The transfer and persistence of environmental trace indicators, and methods for digital data acquisition from photographs and micrographs: applications for forensic science research

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at University College London

Emma Levin

2020

Declaration of Originality

I, Emma Levin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been clearly identified and acknowledged. No part of this thesis contains material previously submitted to the examiners of this or any other university, or any material previously submitted for any other examination.

Signed: .....

2020

Acknowledgements

Although my name is on the title page, this project would not have been possible without a large number of people. Primarily, I need to thank my supervisors, Professors Ruth Morgan and Viv Jones, for their patient assistance over the years. Thanks are also due to the collaborators who have offered their expertise and guidance to the project; I am indebted to Lewis Griffin for his advice on image processing, Michael Kloster for his generous assistance with SHERPA and BIIGLE, Joni Levinkind for his assistance with writing macros, and Kirstie Scott for formative discussions about the field. Massive amounts of gratitude are due to the Geography lab team who provided patient advice, guidance, and resources (especially Janet Hope, Ian Patmore, Kevin Rowe, and Bonnie Atkinson), and the admin team at the Department of Security and Crime Science (especially Amy Clemens and Andrea Kezer) for tolerantly processing all of the expense forms for lab consumables. I would also like to thank my peers at the Department of Security and Crime Science (especially Anna Mattalia, Oli Hutt, Mark Amaral, Sian Smith, Catherine Kerr-Smith, Lauren Horsfall, Nissy Sombatruang, Bev Nutter, Cerys Bradley, and Maryam Hamirani) for making the PhD a genuinely enjoyable endeavour. This research was funded by the Engineering and Physical Sciences Research Council of the United Kingdom through the Security Science Doctoral Training Research Centre (UCL SECReT) based at University College London (EP/G037264/1).

Abstract

Environmental forms of trace evidence (such as mineral grains, pollen grains, algae, and sediment) can offer valuable insights within forensic casework. An issue facing forensic science as a whole, and these environmental indicators specifically, is a relative dearth of empirical research which would underpin the interpretation of such indicators when attempting forensic reconstruction. This thesis aims to address this lacuna, undertaking experiments to: (1) Explore variables which affect the rates of transfer and persistence, with specific focus upon quartz grains (a terrestrial indicator) and diatom valves (an aquatic indicator) upon footwear materials (a substrate that has been under-represented in past studies); (2) Conduct research into the effects of particle size and morphology upon transfer and persistence; (3) Develop and adapt methodologies to undertake this research. Accordingly, the outputs of this thesis are: (1) The creation of new datasets which could inform the interpretation of these trace indicators within forensic investigations and crime reconstruction scenarios and (2) The development of novel methodologies which could be employed in future research to attempt to accelerate data collection and analysis, without compromising on accuracy. This research is interdisciplinary, combining theory from forensic science, analytical techniques from the environmental sciences, and some elements of image processing and analysis.

Impact statement

This thesis has two outputs; datasets relevant to the transfer and persistence of environmental trace evidence, and techniques and methodologies for the generation and analysis of data in such studies. Accordingly, it is hoped that there would be two audiences for the findings of these studies; those within academia, who might undertake further research in the field, and those outside of academia, in the world of forensic casework, who might be able to use the findings of these experiments to underpin inferences made in their own work.

Efforts have been made to ensure that other researchers could easily implement the methodologies described within this thesis. The software used to implement methodologies is either available under a freeware licence (for example, NIH ImageJ), or similar freeware options are available (for example, many digital image processing programmes can perform the same functions as Adobe Photoshop), so cost should not be a barrier to implementation. Similarly, efforts have been made to disseminate findings that would assist casework in an accessible manner; through Open Access publishing in relevant journals (see: Levin *et al.,* 2017; 2019).

Table of Contents

Front Matter

0.1 Front matter

Chapter One

An introduction to the PhD project: its rationale, focus, and structure.

1.1 Introduction

Chapter Two

Contextual literature review

Chapter Three

Quantifying the amount of fluorescent powder in a digital image: a comparison of manual and automated methods of segmentation

Chapter Four

Quantifying the number of quartz grains present in a digital image: Can mathematical image processing be used to locate and count grains?

Chapter Five

Do variations in the activity of a wearer, and the position of sampling, appear to affect the persistence of particulate trace evidence on footwear uppers?

Chapter Six

Does the morphology of quartz grains affect their transfer and/or persistence?

Chapter Seven

Does the morphology of diatom valves affect their transfer and/or persistence?

Chapter Eight

Synoptic discussion, synthesising themes which cut across the chapters

Chapter Nine

Summary of themes, findings, and implications

References

Appendix (SD)

Appendix, containing raw data, statistical analysis of the data (without aggregation), macros, and publications.

List of figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Chapter Seven

List of tables

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Chapter Seven

Chapter Nine

Colophon

This thesis is typeset in Computer Modern Serif,

with Bahnschrift Light used for tables and figures.

1

Introduction

1.1.1 The rationale for this PhD

In the United Kingdom forensic science is, in the words of the Forensic Science Regulator, "lurching from crisis to crisis" (Forensic Science Regulator, 2019, p. 3). In 2019, forensic service providers entered administration[1], deep concerns were raised over the adequacy and sustainability of service provision[2], and a House of Lords Science and Technology Committee inquiry took place into the delivery, quality, and sustainability of both forensic practice and research[3].

The current consensus is that forensic science is facing complex problems, at multiple scales, and with a number of serious consequences for the justice system (e.g. National Academy of Sciences, 2009; Roux *et al.,* 2015; House of Lords, 2019). One key problem is that, over the years, the budgets available to undertake forensic science research have contracted, with the closure of the UK's *Forensic Science Service* (Gonzalez-Rodriguez and Baron, 2018), privatised service providers failing to pursue research and development (ibid.), and the de-prioritisation of forensic science from the Research Councils' remits (Evison, 2018). At the same time, it has been agreed that empirical research is absolutely necessary, to provide foundations for the interpretation of traces within casework (Morgan, 2017; Forensic Science Regulator, 2018), and that for many forensic techniques, the evidence base which is necessary to make informed, nuanced inferences within casework does not currently exist (National Academy of Sciences, 2009; House of Lords, 2019).

This PhD project was a direct response to the Forensic Science Regulator's annual report(s). In each report, the Regulator states priorities for research, and, to quote from the most recent (2017-2018): "From the perspective of improving quality, the priorities identified by the Regulator [include] To provide data and robust interpretation methods to support the effective evaluation of evidential significance. Such data may include, for example [...] structured studies on the transfer and

[1] https://www.thetimes.co.uk/article/police-foot-the-bill-after-collapse-of-forensics-firm-key-forensic-services-limited-bg5nbxkxt
[2] https://www.theguardian.com/science/2019/may/01/forensic-science-labs-are-on-the-brink-of-collapse-warns-report
[3] https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/33302.htm

persistence of trace evidence and the significant factors affecting such transfer" (Forensic Science Regulator, 2019, p. 42).

This PhD therefore presents a series of structured studies into the transfer and persistence of environmental particulate trace evidence, with the aim, as stated above, to "support the effective evaluation of evidential significance" (Forensic Science Regulator, 2019, p. 42).

1.1.2 The focus of this PhD

This PhD was focused upon generating datasets and methodologies for the investigation of environmental or "geoforensic" forms of particulate trace evidence, such as mineral grains and algae (see: Morgan and Bull, 2007; Ruffell and McKinley, 2008). These environmental indicators, which can be widely-distributed within environments (Bull and Morgan, 2006; Jones, 2007), have considerable heritage in associating (or excluding) suspects, victims, and crime scenes (e.g. Ruffell and McKinley, 2005; Bull *et al.,* 2006; Morgan and Bull, 2007). The experiments within this PhD can be broadly divided into three categories:

(1) Exploring variables which affect the rates of transfer and persistence
Conducting experiments to establish whether their influence is observable upon particulate evidence transferred to footwear uppers (Ch. 5).

(2) Exploring the effects of particle morphology upon rates of transfer and persistence
Conducting experiments to establish whether the size and shape of particulates affects their behaviour in observable ways (Ch. 6 and 7).

(3) Developing and adapting methodologies to undertake this research
Conducting experiments to evaluate the performance and potential of existing and novel methodologies, borrowing from image analysis (Ch. 3 and 4).

1.1.3 (Inter)Disciplinary context

This PhD project is interdisciplinary; it adapts techniques from the field of geography (for example, the extraction and analysis of algae and mineral grains), to explore questions relevant to the field of crime reconstruction (for example, how long those traces might persist in a forensically-relevant scenario), in a manner that implements basic developments from the fields of computer science and image processing so as to accelerate the processes of data acquisition and analysis. To these ends, a number of collaborations were pursued:

(1) Professor Lewis Griffin, of the *Department of Computer Science* at University College London, provided invaluable advice and critique with respect to image processing techniques (Griffin *et al.,* 2009; 2012).

(2) Doctor Michael Kloster, of the *Alfred Wegener Institut* in Bremerhaven, Germany, had developed a piece of software capable of segmenting brightfield images of diatoms, and calculating detailed morphometric statistics from those segmentations (Kloster *et al.,* 2014; 2017). This project is heavily indebted to his permission to use the software, his assistance in adapting it for the (slightly unconventional) analysis for which it was used, and his help with arranging access to an online annotation tool, BIIGLE (see: Langenkämper *et al.,* 2017).

(3) Additional assistance, with the writing of macros for the automation of data analysis, was provided by Joni Levinkind, who at the time was based at the *Department of Computer Science* at York University.

1.1.4 Thesis Structure

This thesis is divided into nine chapters, two of which provide preliminary context (1-2), five that outline experiments and the findings drawn from these primary data (3-7), and then the final two, which discuss the ramifications of these findings (8-9) (see Figure 1.1 below).

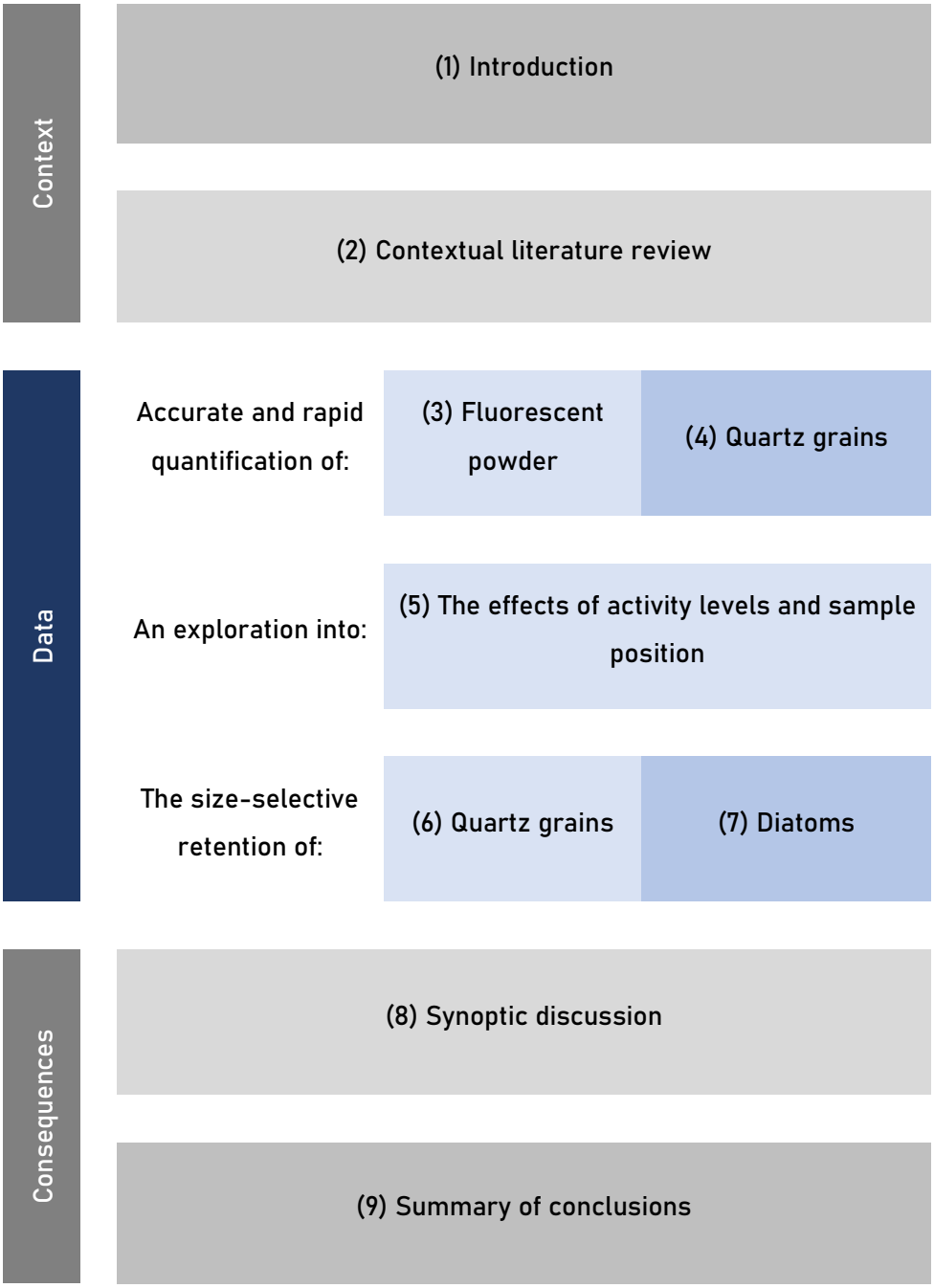


Figure 1.1: A diagram depicting the structure of this thesis

Part One: Situating the research

Chapter One: Introduction

Chapter 1, this chapter, provides an introduction to the project, situating it within the broad, disciplinary context (in which severe and protracted under-funding has left forensic science without a co-ordinated plan for research and development (Forensic Science Regulator, 2018; House of Lords, 2019), and in which one of the highest priorities for research is the field of transfer and persistence, studying trace evidence dynamics to provide the data necessary for the robust and nuanced interpretation of evidence (ibid)). The chapter summarises the foci of experiments, the (inter)disciplinary nature of the work, and provides a list of publications and presentations where elements of the research have been disseminated.

Chapter Two: Contextual literature review

Chapter 2, the Contextual literature review delves further into the context surrounding this research, exploring: (1) Forensic science's heritage as a synthetic discipline, and whether this might be its greatest strength or an often overlooked weakness; (2) What is currently known about the processes of transfer and persistence, which are central to the discipline (but still require much investigation); (3) The sub-discipline of environmental trace evidence, which utilises traces derived from the environment, and more commonly associated with the disciplines of Geography or Ecology than Forensic Science; (4) Specifically, what is known about the transfer and persistence of diatoms and quartz grains, two very common environmental indicators; and, finally, (5) The technological context of this research, undertaken between 2016 and 2019, in that open-source image processing and analysis software were freely available, and represented an immensely valuable resource for researchers.

Part Two: Empirical Research

The first two data chapters (Chapters 3 and 4) consider methodological development. Specifically, the aim was to develop ways of accurately, and rapidly, quantifying either the amount of fluorescent powder (Chapter 3) or the number of quartz grains (chapter 4) present in a digital image. This was necessary to enable later Chapters (e.g. 5)

Chapter Three: Quantifying the amount of fluorescent powder in a digital image: a comparison of manual and automated methods of segmentation

Fluorescent powder is often used as a proxy for particulate trace evidence within studies of transfer and persistence, where it is sampled by macro-photography (e.g. Morgan *et al.,* 2010; French *et al.,* 2012). In order to extract quantitative data from the digital images of the fluorescent material, they must first be 'segmented' – the images must be divided into foreground (the pixels of interest) and background (the pixels that do not represent fluorescent matter) (Sezgin and Sankur, 2004). At present, there is no consensus or convention in terms of the methodology used to segment this sort of fluorescence imagery. In Chapter 3, three experiments were undertaken, to assess the methodological question: 'how might one best segment digital images of fluorescent powder?' Experiment one assessed the extent to which manual segmentation (tracing) was reproducible, finding that observable differences existed between different examiners' estimates of the extent of the foreground (which, when tested with inferential statistics, resulted in $p<0.05$). Experiment two assessed the performance of segmentation by manually-defining a global threshold level, finding that while it would be feasible on the basis of time, it also had problematic reproducibility. The third experiment tested the performance of a suite of global thresholding algorithms, finding that several of the algorithms (Yen, MaxEntropy, and RenyiEntropy) offered comparable levels of accuracy to the manual methods, but at far faster speeds, and with no issues around reproducibility. This chapter was supported by Professor Lewis Griffin, of UCL's Computer Science Department, whose advice (and critique) helped to shape and refine the research, and Joni Levinkind, who assisted with writing macros for ImageJ. Findings from this research were published in the *Journal of Forensic Sciences*, entitled: "A comparison of segmentation methods for

ultraviolet fluorescence imagery: implications for transfer and persistence trace evidence reconstruction studies" (Levin *et al.,* 2019).

Chapter Four: Quantifying the number of quartz grains present in a digital image: Can mathematical image processing be used to locate and count grains? Quartz grains represent a valuable environmental trace indicator – widely distributed in terrestrial environments, and seen to transfer readily between environments and individuals within scenarios of forensic interest. However, many experiments which examine the transfer and persistence of quartz grains involve a tedious and time-consuming method of manually counting the number of quartz grains present in an image. Since quartz grains imaged upon a dark background would represent maxima in the image (i.e. regions where there is a peak in the brightness values), it was thought that it might be possible to use a quantitative image processing function ('find maxima') to locate and count quartz grains within digital images. Accordingly, in Chapter 4 an experiment was undertaken to evaluate the accuracy and reliability of using the "find maxima" function to locate and count quartz grains. Three findings emerged: (1) Using the find maxima function (with an appropriate noise filter value) resulted in a count that was within the range identified by a human examiner (2) Different images required different noise filter values to achieve accurate representations of the image (3) Different examiners using the find maxima to count quartz grains arrived at answers that were not observably different from each other (i.e. when tested with inferential statistics, the p-value produced was >0.05). Accordingly, it was concluded that the find maxima function may be used to accelerate data analysis without compromising on accuracy or reliability. These findings highlight the importance of interdisciplinary research, and the potential benefits that insights from cognate disciplines can offer for improving workflows – for a discipline that has historically been so dependent on microscopic analyses and visual comparison, image processing tools can form a useful tool.

Chapter Five: Do minor variations in the activity of a wearer, and the position of sampling, affect the persistence of particulate trace evidence on footwear uppers?

There is a gap in the empirical evidence base for forensic science, even for concepts as central to the discipline as transfer and persistence. Techniques for digital image processing and analysis, as outlined in Chapters 3 and 4, mean that it is now possible to accelerate the analysis associated with such studies of trace evidence dynamics. Accordingly, in Chapter 5 a series of experiments were carried out in order to investigate whether (1) minor variations in the level of activity (0, 5, 10, or 15 minutes walking per hour) and (2) variation in the position of samples on a shoe (whether they were attached to the heel, inner, outer, or toe) affected the levels of persistence for quartz grains and fluorescent powder, over 8 hours. For quartz grains, no statistically significant differences in the amount of grains retained were observed between activity levels or between sampling positions. For fluorescent powder, statistically significantly differences were not consistently observed. These findings have methodological implications for future studies of transfer and persistence, suggesting that sampling multiple locations on one item of footwear over time is a valid proxy for sampling one location repeatedly. These experiments also demonstrate the importance of undertaking such empirical work – it may seem counterintuitive that, after 8 hours, there is no statistically significant difference between the percentage of particles recovered after walking for 5 minutes per hour or 15 minutes per hour, but that is what this dataset indicates. For nuanced and robust interpretation of forensic evidence within a casework scenario, it is not sufficient to rely on what seems intuitive – it is necessary to test hypotheses.

Chapter Six: The size-selective retention of quartz

Within forensic science, experiments with glass fragments and fibres have demonstrated the propensity for larger traces to persist over shorter timescales. The aim of Chapter 6 was to conduct an experiment to establish whether a similar size-selective retention could be observed with respect to quartz grains, asking: does the size (as measured by cross-sectional area) and morphology (as indicated by the aspect ratio and solidity) of a quartz

grain appear to influence the likelihood of transfer or the likelihood of persistence? Findings were threefold: (1) The largest particles in the population did not seem to transfer (2) The largest particles that did transfer were lost first and (3) The morphological characteristics of solidity and circularity also appeared to vary over time. Together, this research confirmed that morphological characteristics of a particle may be an important consideration when interpreting the presence or absence of quartz grains within a forensic scenario.

Chapter Seven: The size-selective retention of diatoms

Chapter 7 complements Chapter 6, by looking at the problem domain of diatoms rather than quartz grains. Diatoms, a type of unicellular algae, have proven to be valuable trace evidence when attempting forensic reconstruction of crimes committed in aquatic environments (e.g. Scott *et al.,* 2019). While their ecology and biological dynamics are the subject of extensive and ongoing research, relatively few researchers have considered their dynamics as trace evidence. In earlier work it has been identified that the morphological characteristics of diatoms may influence their transfer to clothing fabrics (Scott *et al.,* 2019.) This Chapter sought to investigate this hypothesis, using novel methods of semi-automated diatom analysis, and asking whether the (1) size (measured by Cross Sectional Area), (2) elongation (measured by Width-Height Ratio) and (3) roughness (measured by contour smoothness) of the diatom valves appeared to affect the likelihood of transfer and/or persistence. The work presented in this chapter was greatly assisted by Michael Kloster, who supported the use of automated morphometric software for the morphometric analysis of diatoms, SHERPA (Kloster *et al.,* 2014; 2017). Three findings were established: In this study, (1) All size fractions of diatom found in the population were seen to transfer readily (2) The morphology (i.e. aspect ratio and convexity versus concavity) of the diatom particles seemed to affect their transfer; (3) Over time, the mean values for the morphological metrics changed, suggesting that the size and shape of a particle may indeed affect their likelihood of persisting. Inferential statistical analysis confirmed that these trends may be worth further investigation. Together, these findings reinforce the idea that more empirical

research is needed, in order to understand these trace evidence dynamics and provide a robust underpinning for inferences within forensic casework.

Part Three: Synoptic discussion and Conclusions

Chapter Eight: Synoptic discussion

In Chapter 8, the discussion points from chapters 3, 4, 5, 6, and 7 were revisited, considering the implications of the findings of these experiments for the way that we conduct research and use data. Three main themes emerged; (1) Primarily, that it is necessary to conduct further empirical research into the transfer and persistence of trace evidence (2) Secondly, that borrowing automated techniques from other disciplines can accelerate data acquisition and analysis without sacrificing accuracy; and (3) That many avenues for research in this field remain.

Chapter Nine: Summary of conclusions

Chapter 9 presents the key findings of the thesis as they relate to the key research questions and provides the implications of the thesis for crime reconstruction approaches and forensic science.

1.5 Dissemination

Elements of the research that constitutes this thesis have been presented at national and international conferences, and published as research articles in journals. The following list summarises the most salient, in chronological order:

Published Papers:

- September 2017: Scientific and Technical Article
  Levin, E. A., Morgan, R. M., Scott, K. R., & Jones, V. J. (2017). The transfer of diatoms from freshwater to footwear materials: An experimental study assessing transfer, persistence, and extraction methods for forensic reconstruction. *Science & Justice*, 57(5), 349-360. This paper relates to the material seen in Chapter 7.

- March 2019: Scientific and Technical Article
  Levin, E. A., Morgan, R. M., Griffin, L. D., & Jones, V. J. (2019). A Comparison of Thresholding Methods for Forensic Reconstruction Studies Using Fluorescent Powder Proxies for Trace Materials. *Journal of Forensic Sciences*, *64*(2), 431-442. This paper relates to material seen in Chapter 3.

Papers in progress:

- Scientific and Technical Article
  Morgan, R. M., & Levin, E. A. (drafted, with the intention to submit to *Forensic Science International, Synergy*) "A crisis for the future of forensic science: The importance of epistemology for funding research and development: lessons from the UK." This paper relates to points raised in Chapters 2 and 8.

- Scientific and Technical Article
  Levin, E.A., Kloster, M. Morgan, R.M., Jones, V.J. (drafted, with the intention to submit to submit to either a forensic or interdisciplinary journal). "The transfer and persistence of environmental trace evidence: does the morphology of quartz grains and diatom valves affect their trace evidence dynamics?" This paper relates to the material seen in Chapters 6 and 7.

Conference presentations:

- September 2016: Conference - Poster Presentation
  Australian and New Zealand Forensic Science Society Symposium 2016 (Auckland, NZ) Title of presentation: *Examining diatoms as trace evidence in the context of footwear: an exploratory study on transfer and persistence*
  This presentation related to material conducted as MRes research, which informed the experimental designs of Chapters 5, 6, and 7.

- March 2017: Conference – Oral Presentation
  Bridging the Gap: The Second UCL Geography Conference (London, UK)
  Title of presentation: *The Forensic Analysis of Diatoms.*
  This presentation related to material seen in Chapters 2 and 7.

- June 2017: Conference - Poster Presentation
  The Forensic Archaeology Anthropology and Ecology Symposium 2017 (London, UK). Title of presentation: *"Size does matter: the selective retention of diatoms on footwear".*
  This presentation related to material seen in Chapter 7.

- September 2018: Conference - Poster Preparation
  Australian and New Zealand Forensic Science Society Symposium 2018 (Perth, Australia) Title of presentation: *Using fluorescent powder as a proxy to examine the persistence of trace evidence on footwear*
  This presentation related to material seen in Chapter 3.

- December 2018: Conference - Poster Presentation
  Meeting of the Forensic Geoscience Group of the Geological Society (London, UK) Title of presentation: *The transfer and persistence of particulate environmental trace evidence on footwear uppers and the potential of automated analytical approaches.*
  This presentation related to material seen in Chapter 5.

2

Contextual Literature Review

Abstract

This chapter sets out the intellectual context of the PhD, collecting together the canonical works whose influence continues to ripple through forensic science, and the more recent publications which have informed the experiments undertaken in this project. The literature review is divided into five sections, exploring, in turn: (1) The disciplinary context of forensic science, in which the research was undertaken, (2) The theories of transfer and persistence, which the subsequent experiments explore, (3) The forensic sub-discipline of 'Environmental Trace Evidence', where environmental indicators commonly associated with the discipline of Geography are used within forensic casework, (4) The precedents and rationales for the use of diatoms and quartz grains within forensic casework, and (5) The technological context of the research, in that open-source image processing and analysis software were freely available. Three main arguments are advanced: (1) That forensic science is a synthetic discipline, and that this is both its greatest strength and most challenging weakness, meaning that it incorporates large number of diverse sub-disciplines, and that for many of the techniques that are being used on a routine basis, the datasets which would support nuanced and robust interpretation do not currently exist, (2) That environmental trace indicators can be immensely valuable to investigators due to their ubiquity and specificity, and (3) That this project, undertaken between 2016 and 2019, has the capacity to utilise open-source software, a resource that has great potential to contribute to forensic science research.

2.1 The Forensic Context

2.1.1. Forensic science as a synthetic discipline

Forensic science is commonly defined as "the application of science to address questions related to the law" (Ribaux *et al.,* 2010: 10). As such, 'forensic science' comprises a broad range of disciplines and sub-disciplines, including elements of biology, chemistry, geography, geology, and engineering amongst others, in the pursuit of questions of interest to stakeholders within the justice system (Saferstein and Hall, 2002). Accordingly, it is possible to suggest that forensic science is a *synthetic* discipline; like a magpie, it darts around the cognate subjects picking out the brightest and shiniest advances to take back to its (metaphorical) nest.

This was certainly true during the Victorian Era, when Alphonse Bertillon adopted the nascent medium of photography for recording crime scenes (Tilstone *et al.,* 2006), and it is true today, where advances the advances of Alec Jeffrey and colleagues in population genetics (see: Gill *et al.,* 1985) and profiling (see: Jobling and Gill, 2004) form the basis of many techniques regarded as 'the gold standard' of admissible evidence (Lynch, 2013). What is striking is not just the *breadth* of the knowledge and techniques that are borrowed – with forensic researchers and practitioners drawing upon fields as disparate as psychology and analytical chemistry, (e.g. Nakhaeizadeh *et al.,* 2014; Gherghel *et al.,* 2016) - but also the *speed* with which novel theories and cutting-edge techniques are incorporated (see: Figure 2.1). In the field of serology, there was less than a year between the identification of blood groups in the Twentieth Century and their use in forensic casework (Tilstone *et al.,* 2006). Equally, and perhaps surprisingly, there was also an interval of only one year between the identification of useful polymorphic sequences within human DNA ('Short Tandem Repeats', or STRs) and their use as an analytical technique within casework (Jobling and Gill, 2004).

2.1.2 The missing empirical evidence base

This synthetic approach clearly has some benefits. To return to the magpie analogy, this willingness to share knowledge between, across, and through disciplinary boundaries has meant that forensic science have, historically, had a nest full of shiny things to play with – or, more accurately, cutting-edge techniques with which to approach casework (e.g. Jobling and Gill, 2004). On a practical level, therefore, this porousness of the discipline has been beneficial to the justice system. The first use of DNA for the purposes of identification in a criminal case resulted in the location and conviction of a man guilty of multiple murders (Jobling and Gill, 2004), while the first use of Y-Short Tandem Repeats in a criminal case resulted in the acquittal of a man accused of rape (Roewer and Eppen, 1992; Jobling and Gill, 2004). On an intellectual level, interdisciplinarity is often lauded in other disciplines as it may enable the addressing of complex, multifaceted, and interconnected challenges (e.g. Rhoten and Parker, 2004; Rhoten and Pfirman, 2007), and it is often depicted as a goal inherently worthy of pursuit and funding (Bracken and Oughton, 2006; Rhoten and Pfirman, 2007).

This approach of integrating insights from relevant disciplines is, however, not entirely unproblematic.

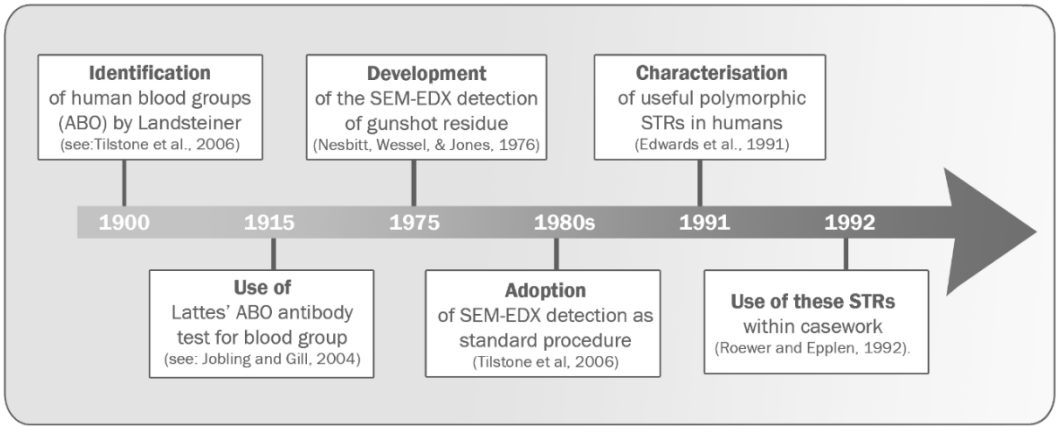


Figure 2.1: Timeline of technological advances and their incorporation into forensic science (after Jobling and Gill 2004). More recent advances include next-generation DNA sequencing, social media analysis, and digital evidence extraction

The main issue concerns research. Or, more specifically, the type of research being undertaken. As a result of forensic science importing techniques, researchers within the forensic arenas tend not to conduct *first principles* research - there being the assumption that the techniques have already been subjected to rigorous and empirical testing within their parent discipline (Morgan and Bull, 2007). It is, however, important to recognise that this leads to a focus on the hypotheses and challenges specific to the parent discipline, which may or may not take the context of forensic applications into account. In order to understand the ramifications of this problem, one can consider one of the canonical theories within forensic science, 'The Hierarchy of Propositions' (Cook *et al.,* 1998).

The hierarchy of propositions states that propositions about evidence can address either: (1) the source of a material, (2) the activity that could have resulted in its deposition, or (3) how it relates to an offence (Cook *et al.,* 1998; Figure 2.2). For example, one can imagine a hypothetical case of commercial burglary where trace DNA was found at the scene and a suspect, Mr. X, has been identified. One could propose (1) at the source level that the DNA cannot be excluded from coming from Mr. X; (2) at the activity level that Mr. X was at some point in the shop, and (3) at the offence level that Mr. X committed the burglary. The problem is that while data generated in other disciplines may provide a firm foundation to interrogate source-level propositions, they are in no way equipped to address the propositions of greater complexity.

The case of DNA provides a helpful example. In order to address the source level proposition outlined above, it would be possible to draw upon the theory of STRs (e.g. methods for their extraction and analysis), and the statistical foundations of genetic analysis (e.g. the Hardy-Weinberg Equilibrium) that are well-developed and well-understood in the arena of population genetics (Balding and Steele, 2015). Crucially, however, in order to address the activity-level proposition a bespoke dataset is required. An examiner would require data pertaining to the behaviour of DNA as a trace. How readily does trace DNA transfer? Could Mr.X's DNA only be transferred directly, or might it have undergone multiple transfers to reach its

destination at the site of the burglary? How readily does trace DNA persist? Could Mr. X have visited the shop a day before, and is that when his DNA was deposited? Could Mr. X have visited the shop a *year* before, and *that* was when his DNA was deposited? These are the sorts of questions that are only just beginning to be addressed (e.g. Meakin and Jamieson, 2013), but for which answers are clearly necessary if one wants to consider activity-level propositions (Jackson and Biederman, 2019).

For many sub-disciplines the datasets which would support nuanced and robust interpretation do not currently exist (NAS, 2009, Mnookin *et al.,* 2011; PCAST, 2016). This sets out a significant challenge, given that it is the activity and offense level propositions in which law enforcement agencies are interested, and these are precisely the propositions that, for many sub-disciplines, forensic science is currently unequipped to address.

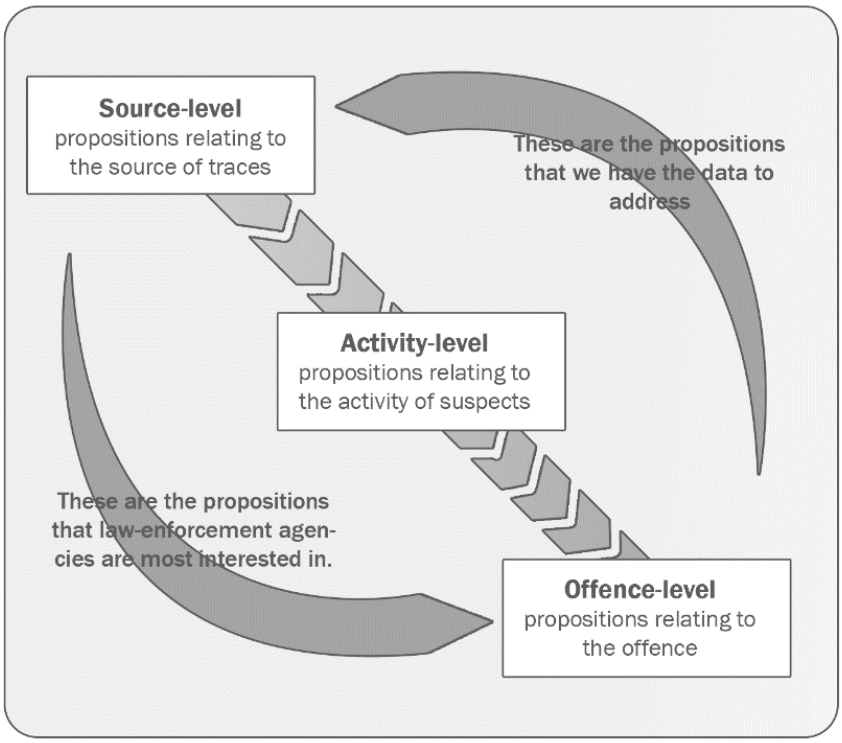


Figure 2.2: The hierarchy of propositions as outlined in Cook et al 1998

2.1.3 The implications for interpretation and validity in forensic science

Evidently, this missing empirical evidence base is problematic. It can be suggested that there are two main dimensions to this problem:

(1) Practically: without this foundation robust interpretation of evidence is not possible

If one considers the process of a forensic investigation (Figure 2.3), there are a series of steps which are followed in sequence. Classification follows collection. Association follows individualisation (Inman and Rudin, 2002). Downstream processes are contingent upon the success of upstream processes, and the progression is often iterative (Morgan, 2017). It is imperative, therefore, to have a firm understanding of the earlier stages of an investigation (e.g. an understanding of the division of matter, its transfer, persistence, and redistribution over time) in order to facilitate effective collection, analysis, and interpretation. It is not possible to have robust forensic reconstruction if there are unknown quantities in these earlier stages (Morgan, 2017).

(2) In terms of legitimacy: These lacunae have not gone unnoticed by regulatory bodies

These lacunae have been addressed by those in regulatory bodies. The landmark report was the 2009 National Academy of Sciences Report ('Strengthening forensic science in the United States: a path forward' (NAS, 2009)) a document that has spawned responses, counter-responses, and editorials warning of 'crisis' (e.g. Kaye, 2010; Roux *et al.,* 2015). The central message was that it is not tenable to continue operating with this dearth of empirical datasets relating to the application of techniques, and that a concerted and co-ordination effort to provide "a research culture" was, and remains, necessary in order to address these issues of validity and admissibility (Mnookin *et al.,* 2011; Margot, 2011; Morgan, 2017). Reports which followed in subsequent years (Forensic Science Regulator, 2015; President's Council of Advisors on Science and Technology, 2016; Chief Scientific Advisor, 2015) reiterated this message, placing emphasis

on the fundamental requirement for such studies to underpin robust interpretation (Morgan, 2017). In the United Kingdom, these reports have included specific calls to action; the 2014-2015 Forensic Science Regulator's Annual Report identified the processes of transfer and persistence as a priority for research (Forensic Science Regulator, 2015), stating that "The personal perspective of the Regulator is that the highest priorities for innovation to drive up the quality of forensic science are the provision of data to support the evaluation of evidential significance" (Forensic Science Regulator, 2015, p.27). This call has been reiterated in subsequent reports (2016, 2018, 2019; see also House of Lords, 2019).

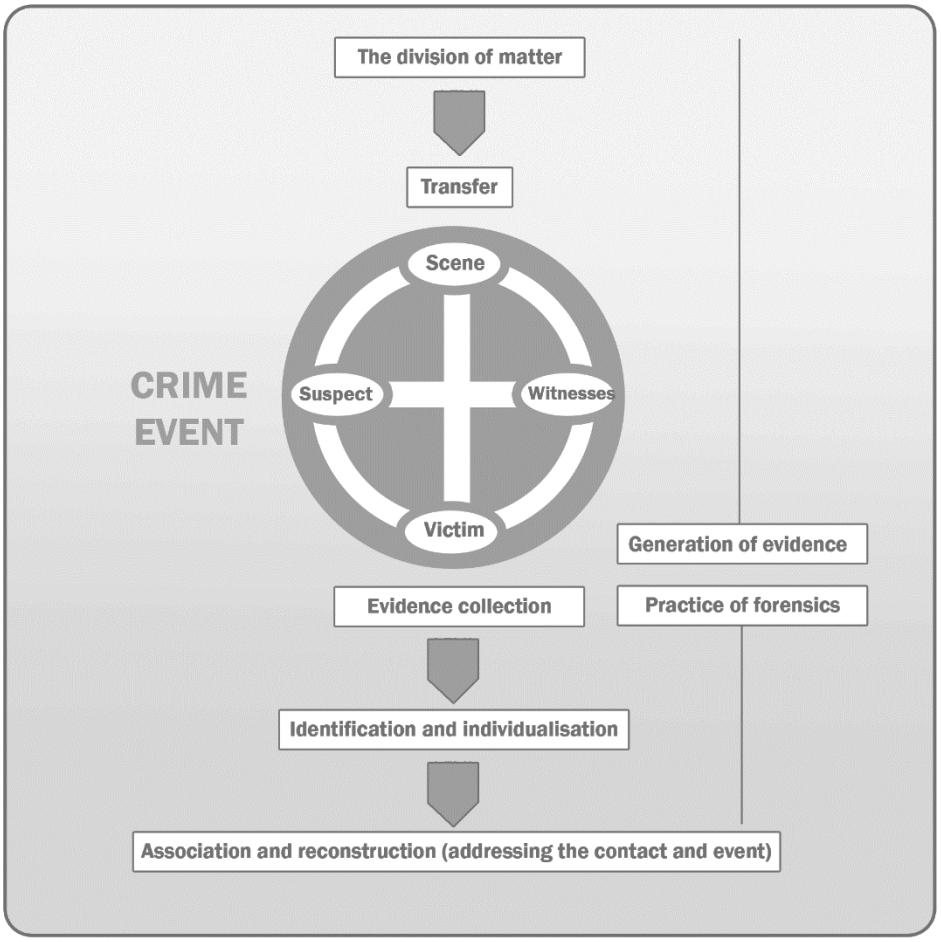


Figure 2.3 A conceptual model of the process of a forensic investigation. This diagram is adapted from Inman and Rudin 2002. Other conceptual models are available (e.g. Ribaux and Talbot Wright, 2014).

2.1.4 A brief note on epistemology

Forensic science might be considered an interesting field in that research happens in both pure and applied settings; findings both inform and emerge from casework (see: Morgan, 2017). While questions of epistemology and ontology are not central to the work in this project, it is important to acknowledge that, within the field, there is a debate surrounding the legitimacy of different forms of knowledge generation (e.g. Cole, 2006). The stance (implicitly) adopted here is that the foundational datasets which are necessary should be tested empirically (echoing Morgan, 2017; Lander, 2017; Edmond and Martire, 2017).

2.1.5 A brief note on austerity

Since 2010, the United Kingdom has been in a period of austerity, with a moratorium or reduction (in real terms) of public spending (O'Hara, 2015). In the last decade, the United Kingdom has seen the closure of a centralised, nationalised facility which undertook forensic research and development (The Forensic Science Service) on economic grounds (Gonzalez-Rodriguez and Baron, 2018), at the same time as "research active Universities have little forensic science research" (Evison, 2018: 282), and private providers of forensic services were unlikely to pursue research and outside of the established services that they were contracted to deliver at volume (House of Lords, 2019). Since 2010 forensic science has not been part of a national Research Council remit and it has been argued that forensic science is a discipline in which research and development has been neglected (House of Lords, 2019).

2.2 Transfer and persistence

2.2.1 The axiom of transfer

There is, therefore, consensus that more data are needed on the 'transfer and persistence' of traces (e.g. Forensic Science Regulator, 2015; 2019). The axiom of transfer is perhaps the most established theory within forensic science. That matter is divisible, and that contact between individuals and environments will result in reciprocal relocations of matter, is a fundamental logical precursor to the analysis of forensic evidence (see: Inman and Rudin, 2002; revisit Figure 2.3). This concept is most commonly (and concisely) expressed as "every contact leaves a trace", and attributed to the French Criminologist Edmund Locard (e.g. Houck, 2001; Morgan *et al.,* 2009; Coyle, 2010). It is, however, expressed more verbosely elsewhere (including in Locard's original paper where the axiom is purported to have appeared for the first time) (Locard, 1920; Rankin, 2010; Figure 2.4).

Figure 2.4: Translations of quotations expressing Locard's principle

While this axiom has become integral to (and oft repeated within) the discipline, it is possible to suggest that the phrase is incomplete, and potentially unhelpful given that: "The strength of evidence, its value to the court, lies in being able to put the findings of these materials into the context of human behaviour [...] Every contact may indeed leave a trace – but what does that trace mean?" (Coyle, 2010: 106). In other words, every contact may leave a trace, but can that trace be recovered, or meaningfully interpreted?

2.2.2 Recovering traces: the concept of persistence

Complementing the concept of transfer is the concept of persistence and 'trace evidence dynamics': that trace evidence will not remain on an article of evidence indefinitely, but will be lost over time (Pounds and Smalldon, 1975; Chisum and Turvey, 2000; Houck, 2001). Five variables which might influence the recovery of evidence have been identified (Houck, 2001), as:

(1) The amount of time that has elapsed since the contact

(2) The texture of the recipient surface

(3) The size and texture of the transferred material

(4) Whether there have been attempts to remove the evidence

(5) The activity levels of the suspect

To date, each of these has been started to be investigated for a number of evidence types:

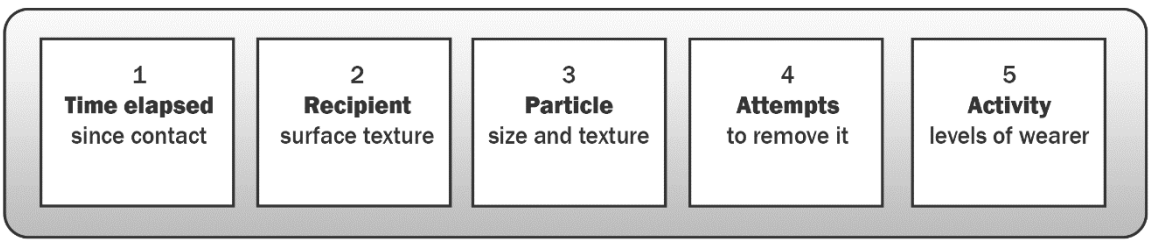


Figure 2.5: The five factors which may influence the recovery of evidence identified by Houck (2001)

(1) The amount of time that has elapsed since the contact

The first systematic studies into persistence generated decay curves for fibre evidence, provided plausible timescales for evidence recovery, and demonstrated that evidence may be lost in two phases – the first very rapid, the second more gradual (Pounds and Smalldon, 1975a; 1975b; 1975c). For a number of decades, research has been skewed towards certain evidence types; glass fragments (e.g. Brewster *et al.,* 1985; Hicks *et al.,* 1996; Allen and Scrannage, 1998), and fibres from clothing, vehicles, and carpets (e.g. Pounds and Smalldon, 1975a; Scott, 1985; Lowrie and Jackson, 1991) – both pertinent to routine investigations (Houck, 2001). Recent efforts have, however, extended the suite of evidence types for which data are available (see: Table 2.1. Recent studies have considered the persistence of human and animal hair (Dachs *et al.,* 2003; Boehme *et al.,* 2009), gunshot residue (e.g. Arndt *et al.,* 2012), and environmental evidence types such as pollen grains (e.g. Bull *et al.,* 2006; Schield *et al.,* 2016) and diatom frustules (Scott *et al.,* 2014; Levin *et al.,* 2017). The vast majority of these studies have considered clothing as the recipient surface (Table 2.1) and very few consider footwear as the context (see: Roux *et al.*, 1999; Morgan *et al.,* 2009; revisit Table 2.1).

(2) The texture of the recipient surface

The effect of the recipient surface has been effectively demonstrated in studies of clothing, suggesting that rougher surfaces entrain greater numbers of particles or fibres (e.g. Scott, 1985; Daly *et al.,* 2012) and retain them for longer (e.g. Pounds and Smalldon 1975a; Bull *et al.,* 2006). This phenomenon has also been demonstrated with respect to footwear materials: in a study comparing the number of diatom valves which could be recovered from swatches of canvas, suede, leather, rubber, and polyurethane immersed in freshwater for three minutes, the numbers of diatom valves recovered from the two woven materials was in the thousands per cm^2 while the leather was in the hundreds, and the rubber and polyurethane in double or single figures (see: Levin *et al.,* 2017).

(3) The size and texture of the transferred material

In the context of glass particles (Brewster *et al.,* 1984; Hicks *et al.,* 1996) and fibres (Pounds and Smalldon 1975b) it has been demonstrated that larger glass fragments and fibres persist for shorter timescales (and, conversely, that smaller particles will persist for longer durations). A paper considering the different size fractions of soil has also demonstrated a loss of the coarser fractions (Chazottes *et al.,* 2004), however research in this field is still relatively nascent, and more research would be beneficial. For forensic examiners who work upon assemblages (compositional data) of e.g. species (such as diatomists) it would be immensely useful for interpretation to know whether absence might be because particles are lost in order of size, not in order of initial abundance (Levin *et al.,* 2017; Scott *et al.,* 2019).

(4) Whether there have been attempts to remove the evidence

There have been a large number of studies considering whether bodily fluids can persist laundering (e.g. Brayley-Morris *et al.,* 2015; Mushtaq *et al.,* 2016; Oldfield *et al.,* 2017). Experiments have also considered the extent to which fibres can persist hand-washing (Hong *et al.,* 2014), and that mineral grains entrained on footwear may persist after going through a washing machine cycle (Morgan *et al.,* 2009), and may also be transferred to other items in the drum. Work has also demonstrated that glass fragments, fibres, and chalk will persist in vehicles after vacuuming (Morgan *et al.,* 2009). With respect to particulate trace evidence, studies have also considered attempts to destroy evidence with fire; suggesting that quartz grains can retain their surface textures with exposure to temperatures of up to 1200°C, so are likely to survive vehicle fires (Morgan *et al.,* 2008). Pollen grains, meanwhile, remain identifiable up to temperatures of 400°C (Morgan *et al.,* 2014). Similar studies have also been undertaken on diatom valves, showing similar persistence, and the potential for identification after a fire (Scott *et al.,* 2017).

(5) The activity levels of the suspect

Experiments with glass fragments have demonstrated that fewer particles remain after 5 minutes of running than five minutes of walking (Morgan *et al.,* 2009). No other published literature on the topic could be found.

Specific Trace	Recipient material	Citation	Dates of first publication
Clothing fibres	Indoor clothing	Robertson and Lloyd, 1984; Grieve *et al.,* 1989 Lowrie and Jackson, 1991	1975
	Outer garments (worn)	Pounds and Smalldon, 1975; Robertson *et al.,* 1982; Krauss, 1995; Akulova *et al.,* 2002;	
	Immersed garments	Lepot *et al.,* 2015; Lepot *et al.,* 2015	
	Skin	Palmer and Burch, 2009; Palmer and Polwarth, 2011; DeBattista *et al.,* 2014	
	Hair	Ashcroft et al., 1988; Salter and Cook, 1996;	
	Pillowcases	Palmer and Banks, 2005	
Glass fragments	clothing fabrics	Brewster *et al.,* 1985 Hicks et al., 1996	1985
gunshot residue (GSR) and firework particles	hair	Zeichner and Levin, 1993	1993
	hands	Jalanti *et al.,* 1999 Linsday *et al.,* 2011 Brożek-Mucha, 2011 Arndt *et al.,* 2012 Lindström *et al.,* 2015 Grima *et al.,* 2014	
Automobile carpet fibres	Shoes' soles	Roux *et al.,* 1999	1999
PU fragments	clothing fabrics	Wiggins *et al.,* 2002	2002
Hairs (human and animal)		Dachs *et al.,* 2003 Boehme *et al.,* 2009	2003
Pollen grains		Bull *et al.,* 2006 Schield *et al.,* 2016	2006
	footwear	Morgan *et al.,* 2009	
Lighter flint	clothing fabrics	Bull *et al.,* 2006	
UV powder		Bull *et al.,* 2006	
	Shoes' soles	Morgan *et al.,* 2009	
Powder based on *Lycopodium* spores	clothing fabrics	Howarth *et al.,* 2009	2009
Diatoms		Scott *et al.,* 2014	2014
	footwear	Levin *et al.,* 2017	
Soils	clothing	Murray *et al.,* 2016	2016
	footwear	Stoney *et al.,* 2016	

Table 2.1: Structured studies into transfer and persistence, organised by evidence type

2.2.3 Interpreting traces: additional considerations

Building on these foundations, researchers have explored additional nuances which might assist more meaningful interpretation. For a number of evidence types, it has been demonstrated that transfers may be multiple and indirect (French *et al.,* 2012; 2014; French and Morgan, 2015; Palmer and Banks*,* 2005); that particles may in fact be redistributed rather than lost (Robertson and Lloyd, 1984; Morgan *et al.,* 2010; Szewcow *et al.,* 2011); and that the material recovered is not likely to represent purely material derived from the crime scene, but also to contain material from before and after the event of forensic interest (Bull *et al.,* 2006; Morgan *et al.,* 2009*).* All of these factors complicate the interpretation of the weight and significance of such materials in a forensic reconstruction, but are necessary to understand the significance of trace evidence in a specific context (Coyle, 2010).

2.2.4 Work in the last decade: extending the suite of evidence types and UV proxies

Since 2005 there have been two main foci for published research. The first involves researchers beginning to extend the suite of evidence types for which we have data (see: Table 2.1). The second is methodological, in that researchers have employed fluorescent proxies to aid in the rapid and non-destructive generation of data to explore the processes of transfer and persistence (e.g. Bull *et al.,* 2006; Morgan *et al.,* 2009; 2013; French *et al.,* 2012; Slot *et al.,* 2017).

The findings from these studies have suggested that the powder is a valid proxy for particulates such as lighter flint particles (Bull *et al.,* 2006), silt-sized mineral particles (Morgan *et al.,* 2009) and pollen grains (Morgan *et al.,* 2013). For more on the specifics of ultra-violet experimentation, please see the literature review in Chapter 3 (section 3.1.1). To summarise, the use of fluorescent proxies and imaging techniques to sample articles non-destructively is seen as a development which might complement more traditional microscopy-based studies in the exploration of hypotheses surrounding the dynamics of trace evidence. The benefits of using fluorescent powder as a proxy are numerous, including the ability to quantify it

though non-destructive photographic sampling with ease, and generating data far more rapidly than traditional microscopic methods. This acceleration of data collection (and, in conjunction with image processing, and potentially analysis) enables experiments to encompass larger datasets, and provides examiners with the opportunity to explore many more variables and contexts within the timescales available for research projects. This may help to facilitate the development of the evidence base that is so needed for the accurate and transparent interpretation of trace evidence.

2.3 Environmental trace evidence

> "The geologist has always been a detective" (Murray and Tedrow, 1975: 3)

2.3.1 Defining environmental trace evidence

Trace evidence can be defined as "materials that, because of their size or texture, transfer from one location to another and persist there for some period of time" (Roux *et al.,* 2015: 2). It is not difficult to see how chemical, physical, and biological elements of the environment might be transferred as trace evidence during an event of forensic interest (see: McCulloch *et al.,* 2017), and the study of such traces has been formalised into its own sub-discipline (see: Morgan and Bull, 2007; Ruffell and McKinley, 2008). It is potentially confusing that the field has multiple, competing designations - if one consults a library one may find almost as many neologisms as books; "forensic geology" (Murray and Tedrow, 1975) "forensic geoscience" (Pye and Croft, 2004; Bull and Morgan, 2007; Bergslien, 2012); "geoforensics" (Ruffell and McKinley, 2008); and "environmental forensics" (Ritz *et al.,* 2008) being just a selection of the published terms. Despite the varying titles, these texts are united by two common themes:

(1) An appreciation that matter which is generally thought of as the realm of geography (for example, soils, sediments, pollen grains and mineral grains) can represent useful and informative trace evidence within forensic investigations, as evidenced by their use over the centuries (Figure 2.6).

(2) Recognition that the practice of 'forensic geoscience' is philosophically discrete from pure geoscience, demanding its own assumptions and practices (Morgan and Bull, 2007; McCulloch *et al.,* 2017; Cheshire *et al.,* 2017). Chief among these are the need to dissociate rather than associate samples, and to recognise the significance that some components of soils are exotic and others ubiquitous – and that the significance of the presence or absence of a component must bear the initial abundance in mind (McCulloch *et al.,* 2017; Cheshire *et al.,* 2017; Morgan and Bull, 2007).

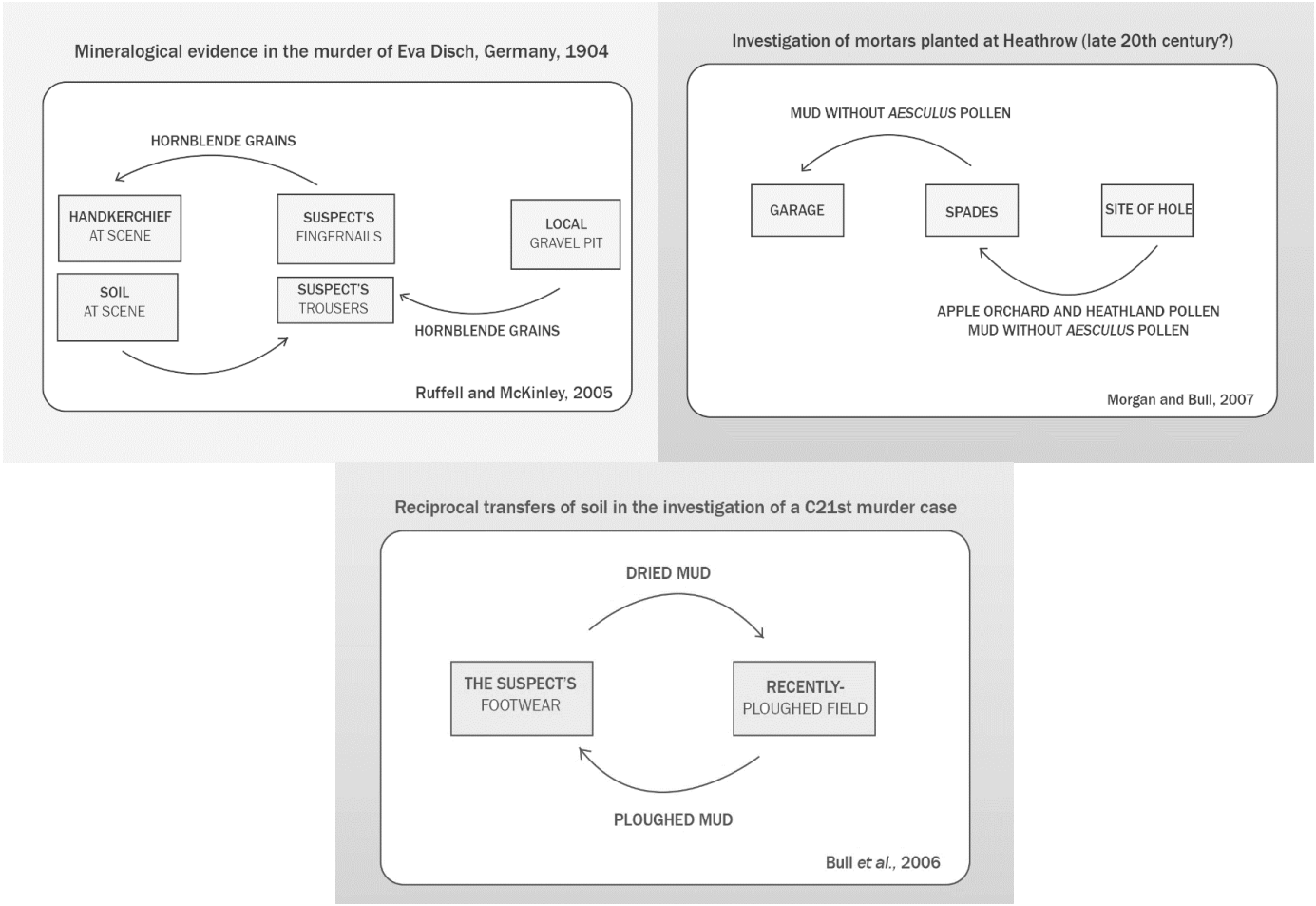


Figure 2.6: Diagrammatic representations of some uses of environmental trace evidence

2.3.2 Environmental forms of trace evidence

There are both exotic and ubiquitous forms of environmental trace evidence (Morgan and Bull, 2007); those which one is unlikely to encounter and those which one is unlikely *not* to encounter during the course of an investigation. Many schemes of classification exist (again, almost as many as there are publications) (e.g. Ruffell and McKinley, 2008; Cheshire *et al.,* 2017). For the sake of simplicity (since taxonomy is tangential to the issues at stake), it will be accepted here that there are evidence types that are terrestrial and aquatic, organic and inorganic, microscopic or macroscopic (Figure 2.7.)

