanner and their standard formatting in matrices is particularly suitable for the development and application of automated comparison algorithms. This is mainly due to the capacity of the acquisition process to fix the size of the data representing each ink sample, independently from the number of dye components, color, intensity, or from the total elution distance of the sample. During the development of the prototype, six algorithms were proposed. These algorithms were designed to investigate the effect of analytical and environmental factors on the comparison of ink samples. These algorithms were also designed to validate some technological choices. These algorithms were not optimized in the first phase of the research. During the implementation of the Digital Ink Library, we were constrained by some user-requirements and we did not fully optimize and validate the chosen algorithm.

Significant improvements are certainly possible in the area of the comparison of samples analyzed by HPTLC. The score is influenced by the variability present between analyses performed on different plates. Artificial neural networks-based comparison algorithms were shown to minimize this problem but were difficult to implement in the final product. A scoring algorithm that would search for the best fit between samples before computing the score could be developed. The signal from the background and the signal from the ink components can also be weighed differently, so that the score would be increased when it measures the difference between dye components, and decreased when it represents the difference between background noises. Finally, the algorithm can be set to take into account the number of components that are being compared: for example, a score computed between inks with three dyes would be weighed differently than a score computed between inks with more dyes.

The scores are also influenced by potential degradations of ink in the questioned specimen due to environmental factors. The use of automated algorithms clearly does not entirely solve the limitations of the current visual side-by-side approach. In fact, current algorithms may have reduced performance when compared to an experienced examiner. The development of heuristic algorithms,

encapsulating this experience needs to be considered. For example, mathematical 'degradation' modeling of the control ink, or mathematical 'reconstruction' of the questioned specimen, may be performed in order to improve the accuracy of the searches.

Nevertheless, the results obtained from this research are promising. The possibility of automatically comparing samples was demonstrated and the performances of the algorithms using industry standard tools were measured. Overall, the performances were found to be very satisfying. As mentioned above, future work will mostly focus on the optimization of the proposed algorithms, or the development of new ones, using the methodology and tools that have been validated here. Changes in the algorithms' parameters, design or training can be directly compared to the level of performance achieved in this study and thus be either validated and implemented or rejected.

The Digital Ink Library is currently undergoing validation testing. Additional refinements and upgrades to the system will be required on a regular basis, as more features and capabilities are needed. Although the analysis time for each plate has increased (due primarily to the change in solvent system and increased time for ink application), the new system has the ability to drastically reduce search times. In addition, the increased control of environmental variables will improve reproducibility. However, larger sample volumes are required for the new analytical protocol. A modified protocol may be required to handle cases where the questioned ink sample is limited. Although there is a significant start-up cost for equipment, the consumables cost is similar. A major advantage of the new system is the fact that the plates have been recorded immediately after analysis and are backed up at regular intervals on a server. There is also a possibility of future exchanges of digital data between different organizations and countries (if the new analytical protocol is followed). A detailed case management system was also included in the DIL to allow for an audit trail and a greater transparency. Multiple cases can be worked at the same time and notes can be recorded. These features are essential for maintaining the laboratory's accreditation under the American Society of Crime Laboratory Directors—Laboratory Accreditation Board International/ISO 17025.

9. Conclusion

The contribution of ink evidence to the criminal and civil justice systems is relatively limited. Some of these limitations are inherent to the evidence type. Ink evidence is not found in cases as frequently as biological, fingerprint or shoemark evidence. Furthermore, the discriminative value of ink evidence is not as high as for these other evidence types. Other limitations arise from the lack of relevant information extracted from ink evidence and/or the inefficient use of this information. These limitations do not appear to have been adequately addressed in the abundant forensic literature on the examination of ink evidence.

A series of technological developments have been proposed for arriving at a standardized method to analyze, record and compare ink samples and to search them in reference libraries. The experiments performed while developing the prototype of the ink library have shown that the method performs well under very stringent conditions. The experiments allowed for gaining a better understanding of the effect of various variables on ink examination. For example, the observation of extensive ink data sets permitted the study of the less well-understood concepts of explainable and genuine differences between ink samples. With the limitations expressed here, mostly in terms of the optimization of the scoring algorithm, it is believed that this research represents an important contribution to the strengthening of the reliability and validity of the examination of ink evidence in forensic science.

It has been described how the method has been further developed in relation to user requirements laid down by the United States Secret Service and implemented into the Digital Ink Library. During this phase of the project, the transition required a move away from the existing paradigm. The substantial benefits in quality assurance, reliability, capability to process significant amount of data, workflow management and interoperability offered by the Digital Ink Library significantly outweigh the challenge of changing the culture in the community. The Digital Ink Library also shows that the development and use of this technology, together with the structured collection of ink data, is not only possible and doable, but also justified economically.