Considering the fields of environmental monitoring and palaeoenvironmental reconstruction, it is a fundamental assumption that trained individuals can extract large amounts of information from small amounts of geological or biological material. From the presence or absence of a species of diatom in a sample of water, an experienced ecologist can draw inferences about pH, salinity, seasonality, flow speed, and the region of the water body from which it was drawn (Smol and Stoermer, 2010). The presence or absence of species of pollen from a soil sample can tell an examiner the likely depth from which it was extracted (Morgan and Bull, 2007). It is a cliché to talk about seeing the world in a grain of sand, but the environmental history of a grain is inscribed upon its face – whether it has been exposed to humic or formic acids (indicative of woodland environments), whether it has been in environments dominated by fluvial, aeolian, or glacial processes (Krinsley and Doornkamp, 1973; Mahaney, 2002; Bull and Morgan, 2006), and an indication as to the extent or intensity of those processes (e.g. Costa *et al.,* 2012).

In this project, experiments will focus upon two forms of environmental trace evidence; diatom valves and quartz grains. These were deliberately chosen to represent both aquatic and terrestrial environments, and are both abundant and widely-distributed (thus contributing towards the relevance of data generated during this project). While the specific qualities of diatoms and quartz grains will be outlined in section 2.4, the key reasons why both diatoms and quartz grains are promising forensic evidence are that their physical properties render them useful for

forensic reconstruction. They are (1) robust (2) widely-distributed and abundant (3) sensitive to their environment, and (4) useful when mixed-provenance samples are addressed (Bull and Morgan, 2006; Scott *et al.,* 2014).

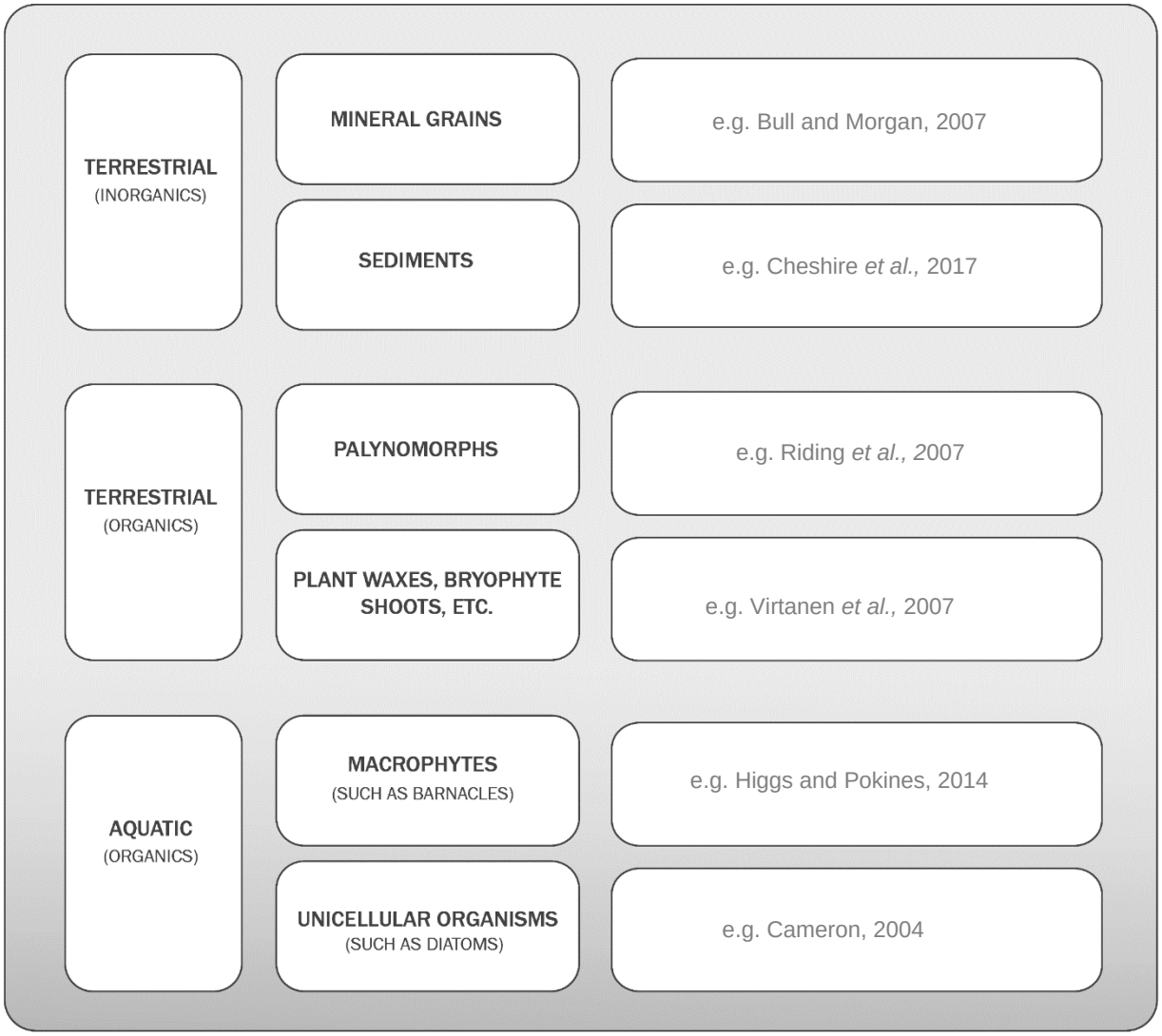


Figure 2.7: A diagram illustrating some forms of environmental evidence types

2.4 Diatoms and quartz grains

2.4.1 What are diatoms?

Diatoms are a type of unicellular autotrophic algae that are widely distributed throughout water bodies (Cameron, 2004; Horton *et al.,* 2006) and moist soils (Cox, 2012). Each cell possesses a robust, porous, external layer of silica, known as a 'frustule', which is composed of two halves ('valves') (Smol and Stoermer, 2010). The ornamentation on the frustule is unique to each species, allowing examiners to identify diatoms under the microscope (Jones, 2007; see Figure 2.8). Diatoms tend to range between 20 and 200 microns in diameter (Kelly, 2000; Cox, 2012) and their examination is primarily through phase contrast microscopy ('PCM') (Jones, 2007), with scanning electron microscopy ('SEM') used to capture high-magnification, high-resolution images (ibid). Despite being very widely distributed as a *class* of organism, most *species* have narrow tolerances for environmental variables, existing in specific alignments of water nutrients, pH, and salinity, as well as exhibiting seasonal variation (Cameron, 2004; Gayle, 2004; Horton *et al.,* 2006; Jones, 2007). Since the presence or absence of particular diatom species within an assemblage can allow examiners to draw inferences about the water body from which the sample was extracted (Horton *et al.,* 2006; Korhola, 2007; Cox, 2012), this environmental specificity has meant that diatom analysis has become a cornerstone of palaeo-environmental reconstruction and environmental monitoring (e.g. Battarbee *et al.,* 2001; Smol and Stoermer, 2010; Mackay *et al.,* 2014).

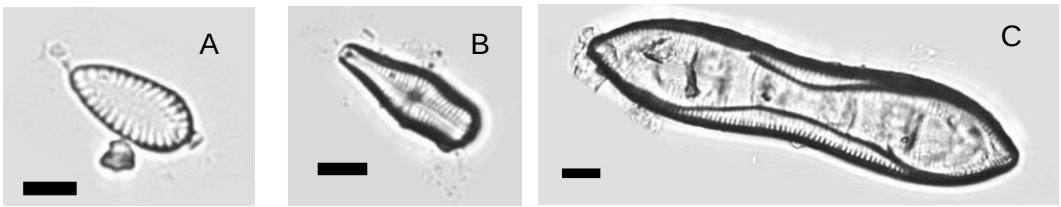


Figure 2.8: Examples of diatom morphology and ornamentation. Brightfield micrographs SEM micrographs featuring the genera: (a) Surirella (b) Gomphonema (c) Cymatopleura. Scalebar is equal to ten microns in length.

2.4.2 How are diatoms forensically useful?

Diatoms have tended to have two uses within forensic investigations:

(1) As a diagnostic test of drowning

The primary use of diatoms within forensic science is within the field of pathology, where they have been used as a diagnostic test for drowning for many years (Cameron, 2004; Cameron and Peabody, 2010). The mechanism for this is quite simple; when a person drowns water enters the lungs, and if this water contains diatoms and the heart is still beating, diatoms will move into the bloodstream and get carried to the well-vasculated organs where they will accumulate (Pollanen, 1998; Peabody, 1999; Cameron, 2004; Fucci *et al.,* 2015). Accordingly, finding large numbers of diatoms within the well-vasculated organs can be used to confirm a diagnosis of death by drowning, and conversely finding few or no diatoms can cast aspersions on a diagnosis – although cases of drowning in freshwater where diatoms were absent from autopsy have been observed (Cameron, 2004).

Subsequent analysis of the species found within the post-mortem can provide additional information. Spatially, the species assemblages have been used to infer the possible location of drowning (Gayle, 2004; Cameron and Peabody, 2010; Kakizaki *et al.,* 2010), whether the environment was freshwater or a domestic water supply (Horton *et al.,* 2006; Ago *et al.,* 2011), and whether a corpse might have been relocated between water bodies (Thakar and Singh, 2010; Verma, 2013). The species assemblages can also offer temporal information; there has been research into the utility of algae to estimate the post-mortem submersion interval of cadavers (e.g. Casamatta and Verb, 2000; Haefner *et al.,* 2004; Zimmerman *et al.,* 2008; Dickson *et al.,* 2011).

(2) As trace evidence to dissociate suspects, victims, and scenes

The second use of diatoms within forensic investigations is as more traditional trace evidence, to attempt to dissociate or fail to dissociate suspects, victims, and crime scenes (e.g. Peabody, 1999; Cameron, 2004; Cameron and Peabody, 2010; Cox, 2012; Scott *et al.,* 2014). Diatoms have proved to be useful evidence in

casework scenarios involving aquatic crime scenes (e.g. Siver et al., 1994; Gayle, 2004; Cameron and Peabody, 2010) and also in cases of safe-cracking – where fossilised diatoms in the form of the mineral 'diatomite' are often used as safe ballast (Peabody, 1999; Cameron, 2004; Cox, 2012).

2.4.3 Strands of current research

There have been several publications relating to diatoms as trace evidence in recent years (e.g. Uitdehaag *et al.,* 2010; Scott *et al.,* 2014; 2017; 2019, Levin *et al.,* 2017), with researchers considering their transfer (Scott *et al.*, 2014), their persistence (Levin *et al.,* 2017), their seasonality (Scott *et al.,* 2017) and how one might adapt existing chemical protocols to extract diatoms from clothing and footwear materials, which may react violently to the reagents used within geographical contexts (Uitdehaag *et al.,* 2010; Scott *et al.,* 2014; Levin *et al.,* 2017).

On the environmental side, some of the most recent papers consider the automation of morphometrics (Kloster *et al.,* 2017) and species-level identification (Kloster *et al.,* 2014). While this drive to automation is not entirely new - following on from laser holography in the 1970s (e.g. Cairns *et al.,* 1977; 1979; 1982), multivariate statistics to describe shape in the 1980s (e.g. Stoermer and Ladewsi, 1982; Mou and Stoermer, 1992), and ADIAC in the early 2000s (e.g. DuBuf *et al.*, 1999; DuBuf and Bayer 2002; see Figure 2.9) this latest tranche of publications and tools is different from previous efforts in that it involves free software and relatively cheap hardware.

Since the manual examination of diatom valves under microscopy is time-consuming, and places a ceiling on the amount of data that can reasonably be collected within a research project, this is a strand of methodological development that will be pursued in this PhD (see Chapter 7).

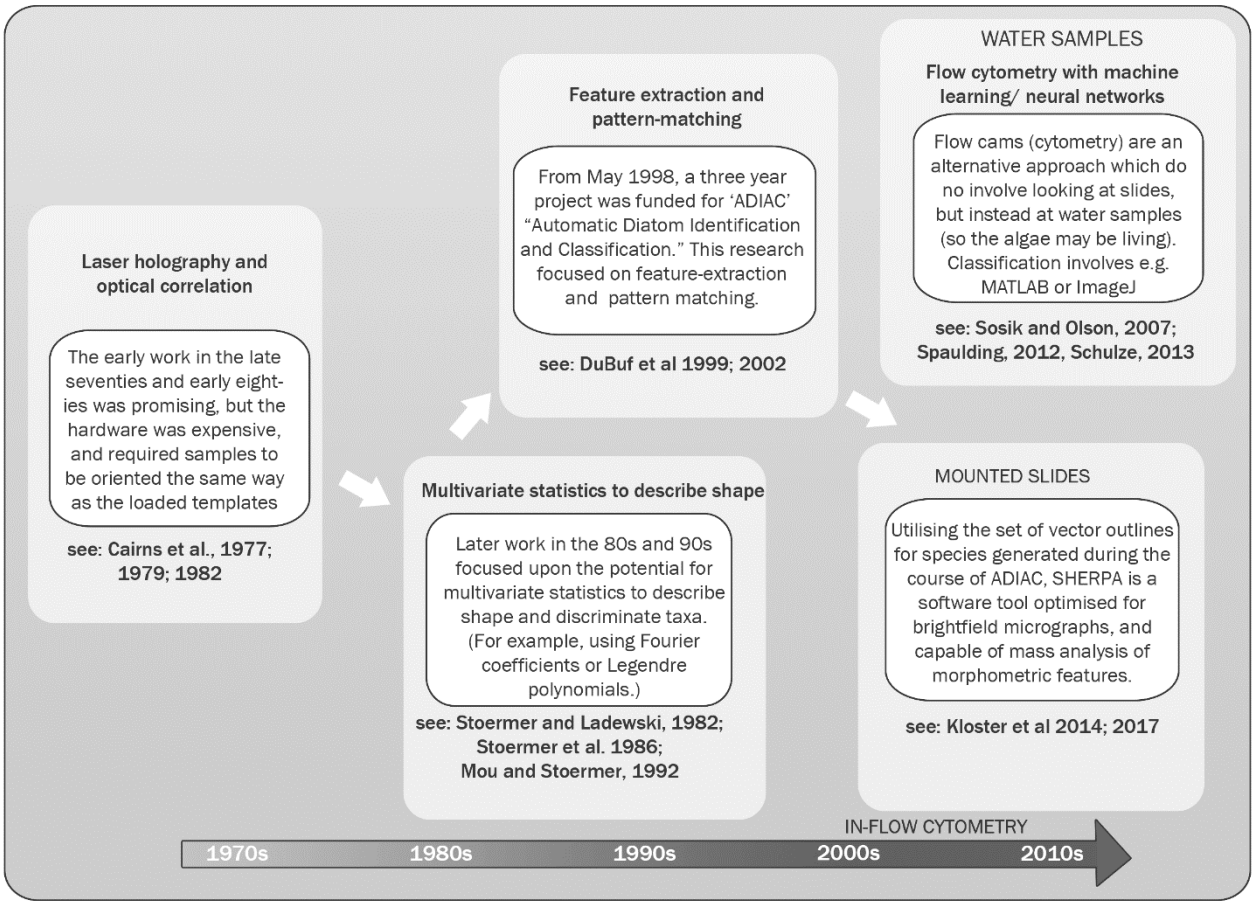
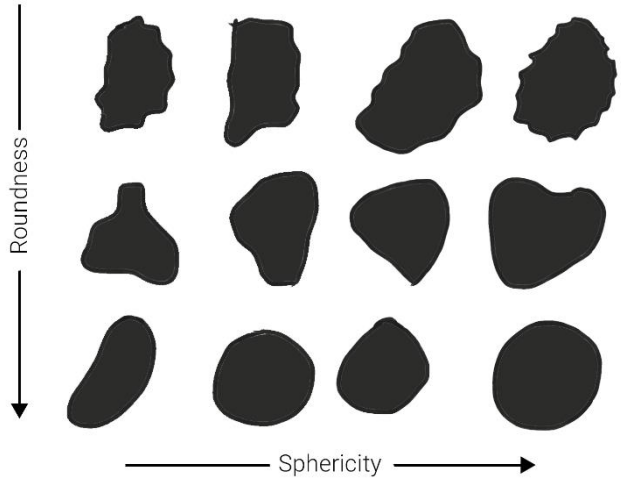


Figure 2.9: A timeline of developments into automated diatom ID and morphometry

2.4.4 What are quartz grains?

Quartz is one of the most abundant minerals on earth (Bull and Morgan, 2006). Its grains, which range between 0.05 and 2 mm in diameter, are almost ubiquitous within soils (ibid). Composed of silicon and oxygen atoms arranged in (SiO_2) tetrahedra, the mineral grains are incredibly robust – meaning that the grains persist in soils over time, while the surface of the grain is modified by environmental and anthropogenic influences (Krinsley and Doornkamp, 1973; Mahaney, 2002; Bull and Morgan, 2006). Analysis of the grain morphology (Figure 2.10) and surface features, which normally involves a scanning electron microscope (Krinsley and Doornkamp, 1973), can be used to draw inferences regarding their environmental history (ibid). For example: angular grains with cleavage plates and striations may be indicative of glacial environments (Krinsley and Doornkamp, 1973), while rounded grains with impact features can be indicative of high-energy fluvial, marine, or aeolian regimes (Bull and Morgan, 2006). This ability for quartz grains to visually reflect their environmental history means that, like diatoms, they have earnt a place in the pantheon of palaeoenvironmental reconstruction techniques (Culver *et al.,* 1983; Mahaney, 2002). And, again like diatoms, it is this ability to reflect the environmental conditions of the location from which they were drawn that renders them valuable in forensic reconstruction scenarios (Bull and Morgan, 2006).

Figure 2.10: Grain morphology: sphericity and roundness (modified from Charpentier et al., 2013)



2.4.5 How are quartz grains forensically useful?

Quartz grains are used as trace evidence to associate suspects, victims, and scenes. They have been used for the prosecution in high profile cases of wildlife crime (Morgan and Bull, 2006; Morgan *et al.,* 2006) and homicide (Bull *et al.,* 2006). The analysis of quartz grain surface textures has various strengths compared to other techniques which might be used to analyse the components of a soil. The analytical technique:

(1) Requires only a small amount of evidence for analysis

While there is some debate over the number of grains required to conduct analysis (Mahaney, 2002; Costa *et al.,* 2012), the authors in the core of the canon all name numbers that are double-figures (see: Krinsley and Doornkamp, 1973; Bull and Morgan, 2006). Since a single teaspoon of quartz would contain around 30, 000 grains (Pye, 2007), there is likely to be sufficient material to conduct analyses.

(2) Can be used even where attempts have been made to remove or destroy evidence

In experimental studies, it has been demonstrated that quartz grains can be recovered from fabric that has been subjected to vacuuming (Morgan *et al.,* 2009), or a washing machine (ibid), while surface textures survive modification by fire in temperatures of many hundreds of degrees (Morgan *et al.,* 2008). This is significant as the soil characteristics of colour, chemical composition, and particle size distribution have all been shown to undergo alteration when heated (e.g. Sertsu and Sanchez, 1978; Sugita and Marumo, 1996; Iglesias *et al.* 1997; Palese *et al.,* 2004).

(3) Does not require homogenisation of the sample

Unlike the analysis of soil colour, particle size, and chemical composition, the analysis of quartz grain surface textures does not require the sample to be homogenised prior to analysis (Morgan *et al.,* 2009a). This means that any layering of evidence can be conserved – and the insights such layering can provide for interpretation can be exploited (ibid.).

2.4.6 Strands of current research

The very latest research on quartz grain surface textures is being carried out in the field of palaeoenvironmental reconstruction, with recent papers for example considering late quaternary lacustrine sediments (Warrier *et al.,* 2016), coasts (Machado *et al.,* 2016), proglacial lakes (Mazumder *et al.,* 2017) and Pleistocene Periglacial Environments (Kalińska-Nartiša *et al.,* 2017). In a similar vein to diatoms, some researchers are currently considering the automation of analyses, in terms of using Fourier analysis to quantify surface roughness coefficients (Charpentier *et al.,* 2017), Basic Image Features (BIFs) to facilitate texture recognition from pre-existing sets of data (Newell *et al.,* 2012), and Fourier analysis to classify grains according to their overall morphology (Curry *et al.,* 2009). Again, like diatoms, there have been precedents for the automation of analysis (again, based on Fourier transforms - see: ARTHUR (Mazullo and Kennedy, 1985); ARTHUR II (Bui *et al.*, 1989), but these were constrained by the technology available at the time. With the increased speed of image processing, this is something that could be pursued with potentially valuable consequences.

2.5 Technology and research

2.5.1 The climate of open source software

Over the course of this literature review, the academic context of the research has been considered (i.e. that there are still lacunae pertaining to transfer and persistence, as a result of the importation of techniques from cognate disciplines), as has the financial context (i.e. that research within forensic science has suffered from a dearth of funding, and a lack of clarity about whose remit funding forensic science should be – whether that should be academia, industry, or a governmental body such as the Home Office). We might also, however, consider the technological context. This research was undertaken between 2016 and 2019, in an era where the internet is both established and accessible. It is relatively common for spatially-distributed communities of individuals with similar interests to connect online, and a

number of enterprises where skilled individuals provide intellectual property under generous licences have emerged (such as the collaborative encyclopaedia *Wikipedia,* the many creative assets released under creative commons licence, or the many coding projects available on Github).

2.5.2 Specific resources

Three pieces of open source software and web applications were used in this project:

(1) ImageJ: an open-source, java-based image processing software that operates across operating systems and with a large number of accepted file formats (Schneider *et al.,* 2012; Arena *et al.,* 2017; Abràmoff *et al.,* 2004). Its real strength comes from its ability to be expanded and specialised through the use of plugins (e.g. Legland *et al.,* 2016) and look-up tables (LUTs) - and an active community of skilled users who contribute enthusiastically (e.g. Nanes*,* 2015; Schindelin *et al.,* 2015). The open-source nature of the endeavour means that many variants of the software now exist, specialised for sub disciplines (e.g. SalsaJ, BoneJ, Fiji) and that a library of plugins and documentation abounds (Schneider *et al.* 2012; Arena *et al* .2017).

(2) BIIGLE: An online environment for the annotation of very large images and collections of images (see: Langenkämper et al., 2017).

(3) SHERPA: an application capable of segmenting brighfield images of diatoms, and calculating a range of morphometric measurements, developed by Michael Kloster at the Alfred Wegener Institut, in Bremerhaven, Germany (see: Kloster *et al.,* 2014).

Since forensic science is a discipline where, to have impact, research findings have to not only be good but be implementable in a lab-based environment where time and cost are critical variables (Cook *et al.,* 1998), it is important to develop methods that can be deployed with powerful free resources.

2.6 Statistical Analysis

2.6.1 The purpose of statistical analysis

In many research disciplines, statistical analysis is a vital stage of the scientific process (e.g. Gardenier and Resnik, 2002; Field and Hole, 2007). Statistical techniques serve a number of purposes, having the potential to provide information that is descriptive, that is to say "producing convenient summaries of data" (Hand, 2008: 3), or inferential, "extracting meaning from data" (Hand, 2008: 2). In other words, statistics can be used to render trends in data legible to the reader – because most readers would struggle to parse vast tables of raw numerical values (Field and Hole, 2007). It is acknowledged that in order for statistical analysis to offer such helpful insights, the statistics used must be appropriate; the type of data must be considered, in addition to the sample size, and the intended purpose of the analysis (Ali and Bhaskar, 2016).

2.6.2 Potential pitfalls in statistical analysis

It is worth acknowledging that the vast majority of scientists who employ statistics within their research, the author of this thesis included, are not statisticians (Gardenier and Resnik, 2002), and that it is therefore possible that statistics are unintentionally mis-used or ab-used when presenting research findings (e.g. Gardenier and Resnik, 2002; Ioannidis, 2005; Goldacre, 2010; Wasserstein *et al.,* 2019). The convention for the reporting of inferential statistical analysis is to provide the *significance level* or '*p-value'* calculated (Dowdy *et al.,* 2011). This number represents the probability of obtaining the observed results when the null hypothesis (H_0) of the research question is true (ibid). The convention within many fields, including the field of experimental forensic science, is to use the significance level of $p<0.05$ as the cut-off for significance in inferential analysis, representing a 1 in 20 chance of obtaining the result when the null hypothesis is true (e.g. Ioannidis, 2019) and to report the p-values obtained when presenting results (as in, for example, Cheshire *et al.,* 2017; McCulloch *et al.,* 2017). In the field of biomedical sciences especially, there has been a concerted effort to problematise this use of p-

values (e.g. Wasserstein et al., 2019), moving away from the use of it in reporting as it can be accidentally (see also: Ziliak and McCloskey, 2008) or deliberately (Hubbard, 2016) misleading. Scholars in the field also argue that the term 'statistical significance' should be avoided (Wasserstein *et al.,* 2019), since the original intention, when the phrase was proposed by Edgeworth (1885) and used by Fisher (1925) was for the tool to connote that a result may warrant further scrutiny, but the meaning has drifted over time, and seen as more of a binary label of whether variables are meaningful or not(Wasserstein et al., 2019). Accordingly, there have been recommendations to avoid the term 'statistically significant' (Hurlbert *et al.,* 2019; Wasserstein *et al.,* 2019), instead substituting the word 'suggestive' for p values between 0.05 and 0.005 (Benjamin and Berger, 2019). The use of $p<0.05$ does, however, remain an understood convention in the field of experimental forensic science (see, for example: Cheshire *et al.,* 2017; McCulloch *et al.,* 2017), and is accordingly a convention that this thesis will employ. It is acknowledged that the term 'statistically significant' is here used to connote results which indicate that trends may be worth further investigation.

2.6.3 The use of statistical analysis in this thesis

This thesis follows in the tradition of theoretical forensic research in using both descriptive and inferential statistics to analyse numerical and categorical datasets, with the significance level of $p<0.05$ used in reporting (as in, for example, Cheshire *et al.,* 2017; McCulloch *et al.,* 2017), so that the results of these experiments are accessible to those who are familiar with this convention of reporting. However, this thesis incorporates the suggestion that 'statistically significant' can be a misleading phrase, and should be referred to as 'suggestive' of difference, and indicative of trends worth further pursuit (Benjamin and Berger, 2019). It is acknowledged that the statistics employed in this thesis were selective in nature. Further details on the statistical analysis performed (including the reasoning behind the use of each statistical test) can be found in Chapters 3, 4, 5, 6, and 7, in sections 3.2.6, 4.2.5, 5.2.8, 6.2.8, and 7.2.9.

3

Chapter Three

Quantifying the amount of fluorescent powder in a digital image: a comparison of manual and automated methods of segmentation

Abstract

Fluorescent powder is often used as a proxy for particulate trace evidence within studies of transfer and persistence, where it is sampled by macro-photography (e.g. Morgan *et al.,* 2010; French *et al.,* 2012). In order to extract quantitative data from the digital images of the fluorescent material, they must first be 'segmented' – the images must be divided into foreground (the pixels of interest) and background (the pixels that do not represent fluorescent matter) (Sezgin and Sankur, 2004). At present, there is no consensus or convention in terms of the methodology used to segment this sort of fluorescence imagery. In Chapter 3, three experiments were undertaken, to assess the methodological question: 'how might one best segment digital images of fluorescent powder?'. Experiment one assessed the extent to which manual segmentation (tracing) was reproducible, finding that differences existed between different examiner's estimates of the extent of the foreground, which yielded p-values of <0.05 when tested with inferential statistics. Experiment two assessed the performance of segmentation by manually-defining a global threshold level, finding that while it would be feasible on the basis of time, it also had problematic reproducibility. The third experiment tested the performance of a suite of global thresholding algorithms, finding that several of the algorithms (Yen, MaxEntropy, and RenyiEntropy) offered comparable levels of accuracy to the manual methods, but at far faster speeds, and with no issues around reproducibility. This chapter was supported by Professor Lewis Griffin, of UCL's Computer Science Department, whose advice (and critique) helped to shape and refine the research, and Joni Levinkind, who assisted with writing macros for ImageJ. Findings from this research have been published (please see: Levin *et al.,* 2019).

3.1 Introduction

3.1.1 Fluorescent powder as a proxy for particulate trace evidence

As discussed in Chapter 2, fluorescent powder has become an established proxy for particulate evidence within forensic experiments which explore the processes of transfer and persistence (e.g. Bull *et al.,* 2006; Morgan *et al.,* 2009; 2013; French *et al.,* 2012; Slot *et al.,* 2017), being used as an analogue for particles such as lighter flint (Bull *et al.,* 2006), pollen grains (Morgan *et al.,* 2013), and fractions of soil (Morgan *et al.,* 2009; Table 3.1*)*.

The reasons for this use are twofold: primarily, that the fluorescent powder bears physical similarities to these particulates, having a diameter of approximately 15 µm (French *et al.,* 2012), and secondly, that its fluorescent properties mean that it is possible to monitor the presence, absence, and distribution of the powder non-destructively – by photography under ultraviolet light (e.g. French *et al.,* 2012; Levin *et al.,* 2019). This is a critical benefit which explains the prevalence and enthusiasm with which fluorescent proxies have been embraced by researchers; with this non-destructive sampling, it becomes possible to monitor the particles *in situ,* without removing them from the experiment, and therefore possible to explore hypotheses about movements of the particles over time (e.g. French *et al.,* 2012). Photographic sampling is also much faster than the traditional sampling of the original evidence types: to take pollen grains as an example, to use fluorescent powder as a proxy, the sampling method is as rapid as taking a photograph with a digital camera (Morgan *et al.,* 2013). In contrast, to sample actual pollen grains might involve electron microscopy, with all the attendant preparation that requires (e.g. Morgan *et al.,* 2014), or the chemical digestion of samples and the production of permanently mounted slides (Riding *et al.,* 2007) – again, a time-consuming process.

Accordingly, the use of a fluorescent proxy for particulate trace evidence types can be considered to accelerate data acquisition and analysis. Since there is a need to conduct transfer and persistence studies to provide relevant datasets to assist in the interpretation of trace evidence (see: Morgan *et al.,* 2017; Forensic Science

Regulator, 2019), the use of a fluorescent proxy might therefore be seen as a promising avenue for research.

Year	Authors	Paper title	Purpose of the fluorescent matter
2006	Bull *et al.*	The transfer and persistence of trace particulates: experimental studies using clothing fabrics	To emulate lighter flint particles
2009	Morgan *et al.*	The Forensic Analysis of Sediments Recovered from Footwear	To explore the movement of silt-sized particles
2010	Morgan *et al.*	The reincorporation and redistribution of trace geoforensic particulates on clothing: an introductory study	To explore the redistribution of particles
2012	French *et al.*	Multiple transfers of particulates and their dissemination within contact networks	To investigate multiple transfers
2013	Morgan *et al.*	The recovery of pollen evidence from documents and its forensic implications	As a proxy for pollen on the surface of documents
2017	Slot *et al.*	Tracers as invisible evidence — The transfer and persistence of flock fibres during a car exchange	Fluorescent flock fibres as a tracer

Table 3.1 Table listing particulates evidence types for which fluorescent proxies have been employed. Reproduced and modified from Levin et al 2019

3.1.2 Image Segmentation

Studies which utilise fluorescent material as a proxy have tended to follow similar methodologies, with three steps followed in sequence (Figure 3.1):

(1) Experimental set-up: At the start of the experiment the fluorescent powder is distributed to the surface of interest, typically by dilution with a non-fluorescent powder and 'sifting' onto the surface (e.g. Bull *et al.,* 2006; Morgan *et al.,* 2013) or by 'flicking' the powder with a stiff bristled brush (e.g. Morgan *et al.,* 2010).

(2) Data acquisition: During the experiment, the surface of interest is repeatedly imaged, using macro-lens photography (e.g. French *et al.,* 2012; Slot *et al.,* 2017), in a darkened room (e.g. Morgan *et al.,* 2012) or bespoke dark box (French *et al.,* 2012) with a source of ultraviolet illumination.

(3) Data analysis: The resultant digital images are then processed and analysed.

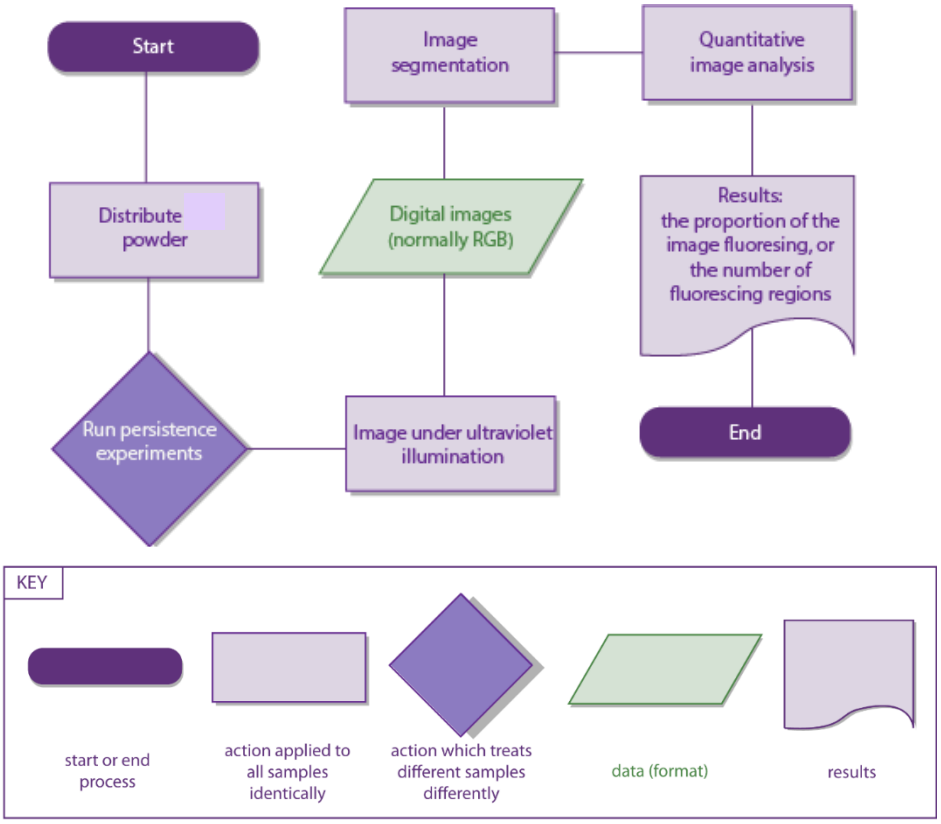


Figure 3.1: Flowchart depicting the typical workflow during a persistence study involving a fluorescent proxy

The first step to enable the analysis of the digital images is a process called 'segmentation'. In the simplest possible terms, this is taking an image, and dividing it into sections that are of interest (i.e. 'the foreground') and regions that are not of interest (i.e. 'the background') (Sezgin and Sankur, 2004). The regions that *are* of interest can then be quantified.

There are numerous ways of segmenting an image, with advantages and disadvantages to each (e.g. Kaur and Kaur, 2014; Figure 3.2). It can be noted, at this early stage, that some methods of segmentation require human input (such as tracing, manually delineating the objects of interest) while others (such as edge-detection) are fully-automated processes which require no human intervention (Kaur and Kaur, 2014). Some methods occupy a grey area – processes which are automated, but require the human definition of parameters (such as global thresholding, where the level of the threshold is manually-defined, or local thresholding where the radius is manually defined (Bankhead, 2014)).

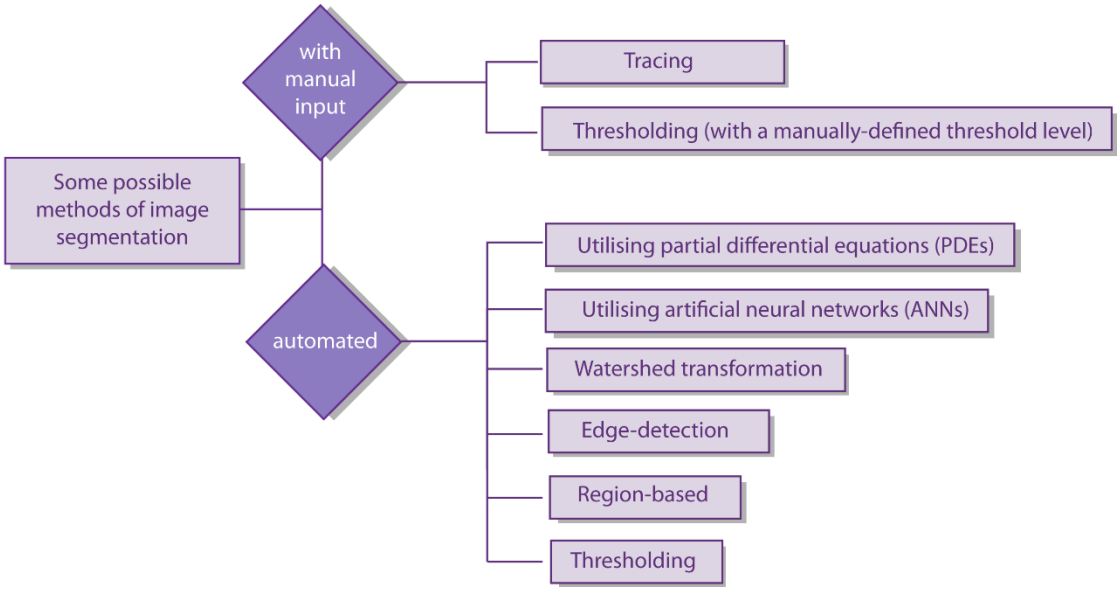


Figure 3.2: Some of the many methods of image segmentation
(figure adapted from Kaur and Kaur 2014, and reproduced from Levin et al., 2019)

Thresholding is one of the most frequently-used methods of segmentation for fluorescence imagery (Oberholzer *et al.,* 1996; Bankhead, 2014). Thresholding coverts a colour or greyscale image, where each pixel may have many possible values for the level of brightness, into a binary image, where each pixel may have one of two levels of brightness (Sezgin and Sankur, 2004).

How it achieves this is by the definition of a threshold value, and then the performance of a mathematical operation upon each pixel – if the pixel's brightness value was higher than the threshold value, then the pixel is converted to white, and if the pixel's brightness value is lower than the threshold value, it is converted to black (see: Figure 3.3; Bankhead, 2014). This results in a binary image, which can look quite different depending upon the level at which the threshold is set (Figure 3.3).

Thresholds can be global (i.e. the same threshold value is applied to each pixel in the image) or local (i.e. a threshold value is calculated for each pixel from the brightness values of the neighbouring pixels in a defined radius (Bankhead, 2014; Zaitoun and Aqel, 2015). Thresholds can also be manual (i.e. a user can define the threshold level) or automated (an algorithm can be used to calculate the optimal threshold value (ibid). It is commonly accepted that global thresholds may struggle with noisy images (Bankhead, 2014), and that different thresholding algorithms may be appropriate for different applications, with most exhibiting varying levels of success when applied to different types of images (e.g. Hoover *et al.,* 1996; Sezgin and Sankur, 2004; López-Leyva *et al.,* 2016).

3.1.3 Methods used to segment fluorescent powder

As mentioned in the previous section, thresholding is one of the most frequently-used methods of segmentation for fluorescence imagery, especially within the context of histology (Oberholzer et al., 1996; Bankhead, 2014). Considering specifically the context of fluorescent proxies within forensic transfer and persistence studies (e.g. Morgan *et al.,* 2009; 2010; 2013; French *et al.*, 2012; Slot *et al.,* 2017) an issue

becomes apparent. There is not a standardised technique for the segmentation of fluorescent powder within these studies (Table 3.2).

While some of the studies openly stated that they used thresholding (e.g. Slot *et al.*, 2017), it was unclear whether the threshold value was defined by an algorithm or the researcher (revisit Table 3.2). This is problematic since there has been some suggestion that the process of manually-defining a threshold level may not be reproducible, and may therefore cast aspersions upon the reliability of a study (Yang *et al.*, 2001; Baveye, 2002; Baveye *et al.,* 2010). Likewise, it is common knowledge in computer science and image processing that the use of different thresholding algorithms will result in the calculation of different threshold values (and sometimes the production of very different output images) – raising questions over the accuracy with which the segmented outputs might reflect the original input images (Hoover *et al.,* 1996; Sezgin and Sankur, 2004; Griffin *et al.,* 2012; Zaitoun and Aqel, 2015; see Figure 3.4).

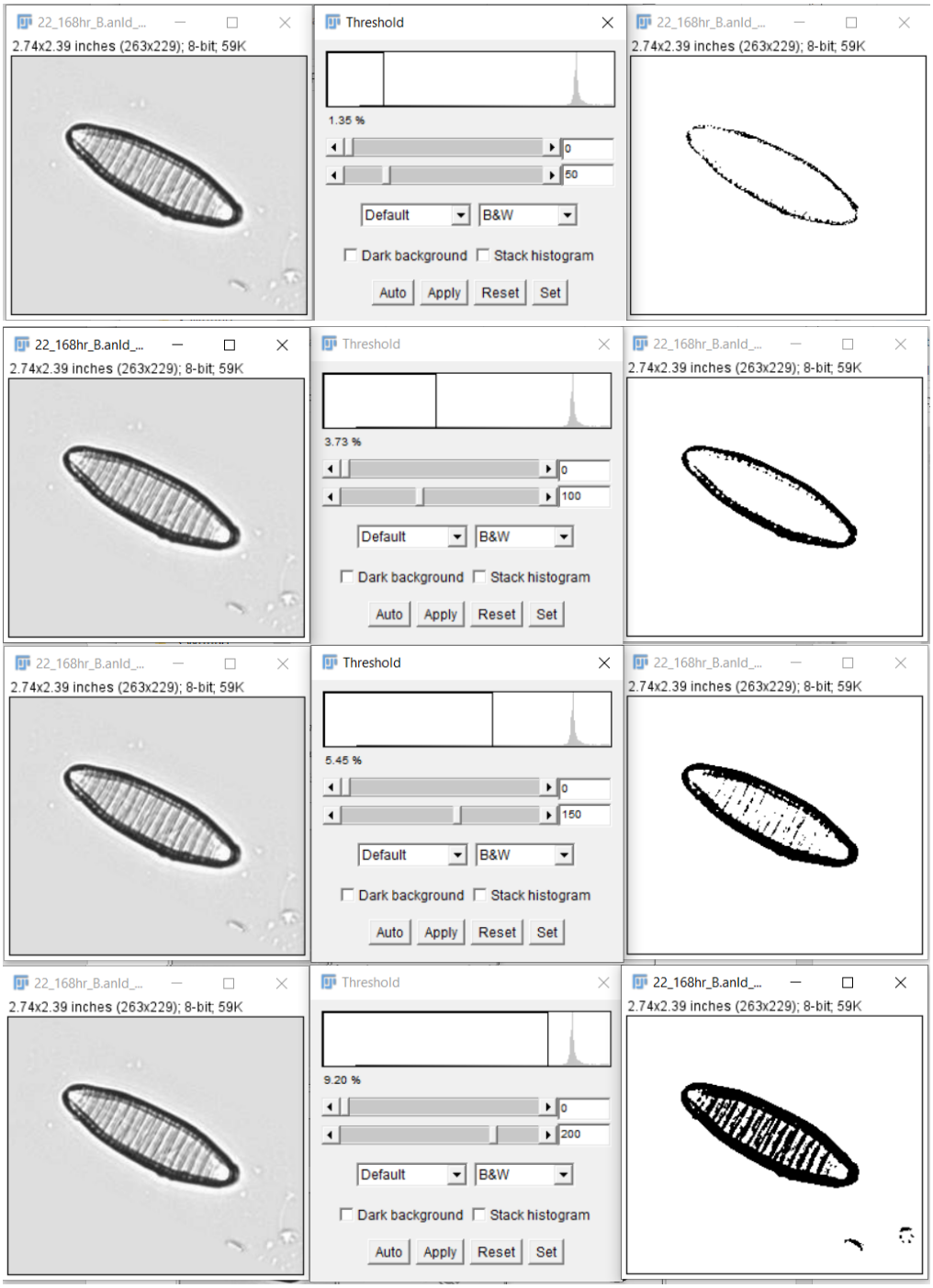


Figure 3.3: Diagram demonstrating how a global threshold works (upon an 8 bit greyscale image of a diatom.) Note that as the threshold value increases, an increasing number of pixels are denoted as foreground (black).

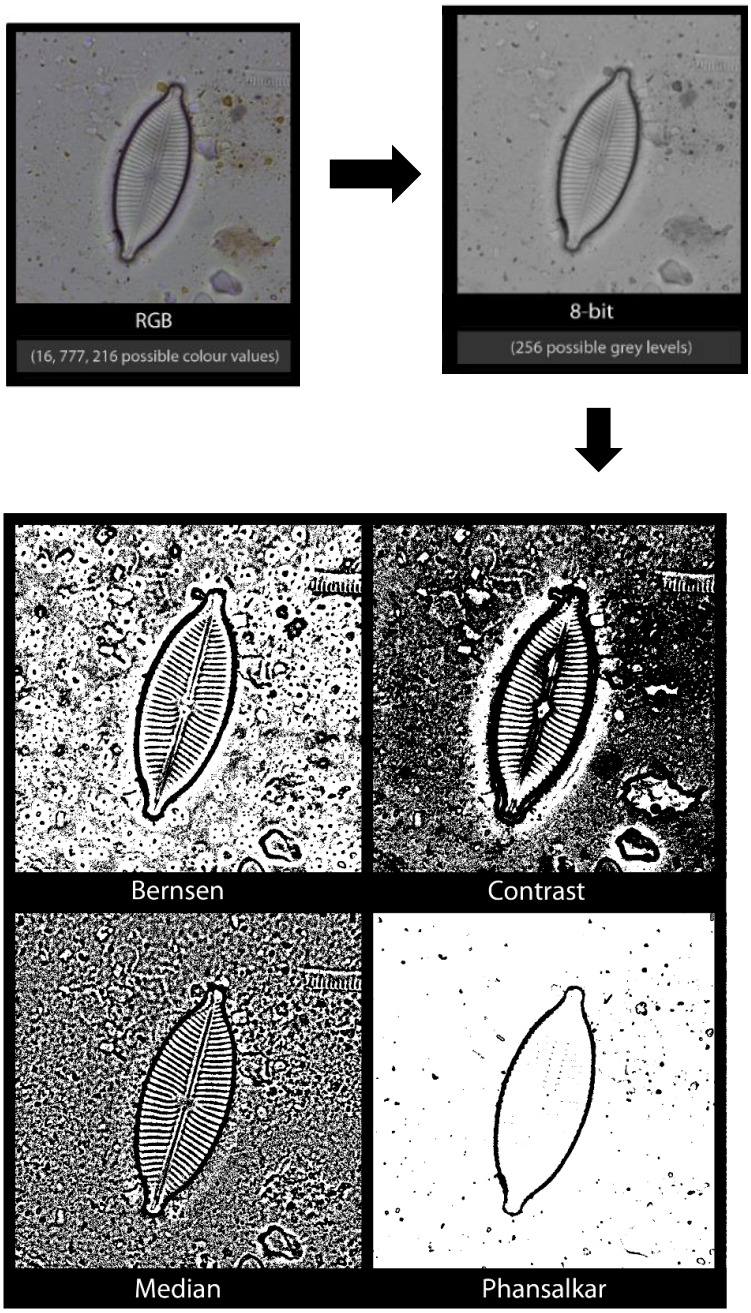


Figure 3.4: The results of applying four different local thresholding algorithms to the same input image (in this case a diatom, a form of environmental trace evidence). Algorithms were applied in ImageJ, with a radius of 15 pixels (from L-R, Bernsen, Contrast, Median, and Phansalkar). Reproduced (and modified) from Levin et al 2019

Year	Authors	Image processing and analysis method (verbatim)
2006	Bull *et al.*	"photographs were pixelated using Corel Photo-Paint 9 and the number of particles in each image were computed (as a function of pixel brightness)" (Bull *et al.,* 2006: 191)
2009	Morgan *et al.*	"This digital image was then pixelated in IDRISI to provide an indication of the amount of silt-sized material remaining on the sole." (Morgan *et al.,* 2009: 9)
2010	Morgan *et al.*	" Presence of fluorescent powder and any observations regarding particular patterns or noteworthy features were recorded and subsequently documented using the macro-zoom feature on a digital camera. Secondly, the designated areas were sampled using aluminium SEM sample stubs coated with black carbon discs (to render any powder visible) and were examined using a light microscope lit by an ultra-violet lamp." (Morgan *et al.,* 2010: 196)
2012	French *et al.*	"The presence of UV powder on the stub was […] quantified using an image rasterisation technique in MATLAB which was specifically adapted for the specifications of this study from Bull et al.[2006]" (French *et al.,* 2012: 34)
2013	Morgan *et al.*	"The digital images taken of each experiment were imported into Coral Photo Paint 11 [sic]. The images were graphically enhanced, pixelated and then 5 sections of 32 × 7 pixels were counted" (Morgan *et al.,* 2013: 377)
2017	Slot *et al.*	"A Matlab (version R2012b) [10] algorithm was created to enable fast automated counting of flock fibres on pictures. The individual pictures were loaded into Matlab and processed automatically. Firstly, the original RGB (red, green, and blue) colour images were converted to a grey value image by extracting the green channel. Subsequently, the foreground (i.e. the flock fibres) was separated from the background (i.e. the target materials) by thresholding. A region of interest (ROI) was selected as well. Next, the fibres were counted. As a result of the varying illumination conditions, the size of one fibre (in pixels) was not the same on all pictures and therefore had to be estimated for each image first. Subsequently, adding up all the foreground pixels and dividing them by the estimated number of pixels per fibre, yielded the amount of fibres on an image." (Slot *et al.,* 2017: 181)

Table 3.2: Examples of the methodologies used to process imagery from persistence studies which employ a fluorescent proxy for trace evidence. Modified from Levin et al., 2019

3.1.4 Aims

The aims of this study were to systematically evaluate the performance of a number of common methods of segmentation upon images of fluorescent powder analogous to those which would be utilised in a study of transfer and persistence. Specifically, the aim was to compare three techniques: (1) Manual tracing (2) Manually-defining a threshold value, and (3) Thresholding where an algorithm calculates the threshold value, and to evaluate three facets of their performance: (1) Accuracy (2) Reliability (3) Efficiency (please see section 3.2.5 for precise definitions).

Three questions were under investigation in three experiments:

1. One: Evaluating the performance of manual segmentation
   In the context of segmenting fluorescent powder from a dark fabric background, does manual segmentation (i.e. 'tracing') create reproducible segmentations? Or do differences in outputs exist when images are segmented by different examiners, or repeatedly by one examiner?

2. Two: Evaluating the performance of manual thresholding
   In the context of segmenting fluorescent powder from a dark fabric background, does segmentation by manually-defining a threshold value result in accurate, reproducible segmentation? Or do differences in outputs exist when images are segmented by different examiners, or repeatedly by one examiner?

3. Three: Evaluating the performance of thresholding algorithms
   Can thresholding algorithms produce segmentations comparable to those produced by manual segmentation (i.e. tracing)? Given the known variation in the performance of thresholding algorithms upon different image types, is there one algorithm (or a suite of algorithms) which offers the best performance for this form of fluorescence imagery?

3.2 Materials & Methods

3.2.1 Materials: Images Used

The images used in these experiments were produced as part of a persistence study (please see Chapter 5 for further details and results of the experiment).

In short: the images depicted green fluorescent powder upon swatches of black cotton fabric, under ultraviolet illumination. The creation of these swatches followed the conventions of forensic experiments using fluorescent proxies; powder with a mean grain diameter of approximately 15 µm (see: French *et al.,* 2012), was transferred to the recipient surface, in this case cotton, by 'flicking' the powder with a stiff bristled brush to distribute it evenly (after Morgan *et al.,* 2010). This powder was a 'thief detection' product, designed to fluoresce green under a wavelength of 385nm[4]. The ultraviolet illumination was supplied by a torch, with a wavelength of 385 nm, and imaging was carried out using a bespoke UV-dark box (after French *et al.,* 2012), which "excluded visible light and held the torch and camera lens a fixed distance from the samples" (Levin *et al.,* 2019: 433; please see section 5.2.4). The image format was 8-bit RGB colour. It is acknowledged that this method of data acquisition relied on trust in the black-box method of the software that generated the images, as is the convention in this field of experimental forensic science, where data acquisition tends to use either DSLRs, or less sophisticated digital cameras (e.g. French *et al.,* 2012; Morgan *et al.,* 2013). Were this research field to be pursued further, or were these experiments to be repeated, it is acknowledged that further image characterisation would be useful if not vital; considering the exact magnification and resolution of the images, having knowledge about the sensors and image quality; for example, considering uniformity, noise, focus, vignetting, compression, pixel resolution, spatial resolution, Bit depth, and the presence of any dead pixels on the sensor. Such information would be essential for the data to be used more generally.

[4] https://uvgear.co.uk/uv-detective/theft-and-tamper-prevention/thief-trapper-invisible-uv-powder-green-blue

It is also acknowledged that this represents the first attempt at conducting this kind of work for the imaging of particles that could be used in forensic science experiments (i.e. fluorescent powder).

Additionally, it is acknowledged that the use of a bespoke UV dark box introduces a number of limitations. Unlike a dark box and light source manufactured by an established forensic brand (for example, Foster and Freeman's CrimeLite series[5]), there was no guarantee of constant light output as battery charge levels changed, no guarantee of the even-ness of illumination (although efforts were made to ensure that UV light was reflected within the box, with the internal walls coated in reflective tinfoil), and although every effort was made to ensure that the camera lens and sample were parallel (i.e. that the images were not taken at an angle, introducing foreshortening to the image), this was not certified. Accordingly, the use of a bespoke box, while a cost-effective and accessible solution, introduces a number of unknowns into the data acquisition.

Two different sets of images were used in this study: for experiments two and three, a suite of ten images (Figure 3.5) was used, while for experiment one, a suite of three images in three rotations was used (Figure 3.6). More information can be found in sections 3.2.2, 3.2.3, 3.2.4, and below.

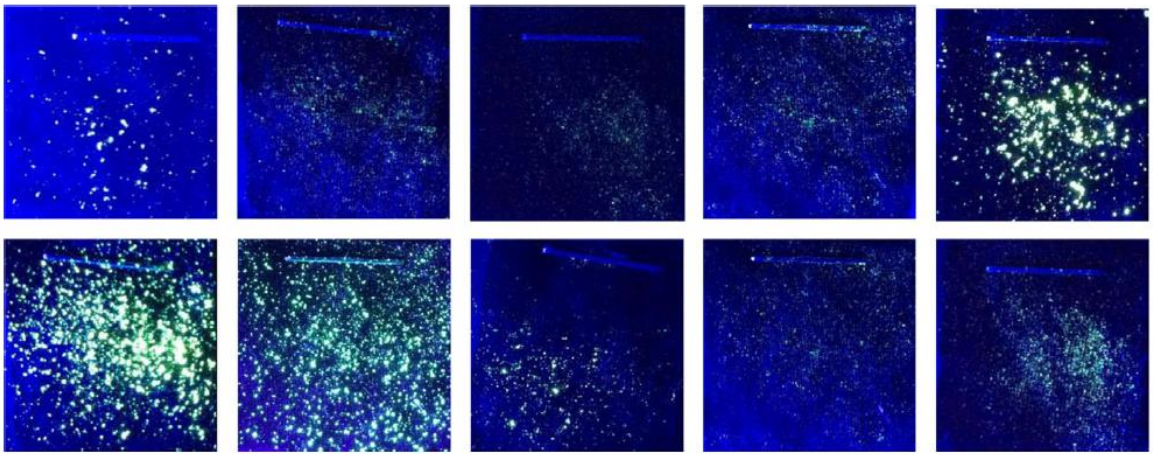


Figure 3.5: The suite of 10 images used in Experiments two and experiment three, figure reproduced from Levin et al 2019

[5] http://www.fosterfreeman.com/index.php/forensic-light-sources/1-crime-liter-82w

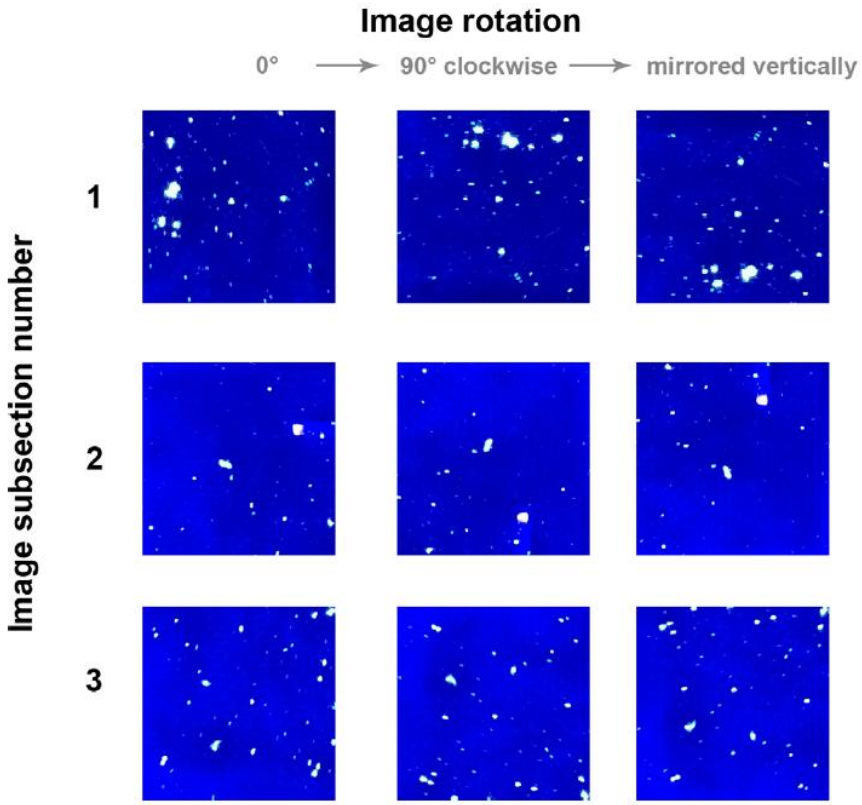


Figure 3.6: The suite of three images (and their rotations) used in experiment one. Figure reproduced from Levin et al. 2019

One phenomenon that has been demonstrated repeatedly within studies which assess the performance of segmentation methods is that different thresholding algorithms can perform with varying levels of success upon different types of images; some perform optimally where images possess unimodal brightness histograms, others show superior performance where the histograms are bimodal (Sezgin and Sankur, 2004; López-Leyva *et al.,* 2016). Similarly, some thresholding algorithms struggle to produce accurate segmentations where images are very noisy (e.g. Bankhead, 2014). Accordingly, since the purpose of this Chapter was to assess the performance of the segmentation methods in the context of a fluorescence imaging for a forensic persistence study, "it was considered important that the thresholding algorithms were tested upon images representative of all of the histogram shapes that one might encounter in a study involving fluorescent powder as a proxy for forensic trace evidence" (Levin *et al.,* 2019: 435).

Consequently, the suite of 10 images used in experiments two and three (Figure 3.5) "were selected by stratified sampling from the available images" (Levin *et al.,* 2019: 435) in order to contain specimens of all of the types of images (i.e. patterns of histograms) observed within the source population of images (the fluorescent persistence study outlined in Chapter 5; see Figure 3.7)

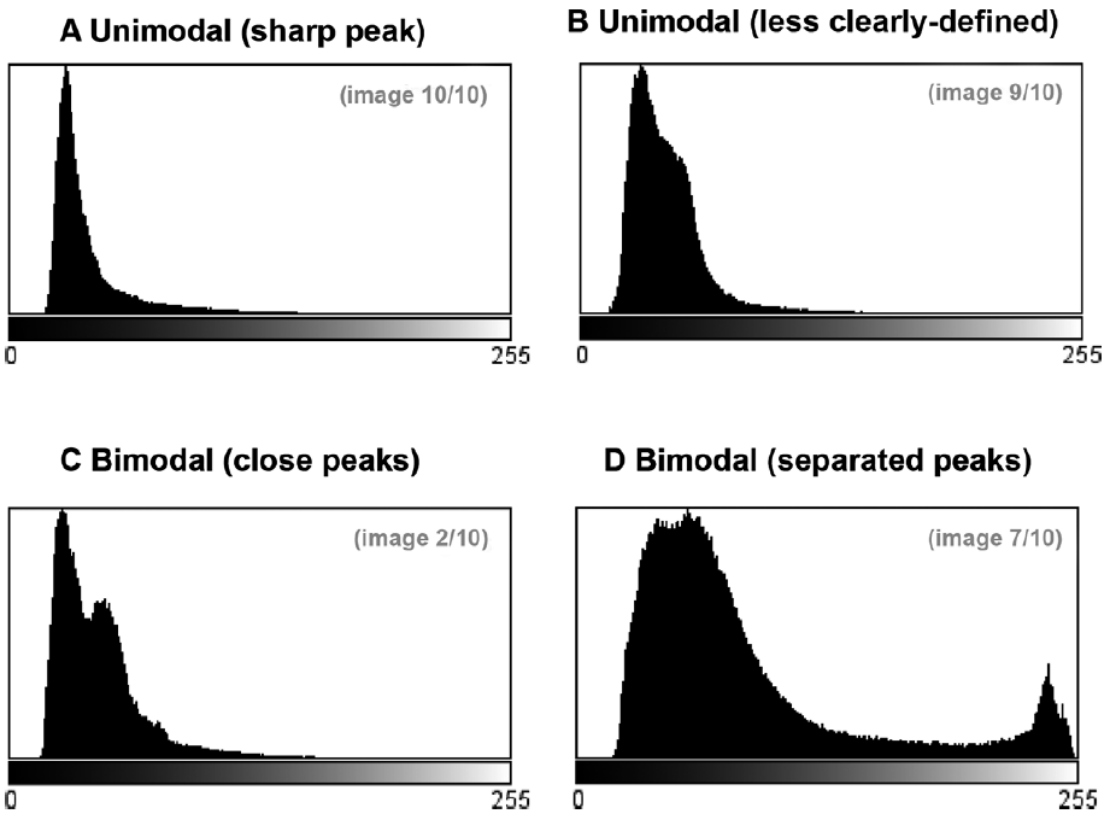


Figure 3.7: Examples of the four different histogram types observed in the images used in Experiments 2 and 3. Figure reproduce from Levin et al., 2019

3.2.2 Methods: Experiment One

As stated in the introduction, the aim of Experiment One was to evaluate the extent to which manual segmentation (tracing) could be considered to result in reproducible segmentations for fluorescence imagery of fluorescent powder on a fabric background. This was considered a valuable experiment, since several authors working in analogous contexts (e.g. Yang *et al.,* 2001; Baveye *et al.,* 2002; 2010) have debated whether manual segmentation (tracing) might actually produce variable outputs, with differences observed between the efforts of different examiners (Baveye *et al.,* 2010), and between in the outputs of a single examiner repeating analysis.

For this experiment, a set of three images was given to three participants (Figure 3.6). These images were of uniform size (300 pixels in width and height), and all of the images depicted relatively sparse amounts of fluorescent powder (i.e. the majority of the image was non-fluorescent background). The three images were given to the examiners in three separate rotations, and presented as a set of nine images (Figure 3.6). The participants were not aware that the set of nine images that they were given contained any duplication.

The participants were instructed to segment each image to the best of their ability, by tracing over the pixels that they believed to represent foreground (i.e. fluorescent matter) in white, and to trace over the pixels that they believed to represent background (i.e. non-fluorescent matter) in black. This tracing was carried out in Adobe Photoshop CS4, using the pencil, brush, and fill tools. The participants, who were computer literate, but not specialists in Photoshop, were given instruction in how the software worked, and encouraged to conduct the tracing by creating a new, separate layer above the image to be traced, and working on this layer with a low opacity value, so that they could see the original underneath. This also gave participants the ability to toggle the visibility of their tracing on and off, to compare it to the original image.

3.2.3 Methods: Experiment Two

The aim of experiment two was very similar to the aim of experiment one; to evaluate the extent to which segmentation by manually-defining a threshold value could be considered to result in reproducible segmentations for fluorescence imagery of fluorescent powder on a fabric background. Again, this experiment was interested in both the reproducibility seen between different examiners, and seen for one examiner repeatedly performing the task. For this experiment, a set of ten images were used (Figure 3.5). In order to test the reproducibility of this type of segmentation between different examiners, twenty participants were asked to manually-define an appropriate threshold value to segment each of the ten images (once per image). In order to test the reproducibility of the technique for one examiner repeating their analysis, one participant was asked to manually-define an appropriate threshold value to segment each of the ten images, twenty times for each image. This analysis was conducted over a number of weeks, so that the examiner could not remember the answers given in prior runs. See Figure 3.8 for a flowchart of this experiment.

Thresholding was conducted in Adobe Photoshop (CS4). Participants were asked to use the pathway Image > Adjustments > Threshold, to open up the global thresholding dialogue box. Participants could then select the threshold level that they felt offered the best division of the image into foreground and background. Using this dialogue box, it was possible to toggle a 'preview' of the segmentation on and off.

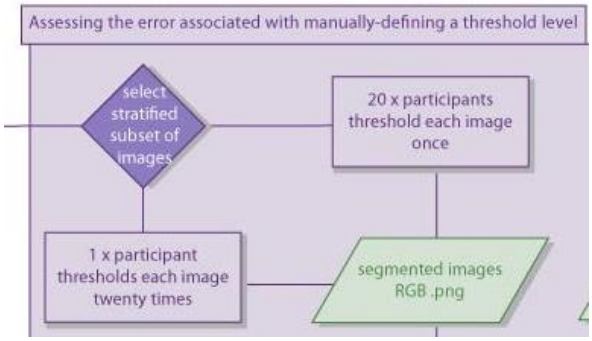


Figure 3.8: Flowchart summarising data generation in experiment two

3.2.4 Methods: Experiment Three

The purpose of Experiment Three was to evaluate the performance of a suite of thresholding algorithms, in order to assess whether they might be a viable and appropriate alternative to a method of segmentation that required manual input for segmenting this sort of fluorescence imagery. Sixteen different global thresholding algorithms were used (Table 3.3).

The thresholding algorithms were applied to the same suite of ten images used in Experiment two. They were applied in NIH ImageJ, a powerful image processing programme distributed under a freeware licence, and supported by an active community which provides both software support, and a constantly updated library of plugins, which can extend and customise the functionality of the software (Schneider *et al.,* 2012; Schulze *et al.,* 2013; Legland *et al.,* 2016; Arena *et al.,* 2017).

ImageJ was deemed a sensible choice for this research, since: (1) Being accessible, that is to say, having a friendly graphical user interface and no financial cost to the user, it would be possible for other researchers within forensic science, who may not have a computer-science background, to implement a thresholding algorithm in their own research; and (2) Despite having a very accessible interface, ImageJ also allows for the use of macros, so it was possible to automate the analysis of very large batches of images (Schneider *et al.,* 2012; https://imagej.nih.gov/ij/developer/macro/macros.html)

Consequently, the thresholding algorithms were applied in ImageJ with the plugin "Auto Threshold v1.15" (https://imagej.net/Auto_Threshold), and "macros were written in order to open, process, and save the outputs of batches of images" (Levin *et al.,* 2019: 435).

Algorithm Name	Further information and mathematical notation available from:
Default	https://imagej.net/Auto_Threshold#Default
Huang	Huang, L-K & Wang, M-J J (1995), "Image thresholding by minimizing the measure of fuzziness", Pattern Recognition 28(1): 41-51
Intermodes	Prewitt, JMS & Mendelsohn, ML (1966), "The analysis of cell images", Annals of the New York Academy of Sciences 128: 1035-1053
IsoData	Ridler, TW & Calvard, S (1978), "Picture thresholding using an iterative selection method", IEEE Transactions on Systems, Man and Cybernetics 8: 630-632
Li	Li, CH & Tam, PKS (1998), "An Iterative Algorithm for Minimum Cross Entropy Thresholding", Pattern Recognition Letters 18(8): 771-776
MaxEntropy	Kapur, JN; Sahoo, PK & Wong, AKC (1985), "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram", Graphical Models and Image Processing 29(3): 273-285
Mean	Glasbey, CA (1993), "An analysis of histogram-based thresholding algorithms", CVGIP: Graphical Models and Image Processing 55: 532-537
MinError(I)	Kittler, J & Illingworth, J (1986), "Minimum error thresholding", Pattern Recognition 19: 41-47
Minimum	Prewitt, JMS & Mendelsohn, ML (1966), "The analysis of cell images", Annals of the New York Academy of Sciences 128: 1035-1053
Moments	Tsai, W (1985), "Moment-preserving thresholding: a new approach", Computer Vision, Graphics, and Image Processing 29: 377-393
Otsu	Otsu, N (1979), "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber. 9: 62-66
Percentile	Doyle, W (1962), "Operation useful for similarity-invariant pattern recognition", Journal of the Association for Computing Machinery 9: 259-267
RenyiEntropy	Kapur, JN; Sahoo, PK & Wong, AKC (1985), "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram", Graphical Models and Image Processing 29(3): 273-285
Shanbhag	Shanbhag, Abhijit G. (1994), "Utilization of information measure as a means of image thresholding", Graph. Models Image Process. (Academic Press, Inc.) 56 (5): 414--419
Triangle	Zack GW, Rogers WE, Latt SA (1977), "Automatic measurement of sister chromatid exchange frequency", J. Histochem. Cytochem. 25 (7): 741–53
Yen	Yen JC, Chang FJ, Chang S (1995), "A New Criterion for Automatic Multilevel Thresholding", IEEE Trans. on Image Processing 4 (3): 370-378

Table 3.3: The Global Algorithms Implemented in this study

3.2.5 Metrics of performance

As a starting point, it is important to establish and define the metrics used to gauge the success of the techniques. In this study, three different facets of performance were considered;

1. The *accuracy* of the outputs produced by the technique
   The accuracy is defined, in this experiment, as the extent to which the output image resembled the input image. In previous studies which evaluate the performance of segmentation methods (e.g Yang *et al.,* 2001; Udupa *et al.,* 2006; Griffin *et al.,* 2012), this comparison was made by comparing the output of the segmentation method to an output generated by manually-tracing the images (which was used as a substitute for a 'ground-truth' value) (Crum *et al.*, 2006). It is, accordingly, not a 'true' truth value, but is (a) the convention and (b) accepted to be likely represent the most accurate representation of the object's extent (see: Udupa *et al.,* 2006). In these experiments, the outputs of the different methods of segmentation were therefore compared to the "mean manual segmentation values" generated (n = 3; see Levin *et al.,* 2019). These 'manual segmentation values' were generated by one examiner 'tracing' each image used in the experiments three times (see: Levin *et al.,* 2019). In terms of producing quantitative comparisons, three calculations were performed:

   a. Absolute Error: (i.e. the residual between the percentage of the segmentation which represented foreground, and the mean percentage of the three manual segmentations which was defined as foreground). The calculation of the percentage of each image which represented foreground was conducted in ImageJ, using the command "Analyse > Analyse Particles" (see: Levin *et al.,* 2019).

   b. Relative Error: (i.e., the percentage mis-estimation compared to the ground truth (mean manual segmentation) value). Since the

different images had varying percentages of foreground, it was felt important to provide this relative measure.

c. The Dice Coefficient of Similarity

"In order to ensure that the images had been segmented with fidelity, it was important to ensure that the segmented outputs were not just quantitatively similar (i.e., that, for example, both classified 5% of the pixels in the image as foreground) but also that they overlapped (i.e., that they identified the same 5% of the pixels as foreground)." (Levin *et al.,* 2019: 436). For these purposes, the Dice Coefficient (a commonly-utilised coefficient for the comparison of the performance of segmentation) was employed (see: Crum *et al.,* 2006). The Dice Coefficient: "consider[s] two images, asking whether each pixel is classified as foreground in either, both, or neither of the images (i.e., the area of the image which represents true and false positives and negatives, with 0 indicating no overlap between the images, and 1 indicating total overlap" (Levin *et al.,* 2019: 436; see Figure 3.9). The calculation of the Dice Index was undertaken with the LIM Tools Similarity Index plugin in ImageJ (available at: https://github.com/HGGM-LIM/limtools, see: Mateos-Pérez, 2014)

For the experiment which evaluated the reproducibility of manual segmentation (tracing), mean similarity index values were calculated conducting comparison of each of the images in turn (i.e. by comparing segmentation 1 to segmentations 2 and 3, and comparing segmentation 2 to segmentation 3), and then calculating the arithmetic mean of those three values. For the experiments which considered the performance of thresholding algorithms, the Dice Coefficient was calculated once, comparing the output of the thresholding algorithm to the first output produced by tracing (i.e. manual segmentation).

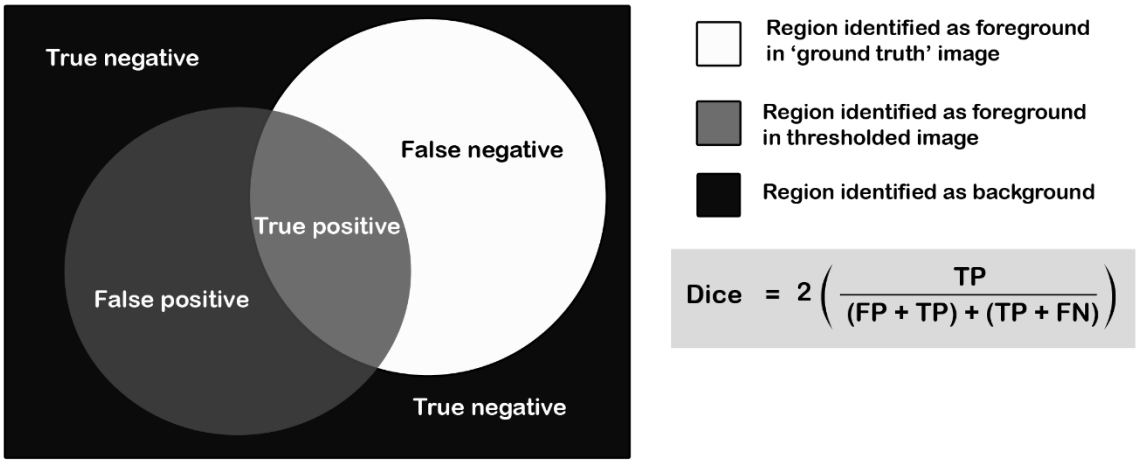


Figure 3.9: An explanation of the calculation of the Dice Coefficient of similarity. FN, false negative; FP, false positive; TN, true negative; TP, true positive. Reproduced and Modified from Levin et al 2019

2. The *reproducibility* of the technique

   The second quality that was evaluated was: "the reproducibility of each segmentation technique (i.e., the extent to which measurements were reliable)" (Levin *et al.,* 2019: 432). The reproducibility of the thresholding algorithms was not an issue: given the same input, and the same parameters, an algorithm will always produce the same output (that is to say, any error will be *deterministic,* determined by the initial states). In contrast, any segmentation method which involves a degree of manual input or decision-making (as in the cases of tracing or manually-defining an appropriate threshold value for the application of a global thresholding operation) may also introduce *stochastic* error, that is a sort of random error that cannot be predicted or modelled from the initial states. In other words, "it is entirely possible that providing an examiner with the same input image repeatedly, or providing multiple examiners with the same input image, might result in dissimilar outputs" (Levin *et al.,* 2019: 432). For future work, it would be ideal to have a segmentation method that could be considered reproducible: i.e. if different experimenters wanted to replicate a study in a different lab,

or to revisit a dataset, they could expect to see results that are similar. Consequently, this analysis considered the ranges of results observed for repeated analysis and analysis by different examiners, and whether the differences between these runs and participants were shown to produce p-values of <0.05 using inferential statistical analysis.

3. The *efficiency* of each technique

   The third metric of performance considered was the amount of time that would be required to implement the technique. This is not a trivial quality; historically, the time-intensive nature of manual visual analysis has constrained the size of datasets and the scope of hypotheses that can be pursued (Du Buf and Bayer, 2002). Accordingly, one should consider the rapidity (and efficiency) of a segmentation technique as an important measure of viability (Udupa *et al.,* 2006). In this study, the metric employed is 'time taken to process one image'.

3.2.6 Statistical Analysis

Two forms of statistical analysis were used in this chapter; descriptive and inferential. The descriptive analysis, conducted for experiments 1, 2, and 3, involved the calculation of the mean (a measure of central tendency) and the standard deviation (a measure of dispersion) for the samples - a combination commonly used to present an effective summary of results (Field and Hole, 2007). For experiment 1, the inferential analysis used was ANOVA (Analysis of Variance), which was appropriate because the data (1) had three or more experimental groups; (2) Different participants were used in each group (i.e. each person contributed one score to the data, which were independent measures), and (3) The aim of analysis was to establish whether there were differences between the groups. (Please see Field and Hole, 2007, p275, for a flowchart which outlines the appropriate tests to use, given this information).

3.3 Results

3.3.1 Results: Experiment One (manual segmentation)

The results of experiment one are displayed in Figure 3.10 (a scatter chart, which depicts the percentage of each image that each examiner decided was foreground (n=3)), Table 3.4, which provides descriptive and inferential statistics for the percentage data, and Table 3.5, which presents the Dice Index values.

Considering the percentage data, two observations can be made. Primarily, it is possible to note that most of the data points produced by each examiner do not overlap (Figure 3.10); that is to say, each time an examiner made an estimate of the extent of the foreground, they mostly arrived at a slightly different value for the extent of the foreground. Secondly, the mean values produced by the three examiners also showed variation (Table 3.4) – so, not only was there a range of values, there was a range of means around which these values were clustered. Inferential statistical analysis was conducted upon these data (ANOVA: see Table 3.4), and it was observed that for all three of the images, the variation in the percentage of the image that was identified as foreground by the three different examiners was seen to be statistically significant at $p < 0.01$ (Table 3.4). In other words, the values for each examiner clustered around a mean value that was statistically significantly different to the mean values identified by the other examiners.

Considering the Dice Coefficient data, another two observations can be made. Primarily, it is possible to observe that the mean Dice Coefficient values obtained when comparing all of the segmentations for all of the participants were relatively high, at 0.80 ± 0.12 (n=27), indicating a good degree of overlap. Secondly, it was possible to observe that the different examiners produced different mean Dice Coefficient values. Examiner 1 produced the lowest mean Dice Coefficient value, of 0.71 ± 0.08 (n=9), Examiner 3 produced a similar mean Dice Coefficient value of 0.75 ± 0.02 (n=9), and Examiner 2 produced the highest mean Dice Coefficient value, of 0.95 ± 0.02 (n=9; Table 3.5).

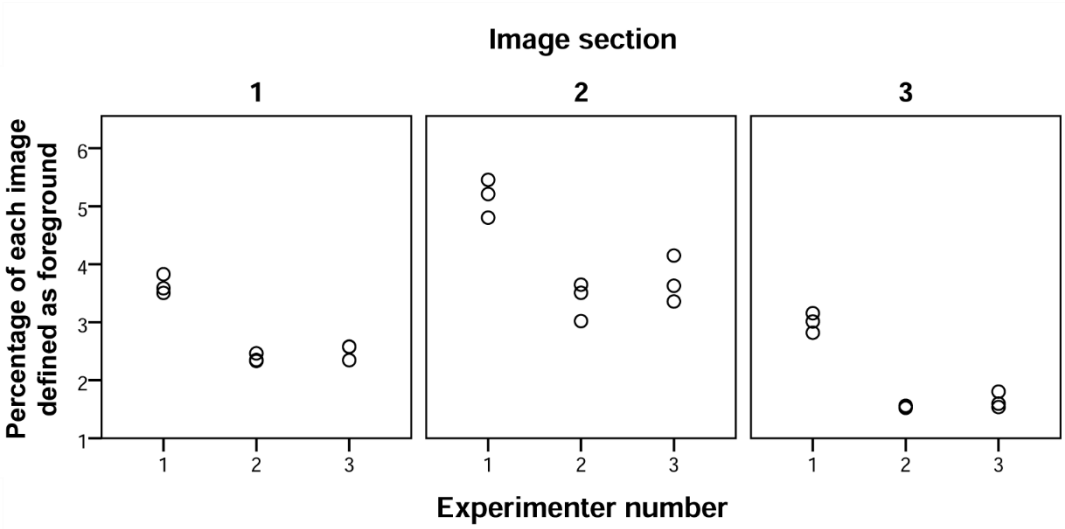


Figure 3.10: Variation in the percentage of each image section defined as foreground (n = 3) in Experiment 1. Reproduced from Levin et al 2019

Image section	Examiner	N	Mean	SD	ANOVA results, comparing examiners		
					f-ratio	p-value	p<0.05?
1	1	3	3.64	0.17	84.817	0.00004	✓
	2	3	2.38	0.07			
	3	3	2.50	0.13			
2	1	3	5.16	0.33	20.986	0.00196	✓
	2	3	3.39	0.33			
	3	3	3.71	0.40			
3	1	3	3.00	0.17	122.184	0.00001	✓
	2	3	1.54	0.02			
	3	3	1.65	0.14			

Table 3.4: Summary table for the percentage of each image defined as foreground. Reproduced from Levin et al 2019

Experimenter number	N	Dice Index value			
		Mean	SD	Min	Max
1	9	0.71	0.07	0.56	0.77
2	9	0.95	0.02	0.90	0.97
3	9	0.75	0.02	0.71	0.79
Total	27	0.80	0.12		

Table 3.5: The mean Similarity Index values for the three examiners' output images, when comparing segmentation attempts 1 and 2, 1 and 3, and 2 and 3. Reproduced from Levin et al 2019

3.3.2 Results: Experiment Two (manual thresholding)

Results for experiment two are presented in Figure 3.11 (Boxplots showing the distributions of (a) the absolute error in the percentage of the image defined as foreground, (b) the relative error in the percentage of the image defined as foreground, and (c) the Dice Coefficient values) and Table 3.6, which provides descriptive statistics.

With respect to accuracy, the technique appeared to perform well *on average*; for both the images segmented by the one examiner and the images segmented by the twenty examiners, when the mean values for the estimated extent of the foreground were compared to the values generated by manual segmentation, there was <1% absolute difference (Figure 3.11A). For the images segmented by one examiner, this translated to a mean residual of -0.96 ± 3.73% (n = 200), and for the images segmented by twenty examiners, the mean residual was +0.95 ± 6.45% (n=200). In terms of the relative mis-estimation, this translated to a mean mis-estimation of -21.06 ± 41.73% (n = 200) and +11.50 ±96.68% (n = 200) respectively, when compared to the mean manual segmentation value.

With respect to reproducibility, both the inter-examiner experiment and intra-examiner experiment produced a much greater range of residuals and Dice Coefficient values than the manual segmentation (Figure 3.11). Multiple outliers were observed, with high values for error or low Dice Coefficient values. The most extreme outliers were seen for the images segmented by the twenty different examiners, which included one overestimation of over 900%, and multiple Dice Coefficient values below 0.2. This suggests that while the averages may approximate the values obtained by manual segmentation, some of the segmentations deviated from the values and distribution of the percentage suggested by manual segmentation.

Considering the efficiency of the technique, performing a global thresholding operation where the threshold value was manually-defined took around 1 minute for image. For comparison, this was much faster than the manual segmentation by tracing, which required 1 hour for each image.

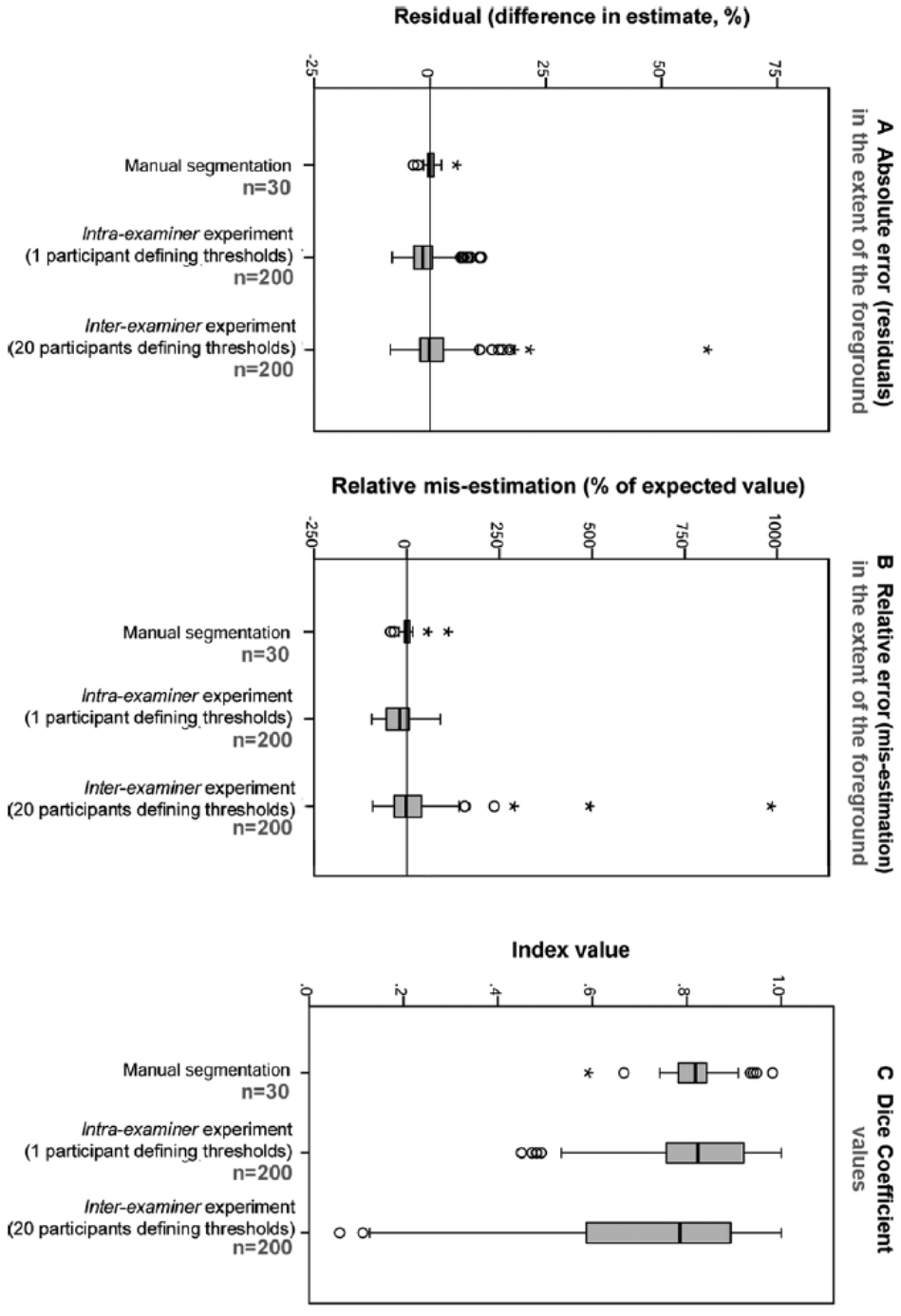


Figure 3.11: Boxplots showing (A) The Absolute error, (B) The Relative error, and (C) The Dice Coefficient values for the manual methods of segmentation. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median). Reproduced from Levin et al., 2019

Segmentation method		(A) Absolute Error (residuals) in the Extent of the Foreground (%)			(B) Relative Error (mis-estimation) in the Extent of the Foreground (%)			(C) Dice Coefficient Values		
		N	M	SD	N	M	SD	N	M	SD
Manual segmentation (tracing)		30	0.11	1.62	30	2.35	26.65	30	0.82	0.08
Manually-defined threshold (1 examiner)		200	-0.96	3.73	200	-21.06	41.73	200	0.82	0.13
Manually-defined threshold (20 examiners)		200	0.95	6.45	200	11.5	96.68	200	0.72	0.23
Global Algorithms	Default	10	3.84	7.74	10	70.48	144.25	10	0.62	0.30
	Huang	10	25.44	21.11	10	567.2	594.79	10	0.33	0.32
	Intermodes	10	-3.22	2.77	10	-51.82	36.62	10	0.74	0.22
	IsoData	10	2.98	6.20	10	55.49	116.19	10	0.63	0.29
	Li	10	9.61	11.37	10	159.87	211.10	10	0.54	0.34
	MaxEntropy	10	-0.77	4.72	10	-22.4	51.96	10	0.79	0.16
	Mean	10	24.26	9.41	10	466.74	329.82	10	0.32	0.30
	MinError	10	53.74	20.58	10	917.85	565.52	10	0.2	0.20
	Minimum	10	-5.09	2.69	10	-73.91	31.45	10	0.33	0.31
	Moments	10	1.92	3.13	10	28.79	42.01	10	0.66	0.28
	Otsu	10	1.44	3.17	10	25.42	51.20	10	0.64	0.28
	Percentile	10	41.94	6.15	10	763.39	491.28	10	0.23	0.23
	RenyiEntropy	10	0.18	5.08	10	-12.75	54.85	10	0.78	0.16
	Shanbhag	10	-3.01	3.76	10	-52.57	46.9	10	0.55	0.31
	Triange	10	4.04	3.24	10	51.82	41.91	10	0.59	0.25
	Yen	10	0.58	5.69	10	-9.99	59.01	10	0.78	0.17

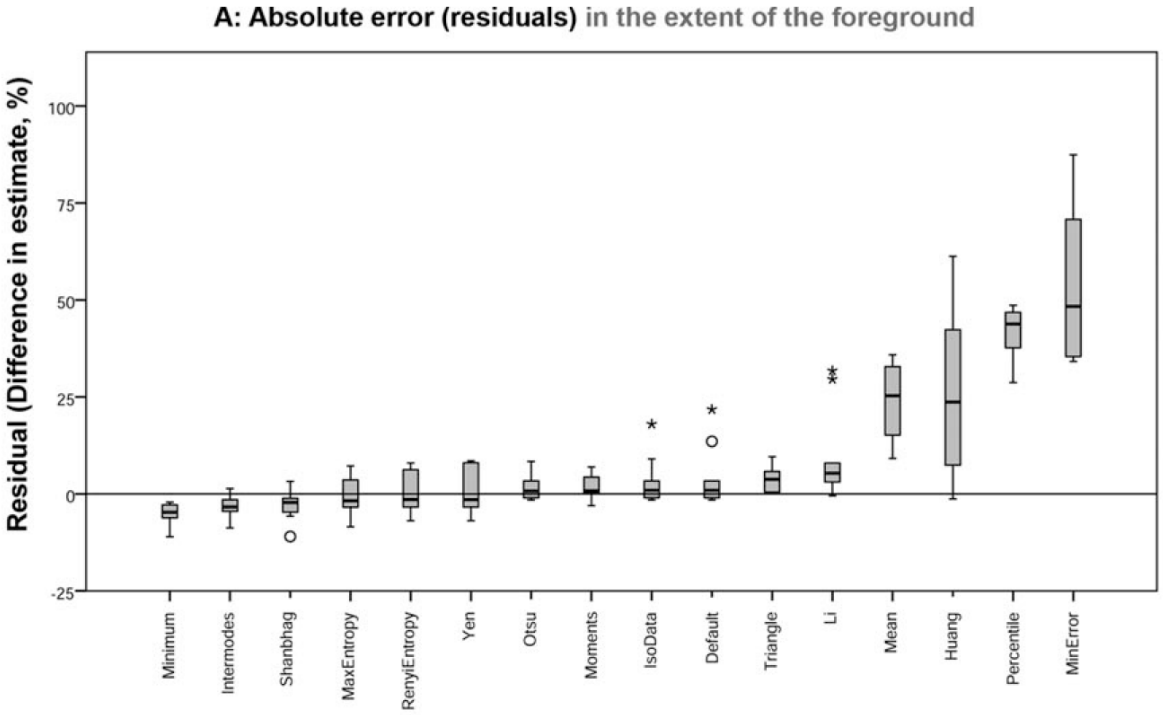
Table 3.6: Descriptive statistics for Experiments 2 and 3. Reproduced and modified from Levin et al 2019. N = number of samples, M = mean, SD = Standard Deviation.

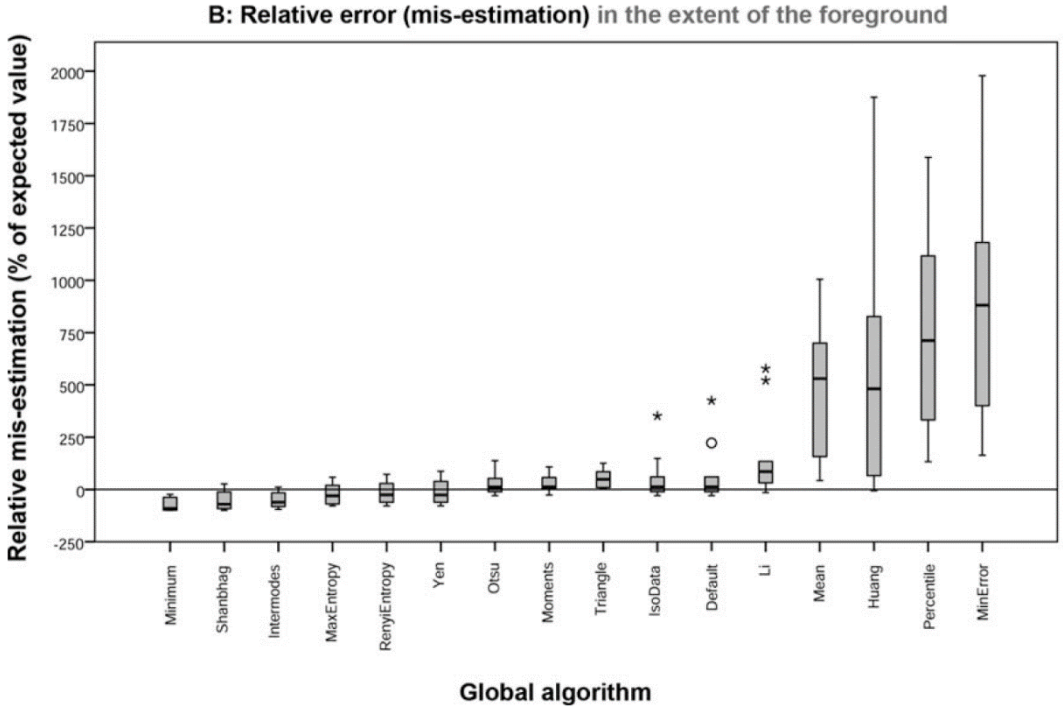
3.3.3 Results: Experiment Three (thresholding algorithms)

The results from experiment three are presented in Table 3.6, which provides descriptive statistics of the three performance measures of (a) the absolute residuals, (b) the percentage mis-estimation, and (c) the Dice Coefficient values, and Figure 3.12 which provides boxplots of the same data.

With respect to accuracy, it is immediately apparent that the different algorithms performed with different degrees of success (Figure 3.12). When compared to the mean value for the extent of the foreground obtained by tracing, several of the algorithms tend to vastly overestimate the extent (e.g. Min Error and Percentile, which both exhibit mean mis-estimates of over 750% of the expected value). At the other end of the spectrum, there were some algorithms which tended to underestimate the extent of the foreground (e.g. Minimum and Shanbhag). Several of the algorithms produced estimates of the extent of the foreground which were relatively close to that produced by manual segmentation. The algorithms which produced the closest estimates were Yen (which produced a mean mis-estimation of -9.99 ± 59.01 % (n=10)), MaxEntropy (which produced a mean mis-estimation of -22.40 ± 51.96% (n = 10)), and RenyiEntropy (which produced a mean mis-estimation of -12.75 ± 54.85% (n = 10)). The mean Dice coefficient values that these algorithms produced were also comparable to the values that the methods with manual input generated. The mean Dice Coefficient value for manual segmentation was 0.82 ± 0.08 (n = 30) For Yen the mean Dice Coefficient value was 0.78 ± 0.17 (n = 10), and similar figures were seen for MaxEntropy (0.79 ± 0.16, n = 10) and RenyiEntropy (0.78 ± 0.16, n = 10). For each of these algorithms, the minimum Dice Coefficient value observed was 0.49 (n = 10).

In terms of reproducibility, as explained in Section 3.2.5, an algorithm will always calculate the same output from the same input. The spread of values seen for the levels of accuracy in Figure 3.12 are a result of the algorithms working with different levels of success upon the different images in the suite. In terms of efficiency, it took less than one second for the macro to open the image, apply the thresholding algorithm, save the image, and close the image.


A: Absolute error (residuals) in the extent of the foreground
Residual (Difference in estimate, %)
100
75
50
25
0
-25
Minimum
Intermodes
Shanbhag
MaxEntropy
RenyiEntropy
Yen
Otsu
Moments
IsoData
Default
Triangle
Li
Mean
Huang
Percentile
MinError


B: Relative error (mis-estimation) in the extent of the foreground
Relative mis-estimation (% of expected value)
2000
1750
1500
1250
1000
750
500
250
0
-250
Minimum
Shanbhag
Intermodes
MaxEntropy
RenyiEntropy
Yen
Otsu
Moments
Triangle
IsoData
Default
Li
Mean
Huang
Percentile
MinError
Global algorithm

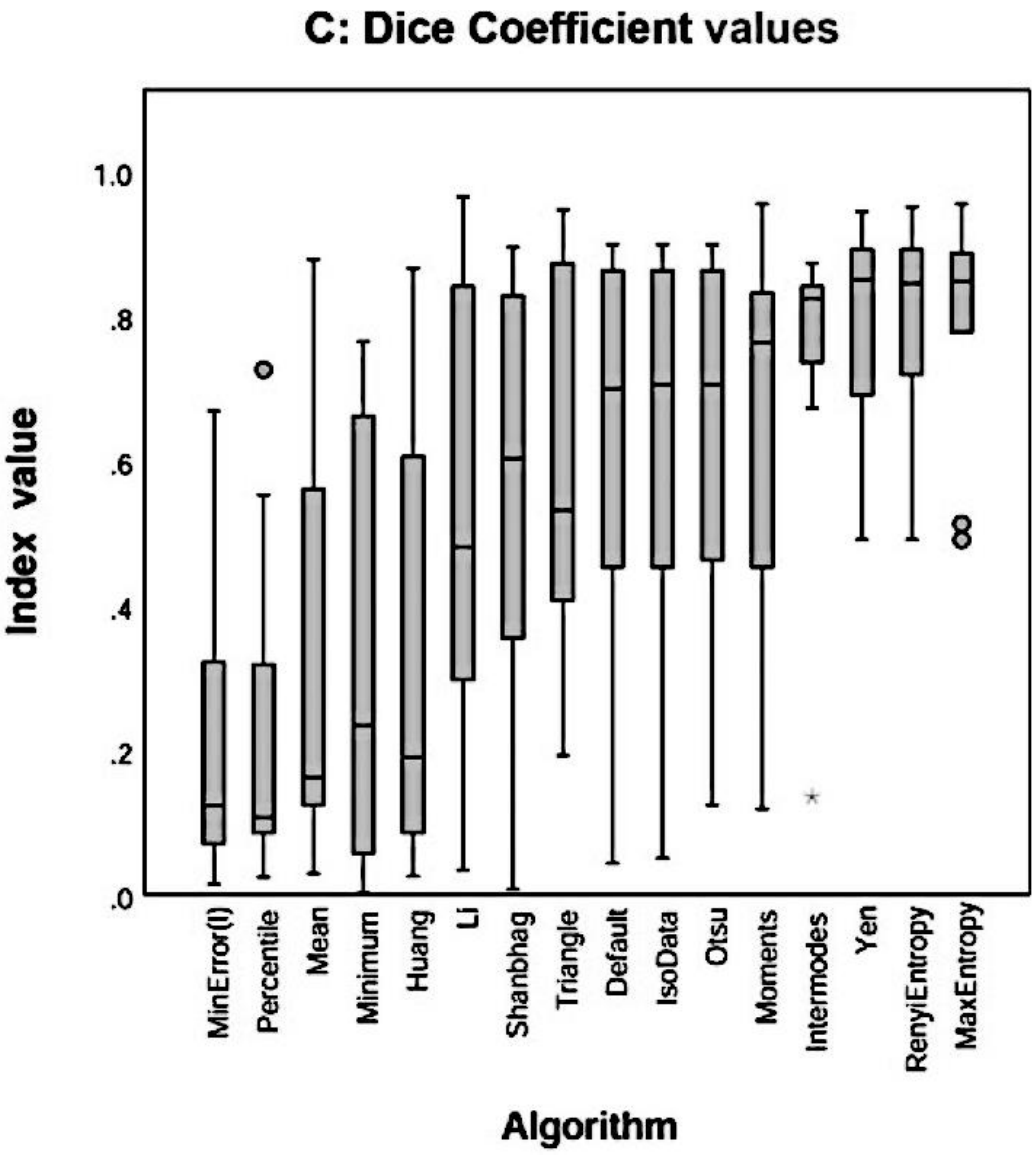


Figure 3.12: Metrics of performance for the global algorithms in order of ascending median. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median). Reproduced from Levin et al 2019

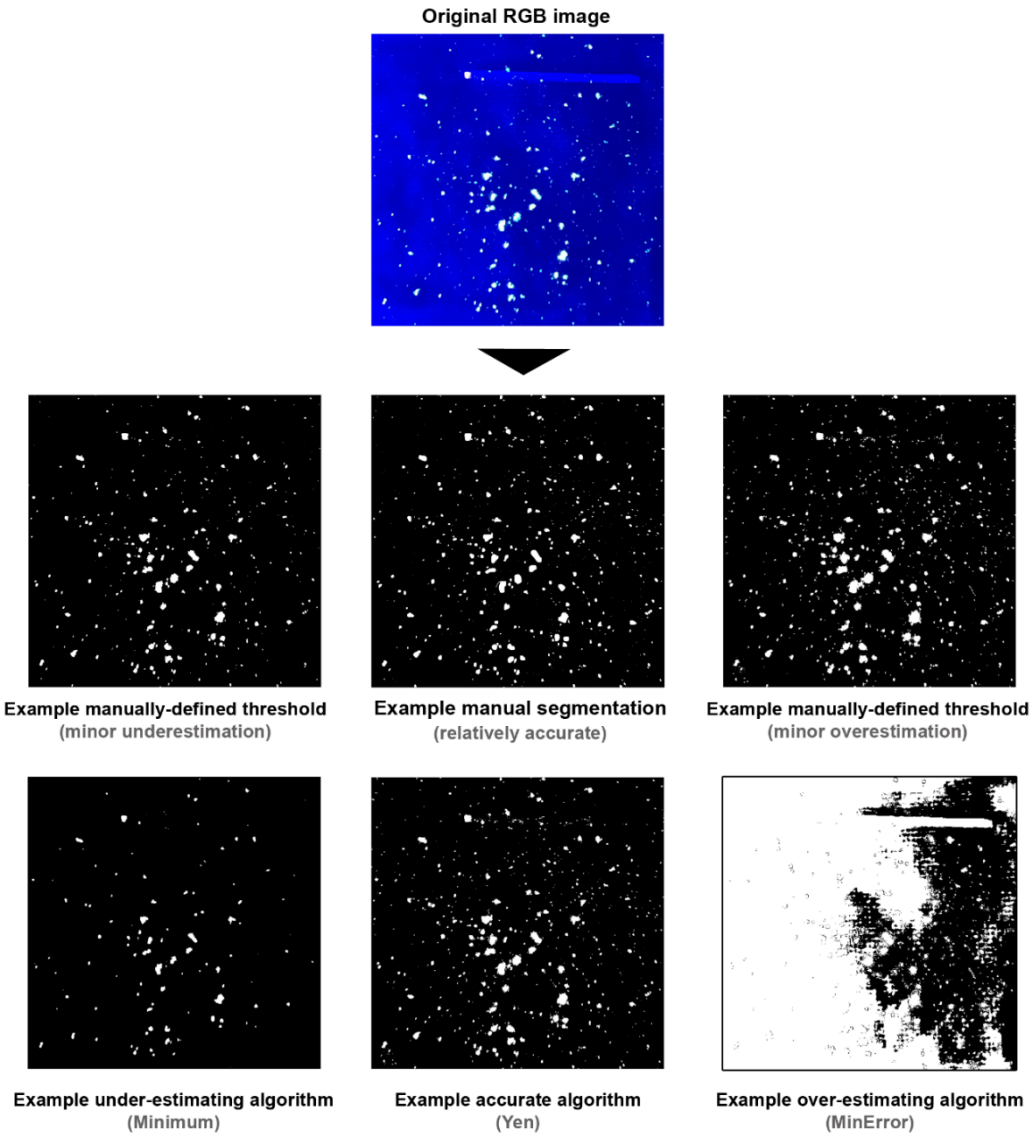


Figure 3.13 –Example outputs for one image (image 1), showing varying levels of accuracy reproduced from Levin et al 2019

3.4 Discussion

3.4.1 Discussion: Experiment One (manual segmentation)

In experiment one, it was seen that all three examiners created a range of values when they repeatedly estimated the extent of the foreground in an image of fluorescent powder, which had been provided to them in three different orientations (Figure 3.10). The mean values for the percentage of the image which was defined as foreground varied between examiners, and statistical analysis confirmed that there were differences between these estimates which may be worth further investigation (Table 3.4).

At first, this may seem a surprising finding, since tracing (manually segmenting an image) is presented within the literature as the "gold standard" of segmentation methods, and often used as a substitute for a ground truth, and comparison point for quality (e.g. Crum *et al.,* 2006; Griffin *et al.,* 2012). These findings are, however, not without precedent: multiple authors have debated whether segmentation methods which involve manual input suffer from "observer-dependent" variation, especially where images do not have clear-cut boundaries between regions of foreground and background (Yang *et al.,* 2001; Baveye *et al.,* 2010). Accordingly, this study supports the idea that methods with manual decision making may introduce variation into a dataset, and different examiners may arrive at different quantitative values to represent the same image. Consequently, this suggests that manual segmentation (tracing) does not offer good levels of *reproducibility.*

Another point suggested by the data was that certain examiners may perform more consistently than others: in this experiment, examiner two produced the highest mean Dice Coefficient values (of 0.95 ± 0.02 ($n = 9$) suggesting that each of their repetitions had a large degree of overlap, while examiner one produced a mean Dice Coefficient value of 0.71 ± 0.08 ($n=9$), suggesting a lower degree of similarity. Consequently, then, if manual segmentation was employed in a study of the persistence of fluorescent powder: "it is therefore possible that the examiner would be unaware how much error the process of manual segmentation (tracing) might be introducing into the dataset" (Levin *et al.,* 2019: 439)

3.4.2 Discussion: Experiment Two (Manual thresholding)

Three findings emerged from experiment two, evaluating the viability of manually-defining a threshold value for processing fluorescence imagery in a persistence study. Primarily, it was seen that on the basis of time taken, manually-defining a threshold value was much more efficient than manually segmenting the image by tracing over the regions of foreground, taking only one minute per image. Consequently, it could be feasible to employ a manually-defined threshold level as part of a fluorescent persistence experiment, depending on the size of the dataset.

Unfortunately, however, the accuracy and reliability of the technique are not quite as promising. The second and third findings were that while the technique of manually-defining a threshold level resulted in a *mean* estimate of the extent of the foreground that was comparable to that achieved by manual segmentation, many outliers were seen, and the estimates occupied a large range of values. Accordingly, while some of the segmentations could be considered accurate, with low residuals and high Dice Coefficient values, others could not. Since an effective segmentation method would not require multiple uses, and the calculation of a measure of central tendency, this consequently suggests that manually defining a threshold level may not be the optimal technique for segmenting this sort of fluorescence imagery.

3.4.3 Discussion: Experiment Three

In these experiments, it was demonstrated that global thresholding algorithms have the potential to segment images of fluorescent powder in a reproducible manner (operating as a fixed calculation), at speed (taking less than a second to process an image, as part of a macro in ImageJ), and if you select an appropriate algorithm, it will produce an accurate segmentation. Three of the algorithms tested seemed to have the best performance: Yen, MaxEntropy, and RenyiEntropy algorithms, which all offered similar mean Dice coefficient values to those obtained by manual segmentation. Consequently, these experiments suggest that an appropriate thresholding algorithm may be the best option for a segmentation technique for large numbers of fluorescence images. An appropriate algorithm can offer high levels

of reproducibility, without having to compromise upon accuracy. An appropriate algorithm, in conjunction with a macro, would be able to deal with very large numbers of images, and perform this repetitive task in a way that does not require manual intervention.

In these experiments, considering images of green fluorescent powder on black fabric, the Yen, RenyiEntropy and MaxEntropy algorithms seemed to be the most appropriate. However, as the performance of a thresholding algorithm is affected by the histogram of the image upon which it is used (Sezgin and Sankur, 2004; López-Leyva *et al.,* 2016), "it would be advisable for researchers to perform an evaluative study on the specific images to be used before employing a particular algorithm within a transfer and persistence experiment." (Levin *et al.,* 2019: 441).

3.4.4 Future work and limitations

Since this is a relatively unexplored area of research, there are many avenues which future researchers could pursue. One option for future work would be to evaluate the performance of these thresholding algorithms under more varied operating conditions (for example, where the lighting is inhomogeneous, where images may not be in focus, or where the image contrast is low). It should be emphasized that the image segmentation methods tested here (i.e. global thresholding methods) are amongst the most basic methods available. Future work may want to consider evaluating the potential of more complex methods of image segmentation within this problem field (i.e. segmenting fluorescent powder). It is acknowledged that the experiments undertaken here are both simple, and small in scale, they are nonetheless foundational for forensic science applications. Future research could repeat the elements of this study which considered the reproducibility of manual methods, with a larger suite of images, a larger number of participants, and, critically, a larger number of replicates. As an exploratory pilot study, the experiments conducted in this chapter offer useful findings and pave the way for dramatically speeding up the analysis of particle traces and increasing the size of datasets it is possible to capture to inform our understanding of the transfer and persistence of traces (see Chapter 5).

3.5 Conclusions

3.5 Conclusions

In summary, this chapter outlined a series of experiments into the performance of different methods of image segmentation upon images of fluorescent powder upon a dark fabric background. The characteristics under investigation included the accuracy of the technique (the extent to which the output quantitatively resembled the input), the reproducibility of the technique (the extent to which the repetition of the technique resulted in similar outputs, whether the repetition was by a single examiner, or multiple different examiners), and the amount of time that the technique required to process images. Four different conclusions were reached:

1. The technique of manually segmenting (i.e. tracing the foreground) cannot be considered reproducible for this context of fluorescence imagery. In experiments, three different examiners produced different mean estimates of the extent of the foreground in an image, and a range of values was seen when comparing successive estimates made by a single examiner.

2. The technique of applying a global threshold, where the value of the threshold is manually-defined, saw similar issues around reproducibility. Although the technique was much faster than manual segmentation, taking around one minute per image, again, a range of estimates were seen, with a number of segmentations representing inaccurate outliers.

3. The technique of applying a global threshold, where the value of the threshold is calculated by an algorithm, has the potential to provide a totally reproducible methodology, in that the algorithms are deterministic. Any examiner using one algorithm on one input will arrive at an identical output. In this experiment, it was suggested that algorithms can also offer improved efficiency, taking less than a second to process an image in conjunction with a macro.

4. In this study, the three algorithms which offered the most accurate estimates of the foreground were Yen, MaxEntropy, and RenyiEntropy. To quote from the published findings: "All three of these global algorithms produced results which mis-estimated the foreground by amounts comparable to a human examiner conducting manual segmentation, but at much faster speeds. It may [therefore] be possible to employ these algorithms in future experiments as part of a rapid and robust image processing workflow" (Levin *et al.,* 2019: 441).

Together, therefore, these experiments suggested that for the segmentation of digital images of fluorescent powder on a dark fabric background, manual segmentation provided accuracy, but not efficiency or reproducibility, manual thresholding offered accuracy *and* efficiency, but still lacked reproducibility, inappropriate algorithms offered inaccurate answers, reproducibly and quickly, while appropriate algorithms matched this speed and determinism with accuracy (see Figure 3.14). Consequently, it is possible to conclude from these experiments that employing an appropriate thresholding algorithm may be advantageous for future forensic research which requires the segmentation of fluorescence imagery.

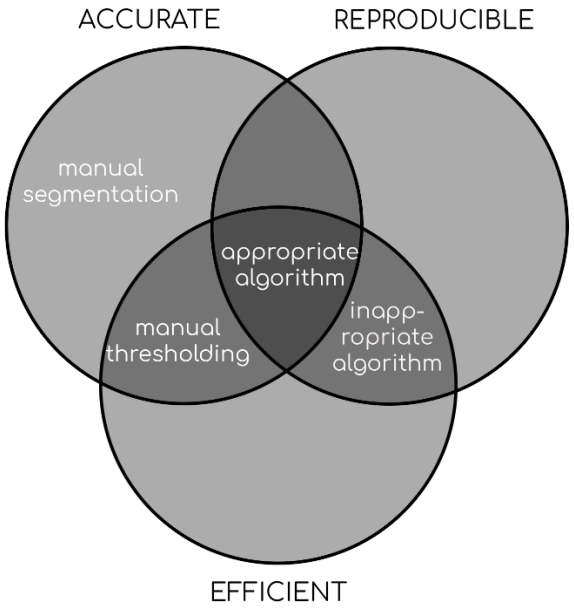


Figure 3.14 A Venn diagram summarising the performance of the different segmentation methods

4

Chapter Four

Quantifying the number of quartz grains present in a digital image:
Can mathematical image processing be used to locate and count grains?

Abstract

Quartz grains represent a valuable environmental trace indicator – widely distributed in terrestrial environments, and seen to transfer readily between environments and individuals within scenarios of forensic interest. However, many experiments which examine the transfer and persistence of quartz grains involve a tedious and time-consuming method of manually counting the number of quartz grains present in an image. Since quartz grains imaged upon a dark background would represent maxima in the image (i.e. regions where there is a peak in the brightness values), it was thought that it might be possible to use a quantitative image processing function ('find maxima') to locate and count quartz grains within digital images. Accordingly, an experiment was undertaken to evaluate the accuracy and reliability of using the "find maxima" function to locate and count quartz grains. Three findings emerged: (1) using the find maxima function (with an appropriate noise filter value) resulted in a count that was within the range identified by a human examiner (2) different images required different noise filter values to achieve accurate representations of the image (3) different examiners using the find maxima to count quartz grains arrived at answers that were not statistically significantly different from each other, at $p<0.05$. Accordingly, it was concluded that the find maxima function may be used to accelerate data analysis without compromising on accuracy or reliability. These findings highlight the importance of interdisciplinary research, and the potential benefits that insights from cognate disciplines can offer for improving workflows – for a discipline that has historically been so dependent on microscopic analyses and visual comparison, image processing can offer a valuable tool in generating data for forensic reconstruction.

4.1 Introduction

4.1.1 Counting in forensic geoscience

As seen in Chapter 2, there are a very wide range of environmental indicators which can find uses within forensic casework (e.g. Ruffell and McKinley, 2008; Cheshire et al., 2017). While the indicators can be quite varied – being either organic such as pollen (see: Riding *et al.,* 2007), inorganic such as minerals (Bull and Morgan, 2006), and found in either aquatic or terrestrial environments (e.g. Scott *et al.,* 2014) – their analysis has one common theme: quite frequently, the method of analysis is to survey what particles are present, counting and classifying each particulate found.

As introduced in Chapter 2 (section 2.4.4. and 2.4.5), quartz grains can represent a useful environmental trace indicator, being both one of the most abundant minerals on earth (Bull and Morgan, 2006), shaped by local environmental conditions (Krinsley and Doornkamp, 1973; Mahaney, 2002), and possessing the ability to transfer between suspects, victims, and their surrounding environment (e.g. Morgan and Bull, 2006; Morgan *et al.,* 2006; Bull *et al.,* 2006). With respect to the study of the transfer and persistence of quartz grains, studies have historically involved an examiner extracting the queried particles from the recipient material and manually counting the particles present under the microscope (e.g. Morgan *et al.,* 2019) This process can be time-consuming and repetitive, and has, historically, placed a low ceiling on the number of data which can feasibly be collected during the course of a study (see: Morgan *et al.,* 2019). Accordingly, the ability to accelerate the counting of quartz grains would be beneficial for future research, enabling the collection of larger datasets and the pursuit of more ambitious hypotheses. Since the task of counting quartz grains is both simple and repetitive, it should be a good candidate for a degree of automation.

As demonstrated in Chapter 3, the first step in enabling any sort of quantitative, automated image analysis is to successfully 'segment' the image, locating the regions of foreground (i.e. the objects of interest) and isolating them from the background (i.e. the regions that are not of interest (Sezgin and Sankur, 2004; Baveye, 2010;

Bankhead, 2014)). In other words: in order to count the number of quartz grains present in an image, it is first necessary to establish which pixels in the image represent quartz grains.

4.1.2 The challenge of segmenting quartz grains

In the previous chapter, Chapter 3, we saw that the use of a thresholding operation was suitable to segment fluorescent powder from a dark fabric background (see: Levin *et al.,* 2009). Unfortunately, the same approach is not viable for images of quartz grains. This can be ascribed to two facts:

(1) Levels of noise: Images of quartz grains on fabric (imaged under visible light) will, in general, tend to be noisier than images of fluorescent powder imaged under ultraviolet illumination, and global thresholding operations tend to produce poor quality segmentations where images are noisy (Bankhead, 2014). An example of an image histogram for quartz grains imaged under visible light, in comparison to fluorescent powder imaged under ultraviolet light, is provided in Figure 4.1. The image of fluorescent powder produces a histogram with two clean peaks, while the image of the quartz grains is both multimodal, rather than bimodal, and much noisier.

(2) The colour (and brightness values) of quartz grains: For pictures of fluorescent matter imaged under the wavelength of light at which the particles fluoresce, the fluorescent particles should all have similar brightness values. Accordingly, it should be possible to isolate a peak on the image histogram which represents those particles alone which should belong to the foreground. In Figure 4.1, this is a small peak at the extreme right-hand side of the graph, approaching 255 for brightness. For quartz grains, this is, however, not the case: not all quartz grains are the same colour (or the same brightness values), and some are glassy and transparent, showing some of the background colour through the grain. It is therefore not as simple as finding a single peak which is likely to represent the particles of interest. An example of the performance of

thresholding algorithms upon an image of quartz grains can be seen in Figure 4.2. The same 16 global thresholding algorithms that were tested in Chapter 3 (see Table 3.3) were applied to the image of quartz grains from Figure 4.1. It is clear that the thresholding operation has struggled to segment the image with fidelity, either including large regions of the fabric in the foreground (as seen in the outputs for the "Default" and "Huang" algorithms) or failing to include the darker quartz grains in the foreground (as seen in the output for the "Intermodes" algorithm).