That said, it is important to emphasize that the use of the proposed technology is not only necessary to construct and manage ink reference collections, but is indispensable when comparing ink samples directly: the quality assurance requirements are to be the same in the identification and comparison processes.

More has to be done to maximize the contribution of ink evidence. New analytical techniques are proposed regularly, which allow for better discrimination between different ink compositions and for better sensitivity of weakly concentrated ink components. In addition, new mathematical algorithms for the post-processing of analytical results and for the objective comparison of ink samples can be implemented. Nevertheless, the framework and ideas developed in this research will support the introduction, optimization and validation of these new analytical techniques and algorithms. The proposed framework and the resulting Digital Ink Library demonstrate that significant improvements in the forensic examination of ink are possible and are being pursued by law enforcement agencies.

In the immediate future, the provision and validation of a quantitative method for assessing the weight of ink evidence is needed. Population studies of inks need to be performed in order to gather statistical information on the frequency of the multiple formulations and their market penetration. The population of the library needs to be formalized in order to reflect this statistical distribution. As demonstrated for other evidence types, such as glass or fiber, this will maximize the usefulness of ink evidence and provide a powerful agent for further transparency. The general approach to the quantitative evaluation of evidence by examiners requires a new paradigm that involves the use of additional information. This is missing from current reporting practices. This research enables the implementation of such methods and the development of training tools to support their introduction in practice.

Ultimately, the introduction of transparency at every level of the examination process is the key for the reduction of variability between individuals and organization. It will support the convergence towards a common examination and reporting standard, in line with the recent recommendations of the National Academy of Sciences, and with the efforts pursued in other fields of forensic science.

Acknowledgements

The author would like to acknowledge the contribution of the following individuals: Dr. Antonio Cantu (retired) from the U.S. Secret Service (Forensic Services Division); Profs. Pierre Margot and Christophe Champod from the University of Lausanne, Switzerland; Dr. Eike Reich from CAMAG AG in Switzerland; and Mr. Shane Cullen from the U.S. Department of Homeland Security.

References

[1] U.S. vs. Kalymon, US 6th Circuit, 07-1965, 2008 http://caselaw.findlaw.com/us-6th-circuit/1076469.html (accessed October 14, 2010).