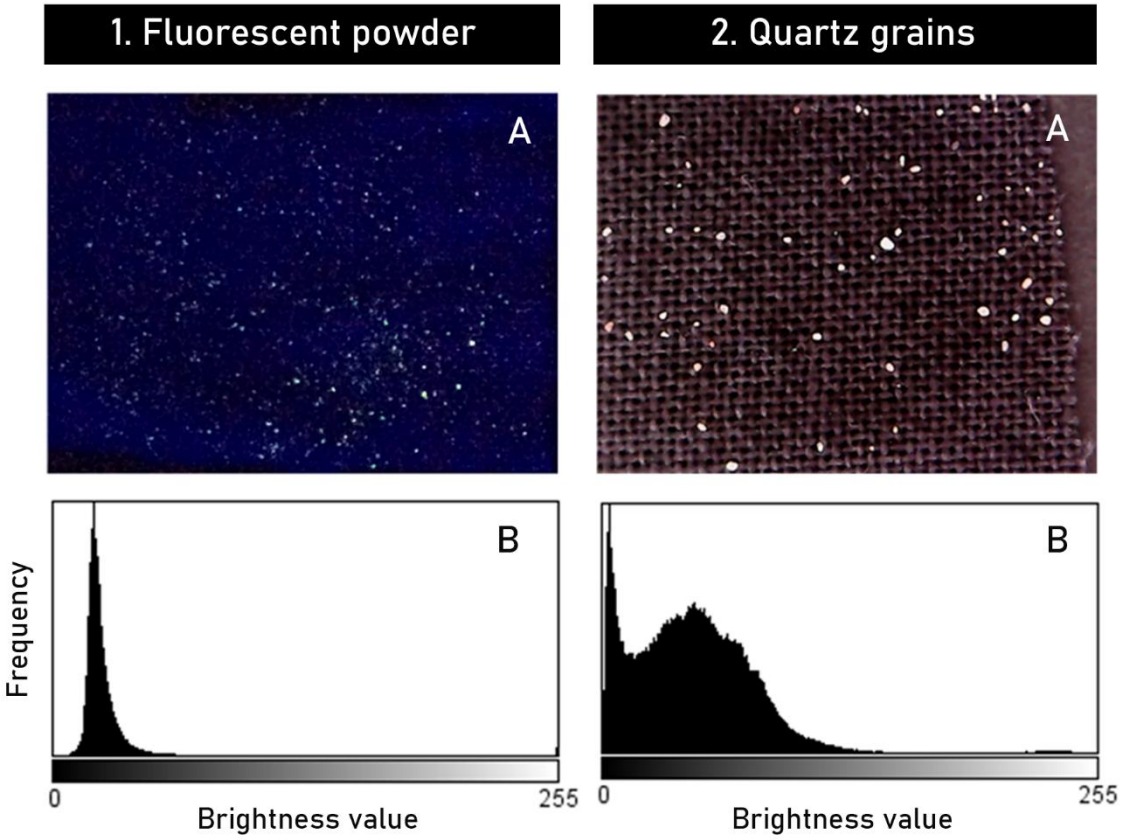


Figure 4.1: Exemplar histograms for (A) fluorescent powder imaged under ultraviolet light, and (B) Quartz grains imaged under visible light

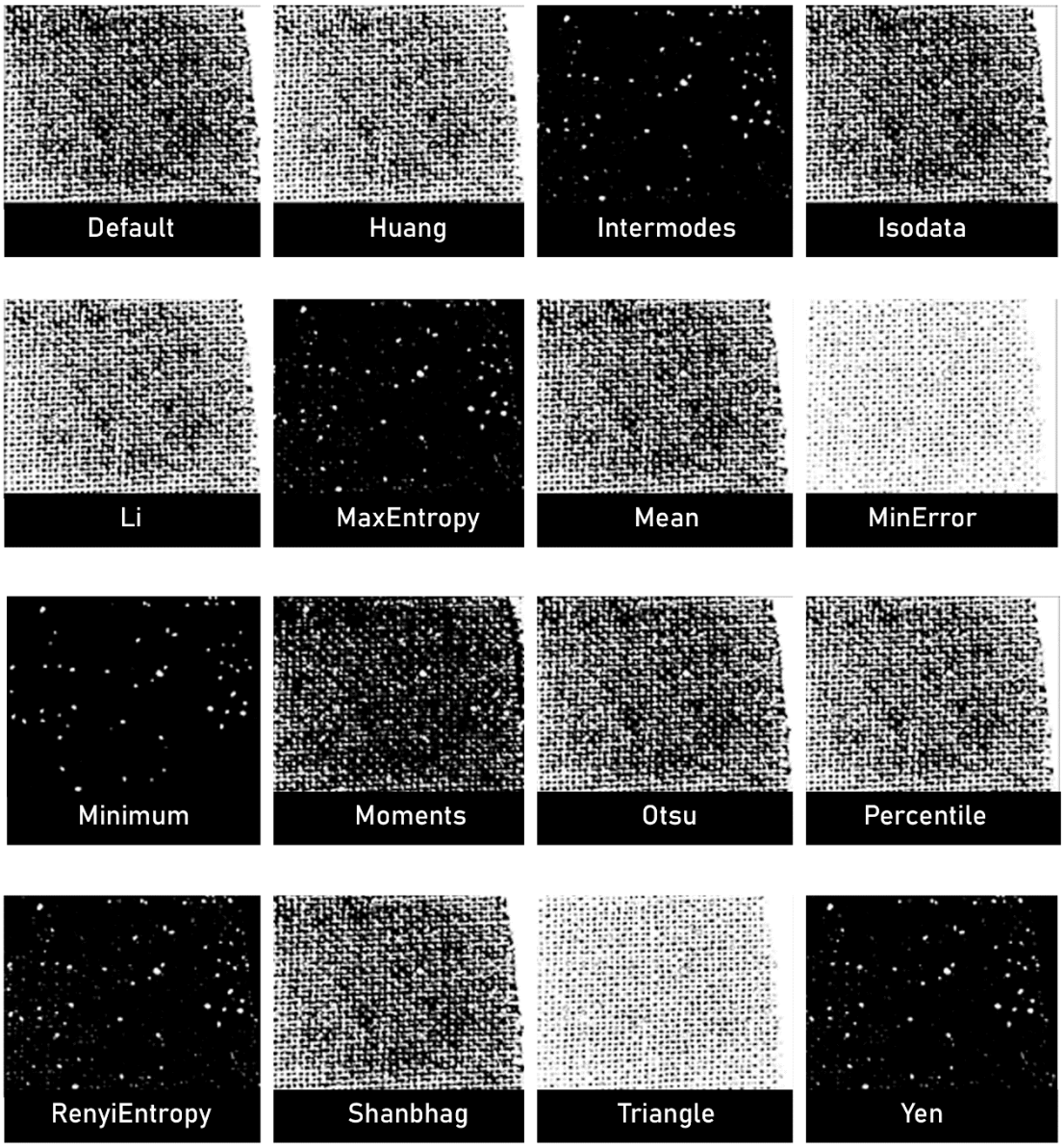


Figure 4.2: The outputs that result when one applies the 16 global thresholding algorithms tested in Chapter 3 to the image of quartz from Figure 4.1. The majority of the algorithms vastly over-estimate the extent of the foreground, including the lighter pixels of the fabric above the threshold.

4.1.3 An alternative method of locating quartz grains within an image

Since a global thresholding operation would be inappropriate for isolating quartz grains from a fabric background in a digital image, an alternative method must therefore be sought.

As mentioned in the preceding two chapters, NIH ImageJ is a free image processing programme, which many researchers in many different disciplines have used for quantitative image analysis (e.g. Schneider *et al.,* 2012; Schindelin *et al.,* 2015; Legland *et al.,* 2016; Arena *et al.,* 2017). The task of locating and counting specific objects within an image is one that many researchers in many different contexts have wanted to achieve (e.g. Grishagin, 2015) and so it should be possible to borrow a method used by researchers in other cognate disciplines, and adapt it to locate and count quartz grains.

The "find maxima" function available in ImageJ (see https://imagej.nih.gov/ij/docs/guide/146-29.html; Grishagin, 2015) has been used for automated counting within many different and varied contexts. It has, for example, been used to locate and count cells in analytical biochemistry (e.g. Meyer dos Santos *et al.,* 2010; White *et al.,* 2012; Grishagin, 2015), and to locate and count specific biological structures such as cell nuclei (e.g. Courtoy *et al.,* 2015; Nguyen *et al.,* 2016), muscle bundles (Ardiel *et al.*, 2017), plant roots (Adu *et al.,* 2015) and parts of fly larvae (Andlauer and Sigrist, 2012). It has also been used to identify macro plant structures – such as counting grains of wheat from top-down photographs of fields (Fernandez-Gallego *et al.,* 2018).

To discuss the way that the "Find Maxima" function works, it might be useful to briefly re-cap exactly what a (raster) digital image is. While a human viewer may see an image as a single graphical object, raster digital images are actually composed of pixels – an image is actually an array, or matrix, of pixels, each of which has numerical values describing its colour and brightness (Bankhead, 2014; see Figure

4.3). The bit-depth of an image, and its colour model determine what channels each pixel has a value for, and the number of values that are in the scale (Gonzalez and Woods, 2010; Bankhead, 2014).

For an RGB image, each pixel will have an intensity value for three channels: red, green, and blue. If the image is 8-bit, then each of these channels will have 2^8 possible values (256 values, expressed as 0-255, because the values are zero-indexed)(Gonzalez and Woods, 2010). There are many different colour spaces available, with varying numbers of channels – common ones include greyscale (where there is one value which denotes the brightness of the pixel), RGB (where three values denote the intensity of red, green, and blue for each pixel), HSB (in which the three values denote the hue, saturation, and brightness of the pixel) and CYMK (a colour format commonly used for printing, where the four values correspond to four colours of ink – cyan, yellow, magenta, and key (or 'black', to those outside the printing profession (Gonzalez and Woods, 2010)).

Accordingly, then, because digital images are actually matrices of numerical values, it is possible to perform mathematical operations upon them (as seen in Chapter 3). The "Find Maxima" function performs a calculation which "determines and counts the local intensity maxima in an image" (Grishagin, 2015), locating regions of high intensity relative to their neighbouring pixels (with specified tolerances for noise). In order to visualise this, one can imagine that the matrix of brightness values for an image could be plotted three-dimensionally, such that they resemble a mountain range, with peaks representing high brightness values, and troughs representing low brightness values (see Figure 4.4).

Applying a global thresholding operation would be the equivalent of drawing a plane through the y-axis, and counting any peaks which break through this plane as a region of interest (i.e. Like measuring the height of a mountain compared to sea level). In contrast, using the Find Maxima function is the equivalent of performing the sort of topographic prominence analysis that one would use to describe the peaks of a mountain range, relative to the height of the surrounding peaks (see:

Gonzalez and Woods, 2010). Accordingly, the Find Maxima function may represent a way of locating and isolating the peaks which correspond to quartz grains from the peaks and troughs which represent the fabric background.

Figure 4.3: Digital images are actually arrays of numerical values. (A) Depicts a greyscale image. (B) Depicts a magnified detail (showing the pixels), and (C) Depicts a magnified detail of the numerical values of those pixels

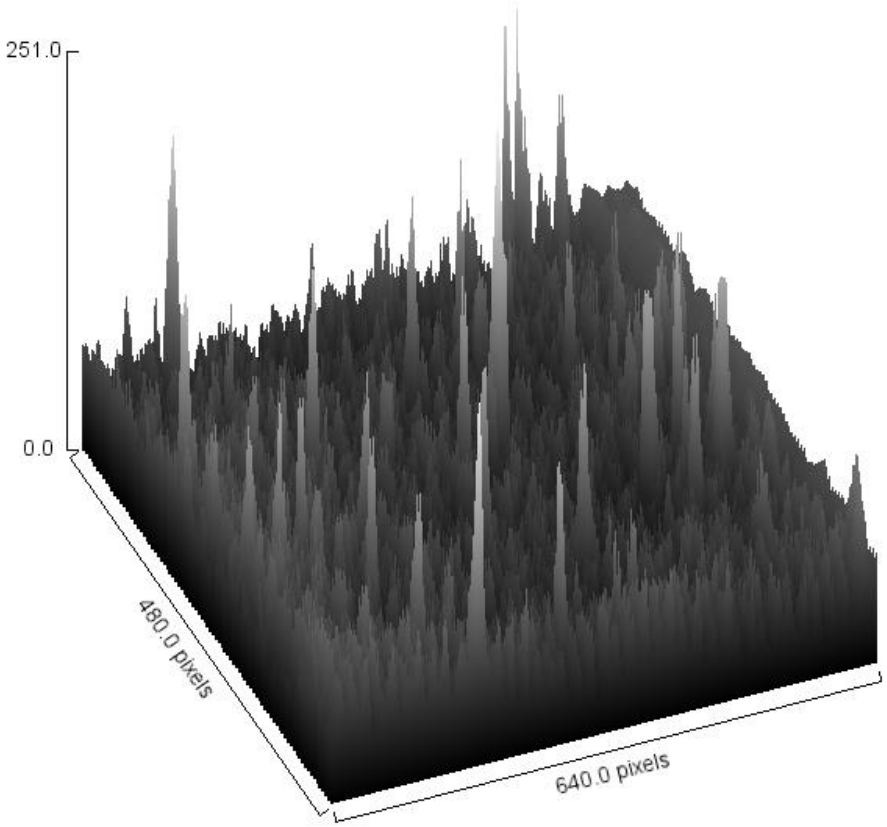


Figure 4.4: A surface plot, showing the peaks and troughs of brightness in the image of quartz grains on fabric shown in Figure 4.1. Created in ImageJ, using the pathway Analyse > SurfacePlot

4.1.4 Aims

Accordingly, the questions under investigation in this chapter relate to the performance of the Find Maxima function upon images of quartz grains on fabric. Broadly, the question under investigation is: *'Is the 'find maxima' processing operation a suitable method of counting the number of quartz grains on a fabric swatch, and can this process be fully-automated through the use of a macro for ImageJ?'* In order to answer this over-arching question, it was necessary to address three sub-questions:

1. Does using 'find maxima' to locate quartz grains result in the same number of quartz grains being identified as a human examiner manually counting the grains?
   (i.e. Can the technique be considered accurate?)

2. Is there one value for noise tolerance (the noise filter) which is appropriate for a batch of images, or is it necessary to define an appropriate noise filter level for each image individually? *(i.e. Can the technique be fully automated and left to run with a macro for a batch of images, or is it necessary for an examiner to remain present and actively involved?)*

3. Will one examiner using the technique repeatedly arrive at the same answer? Will multiple examiners arrive at the same answer? *(i.e. Can the technique of using 'find maxima' be considered reliable/reproducible?)*

Together, the answers to these questions will suggest whether using the 'find maxima' function is capable of forming part of a workflow to enable the rapid analysis of swatches of quartz grains as part of a study exploring the transfer and persistence of forensically relevant trace materials. As with the Chapter 3, three aspects of the method's performance are considered: the *accuracy* (i.e. Will it deliver the same results as a human examiner (which are assumed to be correct)?); the

efficiency (i.e. Can it perform faster than a human examiner, and will this enable the collection of larger datasets and the pursuit of more ambitious hypotheses?); and the *reproducibility* of results (i.e. Can this be considered a repeatable form of analysis? Will multiple examiners using the technique arrive at the same answer? Will one examiner repeating their analysis arrive at the same answer?)

4.2 Materials and Methods

4.2.1 Methods: the structure of the experiments

In order to explore the three questions identified in the introduction, two related, but separate, experiments were undertaken.

Experiment One

The first experiment explored the two questions "Does using 'find maxima' to locate quartz grains result in the same number of quartz grains being identified as a human examiner manually counting the grains?" and "Is there one value for noise tolerance (the noise filter) which is appropriate for a batch of images, or is it necessary to define an appropriate noise filter level for each image individually?" This experiment involved a single participant.

Experiment Two

The second experiment sought to investigate whether the use of the Find Maxima function resulted in consistent counts when used repeatedly by the same examiner, or by multiple different examiners. This experiment involved multiple participants.

4.2.2 Materials: Images

Two different sets of images were used for the two experiments. The set of images for Experiment One consisted of 25 images of quartz grains on fabric, which were generated as part of the persistence experiment outlined in Chapter 5 (see Table 4.1; Figure 4.5). Specifically, the 25 images used represented the five samples taken from the heel of the shoe, imaged at five time intervals (please see section 5.2 and Table 4.1 for more details). Each of the images depicted silver sand grains (a type of quartz sand commonly used in horticulture) upon swatches of black cotton, which had been imaged under visible light using a customised USB microscope set-up (Figures 4.6 and 4.7).

As with the preceding chapter, it is acknowledged that this method of data acquisition relied on trust in the black-box method of the software that generated the images, as is the convention in this field of experimental forensic science (for example, as in French *et al.,* 2012; Morgan *et al.,* 2013). Also as with the preceding chapter, it is acknowledged that further image characterisation would be desirable; considering the exact magnification and resolution of the images, having knowledge about the sensors and image quality; for example, considering uniformity, noise, focus, vignetting, compression, pixel resolution, spatial resolution, Bit depth, and the presence of any dead pixels on the sensor. Such information would be essential for the data to be used more generally.

It is also acknowledged that this represents the first attempt at conducting this kind of work for the imaging of particles that could be used in forensic science experiments (i.e. quartz grains).

Additionally, it is acknowledged that the use of a customised USB microscope set-up introduced a number of limitations. Unlike a laboratory microscope set-up, it is possible that the illumination varied over time (for example, with the natural light in the room changing), and it was possible that the illumination was not uniform across the sample. Although every effort was made to ensure that the camera lens and sample were parallel (i.e. that the images were not taken at an angle, introducing foreshortening to the image), this was not certified, and could have been a very plausible possibility. Accordingly, the use of a bespoke set-up, while a cost-effective and accessible solution for researchers, introduces a number of unknowns into the data acquisition.

Each image contained between 10 and 150 quartz grains, and the scale was approximately 45 pixels per mm.

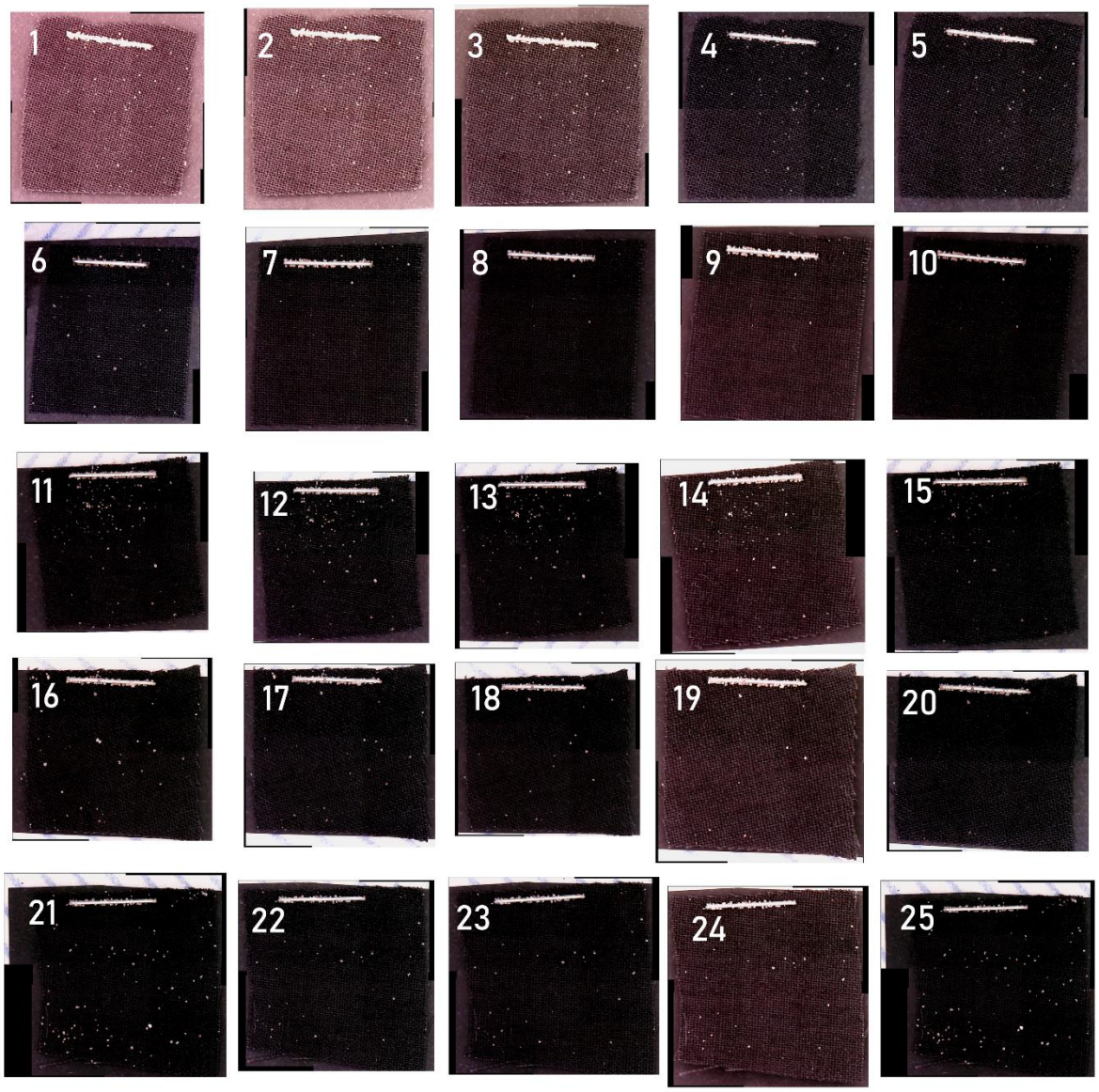


Figure 4.5: The 25 Images used in Experiment 1. Each row represents a different swatch.

Run Number	Time since transfer (hours)				
	0	1	2	4	8
1	1	2	3	4	5
2	6	7	8	9	10
3	11	12	13	14	15
4	16	17	18	19	20
5	21	22	23	24	25

Table 4.1: Descriptive variables for each image – how the images correspond to the study in Chapter 5. All were from the heel of the shoe, with zero activity.

The magnification was such that it was necessary to 'tile' the captured images (i.e. multiple fields of view were stitched together, in order that the entire sample can be contained within one image (see: Preibisch *et al.,* 2009)). This 'tiling' or 'stitching' was undertaken in ImageJ, using the plugin 'Grid/Collection Stitching' (Prebisch *et al.,* 2009). This plugin: " [uses] Fourier Shift Theorem, computes all possible translations between pairs of [...] images, yielding the best overlap in terms of the cross-correlation measure and subsequently finds the globally optimal configuration of the whole group of [...] images" (Prebisch *et al.,* 2009: 1463). Although efforts had been made to control the illumination conditions, there was some variation in the contrast and brightness of the images (Figure 4.5). Apart from stitching, the images were not pre-processed in any way. The second set of images, for Experiment Two, consisted of nine 300 by 300 pixel sub-vignettes extracted from the images, which were in fact three vignettes, presented in three rotations (Figure 4.7).

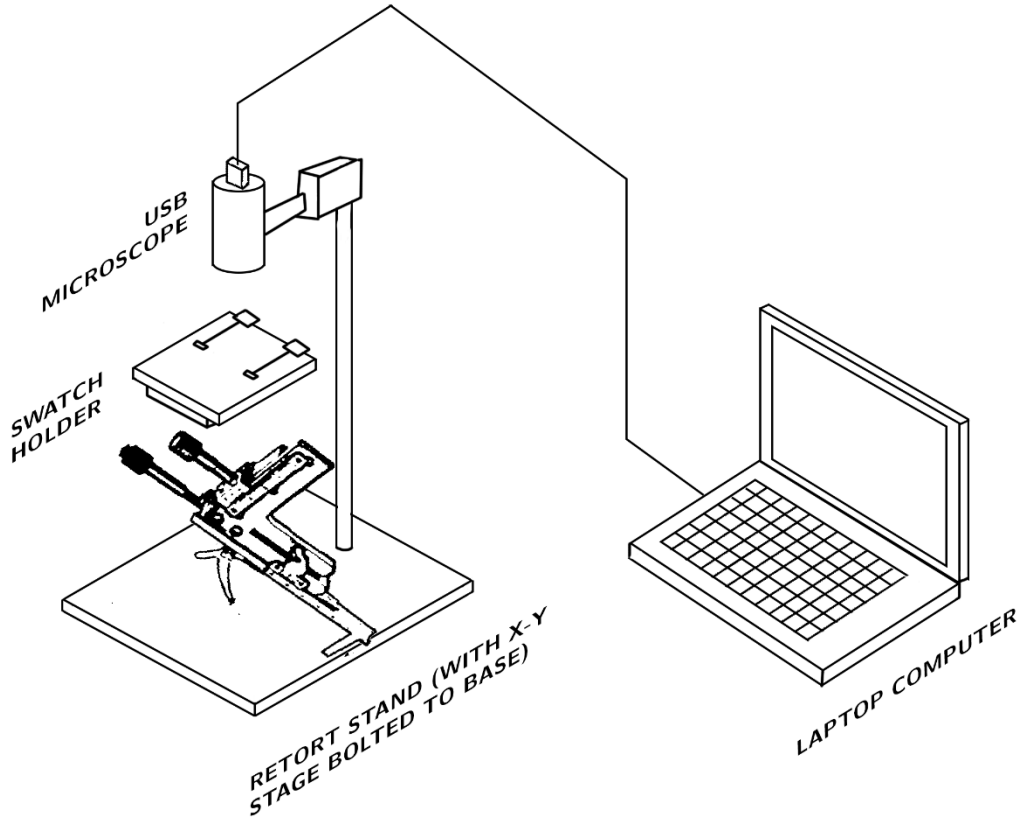


Figure 4.6: Diagram depicting the USB microscope setup, with a customised retort stand to hold the samples, and translate them (move in the X-Y dimensions) relative to the microscope lens.

4.2.3. Methods: Experiment One

The purpose of experiment one was to establish whether using the Find Maxima Function would result in the identification of the same numbers of quartz grains in an image as a human examiner manually counting them, and whether a single noise filter value was appropriate for multiple images (enabling automated batch processing).

The Manual Count

The 25 images in set 1 (Figure 4.5) were opened in Adobe Photoshop (CS4), and using the count tool (accessed through the pathway: Analysis > Count), each quartz grain was counted. This counting was repeated three times for each image. This method of counting was chosen because it meant that each grain was annotated on a layer separate to the underlying image, and this layer of labels could have its visibility toggled on or off. This meant that there was a record of exactly which regions of the image had been identified as quartz grains, and that it was possible to compare discrepancies between different counts.

The Find Maxima Count

The 25 images in set 1 (Figure 4.5) were then opened in ImageJ. Each image was converted from 8-bit RGB colour, to 8-bit greyscale. The region-of-interest (i.e. the fabric swatch) was selected using the 'polygon selection' tool. The 'find maxima' function was then applied using the pathway: Process > Find Maxima. Documentation for this plugin can be found online[6], and the full code for the plugin, which was based on a plugin developed by Michael Schmid, is available on Github[7]. The value for noise tolerance was adjusted until the number of maxima selected accurately reflected the number of quartz grains present in the image (please see Figure 4.8). This was a subjective process achieved by inspecting the image; if there were quartz grains that were not included (false negatives), the value of the noise filter was lowered, if there were too many maxima – (false positives) the value of the

[6] https://imagej.nih.gov/ij/docs/menus/process.html#find-maxima
[7] https://github.com/imagej/imagej1/blob/master/ij/plugin/filter/MaximumFinder.java

noise filter was raised. In the event that there was no noise tolerance value which resulted in the exact number of quartz grains that the examiner could see (or thought they could see), the noise tolerance value which gave the *closest* result was used.

4.2.4 Methods: Experiment Two

The aim of experiment two was to evaluate the repeatability of counting quartz by 'Find Maxima', comparing both inter- and intra- examiner repeatability. Analogous to the experiment in Chapter 3, three examiners were given three 300 pixel by 300 pixel sub-sections of images, at three rotations (and the participants of the experiment were not aware that the images were duplicates (Figure 4.7)). Participants were asked to apply the find maxima function to each image (as described in section 4.2.3), to select the regions which represented quartz grains. They were asked record the noise filter value and count value that they felt was the best fit for each image, neither over nor underestimating the number of quartz grains (see Figure 4.8). ANOVA was used to compare the outputs of the three examiners (see: Field and Hole, 2007).

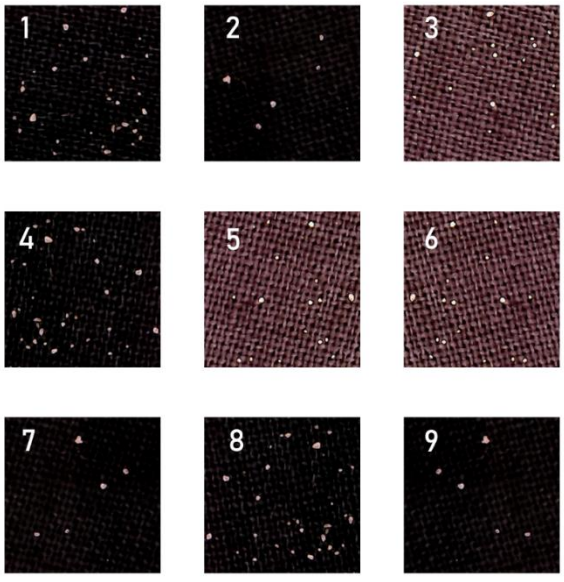


Figure 4.7: The nine vignettes used in Experiment Two. Each is 300 pixels by 300 pixels, and represents three different rotations of three different vignettes.

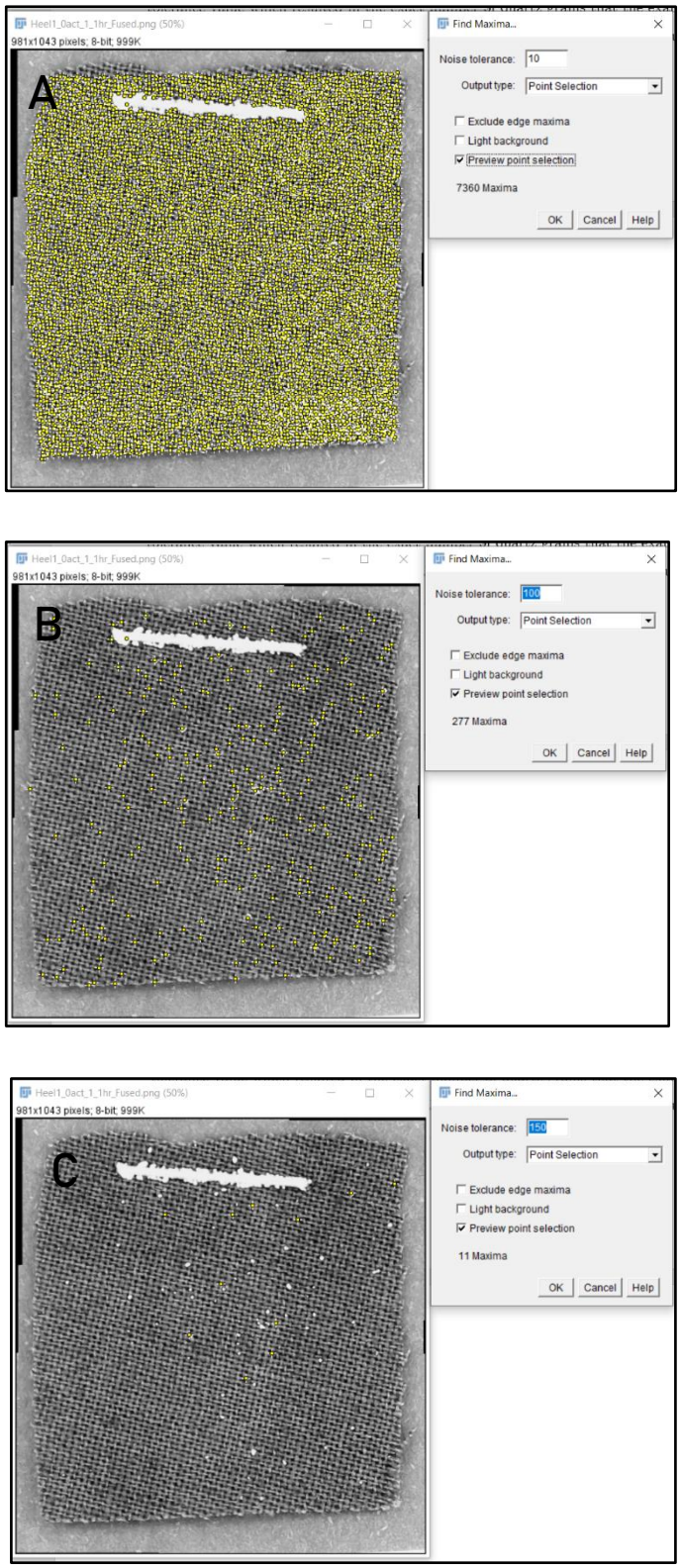


Figure 4.8: Screenshots, showing the user interface for the Find Maxima function. As the noise filter value is raised, fewer maxima are identified. (A) Depicts a noise filter value of 10 (7360 maxima, clearly over-estimating the number of grains), (B) Shows a noise filter value of 100 (277 maxima, over-estimating less extremely), and (C) Shows a noise filter value of 150 (11 maxima, clearly under-estimating the number of quartz grains present)

.

4.2.5 Statistical Analysis

As in Chapter Three, two forms of statistical analysis were used in this chapter; descriptive and inferential. Also as in Chapter Three, the descriptive analysis undertaken involved the calculation of the mean (a measure of central tendency) and the standard deviation (a measure of dispersion), and the inferential analysis undertaken to compare the results obtained by different examiners was ANOVA (Analysis of Variance; see Field and Hole, 2007, for a detailed description). Again, ANOVA was the appropriate statistical tool to use because the data (1) had three or more experimental groups; (2) Different participants were used in each group (i.e. each person contributed one score to the data, which were independent measures), and (3) The aim of analysis was to establish whether there were differences between the groups. (Please see Field and Hole, 2007, p275, for a flowchart which outlines the appropriate tests to use, given this information).

4.3 Results

4.3.1 Question One: does Find Maxima give the same answer as manually counting quartz grains?

The results of Experiment One are summarised in Table 4.2 and Figures 4.9 and 4.10. Two observations can be made. (1) For 21 of the 25 images manually-counting the quartz grains resulted in ranges of values; for 22 of the 25 images, the values obtained by the find maxima function were within these ranges (Figure 4.11); (2) The error seen when comparing the find maxima count to the mean manual count was small and normally distributed (Figure 4.10). The maximum residuals observed were ± 2 grains from the mean manual count, and for 60% of the images there was no difference between the best-value obtained by the find maxima function and the mean of the human counts (Figure 4.10).

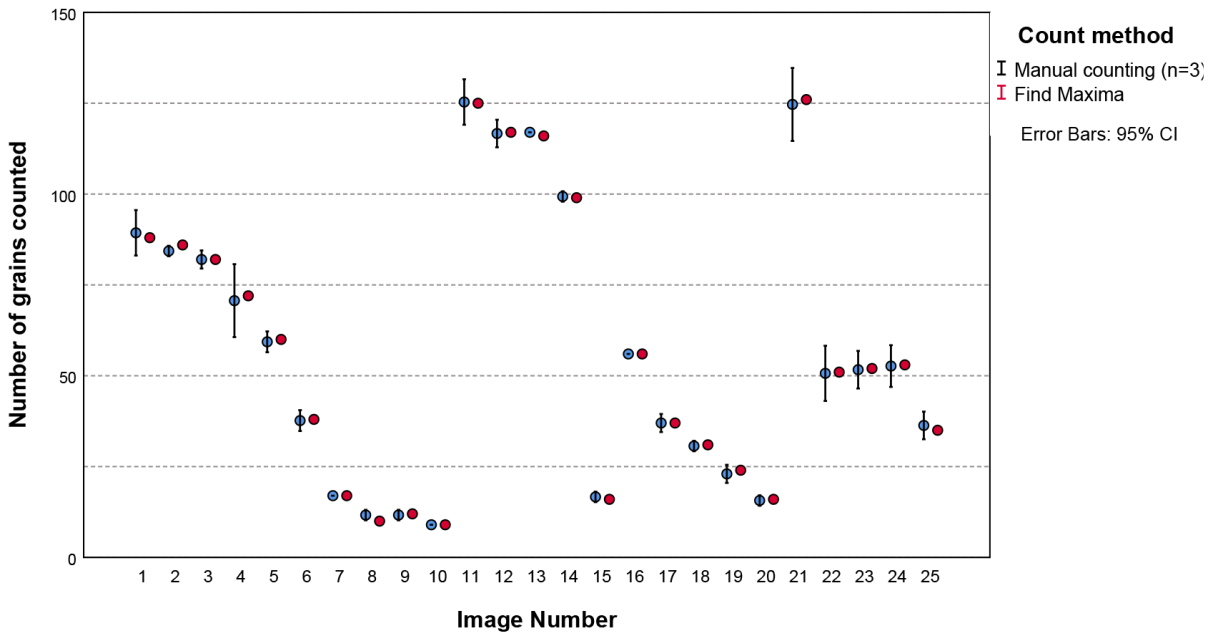


Figure 4.9: Scatterplot comparing the counts produced by manually counting each quartz grain, versus using the Find Maxima function.

Image number	Manual count				Best 'Find Maxima' count	Difference between manual count mean (n=3) and best FM count	Best fit noise filter value
	1	2	3	Mean (n=3)			
1	92	89	87	89	88	-1	121
2	84	85	84	84	86	2	122
3	83	82	81	82	82	0	129
4	73	66	73	71	72	1	143
5	60	58	60	59	60	1	142
6	39	37	37	38	38	0	137
7	17	17	17	17	17	0	130
8	12	12	11	12	10	-2	120
9	11	12	12	12	12	0	120
10	9	9	9	9	9	0	130
11	125	128	123	125	125	0	95
12	118	115	117	117	117	0	112
13	117	117	117	117	116	-1	88
14	99	99	100	99	99	0	115
15	16	17	17	17	16	-1	125
16	56	56	56	56	56	0	100
17	38	37	36	37	37	0	116
18	30	31	31	31	31	0	100
19	24	22	23	23	24	1	120
20	15	16	16	16	16	0	115
21	124	129	121	125	126	1	95
22	54	50	48	51	51	0	123
23	54	51	50	52	52	0	110
24	54	50	54	53	53	0	135
25	38	36	35	36	35	-1	142

Table 4.2: Summary of data from experiment one, showing the manual counts, the best fit noise values for each image, and the resulting 'find maxima' counts

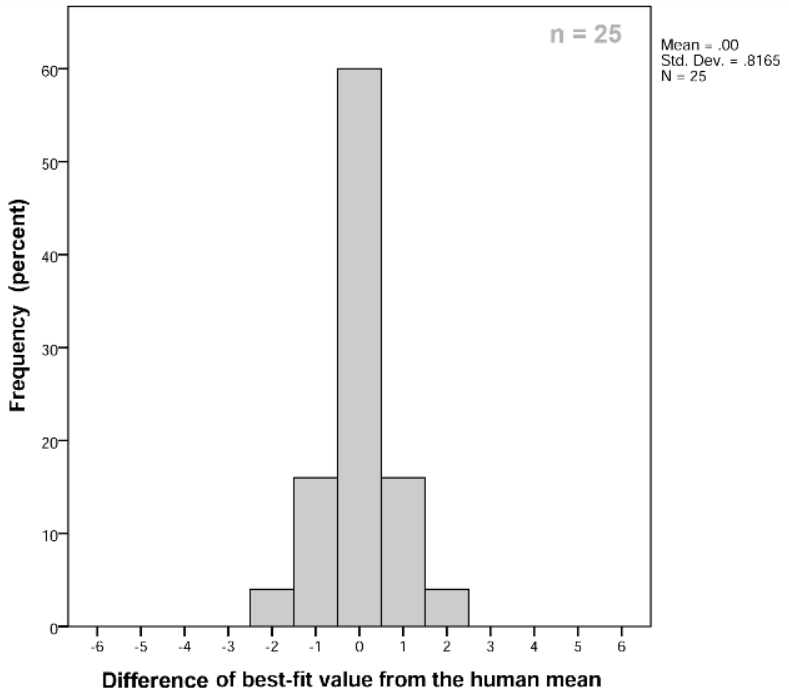


Figure 4.10: Histogram showing the residuals between the mean manual count for each image, and the count produced by using "find maxima" with the best-fit noise filter value (total n=25)

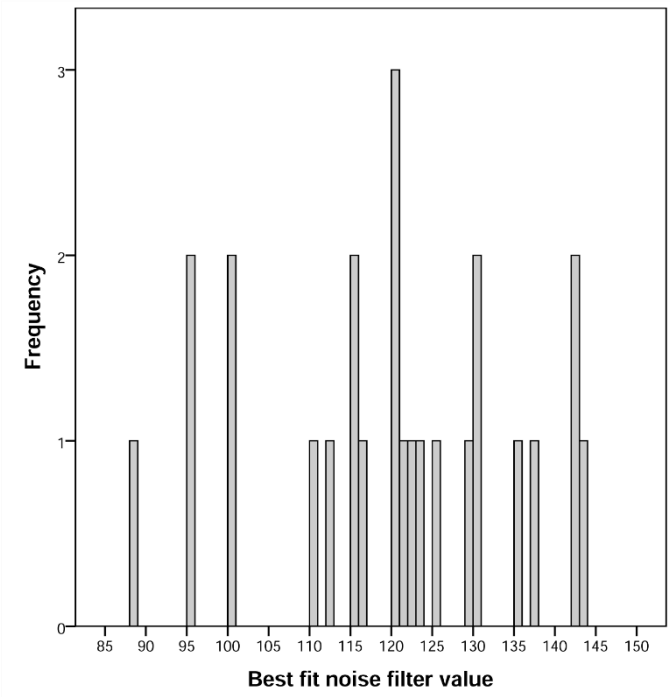


Figure 4.11: Histogram showing the frequency with which each noise filter value was used (total n=25).

4.3.2 Question Two: Is there a single noise filter value which is appropriate for all of the images?

The best-fit solutions were obtained with a range of noise filter values (Table 4.2 Figure 4.11). While there does appear to be a vaguely normal distribution of the values (with a central tendency of around 120), there was no one value which was appropriate for most images: for the 25 images tested, 18 different values were used (Figure 4.11).

4.3.3 Question Three: Do results vary between or within examiners?

For two of the examiners, using the find maxima resulted in identical outputs for the same images (i.e. both examiners produced the same answer three times, and the two examiners agreed on what that answer should be (Figure 4.12; Table 4.3). For the other examiner, some variation was observed (i.e. the identical, but rotated, images resulted in slightly different outputs) (Figure 4.12). Crucially, however, the variation seen between the three examiners was not statistically significant – that is to say, the ANOVA resulted in p values greater than 0.05, which is conventionally taken to denote results which do not indicate a relationship worth further investigation (Table 4.3).

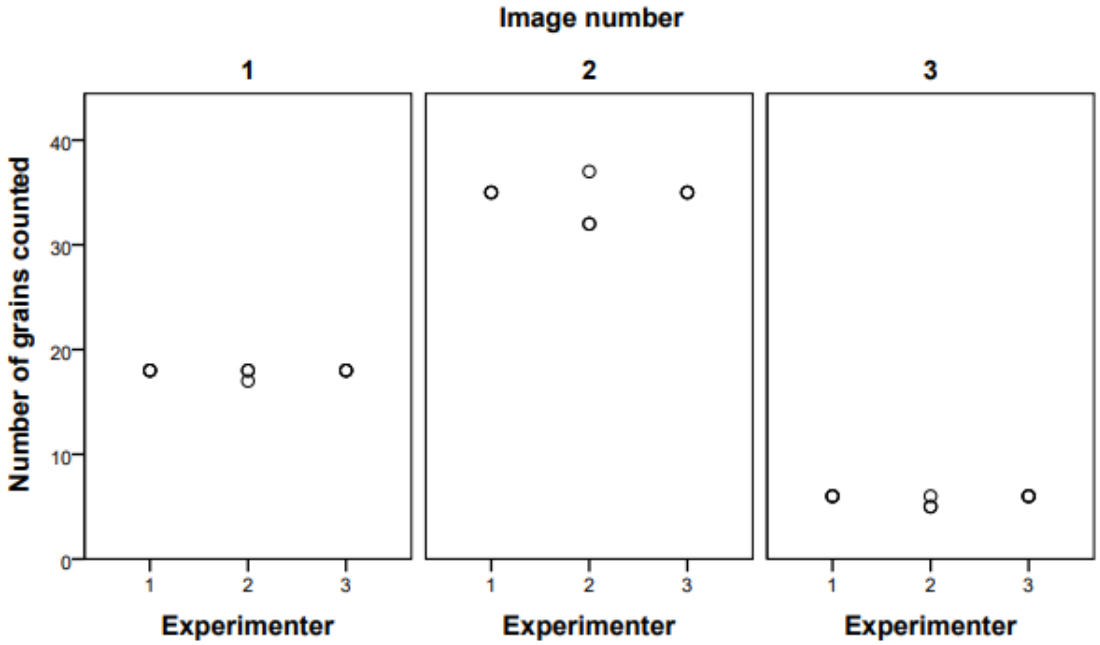


Figure 4.12: Scatter graph showing the number of grains counted by each examiner for each image (n =3)

Image section	Examiner	N	M	SD	ANOVA results, comparing examiners		
					F ratio	p-value	p<0.05?
1	1	3	35	0.00	0.64	.559834	No.
	2	3	34	2.89			
	3	3	35	0.00			
2	1	3	18	0	1	.421875	No.
	2	3	18	0.58			
	3	3	18	0			
3	1	3	6	0	4	0.078717	No.
	2	3	5	0.58			
	3	3	6	0			

Table 4.3: Summary of descriptive and inferential statistics for Experiment Two, showing the mean and standard deviation for the number of quartz grains counted by each examiner, plus ANOVA results, comparing the outputs of each examiner for each image

4.4 Discussion

4.4.1 Comparing 'find maxima' to repeat manual counting

For 22 of the 25 images, the number of quartz grains identified by the find maxima function was within the ranges identified by a human examiner. The maximum residuals observed were ± 2 grains from the manually-defined mean. These results therefore suggest that the find maxima function is consistently capable of delivering the correct answer for the number of quartz grains present on a swatch, or an answer very similar to the one that a human examiner would provide. In other words, in the context of counting quartz grains from light micrographs in the context of a persistence study, utilising a 'find maxima' function could be considered an accurate method, provided that an appropriate noise threshold has been defined.

4.4.2 Noise filter values for batches of images

Regarding appropriate noise threshold values, it was observed that even for a batch of very similar images, there was no one value which was appropriate for the entire set of images: for the training set of 25 images, 18 different noise filter values were used. This suggests that it is not appropriate to set up a fully-automated macro with a single, fixed value for the noise filter to deal with a batch of images. Instead, in order to use this technique to result in an accurate and sensible output, it is necessary for an examiner to pay attention to each image, and manually define an appropriate noise threshold level. It can be argued that this is subjective – but in the data seen in this experiment (see: section 4.3.3) statistically significant differences were not seen between the counts of different examiners using the Find Maxima function.

4.4.3 Reliability between examiners

For two of the three examiners, using the find maxima resulted in identical outputs for the same image. For the remaining examiner, some minor variation was observed, but this was not statistically significant – the ANOVA presented in Table 4.3 suggests that the differences are not a trend worth further investigation. This suggests that the find maxima function can be considered reliable between

examiners (i.e. the same input image is likely to result in similar if not identical outputs) and for one examiner performing analysis repeatedly (which is, again, likely to result in similar if not identical outputs). It does suggest, however, suggest that some examiners might be better at using the technique than others. It might therefore be worth performing a quick preliminary test to see whether the examiner is likely to perform consistently when using the find maxima technique, before undertaking a larger study.

4.4.4 Limitations and Implications

It is worth emphasizing that this study has several limitations, chiefly that the sample sizes were small (only three images were used in Experiment Two) and that the number of participants was also modest (with three participants in Experiment Two). The findings should also be couched in the same caveat that surrounds Chapter 3; images with different histograms and levels of noise and contrast may react differently to the same processing operations, and so it is important that suitable testing is carried out before an image processing technique is used on a larger dataset. Having said this, the findings of this study have three main implications:

1. Feasibility

   Together, these findings suggest that using the find maxima function may be appropriate for locating and counting the number of quartz grains on a dark fabric background, when variation seen in the brightness level of the background would render a thresholding function a poor reflection of the actual extent of the foreground in the image.

2. The expansion of the dataset ceiling

   It is possible to suggest that the ability to automate the counting of quartz grains would enable researchers to collect much larger datasets, exploring more ambitious hypotheses, but also offering more robust explorations of more modest hypotheses. To these ends, the find maxima function will be employed in

Chapter 5, in a persistence study exploring whether minor variations in the level of activity, and the position on the shoe from which samples are taken, influence the persistence of quartz grains on footwear over time.

3. The importance of interdisciplinary research

The subtext of this research project is that using tools that are considered very basic by computer science standards (for example, conceptualising of images as matrices of values for brightness, and applying mathematical operations to these values to locate regions of interest) can accelerate traditionally tedious or slow analyses within forensic science research. This really highlights the importance of interdisciplinary research – of reading around cognate disciplines to borrow advances which can be of benefit (provided that suitable foundational studies are carried out!).

It also suggests the research potential of open-source software to assist in forensic research. Since ImageJ is an accessible tool - free to download, with no cost barrier, and easy to use, with an active community of contributors and a graphical user interface meaning that these sorts of mathematical image processing functions do not require an in-depth knowledge of coding to achieve. (e.g. Schneider *et al.,* 2012; Schindelin *et al.,* 2015; Legland *et al.,* 2016; Arena *et al.,* 2017).

Consequently, it is hoped that other researchers might take advantage of this sort of mathematical operation, and consider that for a discipline that has been so historically entrenched in the analysis of imagery and dependent upon analytical microscopy, there are many possible ways of accelerating analysis without compromising levels of accuracy or reliability. Where tasks are repetitive, automation offers consistent and rapid analysis without fatigue, freeing the analyst to identify the trends from these larger datasets rather than doing the grunt work on the datasets themselves.

4.5 Conclusions

4.5 Conclusions

In summary, a number of conclusions can be drawn:

1. Using the find maxima function to locate and count quartz grains can identify the same number of quartz grains in an image of a fabric swatch as a human examiner, but at a faster rate.
2. There is not one noise filter value which is a one-size-fits all appropriate for dealing with a batch of images of quartz on dark fabric. Instead, if one wished to use 'find maxima' to count quartz grains, it would be necessary to treat each image individually, adjusting the noise filter value until the identified maxima reflect the image contents. This is still much faster than manual counting.
3. When different examiners use the 'find maxima' technique, they are likely to arrive at the same answer, or different answers that are not statistically significantly different from each other (i.e. answers that can be considered 'the same answer'). The same is true of one examiner repeatedly performing the operation on an image.

Together, these findings suggest that the 'find maxima' function can form a fast, accurate, and repeatable method of counting the number of quartz grains in a light micrograph. Since this operation can be performed in open-source software with an intuitive graphical user interface, it is also an accessible method. Accordingly, it is possible to suggest that this method could enable researchers to count quartz grains much more rapidly within the context of a forensic persistence study. This acceleration of the counting process (and the fact that the results are likely to be repeatable) should help to raise the ceiling on the number of data points which can be feasibly collected within such studies. It is worth reiterating here in the conclusion that these mathematical image processing operations are considered very basic (and quite old) by computer science / machine vision / image processing standards. It is possible to suggest that by the standards of theoretical forensic research, however, applying these techniques is considered relatively sophisticated (and quite new).

5

Chapter Five

Do variations in the activity of a wearer, and the position of sampling, appear to affect the persistence of particulate trace evidence on footwear uppers?

Abstract

There is a gap in the empirical evidence base for forensic science, even for concepts as central to the discipline as transfer and persistence. Techniques for digital image processing and analysis, as outlined in Chapters 3 and 4, mean that it is now possible to accelerate the analysis associated with such studies of trace evidence dynamics. A series of experiments were carried out in order to investigate whether (1) minor variations in the level of activity (0, 5, 10, or 15 minutes walking per hour) and (2) variation in the position of samples on a shoe (whether they were attached to the heel, inner, outer, or toe) affected the levels of persistence for quartz grains and fluorescent powder, over 8 hours. For quartz grains, no statistically significant differences in the amount of grains retained were observed between activity levels or between sampling positions. For fluorescent powder, statistically significantly differences were not consistently observed. These findings have methodological implications for future studies of transfer and persistence, suggesting that sampling multiple locations on one item of footwear over time is a valid proxy for sampling one location on the footwear repeatedly. These experiments also demonstrate the importance of undertaking such empirical work – it may seem counterintuitive that, after 8 hours, there is no statistically significant difference between the percentage of particles recovered after walking for 5 minutes per hour or 15 minutes per hour, but that is what this dataset suggests. For nuanced and robust interpretation of forensic evidence within a casework scenario, it is not sufficient to rely on what seems intuitive – it is necessary to test hypotheses.

5.1 Introduction

5.1.1 The need for empirical evidence bases

As stated in Chapter 2, one of the critical challenges facing the modern forensic sciences is the need to develop empirical evidence bases to facilitate the robust interpretation of forensic evidence (e.g. NAS, 2009; Mnookin *et al.,* 2011; Forensic Science Regulator, 2015; PCAST, 2016; Morgan, 2017). Within the United Kingdom, there is particular emphasis upon studies into the transfer and persistence of trace evidence (Forensic Science Regulator, 2015), with recognition that it would be useful to have datasets pertaining to the variables which might affect the initial levels of transfer, subsequent levels of loss, and the redistribution of evidence over time (ibid; Forensic Science Regulator, 2015, 2016, 2018, 2019). An understanding of such processes is fundamental to a robust and nuanced interpretation of trace evidence within forensic reconstruction scenarios (Morgan, 2017).

Despite this very clear rationale for research, not all forms of trace evidence have received equal (or adequate) levels of attention. While robust and detailed work has been carried out in the arenas of DNA persistence (e.g. Meakin and Jamieson, 2013) and the persistence of fibres upon garments (e.g. Lepot, *et al.,* 2015), there is still a relative dearth of studies which consider environmental trace evidence such as sediment, pollen grains, or algae (Morgan *et al.,* 2014). There is a similar scarcity of studies which consider footwear as the substrate. While researchers are aware of variables that might affect the persistence of material (e.g. Houck, 2001) for many environmental forms of trace evidence they have not yet been systematically explored, and where they have been considered sample sizes are small and the number of replicates is low (e.g. Scott *et al.,* 2014; Levin *et al.,* 2017).

5.1.2 Obstacles to empirical research

It is possible to suggest that one of the reasons why so few structured studies of transfer and persistence of environmental trace evidence exist is the time-consuming (and highly-specialised) nature of traditional microscopic analyses. One can illustrate this by considering the analysis of diatoms, a form of robust and widely-distributed unicellular algae which have proven useful in casework (Jones, 2007; Scott *et al.,* 2014). To undertake a persistence study of diatoms on a forensically-relevant substrate such as clothing or footwear would involve considerable laboratory preparatory work, and a very large number of hours on a microscope (see: DuBuf and Bayer, 2002; and the methodology sections of Scott *et al.,* 2014; 2017). This places a relatively low ceiling on the number of samples which can reasonably be prepared and analysed during the course of a research project. The requirement for facilities and personnel (expertise and physical resources to facilitate taxonomic identification of species) also restricts the number of institutions which would be capable of supporting such research.

5.1.3 The potential of proxies (and automation) to accelerate analyses

As seen in Chapters 2 and 3, one proposal to expedite the collection of relevant data is to use a proxy for such environmental evidence types, which can be imaged instantaneously. Fluorescent powder has been employed as exactly such a proxy for silt-sized particles (Morgan *et al.,* 2009) and pollen grains (Morgan *et al.,* 2013), in addition to anthropogenic particles such as lighter flint particles (Bull *et al.,* 2006), and gunshot residue (French *et al.,* 2012). Not only does the use of ultraviolet powder speed up the acquisition of data (since samples can be imaged using photography, rather than requiring removal, chemical treatment, the preparation of slides and then the scanning of slides), but it also has the potential to be coupled with *image processing techniques* to *automate* and *accelerate* data analysis. As has been seen in Chapters 3 and 4, such methods are capable of providing similar

levels of accuracy to human examiners, with improved speed and reliability (see also Kloster *et al.,* 2017). As a result of this increased speed, much larger datasets can feasibly be collected and analysed, and the sort of experiments which were formerly impractical become comfortably feasible.

5.1.4 Aims

As discussed in Chapter 2 (and above), there are significant gaps in our understanding of the transfer and persistence of forensically-relevant environmental particulates, and very few variables have been explored systematically. Where they have been explored, sample sizes have been small (e.g. Pounds and Smalldon, 1975a), and the number of replicates low (e.g. Scott *et al.,* 2014; Levin *et al.,* 2017). The intention of this chapter was to apply the segmentation tools developed in Chapters 3 and 4 to two forms of environmental particulate trace evidence: quartz grains (as discussed in Chapter 4), and fluorescent powder as a proxy for environmental particulate evidence in the region of 15 um diameter (as discussed in Chapter 3).

Specifically, this chapter sought to explore the effects of two variables, asking "Does altering the value of these variables result in a difference in the proportion of particulate evidence retained?" The two variables explored were (1) The location on the footwear that was sampled (the heel, the toe, the inner, and the outer of the shoe); and (2) The level of activity of the shoe's wearer (walking for 0, 5, 10, or 15 minutes per hour).

Both of these questions have theoretical and practical implications for both future research and casework. Considering the location on the shoe that is sampled, if it were demonstrated that there were certain regions of a shoe upper where particulate

trace evidence is more likely to persist over longer timescales, this might guide the locations which might be sampled preferentially during investigations. Likewise, if it is demonstrated that there are differences in the quantity of particulates recovered between different locations on the footwear, this is a variable which must be controlled in future experimentation. Concerning activity, the results will also have the potential to guide future experimental design. If minor variations in activity levels introduce a significant amount of noise into a persistence investigation, it would be a priority to strictly control activity levels in future investigations.

5.2 Materials and methods

5.2.1 Materials: swatches

Swatches were constructed in three layers, bound together with a staple (Figures 5.1 and 5.2). The lowest layer was a thin fabric, which could be easily attached to the shoes of the experimenter with safety pins. The top layer was a 20 x 20 mm swatch of the recipient surface under investigation, which in this experiment was black cotton (as used in Chapters 3 and 4. The intermediate layer was dark card, which acted to ensure that holes in the weave of the fabric would be represented by pixels with a low brightness value when imaged under visible light (i.e.to prevent them from being mistaken for foreground in the quartz element of the experiments). The staple served a dual purpose. It (a) bound the three layers together without introducing adhesives which might have altered the physical properties of the material and (b) provided a landmark that could be used for image registration. It is acknowledged that the cloth swatches used for the experiments may have contained mineral/moisture elements unknown during the process of the experiment set-up.

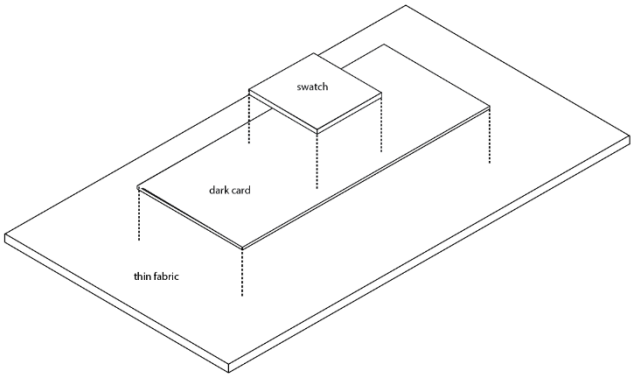


Figure 5.1: Exploded isometric drawing of the swatch design

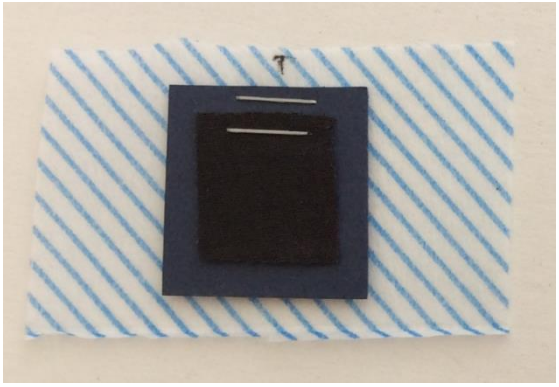


Figure 5.2: Photograph of a black cotton swatch

5.2.2 Materials: Transferred particles

The fluorescent powder used was the same described in Chapter 3, with a mean grain diameter of approximately 15 µm (see: French *et al.,* 2012), distributed across fabric by flicking with a stiff-bristled brush (after: Morgan *et al.,* 2010), and fluorescing green at a wavelength of 385 nm (see: Levin *et al.,* 2019). The quartz grains used were the same as in Chapters 4 and 6, obtained from a bag of horticultural 'silver sand'. It is acknowledged that the silver sand could have contained mineral/moisture elements unknown during the process of the experiment set-up). The quartz grains were distributed onto the swatches by direct contact: the swatch was pressed into a container of quartz grains, as in Chapter 6. The quartz grains ranged in cross-sectional area from 0.002 – 0.630 mm^2, and the size distribution was unimodal, with a mean grain area of 0.098 ± 0.077 mm^2 (n = 500) (see Chapter 6, especially section 6.2.2)

5.2.3 Methods: The persistence study

Swatches were attached to the shoes in four positions: the toe, the heel, the inner, and the outer (Figure 5.3). The shoes were worn for eight hours, and the activity of the wearer was strictly controlled, at either 5, 10, or 15 minutes of walking per hour (with the remaining time spend sedentary). A control run of 'zero' activity was also conducted, where the shoes were not worn (and so the loss of evidence should be down to the influence of gravity and removing the swatch alone). The swatches were detached and imaged at five time intervals: at 0, 1, 2, 4, and 8 hours after the transfer. Five replicates were conducted for each combination of variables.

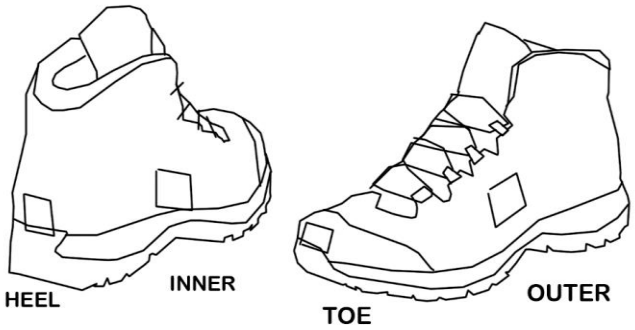


Figure 5.3: Diagram of swatch locations on the shoe

5.2.4 Methods: Imaging Fluorescent powder

Within ultraviolet persistence studies, sampling has tended to involve macro-lens digital photography in either bespoke dark-boxes, darkened rooms, or dedicated imaging suites (see: Chapter 3; French *et al.,* 2012; Levin *et al.,* 2019). The imaging tends to be undertaken either with an SLR or other digital image capturing device (ibid.). It is acknowledged that this method of data acquisition relies on trust in the black-box method of the software that generated the images, as is the convention in this field of experimental forensic science (for example, as in French et al., 2012; Morgan et al., 2013) – which was the case for these experiments. It is also acknowledged, as is the case with chapters 2 and 3, that further image characterisation (such as the exact magnification and resolution of the images, having knowledge about the sensors and image quality; for example, considering uniformity, noise, focus, vignetting, compression, pixel resolution, spatial resolution, Bit depth, and the presence of any dead pixels on the sensor, would be essential for the data to be used more generally.)

For these experiments, a bespoke dark-box was constructed to enable measurements to be taken repeatedly at the same focal distance, with the same illumination, while excluding visible light and containing the ultraviolet light. The box was constructed from opaque continuous cast acrylic (CCA) using a laser-cutter. An ultraviolet torch of 385nm was mounted at one end, and an iPad used as the imaging device, with a hole positioned to allow the lens to be aligned with a sample placed next to a hinged access panel (Figures 5.4 and 5.5). It is acknowledged that the use of a bespoke UV dark box introduces a number of uncertainties. Unlike a dark box and light source manufactured by an established forensic brand (for example, Foster and Freeman's CrimeLite series), there was no guarantee of constant light output as battery charge levels changed, no guarantee of the even-ness of illumination (although efforts were made to ensure that UV light was reflected within the box, with the internal walls coated in reflective tinfoil), and although every effort was made to ensure that the camera lens and sample were parallel (i.e. that the images were not taken at an angle, introducing foreshortening to the image), this was not certified. Accordingly, the use of a bespoke box, while a cost-effective and accessible solution, introduces a number of unknowns into the data acquisition.

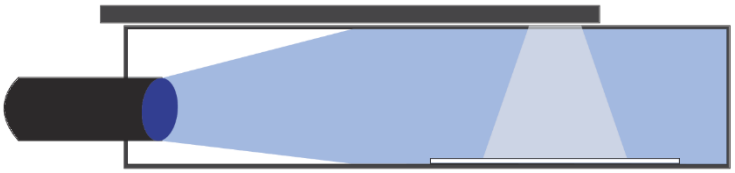


Figure 5.4: Diagram showing the set-up of the dark box, with a UV torch as the light source, and an iPad as the image capturing device

Figure 5.5: Photographs of the UV darkbox

5.2.5 Methods: Quantification of fluorescent powder

The images were cropped in Adobe Photoshop (CS4), so that they contained just the swatch (the region of interest) with no extraneous pixels representing the background of the darkbox. In ImageJ, a macro was employed to convert each RGB image to 8-bit (a requirement of the Yen thresholding algorithm in ImageJ), to segment it by applying the Yen global thresholding algorithm (see Chapter 3; Yen *et al.,* 1995; Levin *et al.,* 2019), and to save this segmented image as a copy, with a suffix. A second macro was employed to quantitatively analyse this segmented image, resulting in a number of measurements (including the percentage of the image that was identified as foreground).

5.2.6 Methods: Imaging quartz grains

The quartz grains were imaged under visible light with a USB microscope (essentially a webcam capable of achieving images at very high magnifications), as described in Chapter 4 (please see Figure 4.6). Again, it is acknowledged that this method of data acquisition relied on trust in the black-box method of the software that generated the images, as is the convention in this field of experimental forensic science (French et al., 2012; Morgan et al., 2013). The retort stand of the microscope was modified, with an x-y mechanical stage installed, and a 'swatch holder' fabricated. This to ensure that the sample could be translated, without rotation, beneath the lens (to enable image stitching using the ImageJ 'Grid/Collection Stitching' plug-in (Preibisch *et al.,* 2009). This was important because the images were taken at sufficient magnification that it was necessary to capture multiple images per swatch, and tile them back together (using the plugin 'Grid/Collection Stitching' (Preibisch *et al.,* 2009)). Each image contained between 10 and 150 quartz grains. Although efforts were made to control the illumination conditions, there was some variation in the contrast and brightness of the images.

5.2.7 Methods: Quantifying quartz grains

Quantitative analysis was conducted in ImageJ, using the 'find maxima' function (please see Chapter 4 for an explanation of the function, and justification of why its use is appropriate within this context). Each stitched image was loaded in turn, and the noise threshold of the find maxima operation was adjusted until no false-positive maxima were observed (i.e. until every point identified as a maximum was actually a quartz grain). For most images, this was an accurate reflection of the number of quartz grains present. For the images where it wasn't (where the examiner could see that one or two quartz grains were not labelled as maxima), the examiner manually added in the false-negatives (i.e. the quartz grains that had been missed by the function). This was far faster than manually counting all grains. All experiments were conducted by a single examiner.

5.2.8 Methods: Statistical Analysis

Inferential analysis was carried out using the Kruskal-Wallis test, a non-parametric test of whether independent samples are drawn from the same distribution (McDonald, 2009). The data generated here fulfil the assumptions of the test, being (a) independent measures (b) having a dependent variable measured at a scale level (better than ordinal level), and (c) having three or more groups (see Field and Hole, 2007: 275).

5.3 Results

5.3.1 Quartz: the effects of altering activity levels

Data for the effects of altering activity levels upon the percentage persistence of quartz grains are summarised in Figure 5.6 and Table 5.1. The three active runs (where the samples had been subjected to 5, 10, or 15 minutes walking per hour) produced typical two-stage decay curves (see: Pounds and Smalldon, 1975a), with an initial steep gradient followed by a subsequent more gradual decline (Figure 5.6). There was no statistically significant difference in the persistence of the quartz grains between these three activity levels for any shoe location sampled, at any time sampled in the study (Table 5.1). For the Kruskal-Wallis test (n=15) the p-value calculated was always greater than 0.05, with a minimum p-value of 0.21 observed (Table 5.1). Those samples which had been treated to 0 minutes activity per hour (i.e. where the shoes had not been worn), showed an almost linear pattern of loss (Figure 5.1). This confirms that activity does have an effect, but here it appears that minor variations in the amount of walking do not alter the extent of that effect.

5.3.2 Quartz: the effects of sample position

Data for the effect of altering the position on the shoe from which the sample was taken upon the persistence of quartz grains are summarised in Figure 5.6 and Table 5.2. Visually inspecting the data, it appears possible that differences exist between the different shoe positions: for the samples where the shoes had not been worn (i.e. with '0 activity' per hour), the toe consistently demonstrated the lowest levels of particulate loss, and the heel the highest, while for the runs with ten minutes of activity, the heel showed the lowest levels of loss (Figure 5.8). The error bars are, however, large, and statistical analysis confirms that there is no statistically significant difference in the median percentage persistence in quartz grains between the four sampling locations for any level of activity, at any time in the experiment (see Table 5.2 for Kurksal-Wallis output. for every combination of variables the n=20 and p-values were greater than 0.1).

5.3.3 Fluorescent powder: the effects of altering activity levels

Data for the effects of altering activity levels upon the percentage persistence of fluorescent powder are summarised in Figure 5.7 and Table 5.1. Similar to the quartz results, the three active ultraviolet runs (with 5, 10, or 15 minutes walking per hour) produced typical decay curves, while the run in which the shoes were static produced a different pattern (Figure 5.7). Conducting a Kruskal-Wallis test to compare the three active runs (i.e. to compare whether altering the activity from 5 to 10 or 15 minutes per hour had an effect on the percentage persistence), statistically significant differences were not consistently seen (Table 5.1). The p-value was less than 0.05 for 2 of the 16 variable combinations, but the p-value was greater (and in most cases, considerably greater) than 0.05 for the remaining 14 (Table 5.1).

5.3.4 Fluorescent powder: the effects of sample position

Data for the effect of altering the position on the shoe from which the sample was taken upon the persistence of fluorescent powder are summarised in Figure 5.7 and Table 5.2. Visually inspecting the data, it appeared that the error bars of the different shoe positions all overlapped (Figure 5.7). Conducting a Kruskal-Wallis test to compare whether altering the location of the swatch had an effect on the percentage persistence, statistically significant differences were not consistently seen (Table 5.2). The p-value obtained was less than 0.05 for 4 of the 16 treatments, but p was greater than 0.05 for the remaining 12 (Table 5.2).

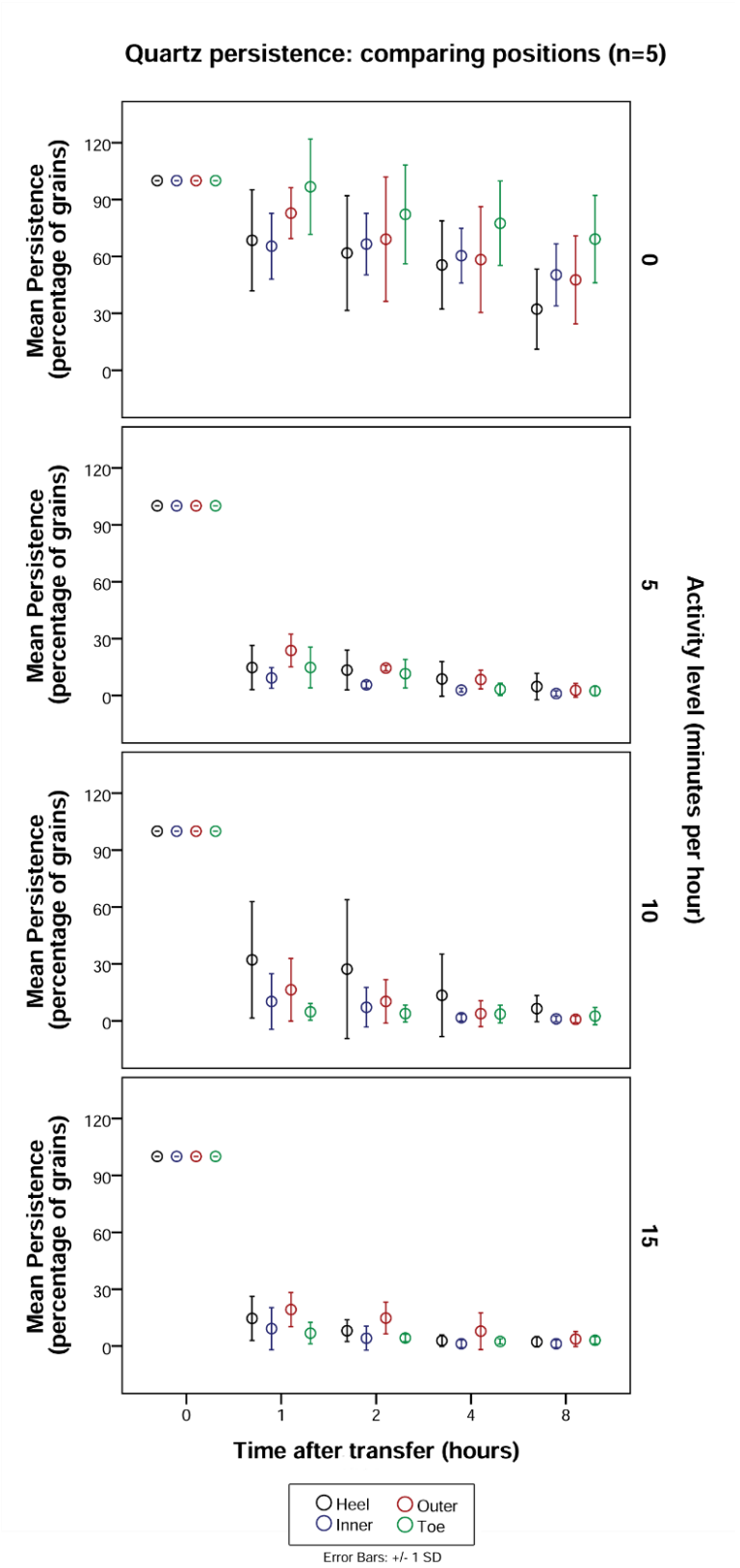


Figure 5.6 Scatter chart showing the mean percentage persistence of quartz grains, for each time interval, level of activity, and shoe position. Error bars indicate ±1SD, (n = 5)

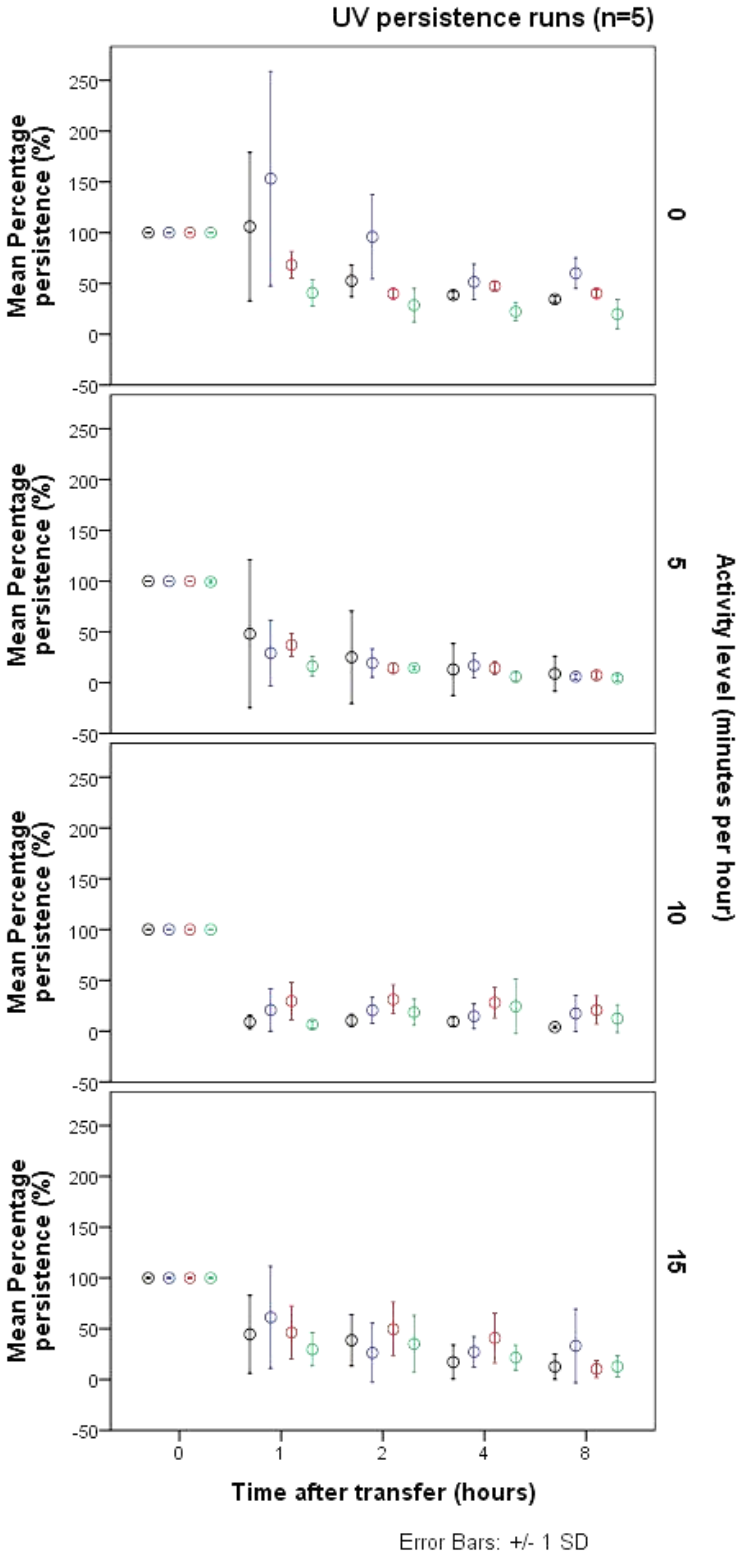


Figure 5.7 Scatter chart showing the mean percentage persistence of fluorescent powder, for each time interval, level of activity, and shoe position. Error bars indicate ±1SD, (n = 5)

Null hypothesis	Combination of variables			Kruskal Wallis p-value	
	Time	Location of sample	N	Quartz	UV powder
"the medians of the percentage persistence are the same across the 5, 10, and 15 minute categories of 'activity'"	1	Heel	15	0.468	0.160
		Inner	15	0.613	0.281
		Outer	15	0.651	0.454
		Toe	15	0.203	0.027
	2	Heel	15	0.512	0.087
		Inner	15	0.219	0.990
		Outer	15	0.697	0.014
		Toe	15	0.210	0.379
	4	Heel	15	0.533	0.403
		Inner	15	0.237	0.336
		Outer	15	0.432	0.061
		Toe	15	0.911	0.141
	8	Heel	15	0.766	0.543
		Inner	15	0.962	0.075
		Outer	15	0.372	0.181
		Toe	15	0.499	0.275

Table 5.1: Kruskal-Wallis results comparing between different activity levels (N=15)

Null hypothesis	Combination of variables			Kruskal Wallis p-value	
	Time	Activity (min/hr)	N	Quartz	UV powder
"the medians of the percentage persistence are the same across all categories of 'location on shoe'."	1	0	20	0.146	0.058
		5	20	0.125	0.415
		10	20	0.129	0.021
		15	20	0.177	0.706
	2	0	20	0.638	0.010
		5	20	0.233	0.385
		10	20	0.247	0.058
		15	20	0.070	0.444
	4	0	20	0.575	0.007
		5	20	0.200	0.099
		10	20	0.512	0.179
		15	20	0.636	0.446
	8	0	20	0.610	0.006
		5	20	0.847	0.185
		10	20	0.423	0.136
		15	20	0.428	0.395

Table 5.2: *Kruskal-Wallis results comparing the medians between different shoe positions (N=20)*

5.4 Discussion

5.4.1 The effects of sampling position

When comparing the effects of taking the samples from the toe, heel, inner, or outer of the shoe upper, for quartz, no statistically significant differences were seen. For fluorescent powder, a statistically significant difference was not consistently seen (i.e. was observed for only 4 of the 16 combinations of variables). Accordingly, this suggests that over the course of eight hours, ultraviolet powder and quartz do not persist preferentially on one region of the shoe upper, nor are they lost more rapidly from one particular region of the shoe upper.

Implications: Evidence recovery

In terms of recovering evidence, this suggests that there is not one particular region on the shoe that might be most useful to sample on the outer of the shoe's upper. This does not, however, contribute as to whether there is variation between the upper and the sole, or comparison between the interior and exterior of the shoe. Some have suggested that the interior of the shoe may represent a useful repository of trace evidence (e.g. Morgan *et al.,* 2009; Stoney *et al.*, 2016;), and also that the material is an important controlling factor (e.g. Levin *et al.,* 2017). Therefore, this indicates that, for evidence recovery, preferentially sampling the woven fabric elements of a shoe, whatever their position on the upper, may represent an approach for optimal particle recovery.

Implications: Methodological development

In terms of methodological development, it might be seen as reassuring that removing swatches at multiple proximate locations results in no statistically significant difference in the percentage persistence for fluorescent powder or quartz. This lack of statistically significant difference suggests that using multiple proximate samples as a substitution for repeatedly sampling one position over time (for example, where sampling is destructive, as in the case of diatom analysis where swatches might be subjected to peroxide digestion) might be considered a methodologically sound substitution.

Redistribution of trace evidence

In the wider literature on transfer and persistence, which generally considers the substrate of clothing, redistribution is often highlighted as a potential issue (e.g. Morgan *et al.,* 2010). While this particular study did not find a statistically significant difference between the percentage persistence for swatches on the toe, heel, inner, and outer, it is possible that further analysis on the position of the foreground pixels (and not just the quantity) would reveal redistribution trends within each swatch (though not necessarily between). It is possible that redistribution may be occurring due to gravity.

Limitations and future avenues for research

One critical limitation of this study is its short timescale; persistence was measured over only 8 hours, and so it is therefore possible that some processes had not yet happened to an observable extent. Other limitations include the relative paucity of intervals for the variables (future research could consider more locations on the shoe, such as inner versus outer, and a comparison of the upper and sole), and the relatively modest size of the dataset – only five replicates were conducted, and all involved the same experimenter. Another possibly important factor is that, in this experiment, the imaging set-up required the removal of swatches, so that they could be imaged beneath, and parallel to, the imaging device. It is possible, if not plausible, that this removal of the swatches from the shoes, and re-attaching of the swatches to the shoes, might have resulted in the loss of some particles. Future research could explore whether this is a significant factor, and possibly to develop a method of imaging the swatches *in situ,* without the need to detach the samples. Future research may also consider the redistribution patterns of the particles (a perfectly possible hypothesis to explore, if one were to compare the images taken over time).

5.4.2 The effects of activity levels

A statistically significant difference between the three activity levels of 5, 10, or 15 minutes of walking per hour was not consistently observed for either quartz grains

or for fluorescent powder. This suggested that minor variations in the amount of activity (e.g. ± 5 minutes of walking per hour) are unlikely to make a statistically significant difference to the percentage persistence of ultraviolet powder during persistence studies.

Implication: Methodological development

This finding implies that, in future persistence studies, although efforts should be made to control the variable of amount of activity, minor variations (on a scale of minutes per hour) would not render data unusable.

Implication: The importance of testing assumptions

These experiments also demonstrate the importance of undertaking such empirical work – it may seem counterintuitive that, after 8 hours, there is no statistically significant difference between the percentage of particulates retained when the wearer of a shoe has been walking for 5 minutes per hour or 15 minutes per hour, but that is what this dataset suggests. For nuanced and robust interpretation of forensic evidence within a casework scenario, it is therefore not sufficient to rely on what seems intuitive – it is necessary to test hypotheses.

Limitations

The limitations outlined above, around modest sample sizes, modest numbers of intervals under investigation, and *in situ* sampling apply. It is also worth noting that in this experiment, only the duration of the activity was investigated. It is possible that the intensity of activity (e.g. running versus walking) might have a statistically significant effect (see: Morgan *et al.,* 2009).

5.5 Conclusions

5.5 Conclusions

In summary, three conclusions can be drawn from this work:

(1) A study was carried out to investigate whether the location at which a shoe was sampled (heel, toe, inner, or outer) affected the percentage persistence of ultraviolet powder on footwear uppers. No statistically significant variation was observed.

(2) A study was carried out to investigate whether minor changes in the level of activity affected the percentage persistence of ultraviolet powder on footwear uppers. Statistically significant variation was not consistently observed.

(3) Together, these findings suggest that in future experiments minor variations in the position of swatches and in the amount of activity of the wearer would not compromise the data generated.

Together, these findings confirm the importance of empirical investigation to explore trace evidence dynamics, and underpin the inferences made within casework (Forensic Science Regulator, 2015; 2019).

6

Chapter Six

Does the morphology of quartz grains affect their transfer and/or persistence?

Abstract

Quartz grains can represent a valuable environmental indicator; they are almost ubiquitous within soils, and can provide an experienced examiner with a rich amount of information about the environment from which they were drawn. Within forensic science, experiments with glass fragments and fibres have demonstrated the propensity for smaller traces to persist over longer timescales. The aim of this chapter was to conduct an experiment to establish whether a similar size-selective retention is observed with respect to quartz grains, asking: does the size (as measured by cross-sectional area) and morphology (as indicated by the aspect ratio and solidity) of a quartz grain appear to influence the likelihood of transfer or the likelihood of persistence? Findings were threefold: (1) The largest particles in the population did not appear to transfer (2) The largest particles that did transfer were lost first and (3) The values observed for the morphological characteristics of solidity and circularity appeared to change over time. Together, this research confirmed that the size and morphology of a particle may be an important consideration when interpreting the presence or absence of quartz grains within a forensic scenario, and that further research into these trends would be of merit.

6. 1 Introduction

6.1.1 Quartz grains as forensic evidence

As mentioned in the introductory literature review (Chapter 2) and revisited in the chapter regarding the segmentation of quartz grains (Chapter 4), quartz grains can represent a valuable form of forensic trace evidence (Bull and Morgan, 2006). The surface textures of the grains can be interpreted in order to make inferences about the environments from which they were drawn (Krinsley and Doornkamp, 1973; Mahaney, 2002; Bull and Morgan, 2006), and comparison between samples derived from suspects, victims, articles of evidence, and crime scenes can be used to exclude (or fail to exclude) contact within forensic casework (Morgan and Bull, 2006; Morgan *et al.,* 2006; Bull *et al.,* 2006).

6.1.2 Size selective trends within trace evidence dynamics

As discussed in the introductory literature review (Chapter 2), it has been suggested that the size and texture of transferred material can influence its persistence (Houck, 2001). Experimental studies have demonstrated that larger particles or fibres seem to persist over shorter timescales, with experiments exploring this hypothesis in the context of glass fragments (Brewster *et al.,* 1984; Hicks *et al.,* 1996), fibres (Pounds and Smalldon, 1975b), and diatom frustules (Levin *et al.,* 2017). It has also been suggested that particles of differing morphologies may have a different likelihood of transferring from a source population to a recipient surface (Scott *et al.,* 2018).

6.1.3 The measurement of particle size and morphology

The measurement of particle size, and in particular of the particle size distributions for soils, is a routine form of analysis in Geography, and one of the preliminary ways in which soils and sediments are characterised (Pye and Blott, 2004). Over the years, many different methods of measuring particle size have been developed

(Mazzoli and Moriconi, 2014). Some methods are contingent upon physical separation (e.g. sieving), others employ gravity-induced settling rates (e.g. sedimentation/ the SediGraph), while others consider x-ray attenuation or electrical impedence (e.g. Coulter Counters), or laser diffraction (e.g. the Malvern Mastersizer), (Vdovíc *et al.,* 2010; Mazzoli and Moriconi, 2014). It is worth noting that "many techniques make the useful and convenient assumption that every particle is a sphere" (Mazzoli and Favoni, 2012: 67). In recent years, many researchers have pursued Image Analysis as an alternative to the physical measurement methods (e.g. Mazzoli and Moriconi, 2014). It can be suggested that this is because it can offer two advantages. Primarily, as mentioned by Mazzoli, it does not assume that the particles are spherical, and is therefore capable of providing multiple descriptors of a particle's shape (Mazzoli and Moriconi, 2014; Mazzoli and Favoni, 2012). Accordingly, it is possible to pursue hypotheses relating to morphological properties beyond their radius and diameter, such as the aspect ratio (whether a particle is relatively square or relatively elongated) or the solidity (whether a particle is convex, or has many regions of concavity). Secondly, and crucially in the domain of forensic geoscience, it can represent a non-destructive technique. Looking at quartz grains under the microscope does not necessarily require the removal of the grains from the surface to which they have been transferred, nor homogenisation of the sample (as would be necessary if one were using, for example, a Mastersizer or Coulter Counter to collect particle size distribution data). Accordingly, if one uses image analysis to derive particle size data, it would be possible to undertake further tests upon the sample which had been imaged, or a longitudinal experiment could be allowed to continue undisturbed. (see: Morgan and Bull 2007, which outlines the philosophy of forensic geoscience, and includes discussion on the importance of non-destructive analysis where limited amounts of sediment are available.)

6.1.4 Image analysis for particle size data

The idea of using image analysis to derive quantitative data about the morphology of quartz grains is not new: attempts have been documented since the 1990s (see:

Kennedy and Mazzullo, 1991). The most recent attempts explore different methods of image capture; Light Microscopy, Scanning Electron Microscopy, and flatbed scanners have all been tested (e.g. Ignathinathane *et al.,* 2009; Cox and Budhu, 2008). The scanning electron microscope seems to be the most prevalently used, despite its relative expense and relatively time-consuming sample preparation protocols (Cox and Budhu, 2009). Crucially, for quartz grains, it has been demonstrated that when comparing the performance of Scanning Electron Microscopy (SEM) to Light Microscopy (LM), light microscopy produces images of a sufficient resolution for morphological analysis and comparison – what one loses is the ability to conduct textural analysis or chemical analysis, as would be possible using an SEM-EDX (Cox and Budhu, 2009). Many of the papers examining the potential to extract morphometric information and particle size analysis from digital images conduct analyses in ImageJ (e.g. Ignathinathane *et al.,* 2008, Cox and Budhu, 2009; Mazzoli and Favoni, 2012; Mazzoli and Moriconi, 2014). A number of plug-ins exist to facilitate such analysis, including MorphoLibJ (https://imagej.net/MorphoLibJ ; Legland et al 2016), Particle8_Plus (see: Cox and Budhu, 2008), and Shape Descriptor (Cox and Budhu, 2008). MATLAB code also exists to enact morphological analysis of particles, including the tools MORPHEO (Charpentier *et al.,* 2013), capable of conducting Fourier analysis, and MORPHEOLV, which can calculate roughness coefficients for grains (Charpentier *et al.,* 2017).

6.1.5 Aims

The overarching aim of this Chapter was to explore hypotheses relating to the transfer and persistence of quartz grains on footwear, over time, employing the image analysis approach for deriving particle size data. In other words, the intention was to respond to the queries raised by Scott *et al.,* 2018 (i.e. "Is particle morphology a significant factor that controls the transfer of evidence from a source population to an article of evidence or from the article of evidence to the

environment as loss?") and to utilise the image-analysis methods developed in recent years, combining light microscopy and ImageJ (after Cox and Buddhu, 2008) to answer the following questions.

Four specific questions were investigated:

1. Do different size fractions of quartz grains transfer preferentially?
   i.e. Comparing the source population and transferred population, are different size fractions represented proportionally, or do some sizes of quartz grains appear to transfer more or less readily from the source population? Are there any size fractions represented in the source population which do not appear to transfer to the fabric?

2. Do different size fractions of quartz grains persist equally over time?
   i.e. Comparing the initially transferred population and populations sampled after time has elapsed, are different size fractions represented proportionally, or do some sizes of quartz grains appear to have been lost more or less readily from the transferred population? Does the mean/median grain size recovered appear to change over time?

3. Do quartz grains with different morphologies transfer preferentially?
   i.e. Comparing the source population and transferred population, are grains with different roundness, aspect ratio, and solidity represented proportionally, or do some morphologies of particle appear to transfer more or less readily from the source population?

4. Do quartz grains with different morphologies persist equally over time?
   i.e. Comparing the initially transferred population and populations sampled after time has elapsed, are grains with different roundness, aspect ratio, and solidity represented proportionally, or do some morphologies of particle appear to have been lost more or less readily from the transferred population?) Does the mean/median roundness, aspect ratio, and solidity appear to change over time?

Together, it was hoped that these questions logically built upon the work from previous chapters (especially Chapter 5). As with previous chapters, the macro-scale intention was to (a) Generate a dataset that is relevant to forensic casework scenarios and would allow for the informed interpretation of evidence (see: Chapter 2), and (b) Explore novel (in this context) techniques of measurement and analysis which could potentially facilitate the faster collection of data in the future.

6. 2 Materials and Methods

6.2.1 Metrics of morphology

As this Chapter considers the influence of particle morphology on the transfer and persistence of quartz grains, it was considered important to use meaningful metrics of morphology, that could (a) meaningfully describe/characterise the morphology of each quartz grain; (b) be compared with the results of other studies; and (c) be undertaken without the need for significant expertise or esoteric, proprietary software. As with so much in forensically-relevant research, the cost of materials and methods is a non-trivial factor, and so a freeware option, with a graphical user interface, would be the ideal.

Particle Size and Shape Analysis

Historically, particle size analysis (which assumes sphericity, see section 6.1.3) has relied upon the metric of size, dividing the soils and sediments up into different size fractions. Where the morphological analysis of individual particles has been undertaken, it traditionally considered the sphericity (Krumbein, 1941) and roundness (Powers, 1953) of grains (see: Cox and Buddhu, 2008; Figure 6.1). Cox and Buddhu neatly summarise the development of other metrics, with the development of Fourier Series and Descriptors in the 1970s (Ehrlich and Weinberg, 1970; Beddow and Vetter, 1977; see Cox and Buddhu, 2008)

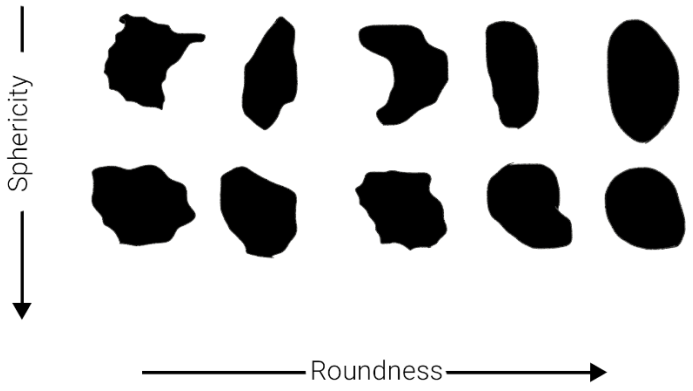


Figure 6.1: Roundness and Sphericity diagrams (after Krumbein, 1941 and Powers, 1953), modified from Chang et al., 2018

It is worth noting here that all of these different metrics broadly deal with three qualities: (1) magnitude (the size of the grains), (2) elongation (does the grain resemble a sphere, or does it have a different aspect ratio?), and (3) the roughness (is the surface smooth, or does it have projections and crevices?). Accordingly, the Metrics employed in this study were:

(1) A measure of size

The grain's cross-sectional area, in mm^2 (see Figure 6.2A)

(2) A measure of elongation

For this, two metrics were calculated: Circularity (Figure 6.2B) and Aspect Ratio (Figure 6.2C). Circularity is defined as $4\pi \times ([Area] / [Perimeter]^2)$. Accordingly, a value of 1.0 indicates a perfect circle, and smaller values are indicative of greater elongation[8]. Aspect Ratio refers to "the aspect ratio of the particle's fitted ellipse, i.e. [Major Axis / Minor axis][9]. Both measurements are provided since, in ImageJ's documentation, it is suggested that circularity "values may not be valid for very small particles" (https://imagej.nih.gov/ij/docs/guide/146-30.html). Additionally, a precedent exists for its use: it appears in both and also it was used by Cox and Buddhu (2008) and Chang *et al.* (2018).

(3) A measure of 'roughness'

The fourth and final shape descriptor employed within this study was 'Solidity' (ImageJ website, see also: citation). This is defined as = [Area]/[Convex area]; so a value of 1 indicates a perfectly convex object (e.g. a sphere), and any concave squiggles in the outline will result in lower solidity values (Figure 6.2D).

All can be calculated in ImageJ, and all can be interpreted intuitively. To give an example, the following shapes were generated (Figure 6.3) and the corresponding shape descriptors can be seen in Table 6.1.

[8] https://imagej.nih.gov/ij/docs/guide/146-30.html
[9] https://imagej.nih.gov/ij/docs/guide/146-30.html

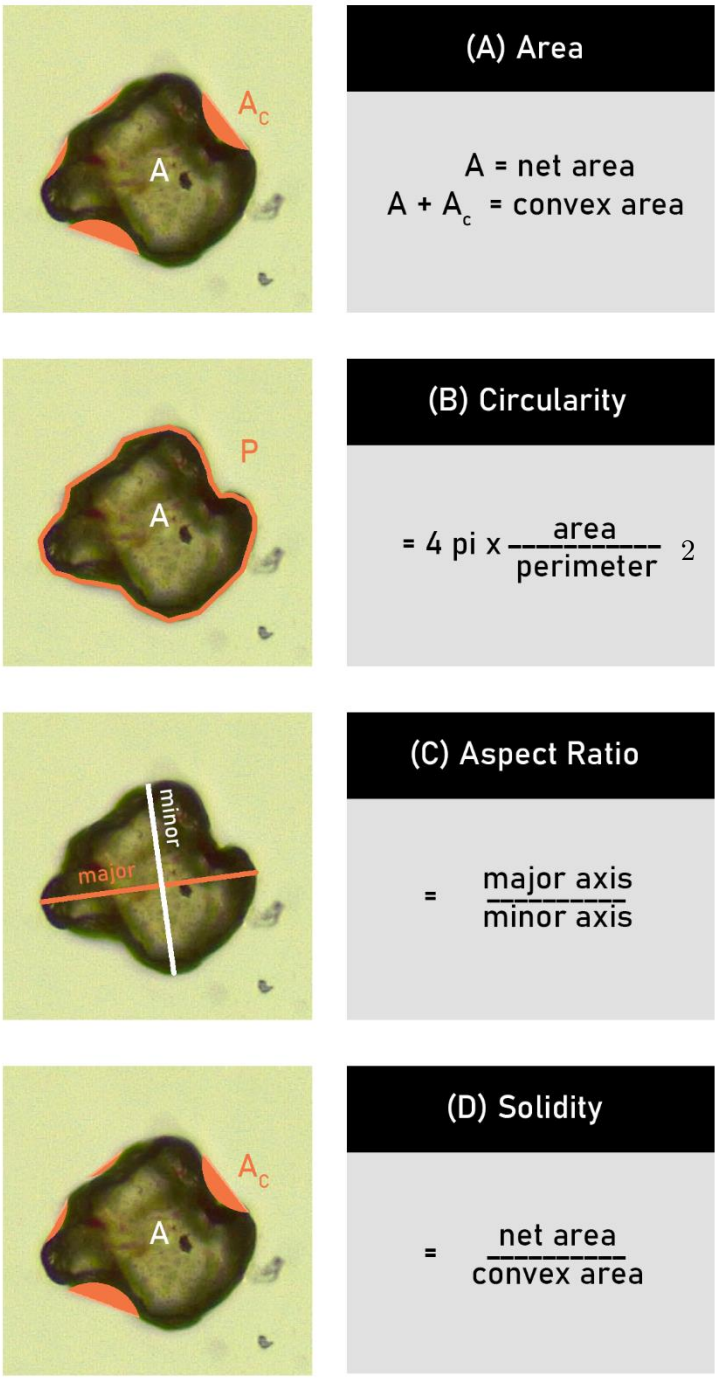


Figure 6.2: Shape descriptors (A) Cross-sectional area (B) Circularity (C) Aspect ratio (D) Solidity.

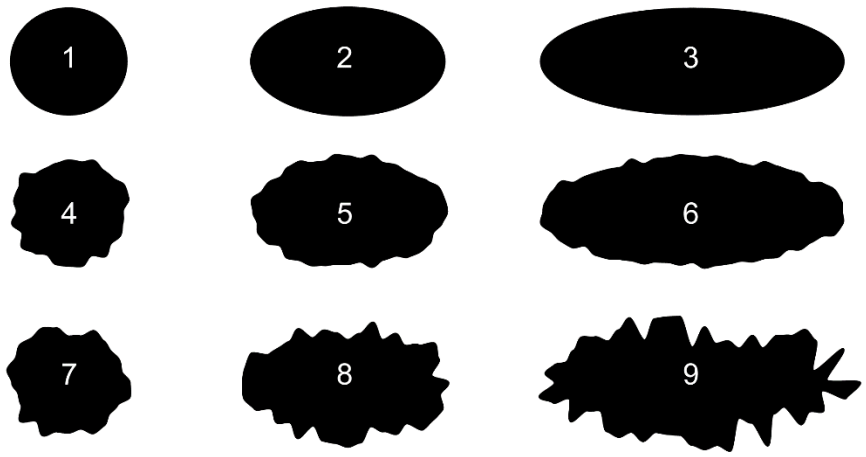


Figure 6.3: Example shape particles (generated in Adobe Illustrator)

Shape number	Circularity	Aspect ratio	Roundness	Solidity
1	0.90	1.09	0.92	0.99
2	0.80	1.78	0.56	0.99
3	0.63	2.84	0.35	0.99
4	0.78	1.10	0.91	0.95
5	0.74	1.76	0.57	0.96
6	0.58	2.80	0.36	0.96
7	0.74	1.13	0.89	0.93
8	0.51	1.86	0.54	0.88
9	0.26	2.56	0.39	0.80

Table 6.1: Example shape descriptors for particles in Figure 6.3

6.2.2 Materials: Quartz

The quartz grains used in this experiment were obtained from a sample of horticultural 'silver sand'[10]. It is acknowledged that the silver sand could have contained mineral/moisture elements unknown during the process of the experiment set-up. The grains ranged in size, as measured by cross-sectional area, from 0.002 – 0.630 mm^2, and the size distribution was unimodal, with a mean grain size of 0.098 ± 0.077 mm^2 (n = 500). Further details characterising the sand can be found in results section 6.3.1. (See especially Table 6.3). A summary description is provided in Figure 6.4. The sample consisted predominantly of rounded rather than elongated grains (mean aspect ratio 1.39 ± 0.28 (n = 500), with a minimum of 1.03 and a maximum of 2.95 observed), and with high levels of solidity (mean solidity = 0.97 ± 0.02, with a minimum of 0.86 and a maximum of 0.99 (n=500)).

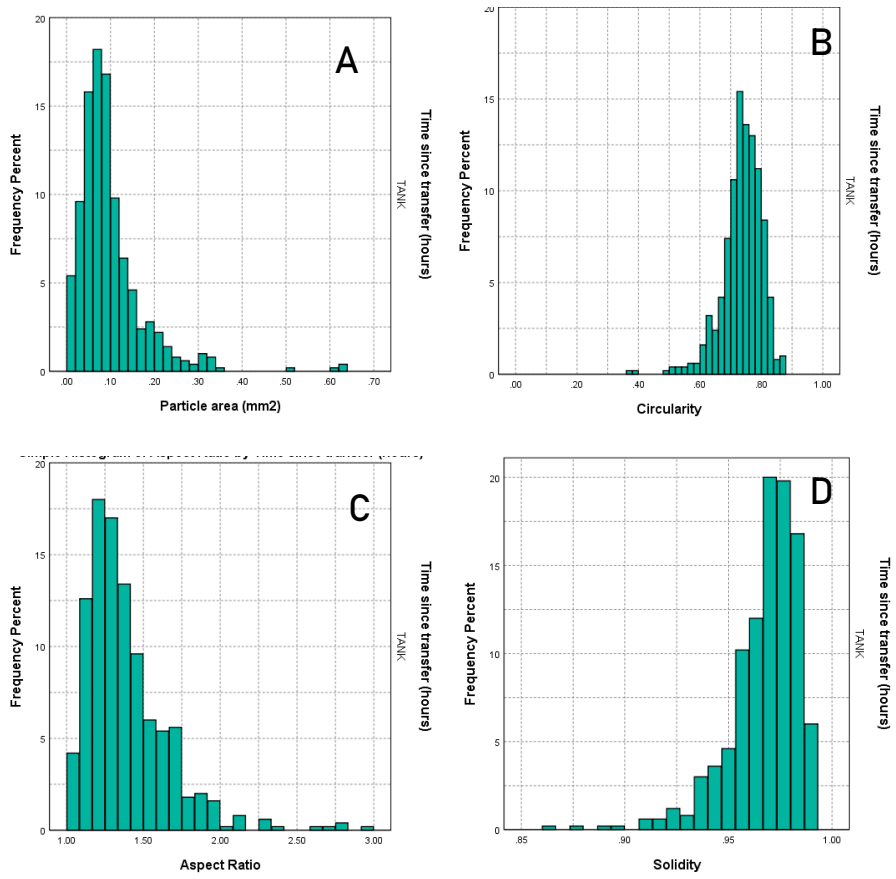


Figure 6.4: The characteristics of the sand grains used (A)cross-sectional area, (B) Circularity, (C) Aspect Ratio, and (D) solidity.

[10] https://www.rhsplants.co.uk/product/_/rhs-horticultural-silver-sand/classid.2000035343/

6.2.3 Materials: Swatches

Using the same method as chapters 3, 4, and 5, swatches of black cotton (2cm by 2cm) were assembled. Each swatch was composed of three layers: a lower soft fabric which could be handled without removing particles from the experiment, a middle layer of card, and an upper layer of the cotton under investigation. These swatches were attached to the sides of shoes with safety pins, in the 'inner' and 'outer' positions described in chapter 5. Each swatch was dampened slightly, with a drop of water, prior to the transfer. It is acknowledged that the cloth swatches used for the experiments may have contained mineral/moisture elements unknown during the process of the experiment set-up.

6.2.4 Methods: Quartz transfer

Quartz was transferred directly to each swatch, by pressing each slightly-damp swatch face-down into a container filled with the silver sand, and inverting to remove loose grains. It is acknowledged that this is not a realistic form of transfer for footwear, and resulted in much higher concentrations of quartz than might be expected in a forensically realistic scenario. This is understood, and assumed not to be problematic; the aim of this experiment was to investigate whether the size and shape of the quartz grains affected their transfer and retention, and no analysis or conclusions were made regarding the number of grains transferred. Five replicates were conducted for each time interval. Due to the slightly stochastic nature of the transfer, this meant that there was some variation in the number of quartz grains in the "0-hour" samples (i.e. the samples removed just after transfer). The N-values for each sample are outlined in Table 6.2.

6.2.5 Methods: Quartz extraction

'Tape lifting' i.e. removal with adhesive tape, is a common and accepted method for the collection of fibres (Lowrie and Jackson, 1991; De Wael et al., 2008) and has also been used for the collection of DNA (Barash et al., 2010; Verdon et al., 2014), gunshot residue (Goleb and Midkiff, 1975; Dockery and Goode, 2003), and traces of pharmaceuticals (Ricci et al., 2006, Rowell *et al.,* 2009) and other chemicals of interest (Brady et al., 2017, Benton et al., 2010). Tape lifting was the method used

here; clear 'Sellotape' (width 24mm) was used to remove the grains from the swatch's surface, and then adhere them to a standard glass slide (instead of using a coverslip). Tapings were taken immediately after the transfer (at "0 hours" after transfer) for five swatches, and then from an additional five swatches at 1, 4, 24 and 168 hours after the transfer. Five samples from the source were also taken by taping, each of 100 grains (see Table 6.1). Accordingly, in this experiment, multiple different samples at slightly different locations on the shoe were taken as a proxy for a single location over time. Previous experiments (see Chapter 5) suggested that this is a methodologically sound substitution, in that there did not appear to be any statistically significant difference between persistence patterns for samples taken from different locations on a shoe's upper for either UV powder or quartz grains (see section 5.3.2 and 5.3.4).

6.2.6 Methods: Image Acquisition

As discussed in the introduction, previous studies on the morphological analysis of images of quartz grains have shown that light microscopy produces images of sufficient magnification and resolution to facilitate successful analysis (Cox and Buddhu, 2008). Accordingly, here imaging was undertaken using a light microscope, the LEICA DM750, with an attached high definition microscope camera, the LEICA ICC50HD. The quartz grains were imaged at 100x magnification, generating images with a resolution of 1325 pixels per mm. Transects were conducted across the entirety of each slide, and all fields of view containing quartz grains were imaged. As with the preceding chapters, it is acknowledged that this method of data acquisition relied on trust in the black-box method of the software that generated the images, as is the convention in this field of experimental forensic science (French et al., 2012; Morgan et al., 2013).

Sample number	Time since transfer (hours)	Run	N	Total N for time interval	Total N across all conditions
1	0	A	320	1291	2410
2		B	186		
3		C	316		
4		D	142		
5		E	327		
6	1	A	30	227	
7		B	71		
8		C	44		
9		D	26		
10		E	56		
11	4	A	11	85	
12		B	21		
13		C	14		
14		D	31		
15		E	8		
16	24	A	29	204	
17		B	41		
18		C	41		
19		D	51		
20		E	42		
21	168	A	33	103	
22		B	25		
23		C	19		
24		D	17		
25		E	9		
26	SOURCE	A	100	500	
27		B	100		
28		C	100		
29		D	100		
30		E	100		

Table 6.2: The number of quartz grains imaged for each sample

6.2.7 Methods: Image Processing and Analysis

Two pieces of software were used; processing was conducted in Adobe Photoshop (CS4) and analysis in NIH ImageJ (please see Figure 6.5).

Processing

In Photoshop, the images were opened, and the quartz grains were cut out from each field of view, and a montage of the grains was created in one document for each sample (i.e. each swatch). For the source population samples, five samples of 100 grains were used (the first 100 counted in transect). Please consult table 6.2 for the number of quartz grains imaged in each sample. Once the montage had been assembled, segmentation was carried out in Photoshop. Firstly, the montage image layer was duplicated. The upper copy was then converted to black and white, with additional brightness on the yellow parts of the image (+300%, relative to the original brightness). This contrast-enhanced image was then thresholded with a global threshold filter, and manual adjustment of the threshold value (Figure 6.5). This resulted in a set of outlines which appeared to accurately represent most of the quartz grains. These outlines were amended by lowering the opacity of the thresholded layer (so that the original images were visible beneath), and using the brush and eraser tools to remove any excesses, close any holes in outlines, and add any omissions (see Figure 6.5).

Analysis

Images were then taken into NIH ImageJ for the analysis. The images were calibrated for scale (1325 pixels per mm), and converted to 8 bit, and thresholded within ImageJ (to select the darker quartz grains as the objects of interest, and the white pixels as representing background). Analysis was conducted using the 'analyse particles' function, enabling shape descriptors, and excluding any particles with a cross-sectional area of less than $0.0001mm^2$ (so that any stray pixels did not get counted as objects of interest).

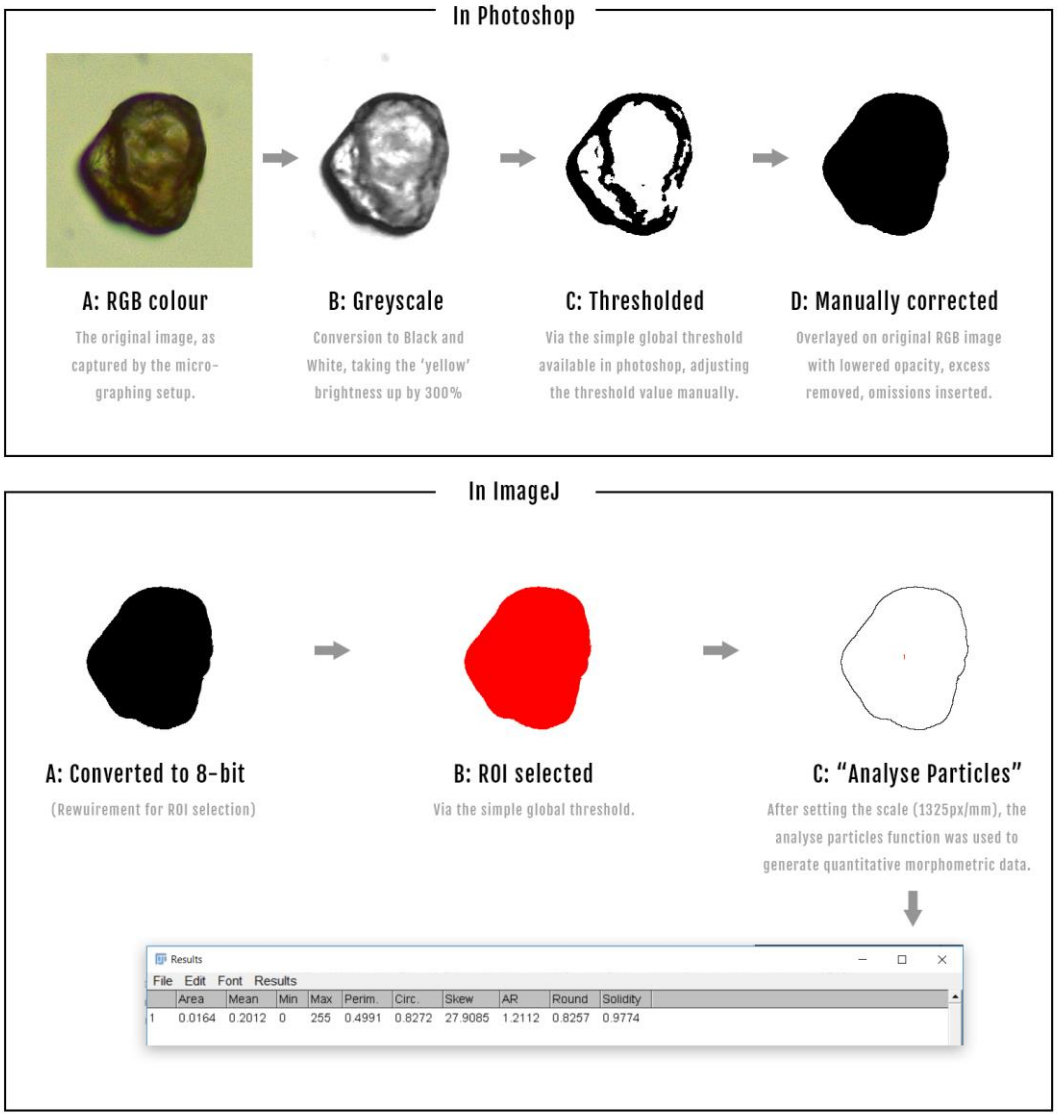


Figure 6.5: Diagram depicting the image processing and analysis workflow

6.2.8 Statistical Analysis

As in the preceding chapters (3, 4, and 5), the statistical analysis undertaken in this chapter consisted of both descriptive and inferential analysis. In addition to the mean and standard deviation (i.e. the measure of central tendency and measure of dispersion used in all preceding chapters), the minimum and maximum values for each morphometric variable (i.e. the particle area, circularity, aspect ratio, and solidity) were calculated and presented, as the maximum particle size has been mooted as a variable which may change throughout the course of a persistence study (Lowrie and Jackson, 1991; Levin *et al.,* 2017).

Guidance was sought as to what the most appropriate inferential statistical treatment of the data in this chapter would be. Since the samples were independent (i.e. no single particle was counted twice), the researcher was advised that the data presented in this chapter are not a true time series – since the purpose of time series statistical analysis is for repeated observations of the same units (i.e. the same particles), where controlling for autocorrelation is important. Accordingly, it was advised that the data collected in this chapter could be treated as independent samples with categories of different treatments – with the category relating to time elapsed. Accordingly, the researcher was advised that 'standard' inferential tests would be the most appropriate.

As with Chapter 5, the Kruskal Wallis test was used as a non-parametric inferential test of whether independent samples are drawn from the same distribution (see: McDonald, 2009). The data for this chapter fulfilled the assumptions of the test, since they were (a) independent measures (b) with a dependent variable measured at a scale level (better than ordinal level), and (c) divided into three or more groups (see Field and Hole, 2007: 275).

6. 3 Results

6.3.1 Checking for linear relationships/correlation

From the findings of previous studies (e.g. Brewster *et al.,* 1984; Hicks *et al.,* 1996), it was perceived as likely that the size of the quartz grains, as measured by the cross-sectional area, would be an influential variable. It was also perceived as plausible that some of the morphological characteristics might covary with particle size (as cross-sectional area). If this were the case, it would be possible that any other trends seen might be merely correlation with particle size, rather than representing an independent and meaningful relationship.

Accordingly, in order to check whether this was a possibility, scatter graphs were created, plotting the cross-sectional area of each quartz grain against circularity (Figure 6.6), aspect ratio (Figure 6.7) and solidity (Figure 6.8). None of the three shape descriptors appeared to covary with particle area; for the linear line of best fit between particle area and circularity the R^2 was 0.018 (N =2410), for the linear line of best fit between particle area and aspect ratio the R^2 was 0.005 (N =2410), and for the linear line of best fit between particle area and solidity, the R^2 was 0.065 (N =2410).

An overview of the entire dataset can be found in Table 6.3, overleaf, while summary boxplots of the four morphometric variables are presented in Figure 6.9.