[2] D.M. Ellen, The Scientific Examination of Documents, 2nd ed., Taylor & Francis, London, 1997.
[3] R.L. Brunelle, R.W. Reed, Forensic Examination of Ink and Paper, C.C. Thomas, Springfield, 1984.
[4] C.H. Breedlove, J. Chem. Educ. 66 (1989) 170.
[5] R.L. Brunelle, in: J. Siegel, G. Knupfer, P. Saukko (Eds.), Encyclopedia of Forensic Sciences, Academic Press, San Diego, 2000, p. 591.
[6] ASTM E444-07, Standard Guide for Scope of Work of Forensic Document Examiners, ASTM International, West Conshohocken, 2007.
[7] Q.Y. Kwan, Inference of identity of source, Berkeley, PhD thesis, Department of Forensic Science, University of California, 1977.
[8] A.A. Cantu, Int. J. Forensic Doc. Exam. 1 (1995) 40.
[9] A.A. Cantu, Int. J. Forensic Doc. Exam. 2 (1996) 192.
[10] V.N. Aginsky, J. Forensic Sci. 38 (1993) 1134.
[11] V.N. Aginsky, Int. J. Forensic Doc. Exam. 2 (1996) 179.
[12] V.N. Aginsky, Int. J. Forensic Doc. Exam. 4 (1998) 214.
[13] C. Weyermann, D. Kirsh, C. Vera, B. Spengler, Forensic Sci. Int. 168 (2006) 119.
[14] L. Brazeau, M. Gaudreau, J. Forensic Sci. 52 (2007) 209.
[15] D.N. Carvalho, Forty Centuries of Ink, Banks Law Pub., New York, 1904.
[16] A.S. Osborn, Questioned Documents, 2nd ed., Boyd Printing Company, New-York, 1929.
[17] E. Locard, Traité de criminalistique (Tome V et VI), Lyon, Desvigne, 1933.
[18] O. Hilton, Scientific Examination of Questioned Documents, Callaghan, Chicago, 1956.
[19] D.A. Crown, J.V.P. Conway, P.L. Kirk, J. Criminal Law Criminol. Police Sci. 52 (1961) 338.
[20] J.W. Brackett, L.W. Bradford, J. Criminal Law, J. Criminal Law Criminol. Police Sci. 43 (1952) 530.
[21] B.B. Coldwell, Analyst 80 (1955) 68.
[22] D. Doud, J. Forensic Sci. 3 (1958) 486.
[23] G.R. Nakamura, C.S. Shimoda, J. Criminal Law, J. Criminal Law Criminol. Police Sci. 56 (1965) 113.
[24] I.R. Tebbett, Forensic Sci. Rev. 3 (1991) 71.
[25] J.A. Zlotnick, F.P. Smith, J. Chromatogr. B 733 (1999) 265.
[26] C. Neumann, P. Margot, Rev. Int. Criminol. Police Tech. 56 (2003) 341.
[27] R.W. Jones, R.B. Cody, J.F. McClelland, J. Forensic Sci. 51 (2006) 915.
[28] J. Buegler, H. Buchner, A. Dallmayer, J. Forensic Sci. 50 (2005) 1209.
[29] C. Weyermann, R. Marquis, W. Mazzella, B. Spengler, J. Forensic Sci. 52 (2007) 216.
[30] Frye v. United States, 293 F. 1013, D.C., 1923.
[31] U.S. v. Bruno, 333 F.Supp. 570, E.D. PA, 1971.
[32] ASTM E1789-04, Standard Guide for Writing Ink Identification, ASTM International, West Conshohocken, 2004.
[33] ASTM E1422-05, Standard Guide for Test Methods for Forensic Writing Ink Comparison, ASTM International, West Conshohocken, 2005.
[34] National Research Council of the National Academies, Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington, DC, 2009.
[35] C. Neumann, P. Margot, J. Forensic Sci. 55 (2010) 1304.
[36] Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579, 1993.
[37] M. Saks, J.J. Koehler, Science 309 (2005) 893.
[38] M. Saks, J.J. Koehler, Vand. Law Rev. 61 (2008) 199.
[39] C. Neumann, P. Margot, Forensic Sci. Int. 192 (2009) 29.
[40] L.A. Foreman, C. Champod, I.W. Evett, J.A. Lambert, S. Pope, Int. Stat. Rev. 71 (2003) 473.
[41] R.S. Ramotowski, E.M. Regen, J. Forensic Sci. 52 (2007) 604.
[42] SDC Colour Index, Society of Dyers and Colourists, 2001.
[43] C. Neumann, New Perspectives in the Use of Ink Evidence in Forensic Science, PhD thesis, University of Lausanne, Switzerland, 2008.
[44] C. Neumann, CAMAG Inc., C. Champod, The International Ink Library of the United States Secret Service: A New and Efficient Way of Managing the Data, DHSARPA, Contract HSHQDC-06-R-00066, 2006.
[45] J.D. Kelly, A.A. Cantu, J. Assoc. Off. Anal. Chem. 58 (1975) 122.
[46] J.A. Lewis, J. Forensic Sci. 41 (1996) 874.
[47] Manuel Suisse des denrees alimentaires colorants pour denrees alimentaires, Chap. 42A, Office central de la sante publique, Bern, 1992.
[48] P. Gill, R. Sparkes, C. Kimpton, Forensic Sci. Int. 89 (1997) 185.
[49] S. Watson, R. Allsop, L. Foreman, Z. Kelsey, P. Gill, Forensic Sci. Int. 115 (2001) 207.
[50] Dean's Analytical Chemistry Handbook, 2nd ed., McGraw-Hill Professional, New-York, 2004.
[51] C. Neumann, P. Margot, Forensic Sci. Int. 185 (2009) 29.
[52] A.H. Stead, R. Gill, T. Wright, J.P. Gibbs, A.C. Moffat, Analyst 107 (1982) 1106.
[53] P. Komarinski, Automated Fingerprint Identification Systems (AFIS), Academic Press, 2004.
[54] S.N. Srihari, S-H. Cha, H. Arora, S. Lee, J. Forensic Sci. 47 (2002) 856.
[55] C. Neumann, P. Margot, Forensic Sci. Int. 185 (2009) 38.
[56] A.J. Mansfield, J.L. Wayman, Best Practices in Testing and Reporting Performance of Biometric Devices v2.0.1, NPL report CMSC 14/02, Center for Mathematics and Scientific Computing, UK National Physical Laboratory, 2002.
[57] J. Gonzalez-Rodriguez, A. Drygajlo, D. Ramos-Castro, M. Garcia-Gomar, J. Ortega-Garcia, Forensic Sci. Int. 155 (2005) 126.
[58] J. Gonzalez-Rodriguez, A. Drygajlo, D. Ramos-Castro, M. Garcia-Gomar, J. Ortega-Garcia, Comput. Speech Lang. 20 (2006) 331.
[59] C.D. Manning, P. Raghavan, H. Shuetze, Introduction to Information Retrieval, Cambridge University Press, 2008, http://nlp.stanford.edu/IR-book/information-retrieval-book.html (accessed October 14, 2010).