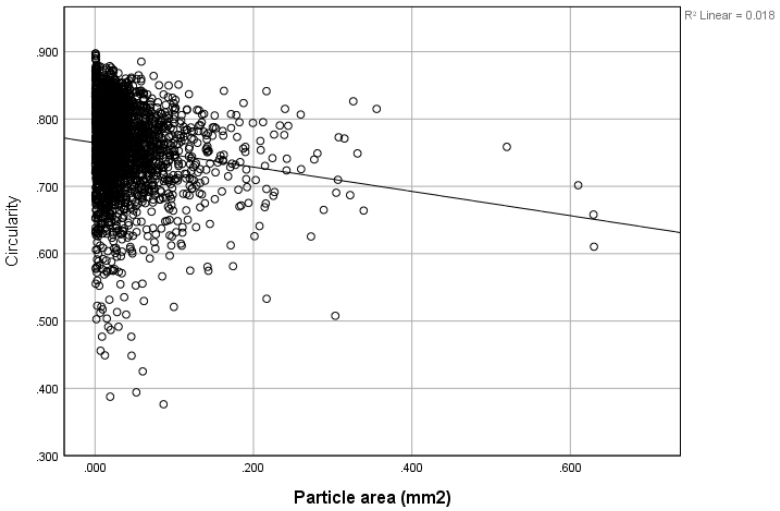


Figure 6.6: Scatter graph of particle area against circularity

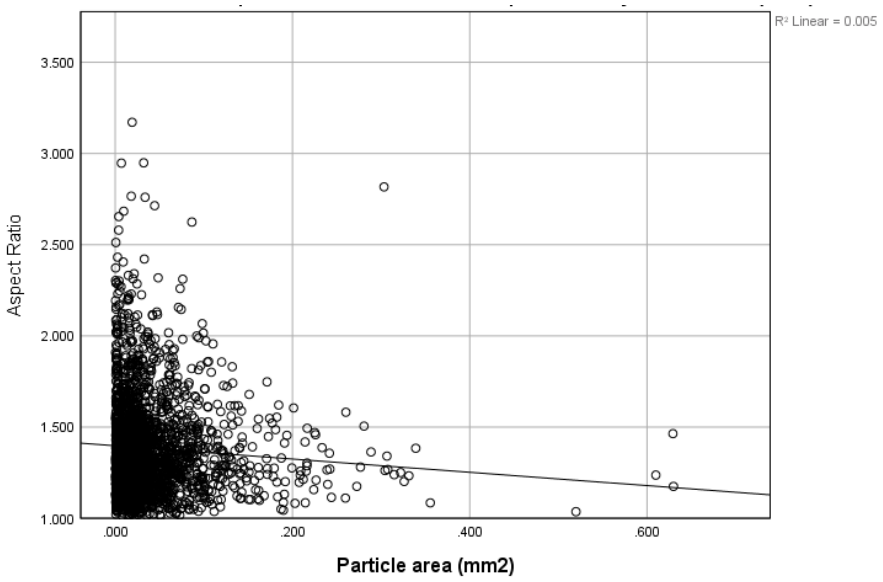


Figure 6.7: Scatter graph of particle area against aspect ratio

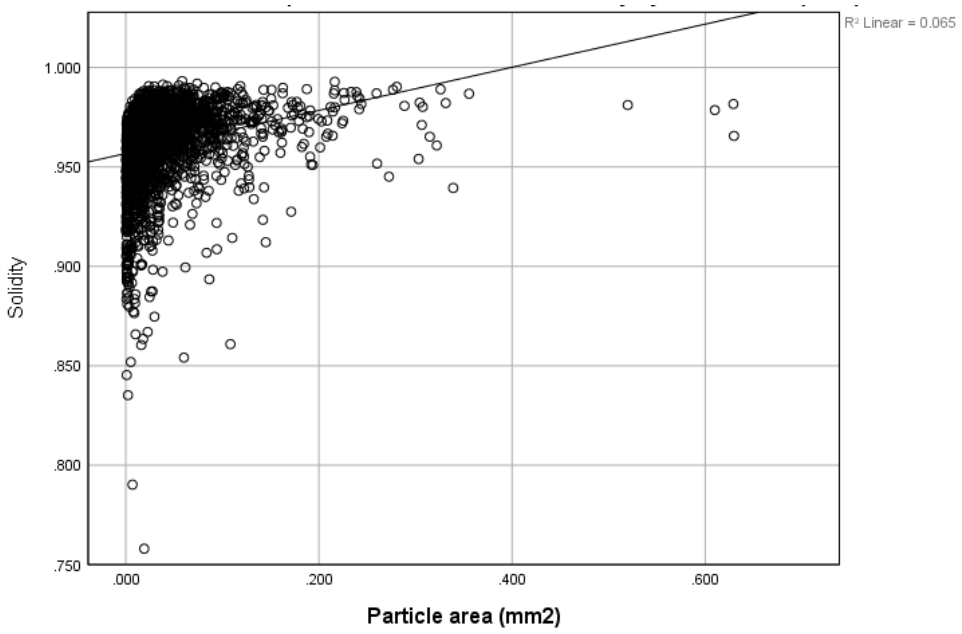


Figure 6.8 Scatter graph of particle area against solidity

Time since transfer (hours)		N	Particle area (mm2)				Circularity				Aspect Ratio				Solidity			
			M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
0	A	320	0.042	0.024	0.005	0.161	0.771	0.053	0.425	0.868	1.343	0.212	1.013	2.406	0.967	0.017	0.854	0.991
	B	186	0.033	0.029	0.001	0.174	0.743	0.085	0.456	0.873	1.408	0.312	1.015	2.947	0.961	0.024	0.790	0.990
	C	316	0.027	0.019	0.003	0.139	0.744	0.072	0.388	0.868	1.392	0.268	1.015	3.171	0.961	0.021	0.758	0.990
	D	142	0.035	0.025	0.006	0.212	0.767	0.055	0.605	0.885	1.362	0.232	1.018	2.022	0.964	0.017	0.892	0.993
	E	327	0.021	0.013	0.002	0.086	0.772	0.059	0.492	0.879	1.387	0.252	1.013	2.285	0.964	0.018	0.875	0.991
	Total	1291	0.031	0.023	0.001	0.212	0.760	0.066	0.388	0.885	1.378	0.255	1.013	3.171	0.964	0.020	0.758	0.993
1	A	30	0.025	0.016	0.003	0.073	0.759	0.062	0.583	0.851	1.322	0.186	1.012	1.825	0.955	0.023	0.901	0.987
	B	71	0.021	0.016	0.001	0.090	0.774	0.056	0.634	0.876	1.347	0.245	1.028	2.290	0.958	0.022	0.901	0.991
	C	44	0.029	0.017	0.004	0.071	0.779	0.041	0.690	0.863	1.307	0.215	1.037	2.104	0.965	0.016	0.913	0.987
	D	26	0.028	0.012	0.010	0.055	0.781	0.050	0.659	0.860	1.328	0.248	1.035	2.230	0.965	0.017	0.921	0.986
	E	56	0.021	0.015	0.002	0.080	0.773	0.054	0.637	0.860	1.353	0.207	1.013	1.892	0.963	0.017	0.903	0.989
	Total	227	0.024	0.016	0.001	0.090	0.774	0.053	0.583	0.876	1.335	0.222	1.012	2.290	0.961	0.019	0.901	0.991
4	A	11	0.031	0.024	0.004	0.075	0.757	0.075	0.629	0.846	1.394	0.237	1.111	1.985	0.955	0.030	0.880	0.982
	B	21	0.012	0.010	0.002	0.037	0.756	0.061	0.606	0.872	1.399	0.218	1.096	1.847	0.949	0.025	0.897	0.984
	C	14	0.010	0.012	0.001	0.039	0.765	0.073	0.629	0.860	1.384	0.315	1.031	2.102	0.945	0.027	0.890	0.991
	D	31	0.009	0.017	0.001	0.086	0.736	0.085	0.477	0.875	1.428	0.319	1.056	2.291	0.938	0.025	0.877	0.978
	E	8	0.006	0.008	0.001	0.025	0.725	0.070	0.577	0.821	1.355	0.179	1.157	1.700	0.931	0.027	0.885	0.965
	Total	85	0.012	0.017	0.001	0.086	0.747	0.074	0.477	0.875	1.402	0.270	1.031	2.291	0.944	0.026	0.877	0.991
24	A	29	0.002	0.001	0.001	0.005	0.791	0.062	0.674	0.880	1.319	0.283	1.036	2.173	0.946	0.017	0.897	0.972
	B	41	0.005	0.005	0.001	0.020	0.757	0.083	0.552	0.878	1.433	0.350	1.061	2.431	0.947	0.026	0.852	0.983
	C	41	0.002	0.002	0.000	0.010	0.746	0.094	0.517	0.869	1.558	0.384	1.044	2.683	0.939	0.028	0.882	0.973
	D	51	0.002	0.002	0.000	0.007	0.761	0.073	0.579	0.895	1.505	0.307	1.022	2.304	0.940	0.024	0.845	0.971
	E	42	0.027	0.020	0.001	0.076	0.754	0.053	0.600	0.851	1.360	0.215	1.059	2.131	0.958	0.017	0.910	0.985
	Total	204	0.008	0.014	0.000	0.076	0.760	0.075	0.517	0.895	1.445	0.323	1.022	2.683	0.946	0.024	0.845	0.985
168	A	33	0.004	0.004	0.001	0.019	0.773	0.065	0.600	0.889	1.375	0.352	1.044	2.654	0.944	0.019	0.895	0.973
	B	25	0.004	0.006	0.001	0.024	0.830	0.043	0.731	0.897	1.281	0.224	1.037	2.086	0.957	0.017	0.901	0.987
	C	19	0.002	0.001	0.001	0.005	0.773	0.089	0.572	0.875	1.469	0.375	1.058	2.580	0.940	0.034	0.835	0.974
	D	17	0.001	0.001	0.000	0.004	0.806	0.065	0.631	0.897	1.394	0.297	1.134	2.373	0.949	0.015	0.918	0.977
	E	9	0.002	0.001	0.001	0.004	0.736	0.061	0.653	0.819	1.477	0.305	1.208	2.194	0.924	0.028	0.881	0.961
	Total	103	0.003	0.004	0.000	0.024	0.789	0.070	0.572	0.897	1.382	0.318	1.037	2.654	0.945	0.024	0.835	0.987
SOURCE	A	100	0.092	0.073	0.002	0.326	0.734	0.082	0.376	0.879	1.431	0.346	1.025	2.949	0.965	0.022	0.877	0.993
	B	100	0.079	0.055	0.021	0.520	0.738	0.059	0.394	0.862	1.436	0.301	1.036	2.760	0.970	0.013	0.918	0.988
	C	100	0.106	0.093	0.005	0.630	0.734	0.068	0.508	0.839	1.350	0.239	1.035	2.817	0.968	0.016	0.909	0.989
	D	100	0.108	0.083	0.002	0.610	0.746	0.054	0.608	0.844	1.338	0.212	1.026	2.311	0.966	0.019	0.861	0.991
	E	100	0.104	0.075	0.019	0.629	0.731	0.059	0.553	0.864	1.397	0.251	1.048	2.319	0.968	0.015	0.912	0.992
	Total	500	0.098	0.077	0.002	0.630	0.737	0.065	0.376	0.879	1.391	0.276	1.025	2.949	0.967	0.017	0.861	0.993

Table 6.3 Overview of the dataset (descriptive statistics)

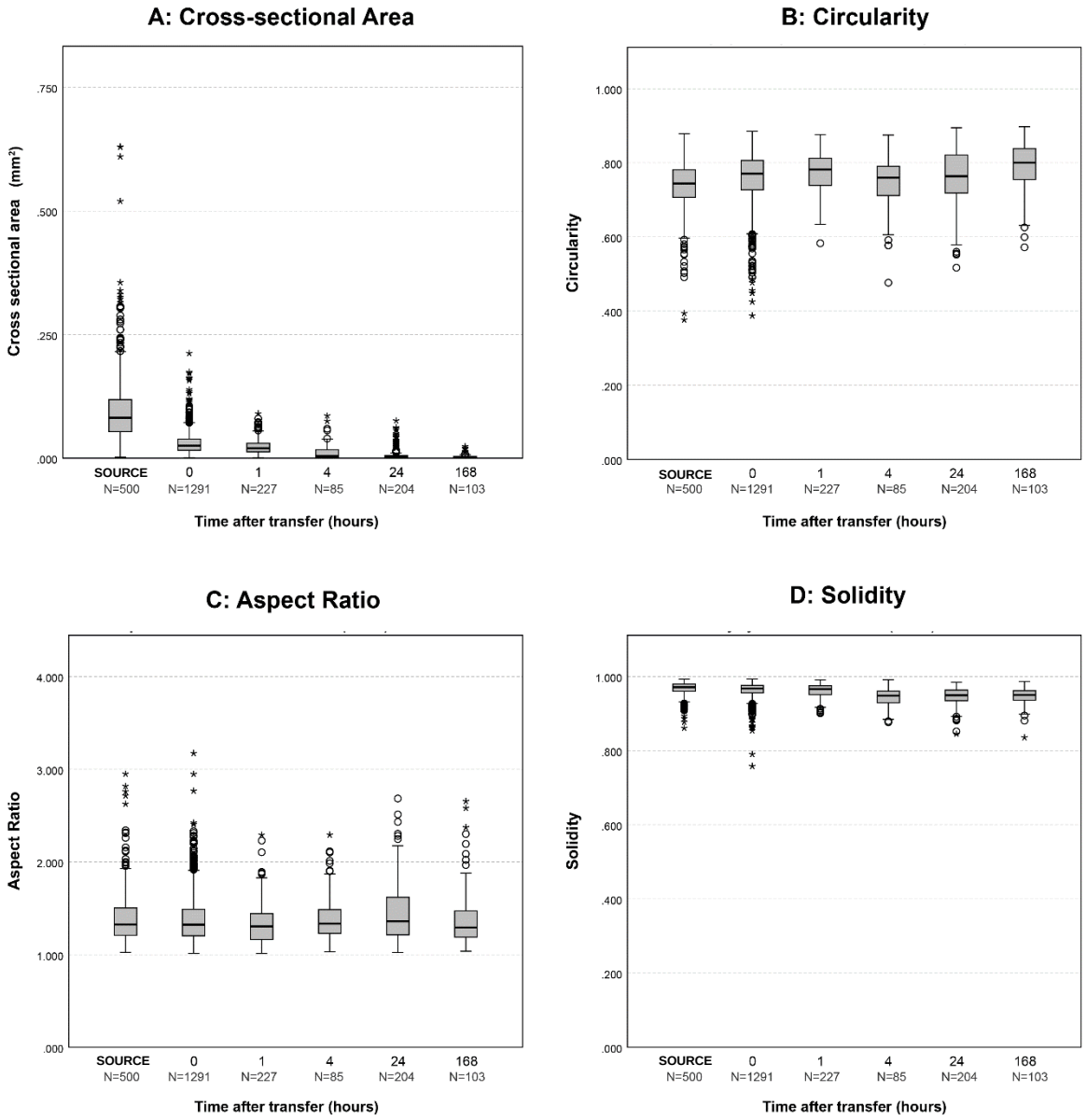


Figure 6.9 Summary boxplots of the four morphometric measurements. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

6.3.2 Grain size and transfer

With respect to particle size, boxplots of the cross-sectional area distributions for the 5 samples of the source population, and the 5 samples collected immediately after transfer (at 0 hours) can be seen in Figure 6.10. Descriptive statistics are provided in Table 6.3, and a boxplot of the aggregated runs can be seen in Figure 6.9. Three observations can be made:

(1) The maxima for cross-sectional area seen in the source population were all larger than the maxima for cross-sectional area observed in the transferred samples (Figure 6.10). Every sample from the source population contained quartz grains with areas in excess of 0.30 mm^2, and no 0-hour sample contained any grains with an area greater than 0.25mm^2 (Table 6.3).

(2) The mean and median values for cross-sectional area were lower for the for the transferred samples than for the source population (Table 6.3; Figure 6.10); aggregating across all samples, the mean particle area for the source population is 0.098 $\pm$ 0.077 mm^2 (n=500), and the mean particle area for the 0-hour sample is 0.031$\pm$ 0.023 mm^2 (n=1291) (Table 6.3).

(3) Conducting statistical analysis suggests that these trends may be suggestive of differences between the population of quartz grains used in the experiment ('source') and the population of quartz grains collected immediately after transfer ('0 hours'), and may be worth further investigation; comparing the distributions with the Kruskal-Wallis test (which considers the distribution of values) p values of <0.05 were obtained. (See Table 6.4).

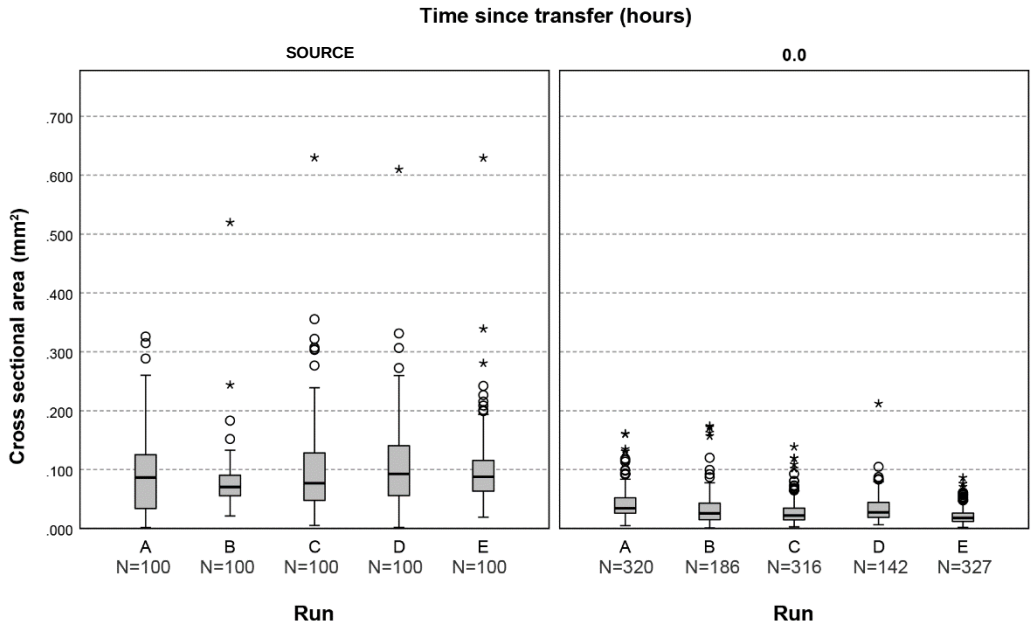


Figure 6.10: Boxplot of particle areas, for the source samples and samples immediately after transfer (0 hours). Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

Null hypothesis	Test	Total N	Test statistic	DF	Significance	Decision
The distribution of particle area (mm2) are the same across the categories of 'source' and '0 hours'	Independent Samples Kruskal Wallis Test	1791	444.555	1	<0.05	Reject the null hypothesis

Table 6.4: Summary of inferential statistics comparing the particle areas of the source population and samples removed immediately after transfer ('0 hours')

6.3.3 Grain size and persistence

Revisiting the data to explore the influence of grain size upon persistence, Figure 6.11 displays the cross-sectional area data generated for the samples over time, while descriptive statistics can be found in Table 6.3, (and summarised in Table 6.5 below). As the time since transfer increased, the mean and maximum cross-sectional area decreased, from a maximum of 0.212mm^2 observed at 0 hours, down to a maximum of 0.024mm^2 observed at 168 hours, and a mean of 0.031 $\pm$ 0.023 mm^2 at 0 hours, to a mean of 0.003 $\pm$ 0.004 mm^2 after 168 hours.

Time since transfer (hours)	N (aggregated across the five replicates)	Cross-sectional area (mm^2)			
		Mean	SD	Min	Max
Source	500	0.098	0.077	0.002	0.63
0	1291	0.031	0.023	0.001	0.212
1	227	0.024	0.016	0.001	0.09
4	85	0.012	0.017	0.001	0.086
24	204	0.008	0.014	0.000	0.076
168	103	0.003	0.004	0.000	0.024

Table 6.5: Cross-sectional area over time, aggregated by time interval

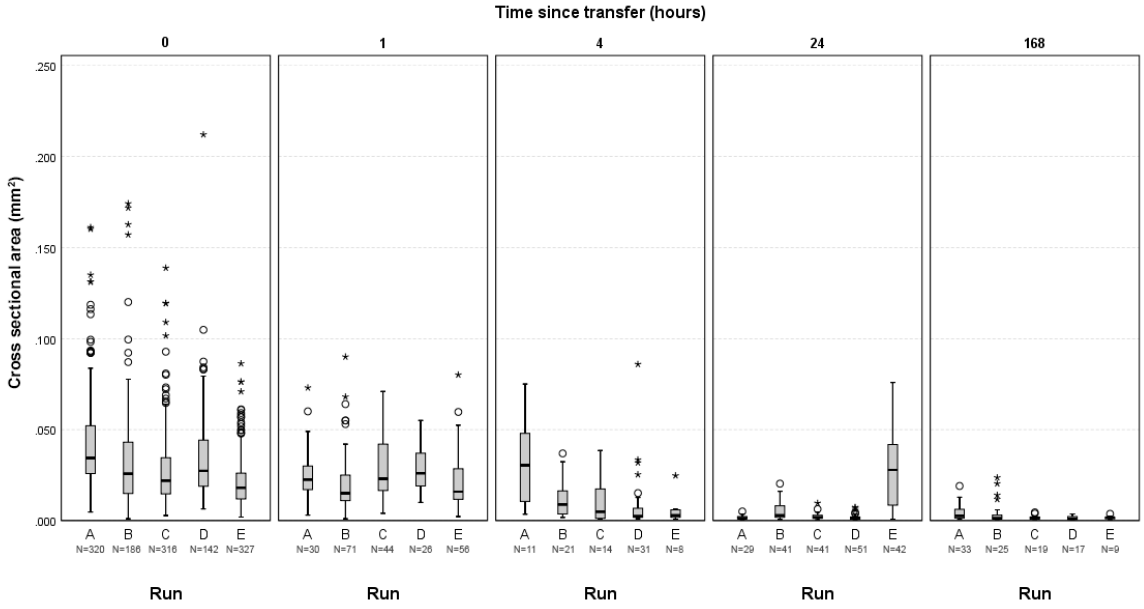


Figure 6.11: Boxplot of cross-sectional area over time (with each replicate plotted individually). Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

When the Kruskal Wallis test was conducted p values of <0.05 were obtained for all of the morphometric variables, suggesting that there may be differences which are worth further investigation (Table 6.6).

Hypothesis test summary: comparing the morphological metrics over time				
Variable	Null hypothesis	Test	Sig	Decision
Cross sectional area (mm^2)	The distribution of "cross sectional area" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
Circularity	The distribution of "circularity" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	.000	Reject the null hypothesis
Aspect Ratio	The distribution of "aspect ratio" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	.016	Reject the null hypothesis
Solidity	The distribution of "solidity" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	.000	Reject the null hypothesis

Table 6.6: Inferential statistical summary comparing the medians and distributions of the morphological metrics between the different time intervals

6.3.4 Grain morphology and transfer

Data for the particle shape descriptors (aspect ratio, circularity, and solidity) are presented in Table 6.3, summarised in Table 6.7, and visualised in figures 6.12, 6.13, and 6.14. Visually inspecting the data, there does not appear to be large variation between the mean values for the shape descriptors seen in the source samples and the zero-hour samples: the mean circularity observed in the source was 0.737 ± 0.065, the mean for the 0 hour samples was 0.760 ± 0.066; the mean aspect ratio observed in the source was 1.391 ± 0.276, while the mean for the 0 hour samples was 1.378 ± 0.255; and the mean solidity for the source samples was 0.967 ± 0.017 and the mean for the 0 hour samples was 0.964 ± 0.020.

Running inferential statistical analysis, in the form of the Kruskal Wallis test (to compare distributions) confirms that there is unlikely to be a relationship between the aspect ratio of a quartz grain and the likelihood of transfer, with a p value of 0.568 obtained. (Table 6.8). The analysis does, however, suggest that there may be differences between the circularity and solidity of the source and transferred populations (Table 6.8).

Time since transfer (hours)	N (aggregated across the five replicates)	Circularity				Aspect Ratio				Solidity			
		Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Source	500	0.737	0.065	0.376	0.879	1.391	0.276	1.025	2.949	0.967	0.017	0.861	0.993
0	1291	0.76	0.066	0.388	0.885	1.378	0.255	1.013	3.171	0.964	0.02	0.758	0.993
1	227	0.774	0.053	0.583	0.876	1.335	0.222	1.012	2.29	0.961	0.019	0.901	0.991
4	85	0.747	0.074	0.477	0.875	1.402	0.27	1.031	2.291	0.944	0.026	0.877	0.991
24	204	0.76	0.075	0.517	0.895	1.445	0.323	1.022	2.683	0.946	0.024	0.845	0.985
168	103	0.789	0.07	0.572	0.897	1.382	0.318	1.037	2.654	0.945	0.024	0.835	0.987

Table 6.7: Shape descriptors summary of descriptive statistics

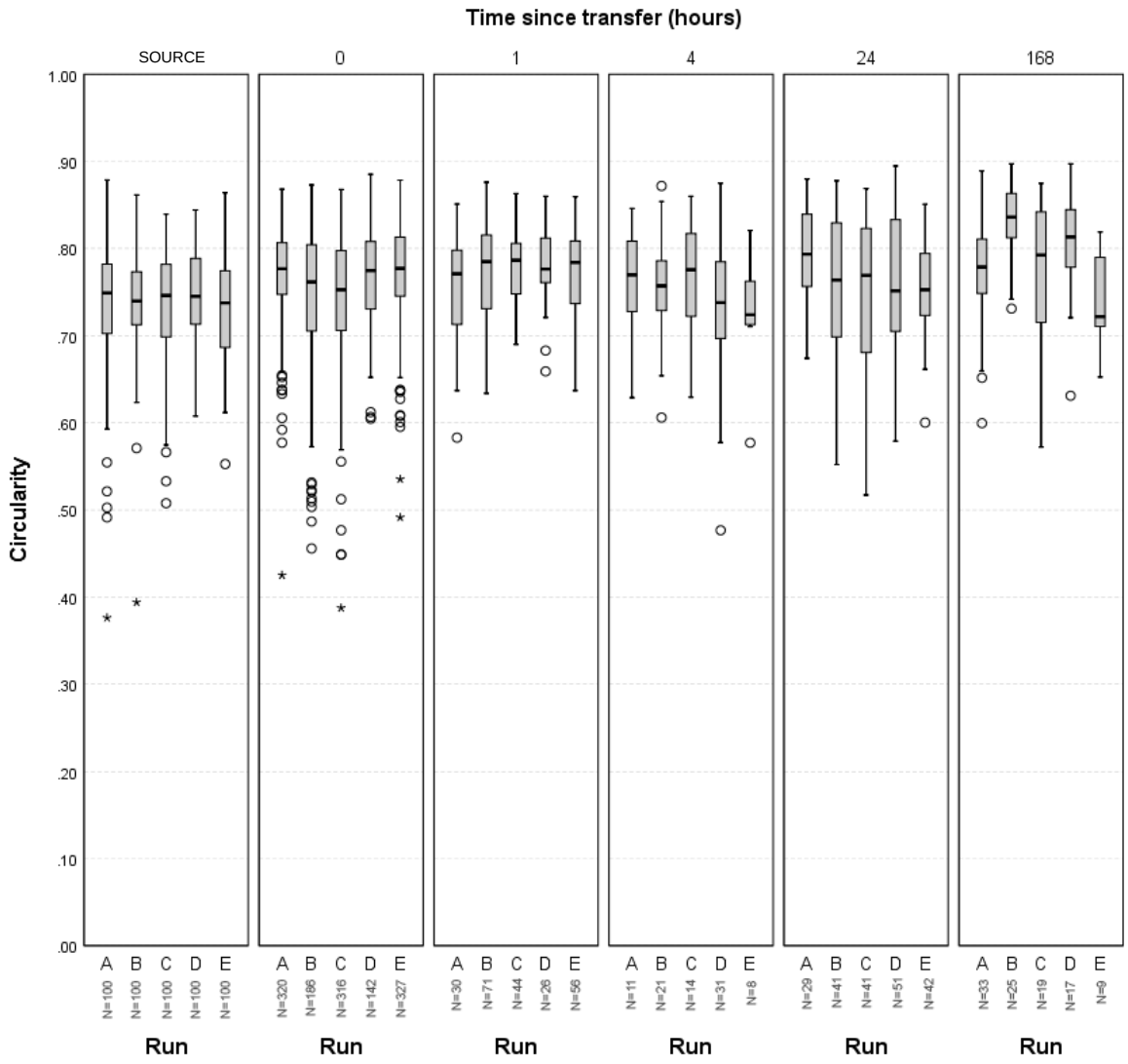


Figure 6.12: Boxplots of Particle Circularity by time interval. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

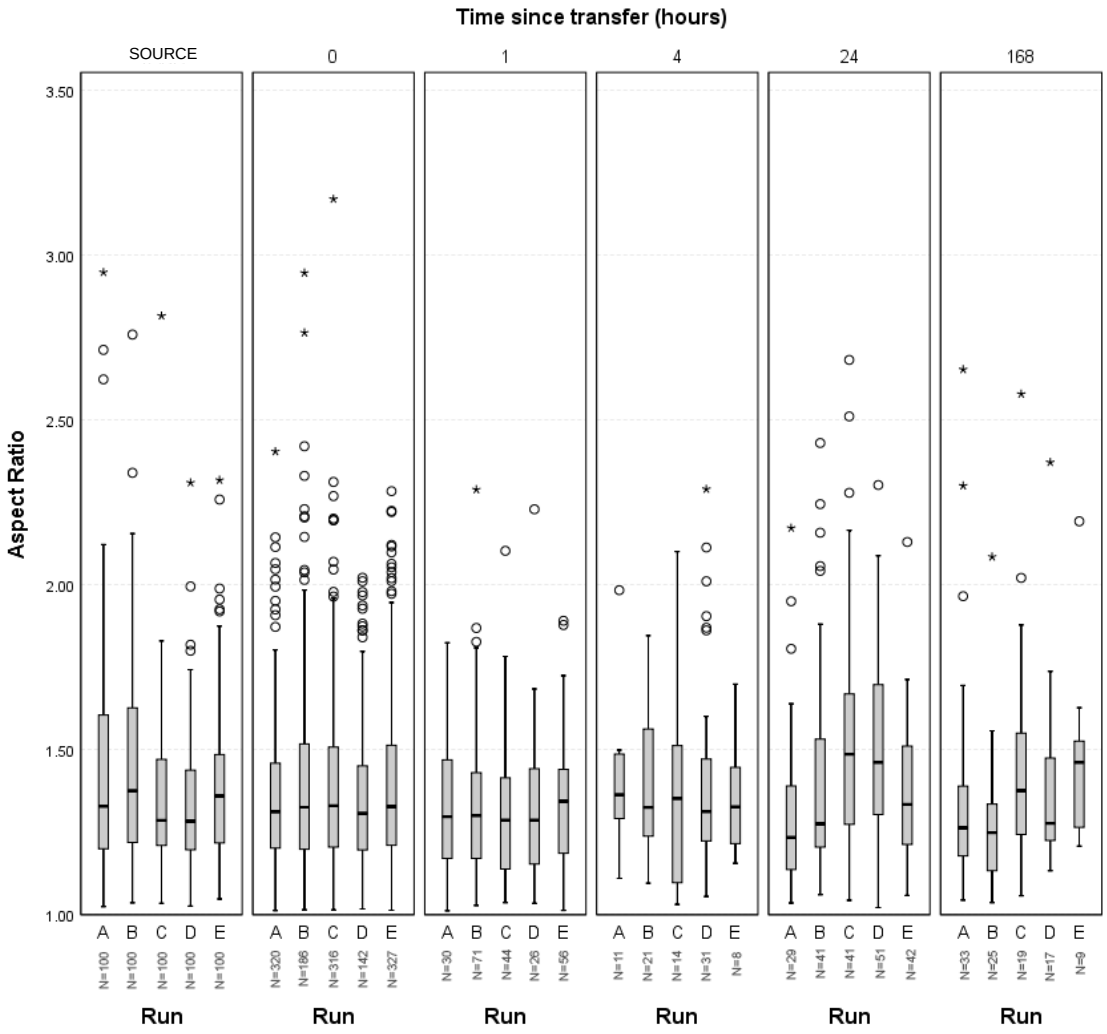


Figure *6.13: Boxplots of Particle Aspect Ratio by time interval. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).*

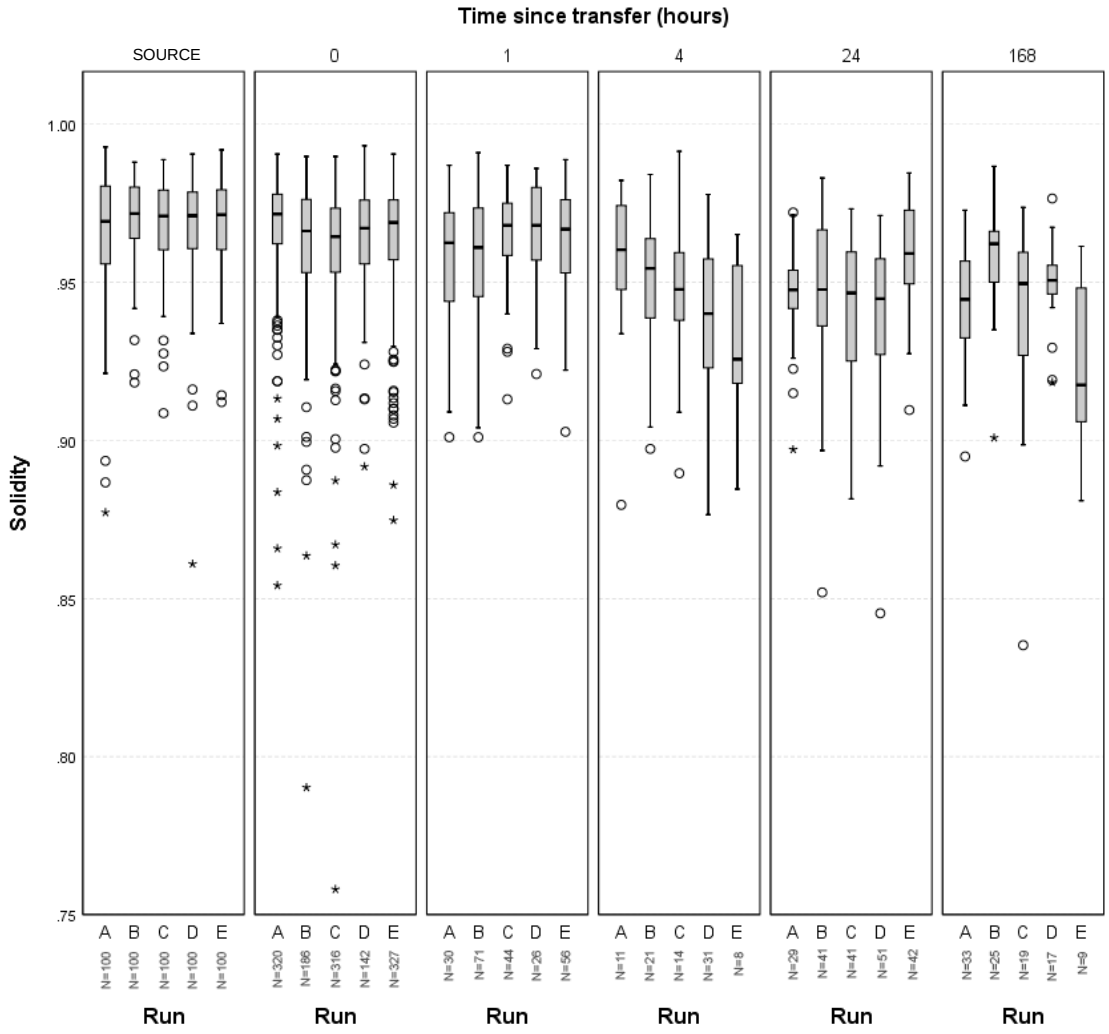


Figure 6.14: Boxplots of particle Solidity by time interval. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

Shape descriptor	Null hypothesis	Test	Total N	Test statistic	DF	Significance	Decision
Circularity	The distribution of circularity values are the same across the categories of 'source' and '0 hours'	Independent Samples Kruskal Wallis Test	1791	63.346	1	p < 0.05	Reject the null hypothesis
Aspect ratio	The distribution of aspect ratio values are the same across the categories of 'source' and '0 hours'	Independent Samples Kruskal Wallis Test	1791	0.325	1	p = 0.568	Retain the null
Solidity	The distribution of solidity values are the same across the categories of 'source' and '0 hours'	Independent Samples Kruskal Wallis Test	1791	19.902	1	p < 0.05	Reject the null hypothesis

Table 6.8 Summary of inferential statistics comparing the source and 0 hours samples

6.3.5 Grain morphology and persistence

The data regarding the shape descriptors for the particles over time are presented in Table 6.3, Table 6.7, and Figures 6.12, 6.13, and 6.14.

Circularity

Most of the quartz grains had relatively high circularity values, with the means for all samples ranging between 0.725 (observed for 4 hours, run D) and 0.830 (observed at 168 hours, run A) (Table 6.3). The maxima observed ranged from 0.819 (168, run E) to 0.897 (observed for 168 run B). The minimum circularity values observed appeared to increase over time; the samples collected immediately after transfer (at "0 hours") contained some grains with very low values for circularity (with a minimum of 0.388 (n = 1291)), while the minimum circularity value seen for the later samples was higher (Table 6.3; Table 6.9; Figure 6.12). After 1 hour, the

minimum was 0.583 (n=227), after 4 hours 0.477 (n=85), after 24 hours was 0.517 (n = 204) and after 168 hours 0.572 (n=103).

The findings of inferential statistical analysis, in the form of the Kruskal Wallis Test (Table 6.6), suggested that differences existed between the distributions of the circularity values over time (Table 6.8). Accordingly, this trend may be worth further investigation.

Aspect Ratio

The range of aspect ratios observed was 1.012 – 3.171 (Table 6.3). All samples produced similar median values for aspect ratio (between 1.25 and 1.50; Figure 6.13). There was not a systematic trend of the minimum or maximum aspect ratio values increasing or decreasing over time (Table 6.7).
Inferential statistical analysis was conducted, in the form of the Kruskal Wallis Test (Table 6.6) , which suggested that this variable may be worth further investigation ($p<0.05$).

Solidity

It appeared that the maximum and mean solidity values declined over time (Table 6.3, Figure 6.14). Inferential statistical analysis was conducted, in the form of the Kruskal Wallis Test (Table 6.6), producing a p-value of <0.05 and suggesting that this variable may be worth further investigation.

6. 4 Discussion

6.4.1 Particle size and transfer

In these experiments the mean and maximum particle areas observed were greater in the source population than on the swatch immediately after transfer, and statistical analysis suggested that this trend may be worth further investigation. These findings suggested that the largest quartz grains in the sample (i.e. those with the greatest cross-sectional area) did not transfer from the population to the swatch in this experiment. This finding is in accordance with studies that have examined other forms of trace evidence, such as fibres, glass fragments, and diatoms (Pounds and Smalldon, 1975b, Brewster *et al.,* 1984, Hicks *et al.,* 1996, Levin *et al.*, 2017). Together, all of these studies suggest that larger particles may not transfer readily. The implications of this finding relate to the interpretation of quartz particle size distributions when comparing samples derived from a source population and from an article of evidence. If a sample derived from an article of evidence is missing a larger size fraction that is present in the source population, it is possible that the larger size fraction did not transfer.

6.4.2 Grain size and persistence

The mean and maximum particle size observed declined over time (Table 6.3), and this variation produced p values for the Kruskal Wallis test, of <0.05 (Table 6.6), and may be worth further investigation. Again, this finding is consistent with the literature for glass fragments, fabric fibres, and diatom valves (Brewster *et al.,* 1984, Hicks *et al.,* 1996, Pounds and Smalldon, 1975b, Levin *et al.*, 2017), and again, it suggests the need for caution and awareness when interpreting particle size data within forensic casework. To build upon the caveat outlined above, in section 6.4.1, it suggests that, when comparing samples derived from a source population and from an article of evidence, nuanced interpretation is required. If a sample derived from an article of evidence is missing a larger size fraction that is present in the source population, it is possible that the larger size fraction failed to persist – i.e. it is possible either that it did not transfer, or that it did transfer but was subsequently lost.

6.4.3 Grain morphology and transfer

No consistent trends were seen when comparing the aspect ratios of the particles in samples from the source to the samples from "0 hours" (immediately after transfer) (Table 6.10), and statistical analysis confirmed that there was no relationship worth further investigation (Table 6.8). However, differences were observed for the variables of circularity and solidity when comparing the source and 0 hour samples, with the transferred samples exhibiting lower solidity (i.e. a higher degree of surface roughness), and a higher circularity (i.e. a lower level of elongation) (Table 6.8). This suggests that it may be possible that particles with a higher degree of roughness are more likely to transfer from the population to an article of evidence, and that particles with a higher degree of circularity are more likely to transfer. Further research is necessary to explore this hypothesis in greater detail. If it is demonstrated that morphological characteristics mean that certain morphologies transfer preferentially from a source population, this may have implications for the interpretation of such indicators.

6.4.4. Grain morphology and persistence

The minimum circularity values observed increased over time; the differences between the samples collected after different times was seen to produce p-values for the Kruskal Wallis test of <0.05, suggesting that these differences may be worth further investigation (Table 6.6). There was not a systematic trend of the minimum or maximum aspect ratio values increasing or decreasing over time (Table 6.7). The maximum and mean solidity values declined over time (Table 6.7, Figure 6.14, which inferential statistical analysis suggests may be worth further investigation (Table 6.6)). As with the preceding sections, this suggests that it is necessary to bear this in mind when interpreting data; if grains are lost selectively according to solidity and circularity, a paucity of smooth and elongated grains may be due to their rapid loss, rather than initial absence.

6.4.5. Limitations and future work

It should be noted, as with all experiments, this work has a number of limitations that should be borne in mind. Primarily, there are two *fundamental* limitations: (1) In these experiments, the quartz grains were imaged under the light microscope, and

the metric of magnitude used was the cross-sectional area. Accordingly, although this experiment did not assume that all of the particles are spherical (as some other measurement techniques would), it is guilty of a similar abstraction – it utilised a two-dimensional representation of a three-dimensional object. Accordingly, this work assumed that measuring one cross-section of the quartz grains was an acceptable measure of size, when it is entirely possible that one of the grain's axes may have been more elongated, angular, or larger than the other. (2) Secondly, and relatedly, this experiment sacrificed realism for control. The method of quartz transfer was not highly realistic - in that contact was direct and under light pressure, and that the quartz was isolated, not contained within a soil, with silt and clay fractions. Accordingly, it is possible to suggest that further, more realistic experiments, might be desirable. Another limitation, or complication, was that it was possible that some of the grains examined in the course of this experiment did not represent the grains that had been deliberately introduced to the fabric, but additional quartz entrained from the environment before or after the initial transfer. As outlined in Morgan and Bull, 2007, material found on a surface can be derived from pre-post-or-syn the forensic event – a significant caveat that must be understood and borne in mind during interpretation (Morgan and Bull, 2007). Future work could consider pursuing the hypotheses pursued in this chapter in further depth, looking at different initial samples of quartz, e.g. some sands with many angular grains, some grains of different sizes, and with respect to other evidence types. As ever, there is a need for empirical research within forensic science, and cognate disciplines which are applied to casework – the data which will allow for nuanced, meaningful interpretation (and the addressing of propositions at the activity level, rather than the source level) are necessary.

6.5 Conclusions

6.5 Conclusions

In conclusion, two inferences were drawn from the datasets:

1. The largest particles in the population don't seem to transfer, and the largest particles that do transfer are lost first. Accordingly, this suggests a need for awareness within interpretation: if a sample derived from an article of evidence is missing a larger size fraction that is present in the source population, it is possible that the larger size fraction either did not transfer, or did transfer and was subsequently lost, rather than signalling a genuine difference between the sample collected from the source population and the sample extracted from fabric.
2. Comparing a source population and samples collected just after transfer, the transferred samples appeared to exhibit lower solidity (i.e. a higher degree of surface roughness), and a higher circularity (i.e. a lower level of elongation), and over time, the mean and maximum solidity appeared to decrease and the minimum circularity appeared to increase. This suggests that future research into these variables may be worthwhile; it would be useful for forensic scientists working to interpret quartz grains as evidence to know whether grains with a higher degree of roughness are more likely to transfer from the population to an article of evidence, and whether particles with a higher degree of circularity are also more likely to transfer. It would also be useful to establish whether rougher grains persist over longer timescales, and grains with lower circularity persist over shorter timescales. Again, if this were established, this would suggest the need for nuanced interpretation when comparing samples derived from clothing and footwear, and samples extracted from the environment.

Together, these findings suggest, again, that nuanced interpretation is vital when analysing environmental indicators within the context of forensic casework. They also reiterate the importance of empirical experimentation, and highlight the potential of semi-automated analysis and image analysis for collecting forensically relevant datasets, with speed.

7

Chapter Seven

Does the morphology of diatom valves affect their transfer and/or persistence?

Abstract

Diatoms, a type of unicellular algae, are valuable trace evidence when attempting the reconstruction of crimes committed in aquatic environments. While their ecology and population dynamics are the subject of extensive and ongoing research, relatively few researchers have considered their behaviour as trace evidence. The characteristics of diatoms may influence their transfer to clothing fabrics (Scott *et al.,* 2019), raising the question of whether the morphology of a diatom valve might affect its transfer and persistence (Scott *et al.,* 2019). This chapter sought to investigate this hypothesis, using novel methods of semi-automated diatom analysis, and asking whether the diatom valve's (1) size (measured by Cross Sectional Area), (2) elongation (measured by Width-Height Ratio) and (3) roughness (measured by contour smoothness) appeared to affect the likelihood of transfer and/or persistence (as was seen in Chapter 6, in the context of quartz grains). The work in this chapter was aided by Michael Kloster, who assisted with the use of automated morphometric software for the morphometric analysis of diatoms, SHERPA (Kloster *et al.,* 2014; 2017). Three findings were established: In this study, (1) All size fractions of diatom found in the population were seen to transfer readily (2) The morphology (i.e. aspect ratio and convexity versus concavity) of the diatom particles seemed to affect their transfer; (3) Differences were observed between the mean values for the morphological metrics over time, which inferential analysis suggested may be a genuine trend worth further investigation, and that the size and shape of a particle may indeed affect their likelihood of persisting. Together, these findings reinforce the idea that morphology appears to be a factor in understanding the transfer and persistence of particulate evidence.

7. 1 Introduction

7.1.1 Diatoms as an environmental indicator

As established in Chapter 2, diatoms are unicellular algae which are widely-distributed in aquatic environments, and possess a durable, silica external shell (or 'frustule') (e.g. Cameron, 2004; Jones, 2007). Individual diatom frustules are not visible to the naked eye; cells tend to range between 20 and 200 µm in diameter (Kelly, 2000; Cox, 2012). Accordingly, analysis tends to take place under the microscope, using either phase contrast microscopy, which magnifies several hundreds of times, or scanning electron microscopy, which magnifies the image thousands of times (Jones, 2007).

Each species of diatom has a visually distinctive cell (or frustule), with a number of different shapes and ornamentation features which mean that a trained examiner can identify a frustule to the genus or species level under the microscope (Jones, 2007). Each species of diatom also has a set of environmental tolerances, existing in specific combinations of water nutrients, pH, and salinity (Cameron, 2004; Gayle, 2004; Horton et al., 2006; Jones, 2007), as well as possessing seasonal trends in abundance (Smol and Stoermer, 2010; Scott et al., 2019). Accordingly, the presence or absence of particular genera or species of diatom within a water sample can allow an examiner to draw inferences about the water body from which the sample was extracted (Horton *et al.,* 2006; Korhola, 2007; Cox, 2012). Consequently, it is possible to use diatoms as an environmental indicator, analysing the species assemblages found upon items of clothing or footwear and at specific sites to attempt to dissociate or fail to dissociate suspects, victims, and crime scenes (e.g. Peabody, 1999; Cameron, 2004; Cameron and Peabody, 2010; Cox, 2012; Scott *et al.,* 2014).

7.1.2 Diatom morphology

In order to ensure that this chapter is accessible to the non-diatomist, there will now follow a brief introduction to diatom morphology. As mentioned in Chapter 2, and above, each diatom cell possesses a robust, porous, external layer of silica, known as a 'frustule' (e.g. Jones, 2007; Smol and Stoermer, 2010). Each frustule is composed of two halves, known as 'valves', which fit together like two halves of a 'pillbox' (see: Barber and Haworth, 1994, Figure 7.1).

The silhouette of the valve can have a number of shapes – for example, it can be very long and thin 'lanceolate', or it can be approximate a circle ('centric') (see: Barber and Haworth, 1994). The silhouettes can have varying levels of complexity, and while they can have rotational symmetry in one or more axes, many valves are asymmetrical in one or more axes (ibid.)

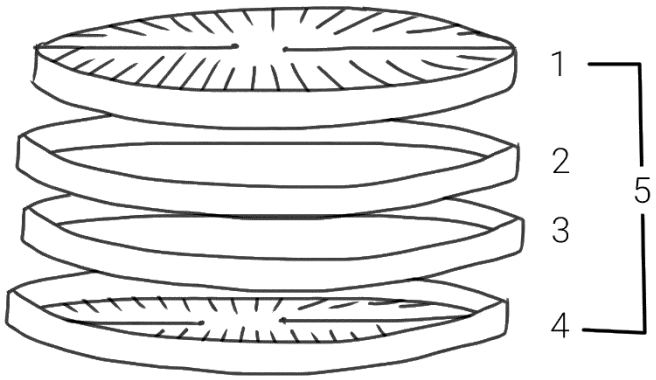


Figure 7.1: Exploded isometric diagram of a diatom valve, reproduced from Barber and Haworth, 1994: p 19. (1) Represents the upper valve (or 'epivalve'), (2) and (3) are girdle bands, which attach to the upper and lower valves. (4) Represents the lower valve, or 'hypovalve'. (5) Represents 'a complete frustule'.

As biological entities, which were once alive, it is important to note that diatoms do not always exist as single cells; some species live as colonies, in strings, or chains (Barber and Haworth, 1994). When one prepares slides of diatoms for analysis, it is possible that the valves remain in a chain form, as is common with the species *Melosira varians* (ibid.) Similarly, it is possible that fragments of a valve can be seen on a slide; although the silica frustule is durable, it can break. Accordingly, when one is surveying diatoms on a slide, the form can be either whole (intact and individual), in a chain (contiguous with other valves of the same species), or fragmented (broken, but identifiable). Another fact which is quite important is that since diatoms are three-dimensional objects, which tend to be mounted in a

permanent slide and imaged in two dimensions, it is possible that a diatom can be oriented towards the microscope either with the face of a valve showing, or in the profile, 'girdle' view (Barber and Haworth, 1994).

7.1.3. Size selective trends within trace evidence dynamics

As discussed in Chapters 2 and 6, it has been suggested that the size and texture of a particle can influence its transfer and persistence behaviours (Houck, 2001). Experimental studies have demonstrated that larger particles or fibres seem to persist over shorter timescales, with experiments exploring this hypothesis in the context of glass fragments (Brewster *et al.,* 1984; Hicks *et al.,* 1996), and clothing fibres (Pounds and Smalldon, 1975b). In Chapter 6, it was demonstrated that quartz grains appear to demonstrate the same trends, with the mean and maximum cross-sectional area observed decreasing over time (Chapter 6, see table 6.5), from a maximum quartz grain cross-sectional area of 0.212mm^2 observed immediately after the transfer, down to a maximum of 0.024mm^2 observed at 168 hours after the transfer, and a mean cross-sectional area of 0.031 ± 0.023 mm^2 observed immediately after the transfer, to a mean of 0.003 ± 0.004 mm^2 after 168 hours had elapsed.

With respect to diatom valves, preliminary work has suggested that similar size-selective persistence may be in operation (Levin et al., 2017), where it was observed that the mean and maximum particle size recovered declined over time, with the larger size fractions of diatoms that were observed in the earlier samples being absent from samples collected after a greater number of hours had elapsed since the transfer. It has also been suggested that, for freshwater diatoms transferring to fabrics, experiments "provisionally suggest that general characteristics including size and shape [...] may support or limit evidential transfer to clothing" (Scott et al., 2019: 304). The aim of this chapter is to test these hypotheses; exploring whether diatom morphology affects their transfer and/or persistence.

7.1.4. Aims

Specifically, this study sought to address five questions (in a complementary approach to the study presented in Chapter 6, addressing quartz grains):

1. Do different size fractions of diatom valves transfer preferentially? i.e. Comparing the source population and transferred population, are different size fractions represented proportionally, or do some sizes of diatom valves transfer more or less readily from the source population? Are there any size fractions represented in the source population which do not transfer to the fabric?

2. Do different size fractions of diatom valves persist equally over time? i.e. Comparing the initially transferred population and populations sampled after time has elapsed, are different size fractions represented proportionally, or do some sizes of diatom valves appear to have been lost more or less readily from the transferred population? Does the mean/median diatom valve size recovered change over time?

3. Do diatom valves with different morphologies transfer preferentially? i.e. Comparing the source population and transferred population, are valves with different roundness, aspect ratio, and solidity represented proportionally, or do some morphologies of particle appear to transfer more or less readily from the source population?

4. Do diatom valves with different morphologies persist equally over time? i.e. comparing the initially transferred population and populations sampled after time has elapsed, are valves with different roundness, aspect ratio, and solidity represented proportionally, or do some morphologies of particle appear to have been lost more or less readily from the transferred population?) Does the mean/median roundness, aspect ratio, and solidity appear to change over time?

5. If size-selective retention is identified, does this affect the presence or absence of specific species or genera?

   i.e. Comparing the source population and samples derived after hours have elapsed, are there some larger, rounder, or more elongated species or genera which cease to be represented?

7.1.5. The measurement of diatom size and morphology

In order to address these hypotheses relating to the morphology of diatom frustules, it is first necessary to quantify the morphologies of the frustules.

Manual measurement

In studies which involve manual inspection of diatoms with a microscope, the pursuit of hypotheses relating to morphology has, by necessity involved categorisation. For morphology, studies have tended to divide all of the valves seen into either 'centric' (rounded, with an aspect ratio approaching 1:1) or 'pennate' (elongated, with higher aspect ratios) categories (e.g. Scott *et al.,* 2019). Similarly, where analysis is manual, size analysis has also involved categories (e.g. classifying each valve as either greater than or less than 10µm in length (Scott *et al.,* 2019)).

Automated measurement

Since "measuring cell dimensions on a light microscope is a time-consuming process." (Spaulding *et al.,* 2012: 882), automated alternatives have been proposed. One method is imaging flow cytometry (e.g. Spaulding et al., 2012, Sosik and Olson, 2007), the other is image analysis (e.g. Kloster et al., 2014; 2017). To these ends, 'SHERPA' – "the tool for "SHapE Recognition, Processing and Analysis" has been developed (Kloster *et al.,* 2014: 3). This software offers a user-friendly, freely available, workflow, which can take batches of images (vignettes containing diatom valves), segment them, and produce morphometric descriptions of the segmented regions (see: Kloster et al., 2014; 2017), while allowing the user to modify the automated segmentations. SHERPA has the ability to accelerate data analysis, and,

crucially: "the accuracy of measurements is expected to be higher [...and] more precisely reproducible, than data obtained by time-consuming manual measurement (whether using an ocular micrometer or software-assisted linear measurement between manually selected points in a digital image)" (Kloster *et al.,* 2017: 11). SHERPA is freely available online, from: http://www.awi.de/en/science/biosciences/polar-biological-oceanography/main-research-focus/hustedt-diatom-study-centre/sherpa.html

7. 2 Materials and Methods

7.2.1 Metrics of morphology

As with Chapter 6, three different morphological characteristics of the particles were considered: (1) the magnitude, (2) the elongation (whether the valve resembles a sphere/square, or whether it resembles a narrower ellipse or rectangle), and (3) the roughness (whether the surface was smooth, or whether it had projections and crevices). The three metrics used to measure these qualities were:

(1) The cross-sectional area (in µm^2)

(2) The aspect ratio (calculated as the width-height ratio)

(3) The "percent concave area fraction" (PCAF),

Explanatory figures are presented in Figures 7.2 and 7.3, and Table 7.1. The "Percent Concave Area Fraction" (PCAF), to quote from the Technical Manual for SHERPA) "describes the ratio between the area of an object and its convex hull as":

$$PCAF = 100 \cdot \frac{A_H - A_O}{A_O}$$

with A_H = convex hull area; A_O = object area

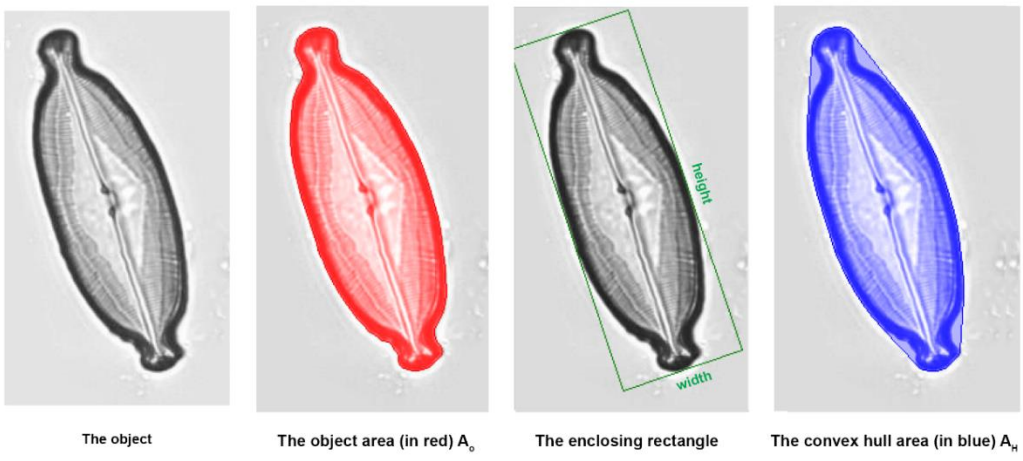


Figure 7.2: Explaining the metrics of morphology (with an example micrograph of a Caloneis)

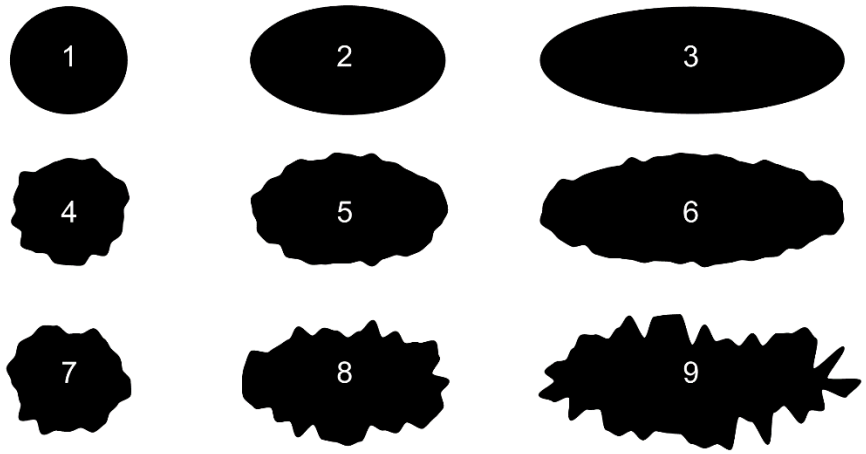


Figure 7.3: Examples of particles, generated in Adobe illustrator

Particle Number	Width-Height Ratio	PCAF
1	1.094	0.77
2	1.775	0.66
3	2.797	0.71
4	1.114	3.74
5	1.727	2.83
6	2.625	3.42
7	1.103	5.5
8	1.646	11.51
9	2.367	20.45

Table 7.1: Examples of the values for the Width-Height ratio and Percentage Concave Area Fraction (PCAF), calculated from the shapes in Figure 7.3

7.2.2 Materials: the diatom suspension

A live freshwater diatom suspension was collected from the River Beane (Hertfordshire, United Kingdom), by wading into the stream, disturbing the sediment, and immersing a container. This live suspension was then decanted into a plastic container for further experiments. This was the same sample site as used in Scott *et al.,* 2019. The water sample contained a relatively diverse mixture of diatom species, with 24 genera identified within the experiment. (Table 7.2). These genera included *Melosira, Navicula*, and *Nitzschia* - all typical of freshwater environments in the United Kingdom (Kelly, 2000; Scott *et al.,* 2019). Tables of micrographs of these genera can be found in the appendix, as can tables outlining descriptive statistics for the morphological measurements. It is worth noting that Table 7.2 provides the species data as percentages of composition, rather than counts. This was a deliberate choice, to enable comparison with other work that has been undertaken not only in the same field, but also using the same fieldwork site (Scott *et al.,* 2019). For full, numerical data, please see the attached SD card containing the data appendix.

Abundance		Genus	Percentage of source samples (N=1489)
>1% of sample	1	Achnanthidium	9.00
	2	Amphora	2.08
	3	Centric	7.52
	4	Cocconeis	7.59
	5	Cymatopleura	1.61
	6	Melosira	7.45
	7	Navicula	8.53
	8	Nitzschia	9.87
	9	Rhoicosphenia	1.95
	10	Surirella	2.96
	11	Synedra	2.22
<1% of sample	12	Caloneis	0.13
	13	Cymbella	0.47
	14	Diatoma	0.67
	15	Eunotia	0.27
	16	Fragilaria	0.67
	17	Gomphonema	0.87
	18	Gyrosigma	<0.01
	19	Hantzschia	0.81
	20	Navicula (with concavity)	0.13
	21	Pinnularia	0.34
	22	Stauroneis	0.54
	23	Staurosirella	0.13
	24	Tryblionella	<0.01

Table 7.2: List of genera identified from the micrographs of the 'source' samples

As outlined in section 7.1.2, three forms of diatom were recovered: whole valves, chains (where multiple cells of the same genus were contiguous), and fragments. The different genera covered a wide range of sizes (Figure 7.4). The genus with the smallest mean cross-sectional area was *Achnanthidium*, with a mean cross sectional area, for a whole valve, of 60.20 ± 28.82 µm^2 , and the largest was *Cymatopleura*, with a mean cross sectional area, for a whole valve, of 2176.47 ± 964.46 µm^2. The only genus to be subdivided into two categories was *Navicula*, a "group" recently split into several genera (Round *et al.,* 1990), where the species presented two different morphologies (i.e. *Navicula capitata* presented regions of concavity, whereas the other species seen in the sample were a simpler shape, lanceolate and convex). In the scanned slides, very few of the diatoms were in girdle view; single figures out of a total number of 8578 vignettes.

7.2.3 Materials: the fabric swatches

The recipient surface used in this experiment was the same woven cotton fabric, that was used in experiments in chapters 3, 4, 5, and 6, and the published paper[s] (Levin *et al.,* 2017, and Levin et al., 2019). The transfer procedure was the same as in Levin *et al.,* 2017 – 4cm^2 swatches of the fabric (2cm by 2cm squares) were immersed in the diatom-containing water for three minutes, and attached to the sides of the footwear using staples. As seen in Chapter 5, the minor variations in positions of the swatches did not introduce statistically significant variations in the number of particulates retained.

Swatches were removed from the footwear at five time intervals; immediately (at 0 hours), and then after 1, 4, 24, and 168 hours. Five swatches (i.e. five replicates) were removed at each time interval. Five samples of the 'source' (diatom suspension) were also removed, with a pipette.

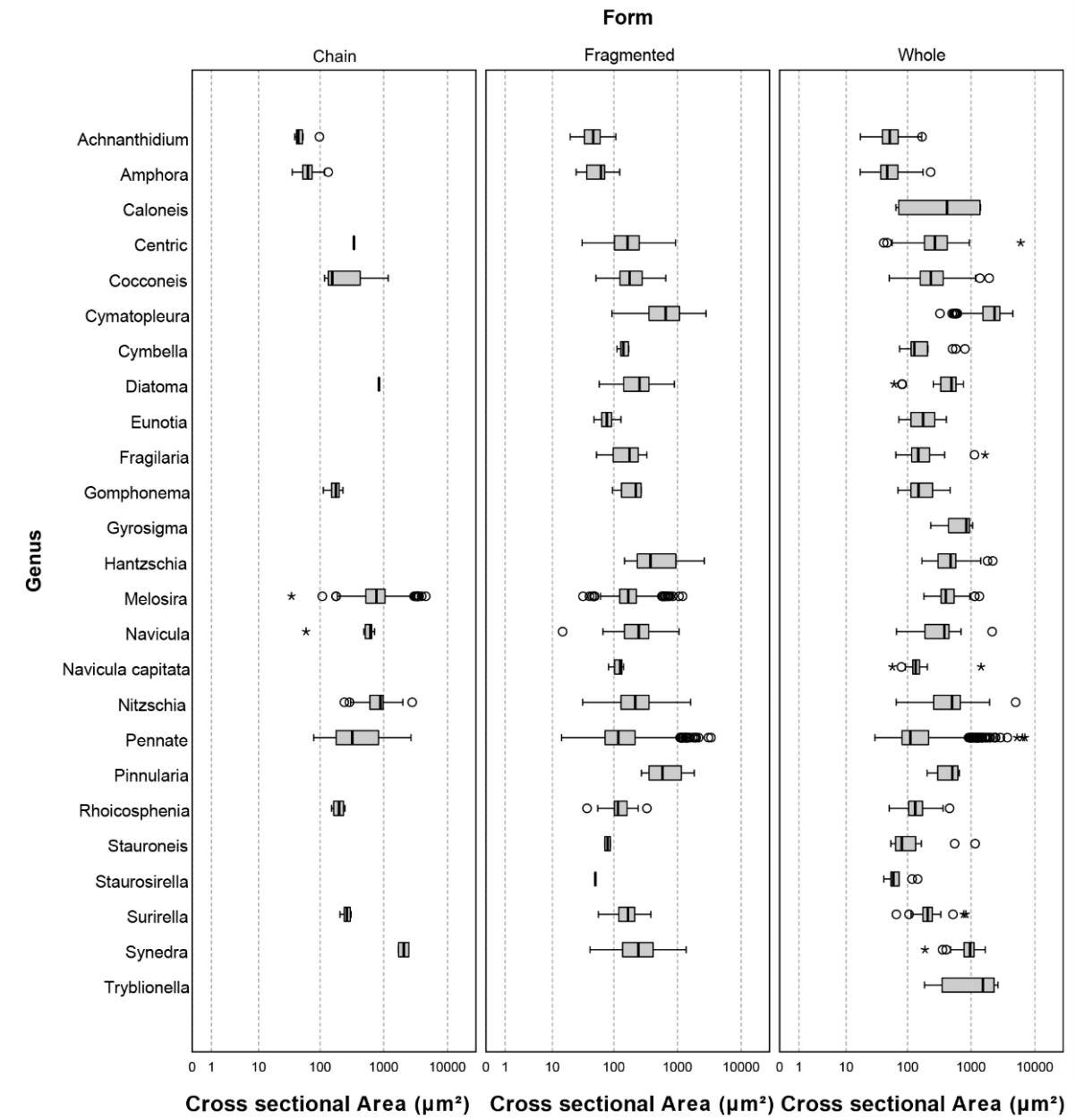


Figure 7.4: Boxplot of the area (μm^2) for all objects (forms of diatom) in the experiment. N = 8578. Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

7.2.4 Method: diatom extraction and slide preparation

Diatoms were extracted from the fabric using the water and agitation method outlined in Levin et al 2017. These diatom suspensions, along with the 'source' samples were then subjected to standard hydrogen peroxide digestion procedure (citation: is it Renberg, 1990; see also Scott *et al.,* 2014). Slides were also prepared according to the standard procedure, with Naphrax used as the mounting medium, a hotplate to drive off the toluene, and a standard 19mm coverslip.

7.2.5 Method: slide imaging and preparation

Slides were imaged at the University College London *Institute of Neurology.* This imaging was undertaken using a LEICA SCN400F , and scans were delivered as very large stitched and stacked files (.scns). The resolution of the highest magnification was 4 pixels per micron (i.e. 1000 pixels = 250 microns). As the system was optimised for histological samples, it was not possible to fully optimise the entirety of each scan: some valves were out of focus, and others were bisected by stitching artefacts. However, for all slides, several hundred valves, chains, or fragments of valves were in focus and not affected by artefacts. As with the preceding chapters, it is acknowledged that this method of data acquisition relied on trust in the black-box method of the software that generated the images, as is the convention in this field of experimental forensic science (French et al., 2012; Morgan et al., 2013).

7.2.6 Methods: digital tools

As stated in Chapter 1, this chapter was a collaboration with Michael Kloster of the AWI, the creator of SHERPA (citations: Kloster *et al.,* 2014; 2017). As such, it was a chance to use some of the cutting-edge systems designed for the analysis (and collaborative analysis) of imagery. Two tools were used. For image annotation, the scanned slides were uploaded to BIIGLE, the "BioImage Indexing, Graphical Labelling and Exploration" (Langenkämper *et al.,* 2017), "a web service for the efficient and rapid annotation of still images" (verbatim quote from description). For image analysis, a customised version of SHERPA was provided, which was able to

deal with very large numbers of manual re-workings of suggested segmentations, without crashing or experiencing memory leakage.

7.2.7 Methods: Image annotation and labelling

Images were annotated within BIIGLE (see Figure 7.5). The "lawnmower" mode was used to navigate the slides; this shows a single field of view at a magnification, and allows the viewer to jump to the next field of view by using the arrow keys, so performing transects across the slide (Langenkämper *et al.,* 2017). Each time a valve, chain of valves, or fragmented valve was encountered, the following flowchart was undertaken (Figure 7.6). Two 'label trees' were created: one for 'form' (whole valve, fragmented valve, or chain of valves), and another, taxonomic label tree. It was decided that, instead of classifying each valve down to species level, genus level classification would be sufficient for this study (especially since, for many of the valves, the image was not high-enough resolution to enable a definitive species-level classification), but it was possible to be confident in a genus-level classification. Initial annotation was using the polygon tool, to isolate the object of interest with a form annotation. A genus label was then attached.

7.2.8 Methods: Morphometric Analysis

Quantitative morphometric analysis was carried out in SHERPA. The vignettes identified in BIIGLE were exported, and these vignettes, which each contained a valve, chain of valves, or fragmented valve, were fed through SHERPA. This resulted in a rough segmentation. This segmentation was manually amended, and then the morphometric data exported, along with the object data (Figure 7.7) While over 25, 000 vignettes were identified during the image annotation phase, it was not possible to conduct morphometric analysis upon them in the scope of this study, due to the time required to manually re-work the outlines for fragmented valves (SHERPA was designed and optimised for whole valves). Therefore, it was decided to analyse 300 per slide. There were fewer than 300 for four slides. The number of vignettes can be seen in Table 7.3.

7.2.9 Statistical Analysis

As in the preceding chapters (3, 4, 5, and 6), the statistical analysis undertaken in this chapter consisted of both descriptive and inferential analysis. In addition to the mean and standard deviation (i.e. the measure of central tendency and measure of dispersion used in all preceding chapters), the minimum and maximum values for each morphometric variable (i.e. the cross-sectional area, width-height ratio, and PCAF) were calculated and presented, as the maximum particle size has been mooted as a variable which may change throughout the course of a persistence study (Lowrie and Jackson, 1991; Levin *et al.,* 2017; Chapter 6).

As with Chapter 6, guidance was sought as to what the most appropriate inferential statistical treatment of the data in this chapter would be. Since the samples were independent (i.e. no single particle was counted twice), the researcher was advised that the data presented in this chapter are not a true time series – since the purpose of time series statistical analysis is for repeated observations of the same units (i.e. the same particles), where controlling for autocorrelation is important. Accordingly, it was advised that the data collected in this chapter could be treated as independent samples with categories of different treatments – with the treatment relating to time elapsed between transfer and sampling. Accordingly, the researcher was advised that, as with Chapters 5 and 6, the Kruskal Wallis test was an appropriate non-parametric inferential test of whether independent samples are drawn from the same distribution (see: McDonald, 2009). The data for this chapter fulfilled the assumptions of the test, since they were (a) independent measures (b) with a dependent variable measured at a scale level (better than ordinal level), and (c) divided into three or more groups (see Field and Hole, 2007: 275). Where the data had two categories rather than three, the Mann Whitney U test was employed (see: Field and Hole, 2007).

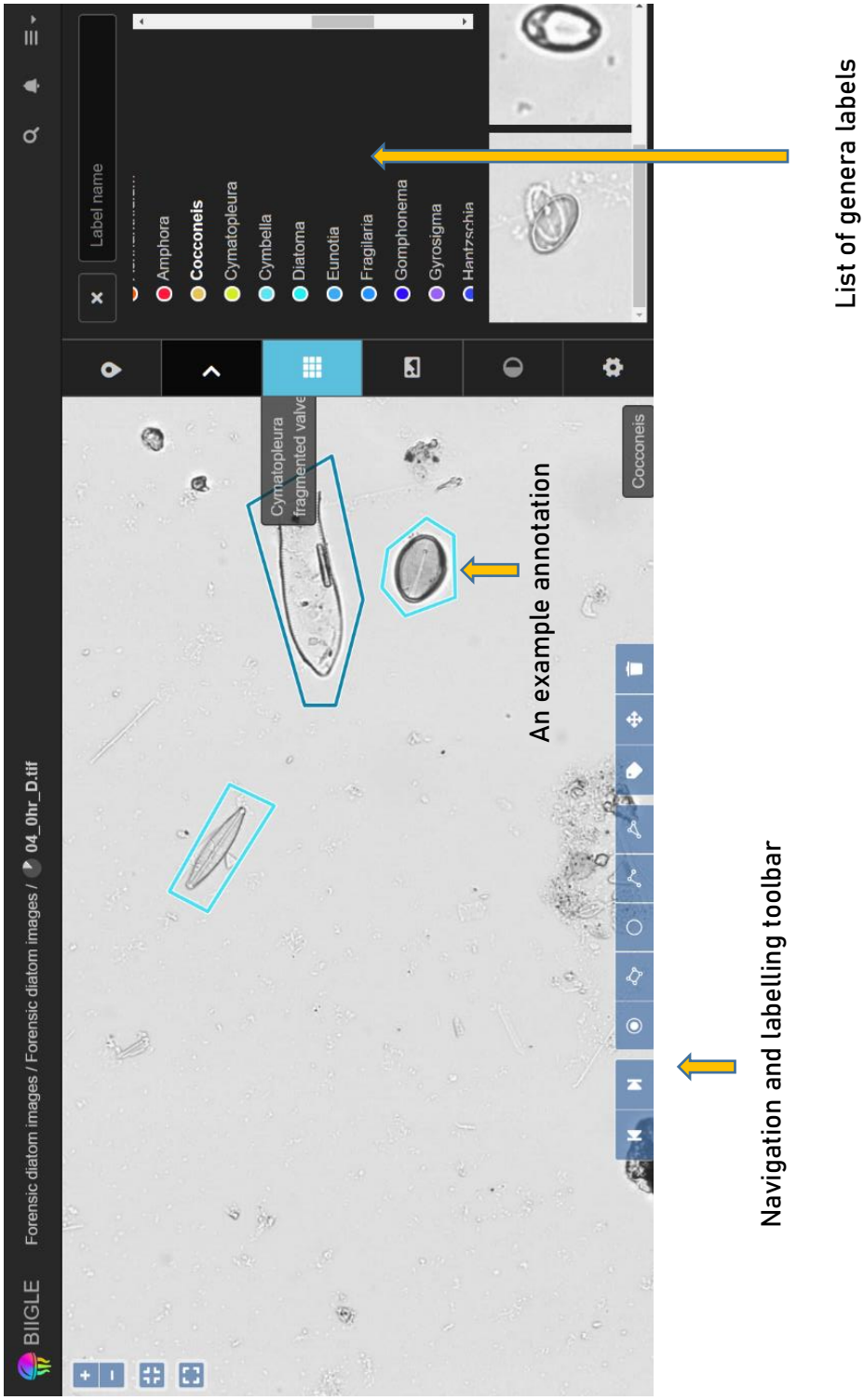


Figure 7.5: An example screenshot of BIIGLE

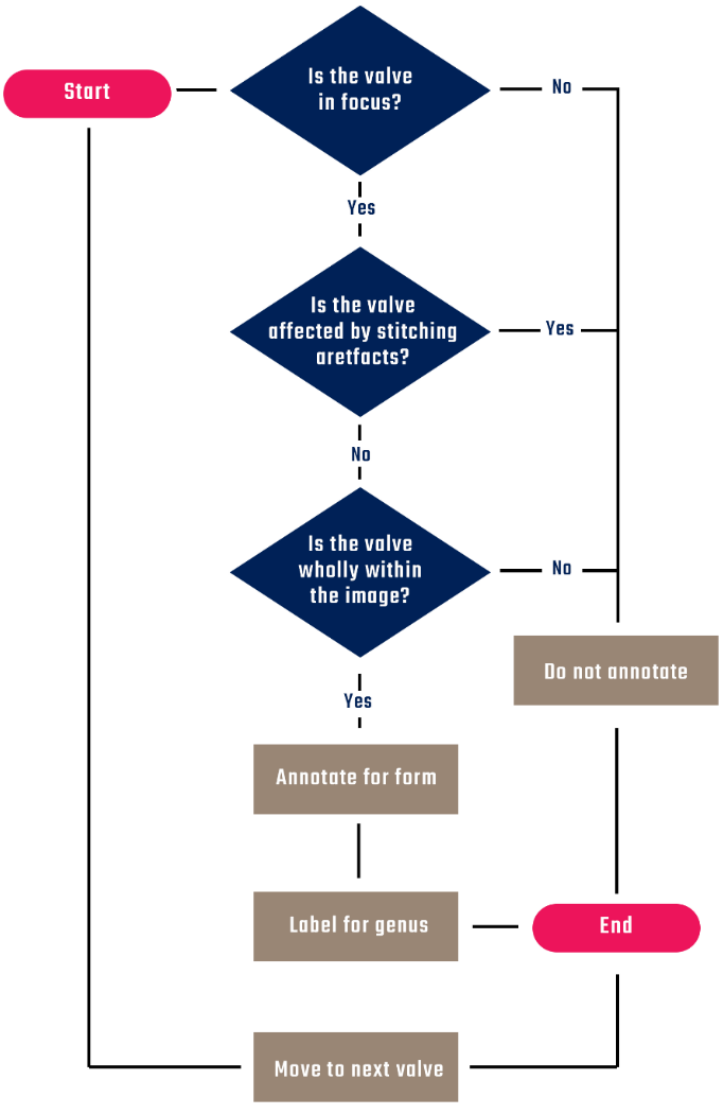


Figure 7.6 Flowchart showing the vignette generation procedure

Time	N					
	Run					Total
	A	B	C	D	E	
0	300	300	300	300	300	1500
1	300	300	197	300	103	1200
4	300	300	300	300	300	1500
24	300	300	300	300	300	1500
168	300	300	300	300	189	1389
Source	300	300	300	289	300	1489

Table 7.3 The number of vignettes analysed for each slide, with the values for each run and time

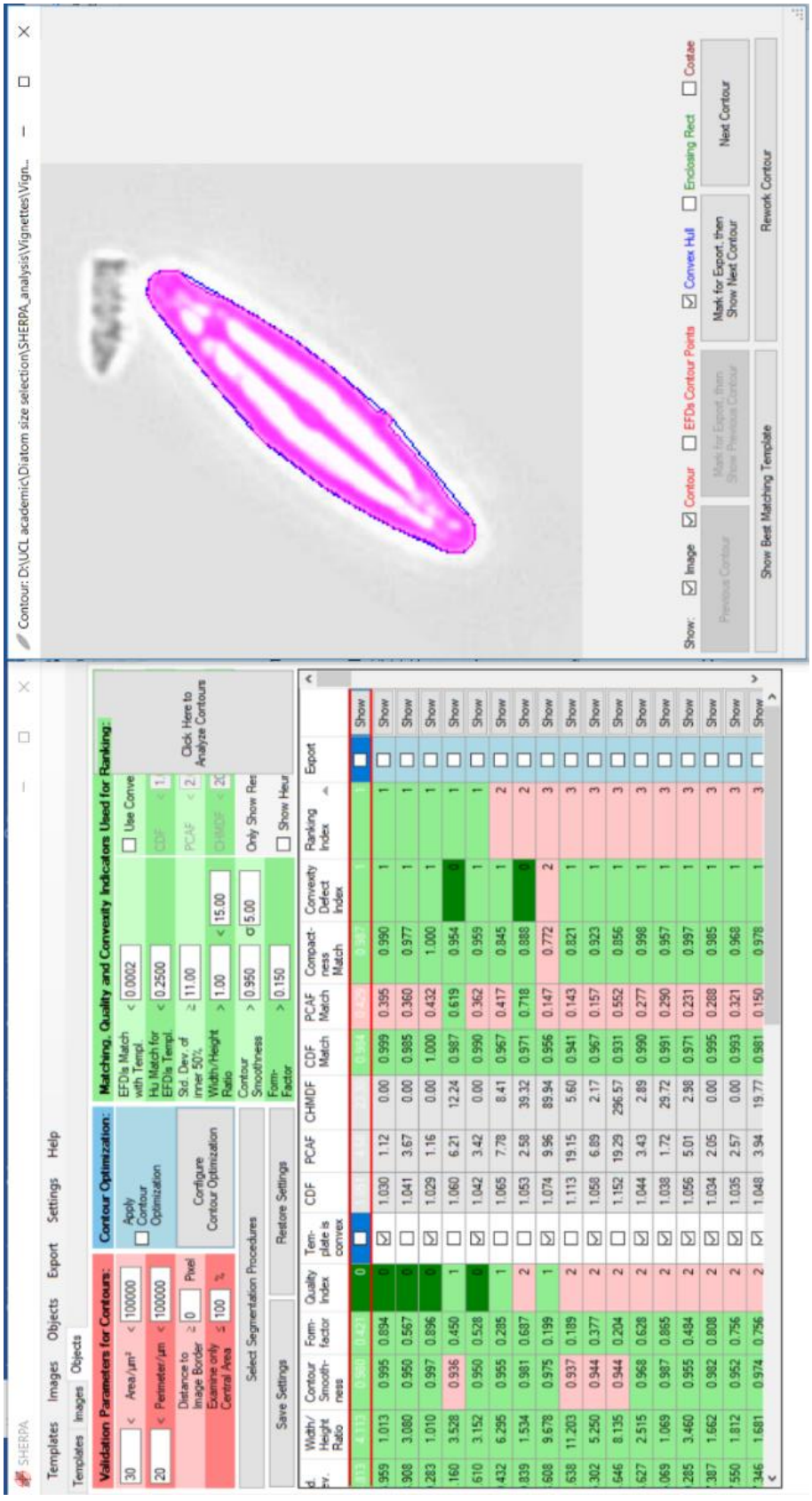


Figure 7.7: An example screenshot of SHERPA, showing a suggested segmentation

7.3 Results

7.3.1. Checking for linear relationships/correlation

In a similar manner to the study addressing quartz grains (see Chapter 6), the first phase of data analysis was to check for covariance between the morphometric variables (Figure 7.8). There was not a linear relationship between the area and width-height ratio (linear $R^2 = 0.049$), nor a relationship between the area and PCAF (linear $R^2 = 0.002$).

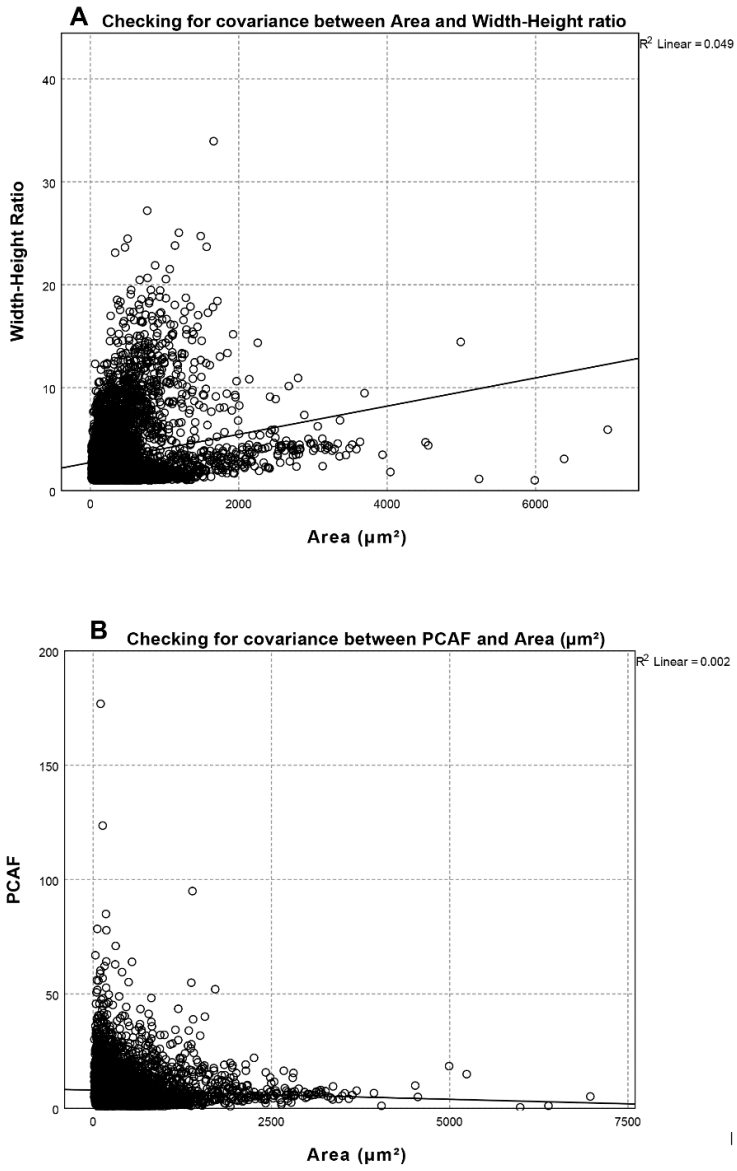


Figure 7.8: Checking for covariance between (A) Cross sectional area and Width-Height Ratio, and (B) Cross sectional area and PCAF

7.3.2. Diatom size and transfer

Boxplots showing the distribution of values for area are presented in Figure 7.9, and descriptive statistics for each run are provided in Table 7.4. When comparing the areas observed between the source population (the 'source' samples), and the particles which were transferred (the '0 hours' samples), it can be observed that both the mean and maximum particle areas observed for the '0 hour' samples were higher than the 'source' samples, and that both samples had a large range in values; the mean particle area for the 0-hour samples was 408.37 ± 513.03 μm^2 (M ± SD, N=1500), while the mean particle area for the source samples was 232.17 ± 299.35 μm^2 (M ± SD, N=1489). Running inferential statistical analysis (Table 7.5) suggested that this trend may be worth further investigation.

7.3.3. Diatom size and persistence

Table 7.4 shows that the mean and median particle areas exhibit a downward trend over time, with the samples collected after 0 hours and 1 hour appearing higher than the subsequent samples removed at 4, 24, and 168 hours. Conducting inferential statistical analysis shows these differences in mean, and distribution, to yield p values of less than 0.05, suggesting that further investigation of the effects of these variables may be worthwhile (Table 7.6).

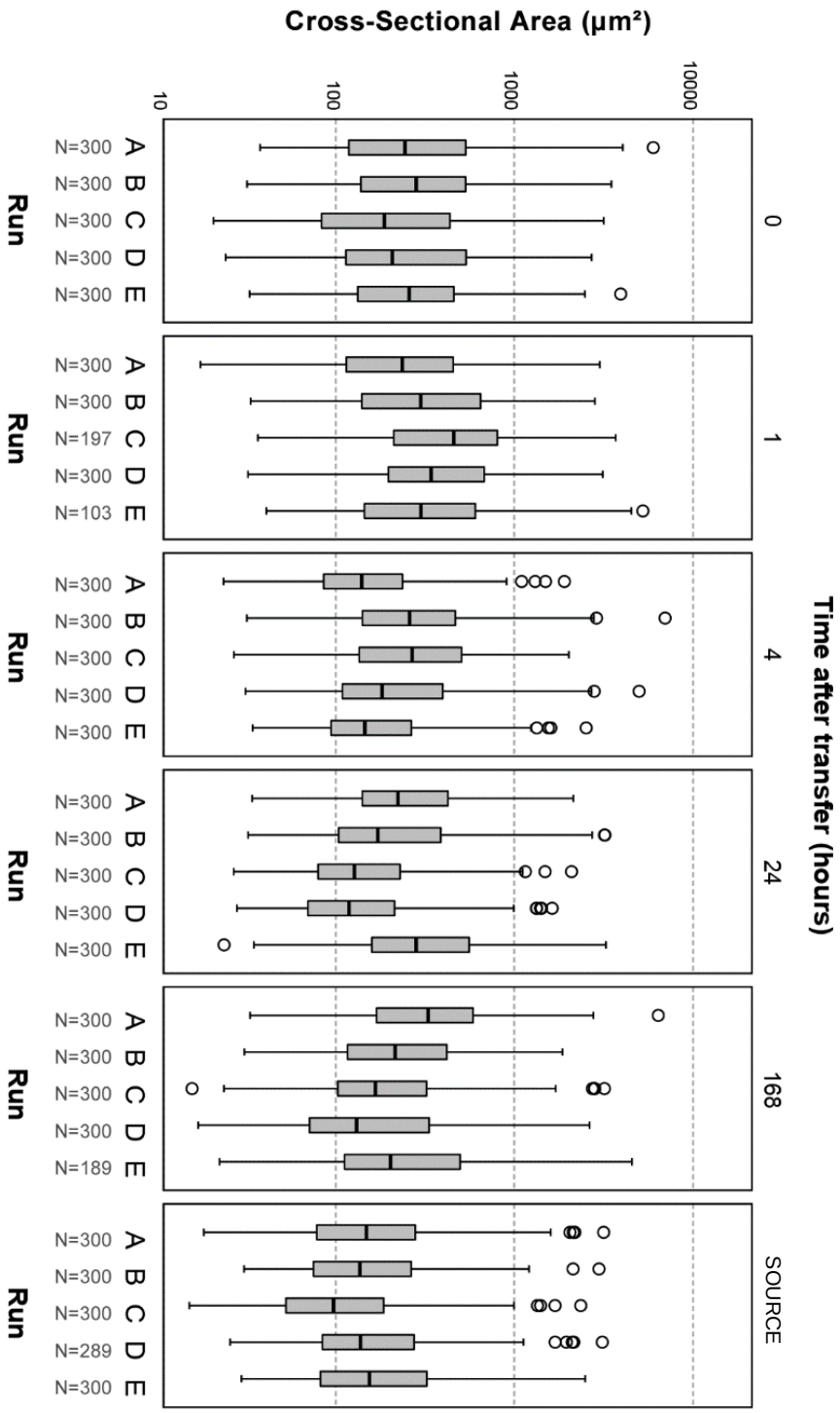


Figure 7.9: Boxplot of particle Cross Sectional Area (μm^2) for each time interval. Scale is logarithmic (base 10). Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

Time	N	Area (µm²)				
		Mean	Std. Deviation	Median	Minimum	Maximum
0	1500	408.3712	513.03399	238.3600	19.94	5988.63
1	1200	482.5406	529.68869	312.7350	16.69	5238.88
4	1500	317.0628	408.08751	188.2650	22.81	6972.25
24	1500	303.9500	372.30000	174.3100	22.97	3258.25
168	1389	347.7078	427.69743	197.8400	14.88	6385.19
Source	1489	232.1706	299.34554	139.3400	14.34	3162.16

Table 7.4: Summary descriptive statistics for particle cross-sectional area over time

Null hypothesis	Test	Total N	Test statistic	DF	Significance	Decision
The distribution of particle area (µm2) are the same across the categories of 'source' and '0 hours'	Independent Samples Mann-Whitney-U Test	2989	784,881	1	<0.05	Reject the null hypothesis

Table 7.5: Inferential comparison between the source and Time = 0 samples, for area

Hypothesis test summary: comparing the morphological metrics over time (total N= 7089)						
Variable	Null hypothesis	Test	Test statistic	DF	Sig	Decision
Cross sectional area (mm^2)	The distribution of "cross sectional area" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	202.794	4	.000	Reject the null hypothesis
Width-Height Ratio	The distribution of ""width-height ratio" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	22.464	4	.000	Reject the null hypothesis
PCAF	The distribution of "PCAF" is the same across categories of "time since transfer" [0, 1, 4, 24, and 168]	Independent Samples Kruskal-Wallis Test	350.056	4	.000	Reject the null hypothesis

Table 7.6: Inferential statistical summary for morphological descriptors

7.3.4. Diatom morphology and transfer

Boxplots of the morphological features are presented in Figures 7.10 and 7.11, and summary descriptive statistics are provided in Table 7.7. Inferential statistics, comparing the '0 hour' samples and 'source' samples are presented in Table 7.8. It can be seen that, comparing the width-height ratio for the source samples and the 0 hour samples, the source samples exhibit a slightly higher mean (of 3.44 ± 2.82) and maximum (24.47) than the 0 hour samples (M = 3.34 ± 2.94, Max = 23.81). The difference between both the medians and distributions of the width-height ratios was seen to be statistically significant at $p<0.05$. Similarly, the mean PCAF was lower for the 0-hour samples (5.87 ± 5.60) than for the source samples (7.83 ± 6.71). Running the Mann Whitney U test for the distributions for the two conditions of 'source' and '0 hours', a p value of <0.05 was calculated, suggesting that this trend may be worth further investigation.(Table 7.8).

Time	N	WidthHeight Ratio				PCAF			
		Mean	SD	Min	Max	Mean	SD	Min	Max
0	1500	3.34	2.94	1.00	23.81	5.87	5.60	0.43	64.21
1	1200	3.19	2.99	1.00	33.94	6.80	7.58	0.80	123.60
4	1500	3.26	2.72	1.00	25.05	8.14	6.65	0.82	84.96
24	1500	3.02	2.30	1.00	24.73	8.86	8.58	0.80	176.85
168	1389	2.93	2.42	1.00	27.19	8.30	7.57	0.85	94.97
Source	1489	3.44	2.82	1.00	24.47	7.83	6.71	0.90	60.26

Table 7.7: Summary of morphological metrics over time

Variable	Null hypothesis	Test	Total N	Test statistic	DF	Significance	Decision
Width Height Ratio	The distribution of width-height ratio (µm2) are the same across the categories of 'source' and '0 hours'	Independent Samples Mann-Whitney-U Test	2989	1182305.500	1	.005	Reject the null hypothesis
PCAF	The distribution of PCAF (µm2) are the same across the categories of 'source' and '0 hours'	Independent Samples Mann-Whitney-U Test	2989	1363292.000	1	.0000	Reject the null hypothesis

Table 7.8: Summary of inferential statistics comparing 0 hour and source samples, morphometrics

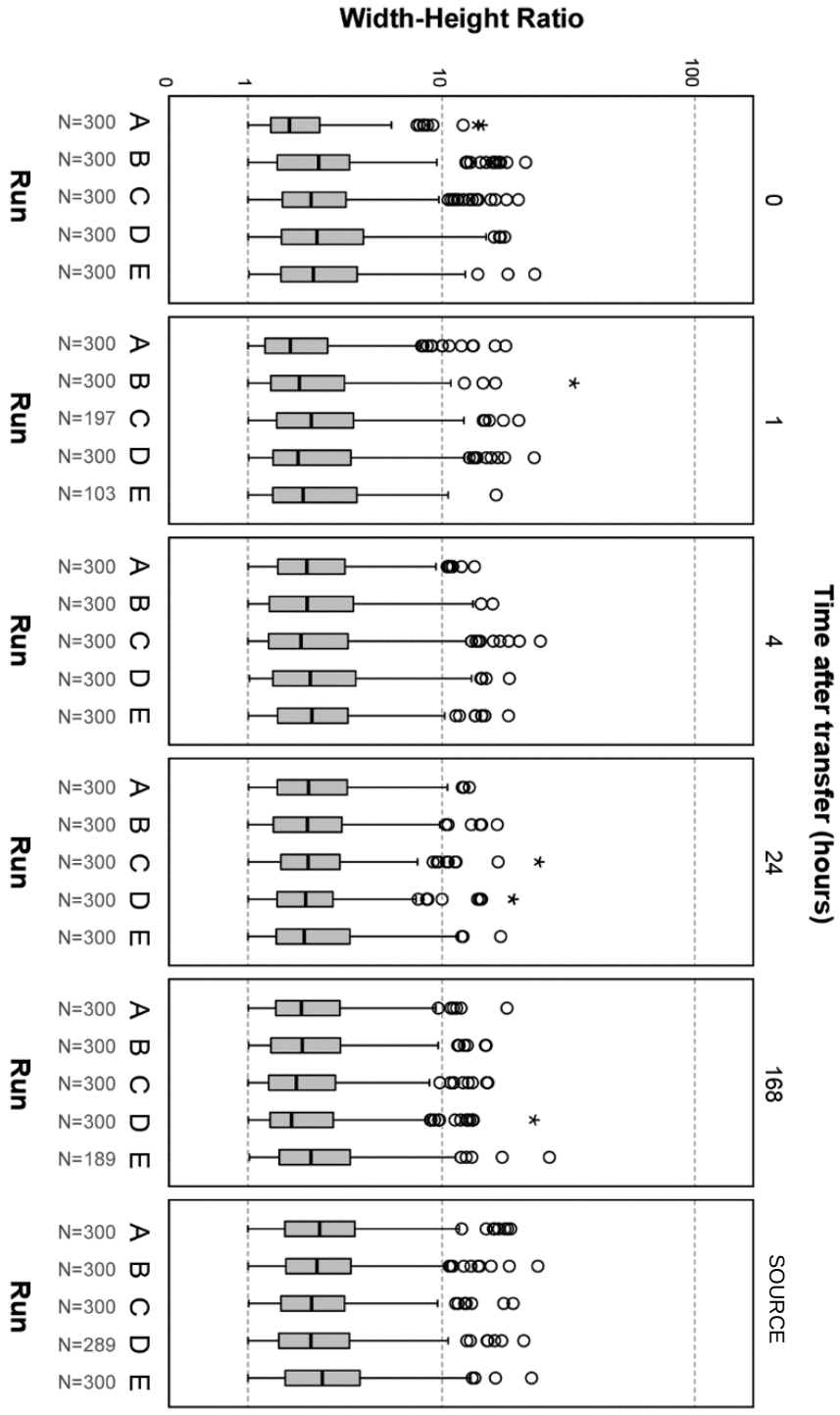


Figure 7.10: Boxplot of particle Width-Height ratio for each time interval. Scale is logarithmic (base 10). Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

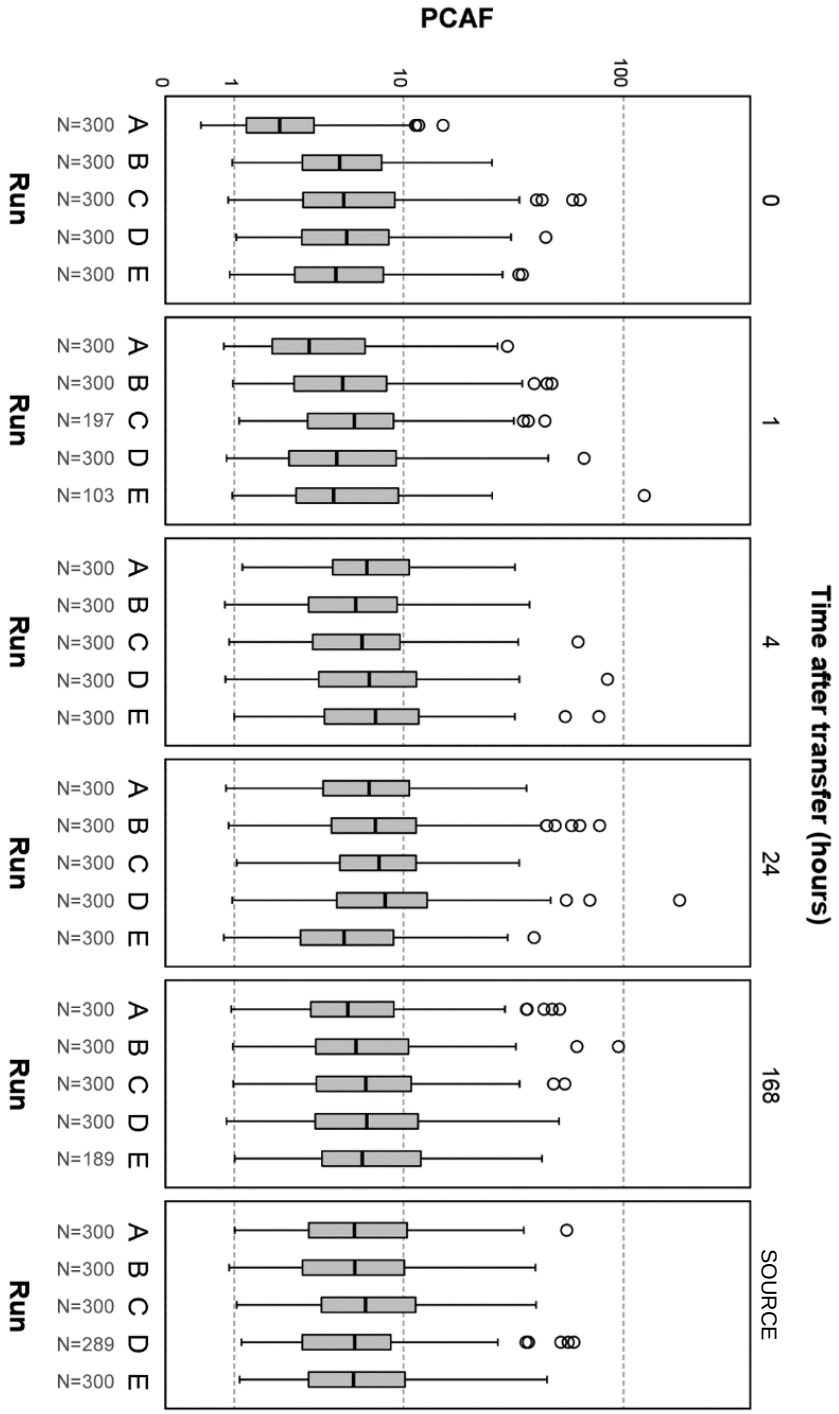


Figure 7.11: Boxplot of particle PCAF for each time interval. Scale is logarithmic (base 10). Circles denote outliers (1.5-3.0 x the interquartile range from the median) while asterisks denote extremes (>3 x the interquartile range from the median).

7.3.5 Diatom morphology and persistence

Inferential statistical summaries can be found in Table 7.6, and boxplots can be seen in Figures 7.10 and 7.11.

For both the Width-Height Ratio and PCAF, the p-values calculated were <0.05, suggesting that further investigation of the effects of these variables may be worthwhile (Table 7.6). For the width-height ratio, the trend was broadly a decline over time, but with a number of outliers, with the mean value declining from 3.34± 2.94 at 0 hours (N = 1500) to 2.93 ±2.42 at 168 hours (N = 1389). The trend for PCAF seemed to be to increase over time, with the mean value increasing from 5.87 ± 5.60 at 0 hours (N = 1500) to 8.30 ± 7.57 at 168 hours (N = 1389) (Table 7.7).

7.3.6 Size selective effects upon genus

The presence or absence of each genus at each time interval is summarised in Table 7.15. There did not seem to be a systematic loss or dropout of larger genera; the largest genus in the experimental diatom solution, *Cymatopleura*, was present in 5 of 5 samples collected 168 hours after the transfer (Table 7.9).

Genus	Number of replicates in which the genus was identified (of 5)					
	Source	0	1	4	24	168
Achnanthidium	5	5	5	5	5	5
Amphora	5	5	4	3	4	5
Centric	5	5	5	5	5	5
Cocconeis	5	5	5	5	5	5
Cymatopleura	5	5	5	5	5	5
Fragilaria	5	5	4	3	3	4
Gomphonema	5	5	4	3	3	3
Hantzschia	5	3	3	4	4	3
Melosira	5	5	5	5	5	5
Navicula	5	5	5	5	5	5
Nitzschia	5	5	5	5	5	5
Pennate	5	5	5	5	5	5
Rhoicosphenia	5	5	5	5	5	4
Surirella	5	5	5	5	5	5
Synedra	5	5	4	5	5	5
Cymbella	4	4	2	0	2	2
Diatoma	4	5	4	5	5	5
Stauroneis	4	3	1	3	0	3
Eunotia	3	1	1	1	0	1
Pinnularia	3	0	0	2	1	0
Navicula capitata	2	3	3	5	3	4
Staurosirella	2	3	2	0	0	1
Caloneis	1	1	2	0	0	1
Gyrosigma	0	2	0	0	0	0
Tryblionella	0	2	2	0	0	1

Table 7.9: The number of replicates in which each genus was recorded

7.4 Discussion

7.4.1. Diatom size and transfer

Comparing the samples collected from the source (Source) and the samples collected immediately after transfer ('0 hours'), the 0-hour samples appeared to have a higher mean and higher maximum particle area. This suggested that the largest diatoms within the source population did transfer readily. For both the 0-hour (M = 408.37 ± 513.03 µm^2 , N = 1500) and source samples (M = 232.17 ± 299.34 µm^2 , N = 1489), the standard deviation of the mean was large (larger than the mean value) implying, as can be seen in the boxplots, that the datapoints were spread out over a large range of values (Table 7.4). This suggests that all sizes of diatoms in the (varied) population were capable of transfer, and that natural populations of diatoms are variable in size.

This finding varied from the findings generated in the preceding chapter, where it was demonstrated that the largest quartz grains in the sample population (i.e. those with the greatest cross-sectional area) did not transfer from the population to the swatch in the experiment. It is worth noting that the sizes of the two particles are several orders of magnitude apart: the maximum cross sectional area seen for the diatom particles was 6972.25 µm^2 , while the maximum cross sectional area seen for quartz grains used in Chapter 6 was 630,000 µm^2 (0.63mm^2).

7.4.2 Diatom size and persistence

The mean and median particle size showed a trend that decreased over time (Table 7.4), and this variation was seen to be confirmed by inferential analysis, in the form of the Kruskal Wallis test, yielding $p<0.05$, as a trend worth further investigation (Table 7.6). However, the maximum particle area observed did not show a similar trend of decline over time, with many (large) outliers visible throughout the data (Figure 7.9). While the maximum cross-sectional area observed at 0 hours was 5988.63 µm^2, the maximum observed at 168 hours after the transfer was 6385.19 µm^2 (Table 7.4).

This finding was not consistent with preliminary research on diatoms or quartz, which suggested that larger size fractions might be absent in samples collected 168 hours after transfer (Levin *et al.,* 2017; see Chapter 6). It was also inconsistent with the variation in findings from the literature for glass fragments and fabric fibres, where the maximum particle size observed was seen to decline over time (Brewster *et al.,* 1984, Hicks *et al.,* 1996, Rupert *et al.,* 2018; Pounds and Smalldon, 1975b). There may be several possible explanations for this deviation. It is possible that the characteristics of the diatoms in this study were genuinely behaving differently to the glass fragments and fibres in previous studies (perhaps due to higher surface roughness, with some species of diatoms extruding an adhesive mucous while living, which may affect how they attach to a recipient medium). Further research is needed to establish whether the nature of the diatoms is a significant factor.

7.4.3. Diatom morphology and transfer

Comparing the samples collected from the diatom suspension ('source') and collected immediately after transfer ('0 hours'), both the median width-height ratio and PCAF were seen to be statistically significantly different (Table 7.10). The 0-hour samples exhibited a lower mean width-height ratio (i.e. they were less elongated; with a mean of 3.34 ± 2.94 (N = 1500), versus a mean of 3.44 ± 2.82 (N=1489)) and a lower mean PCAF (i.e. they had fewer or smaller regions of concavity; with a mean of 5.87 ± 5.60 (N = 1500), versus a mean of 7.83 ± 6.71 (N=1489)). These findings vary from those in Chapter 6, where there was no statistically significant difference observed between the aspect ratios of quartz grains in the population and aspect ratios from the 0 hour samples. They also differ in that, for quartz, the grains which transferred had a lower median solidity (analogous to a higher PCAF). Future research may want to explore this further.

7.4.4 Diatom morphology and persistence

For both the Width-Height Ratio and Percentage Concave Area Fraction (PCAF), statistically significant differences were seen in the median values and distributions over time (at $p<0.05$) (Table 7.6). For the width-height ratio, the trend was broadly to decline over time (from a mean of 3.34 after 0 hours, to a mean of 2.93 after 168

hours), and for the PCAF the trend was to increase over time seemed to be to increase over time (from a mean of 5.87 after 0 hours, to a mean of 8.30 after 168 hours; Table 7.7). This implied that diatoms (or chains of diatoms) with higher aspect ratios were less likely to persist over time, and that diatoms (or chains of diatoms) with higher concave area fractions (i.e. lower levels of smoothness/convexity) seem more likely to persist over time. These findings therefore support the assertion that the morphology of a particle can influence its persistence (as in Houck, 2001).

7.4.5 Size selective effects upon genus

In this experiment, there did not appear to be the loss of any particular genus due to size-selective retention. This suggested that, in the interpretation of an assemblage of diatoms as trace evidence, an examiner may wish to consider the presence or absence of a particular (abundant) genus as salient. What was evident was that, for many genera in this experiment, their presence in the population was rare – of the 24 genera identified, thirteen comprised less than 1% of the original sample (Table 7.2). Accordingly, in a casework scenario, if a specific genus was absent in such a sample, it could be a function of sample size and probability – if few diatom valves are recovered, the rarer species may not be represented in the assemblage.

7.4.6. Limitations of this study, and future work

The main limitation of this study was that working with imagery meant that one is using a two-dimensional representation of three dimensional objects. Future work may wish to repeat the experiment, and see whether the same divergence is seen in terms of size selective trends (to establish whether the maxima decline over time) and with respect to PCAF over time. In a future experiment, it may be beneficial to use different size fractions of diatom; the previous study (in which the maxima were seen to decline over time; Levin *et al.,* 2017) had a maximum particle area of $<1500\mu m^2$, while the maximum particle area of observed within this experiment was $>6000\mu m^2$. There may also be value in running an experiment with particles of known, controlled, and stratified morphological features introduced to a fabric, and

to test whether they decline equally over time, exploring the effects of particle area without the complication of varying initial levels of abundance. Microspheres could be used to achieve this. It is worth noting that this study is limited by the use of genus-level classification rather than species-level classification. It is possible that disaggregating the data to species level may reveal some trends in presence and absence that are not visible when aggregated.

7.5 Conclusions

7.5 Conclusions

In response to the suggestion that "general characteristics including size and shape" may impact the transfer of particulates to clothing (Scott et al., 2019: 304) in the context of the transfer of freshwater diatoms to fabric, two conclusions can be stated:

(1) In this study, all size fractions of diatom found in the population were seen to transfer to the cotton used as a recipient surface (suggesting that size is not a limiting factor with regards to the transfer of freshwater diatoms to fabric).

(2) In this study, the median particle aspect ratio and PCAF (a measure of solidity, the 'percentage concave area fraction') were seen to differ between samples taken from the diatom population and samples taken from the fabric immediately after transfer, (suggesting that the morphological qualities of a diatom may indeed support or limit their transfer). In other words, the findings of this study suggest that the shape and of a diatom valve may matter to its behaviour as trace evidence, and this is the first study which utilised tools that allowed the shape of each valve to be quantified.

In response to the suggestion that the size and shape of particulates can support or limit their persistence upon articles of evidence (e.g. Houck, 2001; Brewster *et al.,* 1984, Hicks *et al.,* 1996, Rupert *et al.,* 2018; Pounds and Smalldon, 1975b), this experiment offered two additional conclusions:

(1) Over time, the mean values for width-height ratio appeared to decline, and PCAF appeared to increase. These differences between time intervals were supported by inferential statistical analysis, supporting the theory that the morphology of a particle can influence its persistence (as in e.g. Houck, 2001), and that these morphological variables merit further investigation. Accordingly, this may create a caveat for the interpretation of evidence.

(2) While these mean values concurred with findings presented in the existing literature (e.g. Pounds and Smalldon, 1975), one of the findings of this study (i.e. that the maximum particle area seen over time did not decline) diverged from the findings of previous studies, both with diatom valves (Levin *et al.,* 2017) and with glass (e.g. Hicks *et al.,* 1996). Together, these findings confirm the need for further empirical research to be undertaken (e.g. Forensic Science Regulator, 2015; 2019), and reassert that such empirical research is necessary if one desires the interpretation of such indicators within casework to be robust.

8

Chapter Eight

Synoptic discussion

8.1 Part One: The importance of empirical experimentation

8.1.1. Findings in this research

The first theme that emerged repeatedly in this research was the importance of undertaking empirical experimentation. This was a theme that was stated in the introduction (see: Morgan, 2017; 2019; Forensic Science Regulator, 2019), revisited in the literature review (e.g. National Academy of Sciences, 2009; e.g. Jackson and Biederman, 2019), and feeds in to a broader point about ensuring the *quality* of inferences drawn within forensic casework. Put simply, in order to make robust inferences, there should be a dataset to back up the assertions that one is making (Forensic Science Regulator, 2018; National Academy of Sciences, 2009; e.g. Jackson and Biederman, 2019).

Crucially, in this research, some of the findings which emerged were:

(1) Unexpected or counter-intuitive

For example, in Chapter 5, it was observed that the location on a shoe's upper that was samples were removed from, the heel, toe, inner or outer edge, did not appear to cause observable variation in the amount of particulate evidence retained (see Section 5.5)

(2) Slightly concerning

For example, in Chapter 3 it was found that different examiners asked to estimate the extent of the foreground in an image of fluorescent powder gave answers which were statistically significantly different (see Section 3.5; Levin *et al.,* 2019).

(3) In contradiction to the findings of previous studies

For example, in Chapter 7, the very largest diatoms in the experimental population were seen to transfer readily to the recipient material surface, in a finding inconsistent with studies of other particulates such as glass

and fibres (e.g. Hicks *et al.,* 1996) and with a similar study of diatoms (Levin *et al.,* 2017; see Section 7.5).

8.1.2. Implications

Accordingly, what these *unexpected* findings suggest is that it is not enough to rely upon what seems intuitive, when dealing with questions of transfer and persistence. Until the research is carried out into the finer points of trace evidence dynamics, activity-level propositions (see: Cook *et al.,* 1998; Section 2.1) cannot be addressed with rigour (Jackson and Biedermann, 2019).

> *"The quality of interpretation of the outcomes of the analysis depends upon the expert's knowledge and expertise in assigning robust probabilities for the transfer and persistence of [the trace evidence] in such circumstances. Of course, this depends on the quality and quantity of relevant data to inform those probabilities" (Jackson and Biedermann, 2019: 39).*

This resonates with the calls from the Forensic Science Regulator for more data to be produced to underpin the evaluative interpretation of traces (e.g. Forensic Science Regulator, 2019). This PhD research illustrates and adds weight to this call and clearly demonstrates that providing such relevant data which can uncover the unexpected, concerning, or contradictory dynamics that would be useful for interpretation remains a priority for research within the field.

For example, in Chapter 6 (see sections 6.4 and 6.5), it was demonstrated that the largest quartz grains which transferred to a swatch of fabric were lost first. Consequently, the caveat for interpretation is that if grain size analysis were conducted in practical forensic casework (e.g. as has been the case Pye and Croft, 2004; Pye and Blott, 2004, Chazottes *et al.,* 2004), one should bear this process in mind. The absence of larger quartz grains in a sample extracted from a fabric may be representative of absence at the point of collection, not absence at the point of transfer.

8.1.3. Complications

Accordingly, given this acknowledgment that empirical experimentation can improve the quality of forensic science (Forensic Science Regulator, 2019; Morgan, 2017; House of Lords, 2019), that the barriers include a currently inadequate funding regime (House of Lords, 2019), and that there is ambiguity over whose remit research and development falls into (Evison, 2018; Home Office, 2019) what is the prognosis for the future?

The recent House of Lords Science and Technology Select Committee Report (2019; "Forensic science and the criminal justice system: a blueprint for change") presented two concrete recommendations:

(1) Primarily, regarding the amount of funding

"We recommend that UK Research and Innovation urgently and substantially increase the amount of dedicated funding allocated to forensic science for both technological advances and foundational research," (House of Lords, 2019: 49)

(2) Secondly, concerning the remit

" We recommend the creation of a National Institute for Forensic Science within the UK Research and Innovation family, to set strategic priorities for forensic science research and development, and to coordinate and direct research and funding." (House of Lords, 2019: 50)

Should both of these recommendations be implemented, the prospects for empirical forensic research would appear much improved.

8.2 Part Two: The potential of borrowed techniques to accelerate data acquisition and analysis

8.2.1. Borrowed techniques can be useful

Another theme that emerged throughout this research was the idea that borrowing very basic advances from cognate digital disciplines such as image processing and computer science has the potential to greatly accelerate the speed at which data can be acquired and analysed, without compromising on accuracy or reliability (see: Kloster *et al.,* 2014; Grishagin, 2015; Levin *et al.,* 2019). In summary, the main findings were that:

(1) Acceleration and automation were possible

In Chapter 3, it was seen that thresholding operations, which are frequently used for the segmentation for fluorescence imagery within the context of histology (Oberholzer *et al.,* 1996; Bankhead, 2014) could be used to successfully quantify the amount of fluorescent powder present in an image. In Chapter 4, it was seen that mathematical processing in order to find the maxima in an image, frequently used to locate biological structures within micrographs (e.g. Meyer dos Santos *et al.,* 2010; White *et al.,* 2012; Courtoy *et al.,* 2015; Grishagin, 2015; Nguyen *et al.,* 2016), could also be used to successfully quantify the number of quartz grains present on a fabric swatch. In Chapter 5, macros were used to run repetitive analyses, and Chapter 7 used the tools of BIIGLE (see: Langenkämper *et al.,* 2017) and SHERPA (citations: Kloster *et al.,* 2014; 2017) were used to generate accurate morphometric measurements for large numbers of diatom valves.

(2) Added benefits beyond acceleration: transparency

Beyond the benefit of accelerated analyses, other advantages were seen. While utilising a macro to run repetitive processes on a batch of images (as in Chapters 3, 4, 5, and 6) does have the primary benefit of freeing the researcher or practitioner from tedious and time-consuming tasks, it also has benefits in terms of reproducibility. With a macro, especially a macro for an open-source

programme such as ImageJ (see: Schneider *et al.,* 2012; Schindelin *et al.,* 2015), it becomes possible for other researchers to replicate precisely the methodologies used (one of the key themes of the burgeoning 'open science' movement (Nosek *et al.,* 2015)). Additionally, if a user learns to read the simple mark-up associated with analyses, the macro becomes not only a tool for analysis, but a record of the processing operations conducted, and a resource for subsequent researchers to audit the precise methodology used, another tenet of the open science movement (ibid).

(3) Added benefits beyond acceleration: determinism

Another theme which emerged in these experiments was that where image processing operations involved a degree of observer-dependent, subjective decision making, then variance was introduced to the data. In Chapter 3, one saw that if one asked multiple participants to define the extent of the foreground in an image of fluorescent powder on a dark background by tracing, one received multiple different answers. In Chapter 4, it was seen that when asking a single examiner to count the number of quartz grains in an image, three times, the same number of grains was not counted each time. Even for processes for which there is a sort of positivist 'correct answer' in existence ('how many grains of sand are in this picture'), the involvement of a human examiner will introduce stochastic error. In contrast, where mathematical processes are used (as in Chapter 3, where an algorithm was used to define an appropriate threshold level to segment images of fluorescent powder), the process is deterministic; repetition will yield an identical answer.

It is worth emphasising that all of the techniques employed within the PhD are viewed as simple or routine in other disciplines. For example, slide scanning, as used in Chapter 7, is routine in pathology and histology (Preibisch *et al.,* 2009), but seen as relatively novel in the analysis of diatoms (e.g. Kloster *et al.,* 2014; 2017), and to date unexplored within studies of transfer and persistence. Likewise, thresholding using an algorithm is common in machine vision (e.g. Gonzales and Woods, 2010)

but novel when applied to fluorescent particle persistence studies for reconstruction in forensic science (e.g. Levin *et al.,* 2019).

8.2.2. Implications

This highlights the importance of ensuring knowledge is not siloed and that cognate disciplines are brought together. Part of the reason that forensic science are in crisis is the historical enthusiasm for borrowing of techniques, without an attendant enthusiasm for conducting first-principles research (see: Section 2.1; Jobling and Gill, 2004; Morgan and Bull, 2007). It is possible that the way out of this predicament is, in fact, more borrowing. If uncritical or hasty interdisciplinarity was the problem, considered, targeted interdisciplinarity may represent a solution.

8.2.3. The identity of forensic science

This ties in with quite an interesting debate within forensic science, about identity. (Morgan, forthcoming). The identity of the discipline has always been mutable – there are no more phrenologists, and possibly in the coming years we will see more of a focus on digital and 'cyber' evidence (House of Lords, 2019). What the findings of this PhD suggest is that collaboration with computer science departments for more traditional microscopic analyses can pay dividends.

8.3 Part Three: The Unknown Unknowns

8.3.1. Findings: Theme

While there were a number of concrete findings in this PhD (please see Chapter 9 for a condensed summary), one theme that emerged in every chapter was that there remain many avenues for future research. In summary, the key areas identified included:

(1) More nuanced hypotheses

In Chapter 5, it was identified that future research may want to consider the *redistribution* patterns of particles, rather than just their loss or retention.

This would be possible, utilising the tools developed over the course of this PhD.

(2) More sophisticated techniques

As mentioned above, in Section 8.2, all of the techniques employed within this PhD were relatively basic, and that more complex methodologies exist (with potential to be harnessed for forensic research). For example, in Chapter 3, it was identified that one could evaluate the performance of more sophisticated methods of image segmentation upon images of fluorescent powder.

(3) Greater realism

The experiments conducted within Chapters 5, 6, and 7 were all very artificial – with staged methods of transfer, and the removal of samples for imaging. Future work could consider pursuing more realistic modes of transfer, and the development of a method of imaging the swatches *in situ,* without the need to detach the samples.

(4) Greater abstraction

Conversely, future research may wish to pursue further abstraction. With respect to the size-selective hypotheses pursued in Chapters 6 and 7, a future experiment could use different, controlled, initial samples. For example: with quartz grains, one could start with known initial populations e.g. grains with known angularity and size. For diatoms, one may want to consider using microspheres, with known concentration.

(5) More ambitious sample sizes and durations

One critical limitation of the research presented here was its modest scale. For example, in Chapter 5, experiments occupied a short timescale; persistence was measured over only 8 hours. Future research could observe trace evidence dynamics over longer durations, and with measurements made at more frequent intervals.

9

Chapter Nine

Synoptic Conclusions

9.1.1 Summary of research themes and questions

In conclusion, this PhD was an interdisciplinary project combining theory from forensic science, techniques from geography, and insights from image processing. There were two overarching objectives: (1) To explore the potential of new (and newly-adapted) techniques to accelerate the collection and analysis of data relevant to the transfer and persistence of particulate environmental trace evidence, and; (2) To generate forensically-relevant datasets exploring the influence of variables upon the rates of transfer and persistence of environmental particles (specifically diatoms, quartz grains, and fluorescent powder as a proxy for other evidence types in the region of 15 µm diameter). These experiments were conducted with a forensically-relevant substrate; footwear material, on the upper of a shoe.

Objective One: Evaluating and new and existing methods

Chapters 3 and 4 considered the first objective, evaluating the performance of existing and new methods for the quantification of two types of particle used in forensically-relevant experiments: fluorescent powder, and quartz grains. The key research questions (as listed in Sections 3.1.4 and 4.1.4), considered:

- Fluorescent powder: evaluating existing techniques (1)
  Can the existing technique used for the segmentation of fluorescent powder of manual segmentation, tracing over the foreground, be considered an accurate, reliable, and efficient method of quantifying the amount of fluorescent powder present in a colour photograph? (Chapter 3)

- Fluorescent powder: evaluating existing techniques (2)
  Can the existing technique used for the segmentation of fluorescent powder of global thresholding, manually defining the level of the threshold, be considered an accurate, reliable, and efficient method of quantifying the amount of fluorescent powder present in a colour photograph? (Chapter 3)

- Fluorescent powder: evaluating a new technique
  Could a thresholding algorithm (i.e. an algorithm which mathematically calculates an appropriate threshold value) be a suitable alternative, offering comparable (or improved) levels of accuracy, reliability, and efficiency in quantifying the amount of fluorescent powder present in a colour photograph? (Chapter 3)

- Quartz grains: evaluating the accuracy of a new technique
  Does using 'find maxima' to locate quartz grains result in the same number of quartz grains being identified as a human examiner manually counting the grains, from a photograph? (Chapter 4)

- Quartz grains: evaluating the efficiency of a new technique
  Exploring the potential for full automation, is there one value for noise tolerance (the noise filter) for a find maxima function which is appropriate for a batch of images, enabling rapid batch processing, or is it necessary to define an appropriate noise filter level for each image individually when trying to quantify the number of quartz grains present in a photograph? (Chapter 4)

- Quartz grains: evaluating the reproducibility of a new technique
  Will multiple examiners, or one examiner using the technique repeatedly, arrive at the same answer for the number of quartz grains present in an image when using a find maxima function? (Chapter 4)

Objective Two: Generating datasets

The remaining questions pertained to trace evidence dynamics and the variables that could influence the transfer and persistence of particles; Chapter 5 considered the influence of variables pertaining to the recipient material surface, while Chapters 6 and 7 considered the characteristics of the particulates. The research questions considered:

- Wearer activity

  Does the level of activity of the shoe's wearer (walking for 0, 5, 10, or 15 minutes per hour) affect the persistence of particulate evidence, represented by quartz grains and fluorescent powder?
  (Chapter 5)

- Position on shoe

  Does the sample position (on the heel, toe, inner, or outer of the shoe) affect the persistence of particulate evidence, represented by quartz grains and fluorescent powder?
  (Chapter 5)

- Particle Size

  Does the size (as measured by cross-sectional area) of quartz grains and diatoms affect their transfer and/or persistence?
  (Chapters 6 and 7)

- Particle Morphology

  Does the morphology (elongation as measured by the aspect ratio, and the roughness as measured by PCAF for diatoms and solidity for quartz grains), affect their transfer and/or persistence?
  (Chapters 6 and 7)

9.1.2 Summary of findings

This PhD sought to generate datasets and methods for the investigation of environmental or "geoforensic" forms of particulate trace evidence. The key findings can be summarised as:

1. Forensic science has a crisis of data

   The heritage of forensic science, as a synthetic discipline with boundaries porous to advances made in cognate disciplines, has resulted in an unusual situation. Forensic science is simultaneously a vibrant discipline, in which researchers and practitioners have access to world-class, cutting-edge scientific techniques, and a discipline where a vast tranche of foundational research is missing (Section 2.1; Jobling and Gill, 2004; Forensic Science Regulator, 2019; House of Lords, 2019).

2. Segmenting fluorescent proxies

   Experiments suggest that an appropriate thresholding algorithm (such as Yen, MaxEntropy, or RenyiEntropy) may be capable of segmenting an image of fluorescent powder with similar accuracy to a human manually undertaking the task by tracing the pixels which represent the foreground of the image, but far faster, and with a guarantee that the same answer will be arrived at each time analysis is undertaken (Section 3.3; Table 9.1; Levin *et al.,* 2019).

3. Locating and counting quartz grains

   Experiments suggest that the 'find maxima' function in ImageJ can locate and count the number of quartz grains in a digital image taken with a dark fabric background, when variation in the brightness level of the background would render a thresholding function a poor reflection of the actual extent of the foreground in the image, in a similar manner to how 'find maxima' has been used to identify biological structures of interest within micrographs (Section 4.3; Table 9.2; Grishagin, 2015).

Chapter	Research Question	Finding
3	Can manual segmentation (tracing) be considered an accurate, reliable, and efficient way of segmenting fluorescent powder from a colour photograph?	Manual segmentation can be considered accurate, but inefficient, with some questions around the repeatability, with statistically significant differences between different examiners' estimates of the foreground (see: 3.3.1).
	Can manual thresholding be considered an accurate, reliable, and efficient way of segmenting fluorescent powder from a colour photograph?	Manual thresholding is much more efficient than manual segmentation (tracing) but experiences similar concerns around the repeatability of results (see: 3.3.2)
	Can a thresholding algorithm be considered an accurate, reliable, and efficient way of segmenting fluorescent powder from a colour photograph?	Global thresholding algorithms operate reproducibly, efficiently, and with varying levels of accuracy. The three most accurate algorithms (with performance similar to manual segmentation) were: Yen, MaxEntropy, and RenyiEntropy (see: 3.3.3)

Table 9.1: Summary of research questions and findings for Chapter 3

Chapter	Research Question	Finding
4	Does using 'find maxima' to locate quartz grains result in the same number of quartz grains being identified as a human examiner manually counting the grains?	Yes; the find maxima function is consistently capable of delivering the same answer as a human examiner for the number of quartz grains present in a photograph. (see: 4.3.1)
	Is there one value for noise tolerance (the noise filter) for a find maxima function which is appropriate for a batch of images, enabling rapid batch processing,	No; it is necessary to define an appropriate noise filter level for each image individually. (see: 4.3.2)
	Will multiple examiners, or one examiner using the technique repeatedly, arrive at the same answer for the number of quartz grains present in an image when using a find maxima function?	Yes; statistically significant differences were not seen between examiners' results, or when one examiner performed the function repeatedly. (see: 4.3.3)

Table 9.2: Summary of research questions and findings for Chapter 4

4. Examining variables that might affect persistence

   Experiments suggested that the position on the shoe from which a sample was taken (heel, toe, inner or outer) did not statistically significantly affect the percentage persistence of either quartz grains or fluorescent powder – so taking samples from a shoe upper at multiple locations is a valid proxy for taking samples from one location over time (which may not be possible, if a sampling technique is destructive). Experiments also suggested that minor variations in duration of walking per hour (5, 10, or 15 minutes) did not consistently affect the percentage persistence, implying that minor variation could be permitted in future experimental design, and confirming that it is necessary to undertake empirical experiments to explore trace evidence dynamics (Section 5.3; Table 9.3; Forensic Science Regulator, 2019; Jackson and Biedermann, 2019).

5. The morphology of quartz and transfer and persistence

   Experiments suggested that the largest quartz grains in a population don't seem to transfer readily to footwear fabric, and the largest particles that did transfer persisted over shorter timescales than the smaller particles, in accordance with previous studies of particulate evidence (Section 6.3; Table 9.4; Brewster *et al.,* 1984; Hicks *et al.,* 1996; Pounds and Smalldon, 1975b; Levin *et al.,* 2017). Over time, the median solidity of particles in the samples declined (implying a higher degree of surface roughness in later samples), and median circularity increased (implying that very elongated particles do not persist). Together, these findings suggested that the morphology of a quartz grain can affect both its transfer and its persistence (as in e.g. Houck, 2001).

Chapter	Research Question	Finding
5	Does the level of activity of the shoe's wearer (walking for 0, 5, 10, or 15 minutes per hour) affect the persistence of particulate evidence, represented by quartz grains and fluorescent powder?	No; For quartz, no statistically significant differences were seen when comparing the effects of sample position. For fluorescent powder, a statistically significant difference was not consistently seen. (see: 5.3.1 and 5.3.3)
	Does the sample position (on the heel, toe, inner, or outer of the shoe) affect the persistence of particulate evidence, represented by quartz grains and fluorescent powder?	No; A statistically significant difference between the three activity levels of 5, 10, or 15 minutes of walking per hour was not consistently observed for either quartz grains or for fluorescent powder. (see: 5.3.2 and 5.3.4)

Table 9.3: Summary of research questions and findings for Chapter 5

Chapter	Research Question	Finding
6	Does the size (as measured by cross-sectional area) of quartz grains affect their transfer and/or persistence?	Yes; The largest quartz grains in the population did not transfer, and the largest quartz grains that did transfer were lost first. (see: 6.3.2 and 6.3.3)
	Does the morphology (elongation as measured by the aspect ratio, and the roughness as measured by solidity affect their transfer and/or persistence?	Yes; quartz grains with a higher degree of roughness seemed to transfer preferentially, and rougher quartz grains persisted over longer timescales. Particles with a higher degree of circularity seemed more likely to transfer and to persist. (see: 6.3.4 and 6.3.5)

Table 9.4: Summary of research questions and findings for Chapter 6

6. The morphology of diatoms and transfer and persistence

   Experiments involving diatoms suggested that all size fractions of diatom found in the population transferred readily to the cotton used as a recipient surface, but that the median particle aspect ratio and concavity were seen to differ between the population and the transferred samples, and over time (Table 9.5). Together, these findings supported the theory that the morphology of a particle can influence its persistence (as in Houck, 2001; Scott *et al.,* 2019), but diverged from a previous study of diatom persistence, which suggested that the maximum particle size observed may decline over time (Levin *et al.,* 2017).

Chapter	Research Question	Finding
7	Does the size (as measured by cross-sectional area) of diatom valves affect their transfer and/or persistence?	Transfer was not affected; all size fractions of diatom found in the population transferred readily to the cotton used as a recipient surface. Persistence was affected; the mean and median particle areas exhibited a downward trend over time. (see: 7.3.2 and 7.3.3)
	Does the morphology (elongation as measured by the aspect ratio, and the roughness as measured by PCAF affect their transfer and/or persistence?	Yes; Both the median width-height ratio and PCAF were seen to be statistically significantly different between the source and transferred samples, and to vary statistically significantly over time. (see: 7.3.4 and 7.3.5)

Table 9.5: Summary of research questions and findings for Chapter 7

7. The interdisciplinary problem might have an interdisciplinary solution

   Through the discussion of the experiments in this dissertation, a recurring theme was that insights and techniques from other disciplines (such as image processing and computer science) can assist in generating the empirical datasets which forensic science so badly needs by accelerating data acquisition and analysis, and automating repetitive and time-consuming

tasks. Accordingly, in a circular manner, the solution to the problem of ad-hoc interdisciplinarity may be conscientious interdisciplinarity.

8. The continued need for empirical research to underpin inferences

   A second theme which emerged from the experiments was the ongoing importance of empirical research to underpin the interpretations made within forensic casework – the impetus for this PhD (Forensic Science Regulator, 2015; 2016; 2018; 2019) but also a conclusion reached in the recent House of Lords Science and Technology Select Committee hearings (House of Lords, 2019).

9. The continued potential for research in the field

   A final conclusion in every data chapter (Chapters 3, 4, 5, 6, and 7) was that there remained unexplored areas with respect to trace evidence dynamics. If, as was recommended by the recent House of Lords Science and Technology Select Committee (House of Lords, 2019), the amount of funding made available for forensic research in the United Kingdom is increased, it is hoped that the techniques developed in this thesis may assist in the acceleration of data acquisition and analysis for foundational research.

9.1.3 Summary of implications

Tools and methods

It is hoped that that the tools and methods tested and validated within this thesis can be used productively by researcher and practitioners (Please see Tables 9.1 and 9.2 for a summary of the main methodological findings).

Considering the segmentation of fluorescent powder from digital photographs, Chapter 3 offers suggestions for appropriate thresholding algorithms, the appendix contains relevant macros for use in ImageJ, and the publication (Levin *et al.,* 2019) offers a validation study for the method, presenting findings that this automated method of segmenting fluorescent powder is as accurate, and more reliable, than a human examiner. Considering the quantification of quartz grains, this thesis proposed the 'find maxima' function as a reliable, efficient, and accurate alternative to time-consuming manual counting, and also, in Chapter 6, outlined how accessible freeware can be used to extract precise, quantitative morphometric information about quartz grains from photographs or micrographs.

These methods of quantification could be used in further, future forensic studies, and this will create significant opportunities to form datasets that researchers have not previously been able to produce (hampered by time constraints). This may contribute to the potential for ensuring that reconstruction practices are evidence-based, and that inferences can be more transparent (because the data exists to demonstrate how the inference that has been achieved can be created). These quantification techniques may also be adapted for analogous cases (for example, the rapid counting and morphological characterisation of quartz grains from photographs or micrographs may be useful for researchers within Geography).

It is hoped that the inclusion of SHERPA and BIIGLE may introduce these tools to a forensic audience who might not have encountered these approaches previously (as they are predominantly used in marine science and (palaeo)ecology).

Datasets

The datasets and findings generated over the course of this PhD (please see especially Tables 9.3 – 9.5 for a summary) are relevant to the interpretation of environmental traces within forensic reconstruction scenarios.

These findings include the suggestion that minor variation in the activity of a wearer, and the position of the sample on a shoe upper, may not influence the percentage persistence, but that particle size and morphology can have observable effects on the transfer and persistence of environmental trace evidence types. Accordingly, if one is attempting to collect environmental particulate trace evidence for forensic reconstruction, persistence trends are consistent regardless of the spatial distribution of the trace on the footwear (so a practitioner could interpret the presence of a trace collected from any location on the shoe upper without having to consider any spatially-specific persistence trend), but the morphology of the trace does impact that trend (so it is really important that the morphology of particles is considered when reconstructing events).

It is hoped that there would be two audiences for the findings of these studies; those within academia, who might undertake further research in the field (further enabling the nuanced and evidence-based interpretation of forensically-relevant traces), and those outside of academia, in the world of forensic casework, who might be able to use the findings of these experiments to underpin inferences made in crime reconstructions. That is to say, it is possible to produce viable datasets to act as an empirical evidence base which increases the transparency and reproducibility of the inferences being made in reconstruction approaches.

References

A

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International, 11(7),* 36-42.

Adu, M. O., Wiesel, L., Bennett, M. J., Broadley, M. R., White, P. J., & Dupuy, L. X. (2015). Scanner-based time-lapse root phenotyping. *Bio-protocol*, *5*(6), e1424.

Ago, K., Hayashi, T., Ago, M., & Ogata, M. (2011). The number of diatoms recovered from the lungs and other organs in drowning deaths in bathwater. *Legal Medicine, 13(4),* 186-190.

Akulova, V., Vasiliauskiené, D., & Talaliené, D. (2002). Further insights into the persistence of transferred fibres on outdoor clothes. *Science & Justice, 42(3),* 165-171.

Allen, T. J., & Scranage, J. K. (1998). The transfer of glass? Part 1: Transfer of glass to individuals at different distances. *Forensic Science International, 93(2?3),* 167-174.

Ali, Z., & Bhaskar, S. B. (2016). Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia, 60(9),* 662.

Andlauer, T. F., & Sigrist, S. J. (2012). Quantitative analysis of Drosophila larval neuromuscular junction morphology. *Cold Spring Harbor Protocols*, *2012*(4), pdb-prot068601.

Arena, E. T., Rueden, C. T., Hiner, M. C., Wang, S., Yuan, M., & Eliceiri, K. W. (2017). Quantitating the cell: Turning images into numbers with ImageJ. *Wiley Interdisciplinary Reviews: Developmental Biology, 6(2).* e260.

Ardiel, E.L., Kumar, A., Marbach, J., Christensen, R., Gupta, R., Duncan, W., Daniels, J.S., Stuurman, N., Colón-Ramos, D. and Shroff, H., 2017. Visualizing calcium flux in freely moving nematode embryos. *Biophysical journal*, *112*(9), pp.1975-1983.

Arndt, J., Bell, S., Crookshanks, L., Lovejoy, M., Oleska, C., Tulley, T., & Wolfe, D. (2012). Preliminary evaluation of the persistence of organic gunshot residue. *Forensic Science International, 222(1),* 137-145.

Arthur, R. M., Humburg, P. J., Hoogenboom, J., Baiker, M., Taylor, M. C., & de Bruin, K. G. (2017). An image-processing methodology for extracting bloodstain pattern features. *Forensic Science International. 277,* 122-132.

Ashcroft, C. M., Evans, S., & Tebbett, I. R. (1988). The persistence of fibres in head hair. *Journal of the Forensic Science Society, 28(5-6),* 289-293.

B

Balding, D. J., & Steele, C. D. (2015). Weight-of-evidence for forensic DNA profiles. Chichester: John Wiley & Sons.

Bankhead, P. (2014) Analyzing Fluorescence Microscopy Images with Image J. Belfast: Queen's University, Belfast. Retrieved from: https://sydney.edu.au/medicine/bosch/ facilities/advanced-microscopy/user-support/ImageJ_FL_Image_Analysis.pdf

Barash, M., Reshef, A., & Brauner, P. (2010). The use of adhesive tape for recovery of DNA from crime scene items. *Journal of Forensic Sciences, 55(4),* 1058-1064.

Barber, H. G., & Haworth, E. Y. (1994). A guide to the morphology of the diatom frustule: with a key to the British freshwater genera. Freshwater Biological Association.

Battarbee, R., Jones, V., Flower, R., Cameron, N., Bennion, H., Carvalho, L., & Juggins, S. (2001). Diatoms. In 'tracking environmental change using lake sediments. Volume 3: Terrestrial, algal, and siliceous indicators'.(eds Smol, J.P, Birks, H.J., & Last., W.M) pp. 155–202. Dordrecht, Netherlands: Kluwer Academic Publishers.

Baveye, P. (2002). Comment on "evaluation of biofilm image thresholding methods". *Water Research, 36(3),* 805-806.

Baveye, P. C., Laba, M., Otten, W., Bouckaert, L., Sterpaio, P. D., Goswami, R. R., & Liu, J. (2010). Observer-dependent variability of the thresholding step in the quantitative analysis of soil images and x-ray microtomography data. *Geoderma, 157(1),* 51-63.

Beddow, J. K., & Vetter, A. F. (1977). A note on the use of classifiers in morphological analysis of particulates. *Journal of Powder and Bulk Solids Technology, 1,* 42-44.

Benjamin, D. J., & Berger, J. O. (2019). Three Recommendations for Improving the Use of p-Values. *The American Statistician 73(sup1),* 186-191.

Benton, M., Chua, M. J., Gu, F., Rowell, F., & Ma, J. (2010). Environmental nicotine contamination in latent fingermarks from smoker contacts and passive smoking. *Forensic Science International*, *200*(1-3), 28-34.

Bergslien, E. (2012). An Introduction to Forensic Geoscience. Chichester: John Wiley & Sons.

Bernsen, J. (1986). Dynamic thresholding of grey-level images. Paper presented at the Proc. 8th Int. Conf. on Pattern Recognition, 1986.

Boehme, A., Brooks, E., McNaught, I., & Robertson, J. (2009). The persistence of animal hairs in a forensic context. *Australian Journal of Forensic Sciences, 41(2),* 99-112.

Bracken, L. J., & Oughton, E. A. (2006). 'What do you mean?' The importance of language in developing interdisciplinary research. *Transactions of the Institute of British Geographers, 31(3),* 371-382.

Brady, K., Stilley, B., Olds, M., O'neill, T., Egan, J., & Durnal, E. (2017). Tape lift sampling of chemical threat agents. *Journal of forensic sciences*, *62*(4), 1015-1021.

Brayley-Morris, H., Sorrell, A., Revoir, A. P., Meakin, G. E., Court, D. S., & Morgan, R. M. (2015). Persistence of DNA from laundered semen stains: Implications for child sex trafficking cases. *Forensic Science International: Genetics, 19*, 165-171.

Brewster, F., Thorpe, J. W., Gettinby, G., & Caddy, B. (1985). The retention of glass particles on woven fabrics. *Journal of Forensic Science, 30(3),* 798-805.

Brożek-Mucha, Z. (2011). Chemical and morphological study of gunshot residue persisting on the shooter by means of scanning electron microscopy and energy dispersive x-ray spectrometry. *Microscopy and Microanalysis, 17(06),* 972-982.

Bui, E., Mazzullo, J., & Wilding, L. (1989). Using quartz grain size and shape analysis to distinguish between aeolian and fluvial deposits in the Dallol Bosso of Niger (West Africa). *Earth Surface Processes and Landforms, 14(2),* 157-166.

Bull, P., Magee, A., & Whalley, W. B. (1986). An annotated bibliography of environmental reconstruction by SEM, 1962-1985. London: British Geomorphological Research Group.

Bull, P., & Morgan, R. (2006). Sediment fingerprints: A forensic technique using quartz sand grains. *Science & Justice, 46(2),* 107-124.

Bull, P. A., & Morgan, R. M. (2007). Sediment fingerprints: A forensic technique using quartz sand grains ? A response. *Science & Justice, 47(3),* 141-144.

Bull, P. A., Morgan, R. M., & Freudiger-Bonzon, J. (2008). A critique of the present use of some geochemical techniques in geoforensic analysis. *Forensic Science International, 178(2?3),* e35-e40.

Bull, P. A., Morgan, R. M., Sagovsky, A., & Hughes, G. J. A. (2006). The transfer and persistence of trace particulates: Experimental studies using clothing fabrics. *Science & Justice, 46(3),* 185-195.

Bull, P. A., Parker, A., & Morgan, R. M. (2006). The forensic analysis of soils and sediment taken from the cast of a footprint. *Forensic Science International, 162(1?3),* 6-12.

C

Cairns, J., Dickson, K., Pryfogle, P., Almeida, S., Case, S., Fournier, J., & Fujii, H. (1979). Determining the accuracy of coherent optical identification of diatoms. *JAWRA Journal of the American Water Resources Association, 15(6),* 1770-1775.

Cairns, J., Dickson, K., & Slocomb, J. (1977). The abc's of diatom identification using laser holography. *Hydrobiologia, 54(1),* 7-16.

Cairns Jr, J., Almeida, S. P., & Fujii, H. (1982). Automated identification of diatoms. *BioScience, 32(2),* 98-102.

Cameron, N. (2004). The use of diatom analysis in forensic geoscience. Geological Society, London, Special Publications, 232(1), 277-280.

Cameron, N., & Peabody, A. (2010). Forensic science and diatoms. In J. Stoermer & E. Smol (Eds.), The diatoms: Applications for the environmental and earth sciences. Cambridge: CUP.

Casamatta, D. A., & Verb, R. G. (2000). Algal colonization of submerged carcasses in a mid-order woodland stream. *Journal of Forensic Science, 45(6),* 1280-1285.

Charpentier, I., Sarocchi, D., & Sedano, L. A. R. (2013). Particle shape analysis of volcanic clast samples with the MATLAB tool MORPHEO. *Computers & Geosciences, 51,* 172-181.

Charpentier, I., Staszyc, A. B., Wellner, J. S., & Alejandro, V. (2017). Quantifying grain shape with MORPHEOLV: A case study using holocene glacial marine sediments. Paper presented at the EPJ Web of Conferences.

Chazottes, V., Brocard, C., & Peyrot, B. (2004). Particle size analysis of soils under simulated scene of crime conditions: The interest of multivariate analyses. *Forensic Science International, 140(2),* 159-166.

Cheshire, K., Morgan, R., & Holmes, J. (2017). The potential for geochemical discrimination of single-and mixed-source soil samples from close proximity urban parkland locations. *Australian Journal of Forensic Sciences, 49(2),* 161-174.

Chief Scientific Advisor, Government Office for Science (2015). Annual report of the government chief scientific adviser 2015 - Forensic science and beyond: Authenticity, provenance and assurance. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/506461/gs-15-37a-forensic-science-beyond-report.pdf

Chisum, W. J., & Turvey, B. (2000). Evidence dynamics: Locard's exchange principle & crime reconstruction. *Journal of Behavioral Profiling,* 1(1), 1-15.

Cole, S. A. (2006). "Implicit testing": Can casework validate forensic techniques? *Jurimetrics,* 117-128.

Cook, R., Evett, I. W., Jackson, G., Jones, P., & Lambert, J. (1998). A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice, 38(4),* 231-239.

Costa, P., Andrade, C., Freitas, M. d. C., Oliveira, M., Lopes, V., Dawson, A., & Jouanneau, J. (2012). A tsunami record in the sedimentary archive of the central Algarve coast, Portugal: Characterizing sediment, reconstructing sources and inundation paths. *The Holocene, 22(8),* 899-914.

Courtoy, G. E., Donnez, J., Marbaix, E., & Dolmans, M. M. (2015). In vivo mechanisms of uterine myoma volume reduction with ulipristal acetate treatment. *Fertility and Sterility, 104*(2), 426-434.

Cox, E. J. (2012). Diatoms and forensic science. In Márquez-Grant, N., & Roberts, J. (Eds.). *Forensic Ecology Handbook: From Crime Scene To Court* (pp. 141-151). Chichester: John Wiley & Sons.

Cox, M. R., & Budhu, M. (2008). A practical approach to grain shape quantification. *Engineering Geology,96(1),* 1-16.

Coyle, B. (2010). Trace and contact evidence. In P. White (Ed.), Crime scene to court fourth edition: The essentials of forensic science (4th Ed. ed., pp. 106-127): London: Royal Society of Chemistry.

Crum, W. R., Camara, O., & Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging, 25(11),* 1451-1461.

Culver, S., Bull, P., Campbell, S., Shakesby, R., & Whalley, W. (1983). Environmental discrimination based on quartz grain surface textures: A statistical investigation. *Sedimentology, 30(1),* 129-136.

Curry, A., Porter, P., Irvine-Fynn, T., Rees, G., Sands, T., & Puttick, J. (2009). Quantitative particle size, microtextural and outline shape analyses of glacigenic sediment reworked by paraglacial debris flows. *Earth Surface Processes and Landforms, 34(1),* 48-62.

D

Dachs, J., McNaught, I. J., & Robertson, J. (2003). The persistence of human scalp hair on clothing fabrics. *Forensic Science International, 138(1-3),* 27-36.

Dahiru, T. (2008). P-value, a true test of statistical significance? A cautionary note. *Annals of Ibadan postgraduate medicine,* 6(1), 21-26.

Daly, D. J., Murphy, C., & McDermott, S. D. (2012). The transfer of touch DNA from hands to glass, fabric and wood. *Forensic Science International: Genetics, 6(1),* 41-46.

DeBattista, R., Tidy, H., Thompson, T. J. U., & Robertson, P. (2014). An investigation into the persistence of textile fibres on buried carcasses. *Science & Justice, 54(4),* 288-291.

De Wael, K., Gason, F. G., & Baes, C. A. (2008). Selection of an adhesive tape suitable for forensic fiber sampling. *Journal of Forensic Sciences, 53(1),* 168-171.

Dickson, G. C., Poulter, R. T. M., Maas, E. W., Probert, P. K., & Kieser, J. A. (2011). Marine bacterial succession as a potential indicator of postmortem submersion interval. *Forensic Science International, 209(1–3),* 1-10.

Dockery, C. R., & Goode, S. R. (2003). Laser-induced breakdown spectroscopy for the detection of gunshot residues on the hands of a shooter. *Applied Optics, 42(30),* 6153-6158.

Dowdy, S., Wearden, S., & Chilko, D. (2011). Statistics for research (Vol. 512). Chichester: John Wiley & Sons.

Doyle, W. (1962). Operations useful for similarity-invariant pattern recognition. *Journal of the ACM (JACM),* 9(2), 259-267.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International, 156(1),* 74-78.

Du Buf, H., Bayer, M., Droop, S., Head, R., Juggins, S., Fischer, S., & Pech-Pacheco, J. (1999). Diatom identification: A double challenge called ADIAC. Paper presented at the Image Analysis and Processing, 1999 International Conference.

Du Buf, H., & Bayer, M. M. (2002). Automatic Diatom Identification. London: World Scientific.

E

Edgeworth, F. Y. (1885), Methods of Statistics, *Journal of the Statistical Society of London, Jubilee Volume*, 181–217.

Edmond, G., & Martire, K. A. (2017). Antipodean forensics: A comment on ANZFSS's response to PCAST. *Australian Journal of Forensic Sciences, 50*(2), 140-151.

Edwards, A., Civitello, A., Hammond, H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics, 49(4),* 746.

Ehrlich, R., & Weinberg, B. (1970). An exact method for characterization of grain shape. *Journal of Sedimentary Research, 40(1),* 205-212.

Evison, M. P. (2018). Forensic science policy and the question of governmental University research quality assessment. *Forensic science international*, *290*, 279-296.

F

Fernandez-Gallego, J. A., Kefauver, S. C., Gutiérrez, N. A., Nieto-Taladriz, M. T., & Araus, J. L. (2018). Automatic wheat ear counting in-field conditions: simulation and implication of lower resolution images. *Proceedings of the Society of Photo-Optical Instrumentation Engineers 10783.*, p.107830M

Field, A., & Hole, G. (2002). How to design and report experiments. London: Sage.
Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.

Forensic Science Regulator (2015). Forensic Science Regulator Annual Report November 2014 – November 2015. London: Forensic Science Regulator. https://www.gov.uk/government/publications/forensic-science-regulator-annual-report-2015

Forensic Science Regulator (2016). Forensic Science Regulator Annual Report November 2015 – November 2016. London: Forensic Science Regulator. https://www.gov.uk/government/publications/forensic-science-regulator-annual-report-2016

Forensic Science Regulator (2018). Forensic Science Regulator Annual Report

November 2016 – November 2017. London: Forensic Science Regulator. https://www.gov.uk/government/publications/forensic-science-regulator-annual-report-2017

Forensic Science Regulator (2019). Forensic Science Regulator Annual Report 17 November 2017 – 16 November 2018. London: Forensic Science Regulator. https://www.gov.uk/government/publications/forensic-science-regulator-annual-report-2018

French, J., & Morgan, R. (2015). An experimental investigation of the indirect transfer and deposition of gunshot residue: Further studies carried out with SEM–EDX analysis. *Forensic Science International, 247,* 14-17.

French, J., Morgan, R., & Davy, J. (2014). The secondary transfer of gunshot residue: An experimental investigation carried out with SEM-EDX analysis. *X-Ray Spectrometry, 43(1),* 56-61.

French, J. C., Morgan, R. M., Baxendell, P., & Bull, P. A. (2012). Multiple transfers of particulates and their dissemination within contact networks. *Science & Justice, 52(1),* 33-41.

Fucci, N., Pascali, V. L., Puccinelli, C., Marcheggiani, S., Mancini, L., & Marchetti, D. (2015). Evaluation of two methods for the use of diatoms in drowning cases. *Forensic Science, Medicine, and Pathology, 11(4),* 601-605.

G

Gardenier, J., & Resnik, D. (2002). The misuse of statistics: concepts, tools, and a research agenda. *Accountability in Research: Policies and Quality Assurance, 9(2),* 65-74.

Gayle, T. (2004). Something in the water. *Law and Order, (June),* 92-93.

Gherghel, S., Morgan, R. M., Blackman, C. S., Karu, K., & Parkin, I. P. (2016). Analysis of transferred fragrance and its forensic implications. *Science & Justice, 56(6),* 413-420.

Gill, P., Jeffreys, A. J., & Werrett, D. J. (1985). Forensic application of DNA 'fingerprints'. *Nature, 318(6046),* 577-579.

Glasbey, C. A. (1993). An analysis of histogram-based thresholding algorithms. *CVGIP: Graphical models and image processing, 55(6),* 532-537.

Goldacre, B. (2010). Bad science: Quacks, hacks, and big pharma flacks. London: McClelland & Stewart.

Goleb, J., & Midkiff, C. (1975). Firearms discharge residue sample collection techniques. *Journal of Forensic Sciences, 20*(4), 701-7.

Gonzalez, R., & Woods, R. (2010). Digital Image Processing (3rd Edition ed.). Upper Saddle River NJ: Pearson Prentice Hall.

Gonzalez-Rodriguez, J., & Baron, M. G. (2018). Forensic Science in UK. Part I: Historical Development and Current Status. *Forensic Science Review, 30*(1), 2-5.

Grieve, M. C., Dunlop, J., & Haddock, P. S. (1989). Transfer experiments with acrylic fibres. *Forensic Science International, 40(3),* 267-277.

Griffin, L. D., Elangovan, P., Mundell, A., & Hezel, D. C. (2012). Improved segmentation of meteorite micro-CT images using local histograms. *Computers & Geosciences, 39,* 129-134.

Griffin, L. D., Lillholm, M., Crosier, M., & van Sande, J. (2009). Basic image features (bifs) arising from approximate symmetry type. In *International Conference on Scale Space and Variational Methods in Computer Vision* (pp. 343-355). Springer, Berlin, Heidelberg.

Grima, M., Hanson, R., & Tidy, H. (2014). An assessment of firework particle persistence on the hands and related police force practices in relation to GSR evidence. *Forensic Science International, 239,* 19-26.

Grishagin, I. V. (2015). Automatic cell counting with ImageJ. *Analytical biochemistry*, *473*, 63-65.

H

Haefner, J. N., Wallace, J. R., & Merritt, R. W. (2004). Pig decomposition in lotic aquatic systems: The potential use of algal growth in establishing a postmortem submersion interval (PMSI). *Journal of Forensic Science, 49(2),* 1-7.

Hand, D. J. (2008). Statistics: a very short introduction (Vol. 196). Oxford: Oxford University Press.

Hicks, T., Vanina, R., & Margot, P. (1996). Transfer and persistence of glass fragments on garments. *Science & Justice, 36(2),* 101-107.

Higgs, N. D., & Pokines, J. T. (2014). Marine environmental alterations to bone. In Pokines, J. & Symes, S. (Eds.) *Manual of Forensic Taphonomy*, 143-179. Boca Raton: CRC Press.

Hong, S., Han, A., Kim, S., Son, D., & Min, H. (2014). Transfer of fibres on the hands of living subjects and their persistence during hand washing. *Science & Justice, 54(6),* 451-458.

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P. J., Bunke, H., Goldgof, D. B., & Fisher, R. B. (1996). An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18(7), 673-689.

Horton, B. P., Boreham, S., & Hillier, C. (2006). The development and application of a diatom-based quantitative reconstruction technique in forensic science. *Journal Of Forensic Sciences, 51(3),* 643-650.

Houck, M. M. (2001). Mute witnesses: Trace evidence analysis. Cambridge: Academic Press.

House of Lords, Science and Technology Committee (2019). Forensic science and the criminal justice system: a blueprint for change. 3rd Report of session 2017-2019 HL Paper 333 (2019) available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/333.pdf

Howarth, J., Coulson, S., & Newton, A. (2009). Simulating transfer and persistence of a chemical marker powder for *Lycopodium clavatum* spores. *Forensic Science International, 192(1-3),* 72-77.

Hubbard, R. (2016), Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science, Thousand Oaks, CA: Sage.

Hurlbert, S., Levine, R., and Utts, J. (2019), Coup de Grâce for a Tough Old Bull: 'Statistically Significant' Expires, *The American Statistician, 73(sup1),* 352-357.

I

Iglesias, T., Cala, V., & Gonzalez, J. (1997). Mineralogical and chemical modifications in soils affected by a forest fire in the Mediterranean area. *Science of The Total Environment, 204(1),* 89-96.

Igathinathane, C., Pordesimo, L. O., Columbus, E. P., Batchelor, W. D., & Sokhansanj, S. (2009). Sieveless particle size distribution analysis of particulate materials through computer vision. *Computers and Electronics in Agriculture, 66(2),* 147-158.

Inman, K., & Rudin, N. (2002). The origin of evidence. *Forensic Science International, 126(1),* 11-16.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2(8),* e124.

Ioannidis, J. P. (2019). What Have We (Not) Learnt from Millions of Scientific Papers with P Values?. *The American Statistician*, *73*(sup1), 20-25.

J

Jackson, G., & Biedermann, A. (2019). "Source" or "activity" What is the level of issue in a criminal trial? *Significance*, *16*(2), 36-39.

Jalanti, T., Henchoz, P., Gallusser, A., & Bonfanti, M. (1999). The persistence of gunshot residue on shooters' hands. *Science & Justice, 39(1),* 48-52.

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics, 5(10),* 739-751.

Jones, V. (2007). Diatom introduction. In Elias, S. (Eds). *Encyclopedia of Quaternary Science* (pp. 476-484). Oxford: Elsevier.

K

Kakizaki, E., Kozawa, S., Matsuda, H., Muraoka, E., Uchiyama, T., Sakai, M., & Yukawa, N. (2010). Freshwater bacterioplankton cultured from liver, kidney and lungs of a decomposed cadaver retrieved from a sandy seashore: Possibility of drowning in a river and then floating out to sea. *Legal Medicine, 12(4),* 195-199.

Kalińska-Nartiša, E., Woronko, B., & Ning, W. (2017). Microtextural inheritance on quartz sand grains from Pleistocene periglacial environments of the Mazovian Lowland, central Poland. *Permafrost and Periglacial Processes, 28(4),* 741-756.

Kapur, J. N., Sahoo, P. K., & Wong, A. K. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing, 29(3),* 273-285.

Kaye, D. H. (2010). The good, the bad, the ugly: The NAS report on strengthening forensic science in America. *Science & Justice, 50(1),* 8-11.

Kelly, M. (2000). Identification of common benthic diatoms in rivers. Shrewsbury: Field Studies Council.

Kennedy, S., & Mazzullo, J. (1991). Image analysis method of grain size measurement. In J. Syvitski (Ed.), *Principles, Methods and Application of Particle Size Analysis* (pp. 76-87). Cambridge: Cambridge University Press.

Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition, 19(1),* 41-47.

Kloster, M., Esper, O., Kauer, G., & Beszteri, B. (2017). Large-scale permanent slide imaging and image analysis for diatom morphometrics. *Applied Sciences, 7(4),* 330.

Kloster, M., Kauer, G., & Beszteri, B. (2014). Sherpa: An image segmentation and outline feature extraction tool for diatoms and other objects. *BMC bioinformatics, 15(1),* 218.

Korhola, A. (2007). Diatom methods | data interpretation. In Elias, S. (Ed.) Encyclopedia of Quaternary Science (pp. 494-507). Oxford: Elsevier.

Krauss, W., & Hildebrand, U. (1995). Fibre persistence on garments under open-air conditions. Paper presented at the Proceedings of European Fibres Group Meeting, Linkoping (Sweden).

Krinsley, D. H., & Doornkamp, J. C. (1973). Atlas of quartz sand surface textures: Cambridge University Press: Cambridge.

Krumbein, W. C. (1941). Measurement and geological significance of shape and roundness of sedimentary particles. *Journal of Sedimentary Research*, *11*(2), 64-72.

L

Lander, E. S. (2017). Response to the ANZFSS council statement on the president's council of advisors on science and technology report. *Australian Journal of Forensic Sciences*, *49*(4), 366-368.

Langenkämper, D., Zurowietz, M., Schoening, T., & Nattkemper, T. W. (2017). Biigle 2.0-browsing and annotating large marine image collections. *Frontiers in Marine Science*, *4*, 83. 1-10

Legland, D., Arganda-Carreras, I., & Andrey, P. (2016). Morpholibj: Integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics, 32(22),* 3532-3534.

Lepot, L., & Vanden Driessche, T. (2015). Fibre persistence on immersed garment — influence of water flow and stay in running water. *Science & Justice, 55(6),* 431-436.

Lepot, L., Vanden Driessche, T., Lunstroot, K., Gason, F., & De Wael, K. (2015). Fibre persistence on immersed garment—influence of knitted recipient fabrics. *Science & Justice, 55(4),* 248-253.

Levin, E. A., Morgan, R. M., Scott, K. R., & Jones, V. J. (2017). The transfer of diatoms from freshwater to footwear materials: An experimental study assessing transfer, persistence, and extraction methods for forensic reconstruction. *Science and Justice, 57(5),* 349-360.

Levin, E. A., Morgan, R. M., Griffin, L. D., & Jones, V. J. (2019). A Comparison of Thresholding Methods for Forensic Reconstruction Studies Using Fluorescent Powder Proxies for Trace Materials. *Journal of Forensic Sciences*, *64(2),* 431-442.

Li, C., & Tam, P. K.-S. (1998). An iterative algorithm for minimum cross entropy thresholding. *Pattern Recognition Letters, 19(8),* 771-776.

Lindsay, E., McVicar, M. J., Gerard, R. V., Randall, E. D., & Pearson, J. (2011). Passive exposure and persistence of gunshot residue (GSR) on bystanders to a shooting:

Comparison of shooter and bystander exposure to GSR. *Canadian Society of Forensic Science Journal, 44(3),* 89-96.

Lindström, A. C., Hoogewerff, J., Athens, J., Obertova, Z., Duncan, W., Waddell, N., & Kieser, J. (2015). Gunshot residue preservation in seawater. *Forensic Science International, 253,* 103-111.

Locard, E. (1920). L'enquête criminelle et les méthodes scientifiques. Paris: E. Flammarion.

Locard, E. (1930). The analysis of dust traces (in three parts). *American Journal of Police Science, 1(4),* 401-418.

López-Leyva, R., Rojas-Domínguez, A., Flores-Mendoza, J. P., Casillas-Araiza, M. Á., & Santiago-Montero, R. (2016). Comparing threshold-selection methods for image segmentation: Application to defect detection in automated visual inspection systems. Paper presented at the Mexican Conference on Pattern Recognition.

Lowrie, C., & Jackson, G. (1991). Recovery of transferred fibres. *Forensic Science International, 50(1),* 111-119.

Lynch, M. (2013). Science, truth, and forensic cultures: The exceptional legal status of DNA evidence. Studies in History and Philosophy of Science Part C: *Studies in History and Philosophy of Biological and Biomedical Sciences, 44(1),* 60-70.

M

Machado, G.M.V., Albino, J., Leal, A.P., & Bastos, A.C. (2016). Quartz grain assessment for reconstructing the coastal palaeoenvironment. *Journal of South American Earth Sciences, 70,* 353-367.

Mackay, A., Jones, V., & Battarbee, R. (2014). Approaches to Holocene climate reconstruction using diatoms. In A. Mackay, R. Battarbee, J. Birks, & F. Oldfield (Eds.), Global change in the holocene (pp. 294-309). Chichester: Routledge.

Mahaney, W. C. (2002). Atlas of sand grain surface textures and applications: Oxford University Press, USA.

Margot, P. (2011). Forensic science on trial-what is the law of the land? *Australian Journal of Forensic Sciences, 43(2-3),* 89-103.

Mazumder, A., Govil, P., Kar, R., & Gayathri, N. M. (2017). Paleoenvironments of a proglacial lake in Schirmacher Oasis, East Antarctica: Insights from quartz grain microtextures. *Polish Polar Research, 38(1),* 1-19.

Mazzoli, A., & Favoni, O. (2012). Particle size, size distribution and morphological evaluation of airborne dust particles of diverse woods by scanning electron microscopy and image processing program. *Powder Technology, 225,* 65-71

Mazzoli, A., & Moriconi, G. (2014). Particle size, size distribution and morphological evaluation of glass fiber reinforced plastic (GRP) industrial by-product. *Micron, 67,* 169-178.

Mazzullo, J., & Kennedy, S. K. (1985). Automated measurement of the nominal sectional diameters of individual sedimentary particles. *Journal of Sedimentary Research, 55(4)* 593-595.

McCulloch, G., Morgan, R., & Bull, P. (2017). High performance liquid chromatography as a valuable tool for geoforensic soil analysis. *Australian Journal of Forensic Sciences, 49(4),* 421-448.

McDonald, J. H. (2009). *Handbook of Biological Statistics* Baltimore, MD: Sparky House Publishing.

Meakin, G., & Jamieson, A. (2013). DNA transfer: Review and implications for casework. *Forensic Science International: Genetics, 7(4),* 434-443.

Meyer dos Santos, S., Klinkhardt, U., Schneppenheim, R., & Harder, S. (2010) Using ImageJ for the quantitative analysis of flow-based adhesion assays in real-time under physiologic flow conditions, *Platelets, 21(1),* 60-66.

Miller, N. A., & Henderson, J. J. (2011). Correlating particle shape parameters to bulk properties and load stress at two water contents. *Agronomy Journal*, *103*(5), 1514-1523.

Mnookin, J., Cole, S. A., Dror, I., Fisher, B. A., Houk, M., Inman, K., & Risinger, D. M. (2011). The need for a research culture in forensic science. *UCLA Law Review,* 58, 725–779.

Morgan, R. (2017). Conceptualising forensic science and forensic reconstruction; part i: A conceptual model. *Science & Justice*, *57*(6), 455-459.

Morgan, R., Allen, E., King, T., & Bull, P. (2014). The spatial and temporal distribution of pollen in a room: Forensic implications. *Science & Justice, 54(1),* 49-56.

Morgan, R., Davies, G., Balestri, F., & Bull, P. (2013). The recovery of pollen evidence from documents and its forensic implications. *Science & Justice, 53(4),* 375-384.

Morgan, R., Flynn, J., Sena, V., & Bull, P. (2014). Experimental forensic studies of the preservation of pollen in vehicle fires. Science & Justice, 54(2), 141-145.

Morgan, R., Little, M., Gibson, A., Hicks, L., Dunkerley, S., & Bull, P. (2008). The preservation of quartz grain surface textures following vehicle fire and their use in forensic enquiry. *Science & Justice, 48(3),* 133-140.

Morgan, R. M., & Bull, P. A. (2007). The philosophy, nature and practice of forensic sediment analysis. *Progress in Physical Geography, 31(1),* 43-58.

Morgan, R. M., & Bull, P. A. (2007). The use of grain size distribution analysis of sediments and soils in forensic enquiry. Science & Justice, 47(3), 125-135.

Morgan, R. M., Cohen, J., McGookin, I., Murly-Gotto, J., O'Connor, R., Muress, S., Bull, P. A. (2009). The relevance of the evolution of experimental studies for the interpretation and evaluation of some trace physical evidence. *Science & Justice, 49(4),* 277-285.

Morgan, R. M., French, J. C., O'Donnell, L., & Bull, P. A. (2010). The reincorporation and redistribution of trace geoforensic particulates on clothing: An introductory study. *Science & Justice, 50(4),* 195-199.

Morgan, R. M., Freudiger-Bonzon, J., Nichols, K. H., Jellis, T., Dunkerley, S., Zelazowski, P., & Bull, P. A. (2009). The forensic analysis of sediments recovered from footwear. In Ritz, K., Dawson, L., & Miller, D., Criminal and Environmental Soil Forensics (pp. 253-269). Berlin: Springer.

Morgan, R. M., Wiltshire, P., Parker, A., & Bull, P. A. (2006). The role of forensic geoscience in wildlife crime detection. *Forensic Science International, 162(1-3),* 152-162.

Morgan, R. M., Scott, K. R., Ainley, J., & Bull, P. A. (2019). Journey history reconstruction from the soils and sediments on footwear: An empirical approach. *Science & Justice*, *59*(3), 306-316.

Mou, D., & Stoermer, E. F. (1992). Separating tabellaria (bacillariophyceae) shape groups: A large sample approach based on fourier descriptor analysis. *Journal of Phycology, 28,* 386-395.

Murray, K. R., Fitzpatrick, R. W., Bottrill, R. S., Berry, R., & Kobus, H. (2016). Soil transference patterns on bras: Image processing and laboratory dragging experiments. *Forensic Science International, 258,* 88-100.

Murray, R. C., & Tedrow, J. C. (1975). Forensic geology: Earth sciences and criminal investigation. New Brunswick: Rutgers University Press.

Mushtaq, S., Rasool, N., & Firiyal, S. (2016). Detection of dry bloodstains on different fabrics after washing with commercially available detergents. *Australian Journal of Forensic Sciences, 48(1)*, 87-94.

N

Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice, 54(3),* 208-214.

Nanes, B. A. (2015). Slide set: Reproducible image analysis and batch processing with ImageJ. *Biotechniques, 59(5),* 269.

National Academy of Sciences (2009). Strengthening Forensic Science in the United States: A path forward. Washington, D.C: National Academy Press. Available online at: https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf

Nesbitt, R., Wessel, J., & Jones, P. (1976). Detection of gunshot residue by use of the scanning electron microscope. *Journal of Forensic Science, 21(3),* 595-610.

Newell, A. J., Morgan, R. M., Griffin, L. D., Bull, P. A., Marshall, J. R., & Graham, G. (2012). Automated texture recognition of quartz sand grains for forensic applications. *Journal of Forensic Sciences, 57(5),* 1285-1289.

Nguyen, A., Beyersdorf, J., Riethoven, J. J., & Pannier, A. K. (2016). High-throughput screening of clinically approved drugs that prime polyethylenimine transfection reveals modulation of mitochondria dysfunction response improves gene transfer efficiencies. *Bioengineering & Translational Medicine*, *1*(2), 123-135.

Niblack, W. (1985). An Introduction to Digital Image Processing. Birkeroed, Denmark: Strandberg Publishing Company.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G. and Contestabile, M., (2015). Promoting an open research culture. *Science, 348(6242),* pp.1422-1425.

O

O'Hara, M. (2015). *Austerity bites: A journey to the sharp end of cuts in the UK.* Bristol: Policy Press.

Oberholzer, M., Östreicher, M., Christen, H., & Brühlmann, M. (1996). Methods in quantitative image analysis. *Histochemistry and Cell Biology, 105(5),* 333-355.

Oldfield, C., Morgan, R., Miles, H., & French, J. (2017). The efficacy of luminol in detecting bloodstains that have been washed with sodium percarbonate and exposed to environmental conditions. *Australian Journal of Forensic Sciences 50(4),* 345-354.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, And Cybernetics, 9(1),* 62-66.

P

Palese, A., Giovannini, G., Lucchesi, S., Dumontet, S., & Perucci, P. (2004). Effect of fire on soil C, N and microbial biomass. *Agronomie, 24(1),* 47-53.

Palmer, R., & Banks, M. (2005). The secondary transfer of fibres from head hair. *Science & Justice, 45(3),* 123-128.

Palmer, R., & Burch, H. J. (2009). The population, transfer and persistence of fibres on the skin of living subjects. *Science & Justice, 49(4),* 259-264.

Palmer, R., & Polwarth, G. (2011). The persistence of fibres on skin in an outdoor deposition crime scene scenario. *Science & Justice, 51(4),* 187-189.

Peabody, A. J. (1999). Forensic science and diatoms. In Smol, J.P., and Stoermer, E.F., (Eds.) The Diatoms: Applications for the Environmental and Earth Sciences. 389-401. Cambridge: Cambridge University Press.

Phansalkar, N., More, S., Sabale, A., & Joshi, M. (2011). Adaptive local thresholding for detection of nuclei in diversity stained cytology images. Paper presented at the Communications and Signal Processing (ICCSP), 2011 International Conference.

Pollanen, M. S. (1998). Diatoms and homicide. *Forensic Science International, 91(1),* 29-34.

Pounds, C. A., & Smalldon, K. W. (1975a). The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part iii: A preliminary investigation of the mechanisms involved. *Journal of the Forensic Science Society, 15(3),* 197-207.

Pounds, C. A., & Smalldon, K. W. (1975b). The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part i: Fibre transference. *Journal of the Forensic Science Society, 15(1),* 17-27.

Pounds, C. A., & Smalldon, K. W. (1975c). The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part ii: Fibre persistence. *Journal of the Forensic Science Society, 15(1),* 29-37.

Powers, M. C. (1953). A new roundness scale for sedimentary particles. *Journal of Sedimentary Petrolrology, 23(2),* 117–119.

Preibisch, S., Saalfeld, S., & Tomancak, P. (2009). Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics, 25(11),* 1463-1465.

President's Council of Advisors on Science and Technology (2016). Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. Executive Office of The President's Council of Advisors on Science and Technology, Washington DC.

Prewitt, J., & Mendelsohn, M. L. (1966). The analysis of cell images. *Annals of the New York Academy of Sciences, 128(1),* 1035-1053.

Pye, K. (2007). Sediment fingerprints: A forensic technique using quartz sand grains—a comment. *Science & Justice, 47(1),* 34-36.

Pye, K., & Blott, S. J. (2004). Particle size analysis of sediments, soils and related particulate materials for forensic purposes using laser granulometry. *Forensic Science International*, *144*(1), 19-27.

Pye, K., & Croft, D. J. (2004). Forensic geoscience: Principles, techniques and applications. London: Geological Society of London.

R

Rankin, B. (2010) Forensic Practice in White, P. (Ed.). *Crime Scene to Court: The Essentials of Forensic Science.* 1-24 London: Royal Society of Chemistry.

Renberg, I. (1990). A procedure for preparing large sets of diatom slides from sediment cores. *Journal of Paleolimnology, 4(1),* 87-90.

Rhoten, D., & Parker, A. (2004). Risks and rewards of an interdisciplinary research path. *Science, 306(5704),* 2046-2046.

Rhoten, D., & Pfirman, S. (2007). Women in interdisciplinary science: Exploring preferences and consequences. *Research Policy, 36(1),* 56-75.

Ribaux, O., Baylon, A., Roux, C., Delémont, O., Lock, E., Zingg, C., & Margot, P. (2010). Intelligence-led crime scene processing. Part I: Forensic intelligence. *Forensic Science International*, *195(1-3),* 10-16.

Ribaux, O., & Wright, B. T. (2014). Expanding forensic science through forensic intelligence. *Science & Justice, 54(6),* 494-501.

Ricci, C., Chan, K.L., & Kazarian, S. G. (2006). Combining the Tape-Lift Method and Fourier Transform Infrared Spectroscopic Imaging for Forensic Applications. *Applied Spectroscopy, 60(9),* 1013–1021.

Riding, J. B., Rawlins, B. G., & Coley, K. H. (2007). Changes in soil pollen assemblages on footwear worn at different sites. *Palynology*, *31*(1), 135-151.

Ridler, T., & Calvard, S. (1978). Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man, and Cybernetics, 8(8),* 630-632.

Ritz, K., Dawson, L., & Miller, D. (2008). Criminal and environmental soil forensics. Berlin: Springer.

Robertson, J., Kidd, C. B. M., & Parkinson, H. M. P. (1982). The persistence of textile fibres transferred during simulated contacts. *Journal of the Forensic Science Society, 22(4),* 353-360.

Robertson, J., & Lloyd, A. (1984). Observations on redistribution of textile fibres. *Journal of the Forensic Science Society, 24(1),* 3-7.

Roewer, L., & Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by pcr amplification of polymorphic simple repeat sequences in case work. *Forensic Science International,* 53(2), 163-171.

Round, F. E., Crawford, R. M., & Mann, D. G. (1990). *Diatoms: biology and morphology of the genera.* Cambridge. Cambridge University Press.

Roux, C., Langdon, S., Waight, D., & Robertson, J. (1999). The transfer and persistence of automotive carpet fibres on shoe soles. *Science & Justice, 39(4),* 239-251.

Roux, C., Talbot-Wright, B., Robertson, J., Crispino, F., & Ribaux, O. (2015). The end of the (forensic science) world as we know it? The example of trace evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1674),* 20140260.

Rowell, F., Hudson, K., & Seviour, J. (2009). Detection of drugs and their metabolites in dusted latent fingermarks by mass spectrometry. *Analyst*, *134*(4), 701-707.

Ruffell, A., & McKinley, J. (2005). Forensic geoscience: applications of geology, geomorphology and geophysics to criminal investigations. *Earth-Science Reviews*, *69*(3-4), 235-247.

Ruffell, A., & McKinley, J. (2008). Geoforensics: John Wiley & Sons: Chichester, UK.

Rupert, K., Ho, M., & Trejos, T. Study of Transfer and Persistence of Glass in a Mock Kidnapping Journal of the American Society of Trace Evidence Examiners, 8(1) 16-33

S

Saferstein, R., & Hall, A. B. (Eds.). (2002). *Forensic Science Handbook*. Upper Saddle River: Prentice Hall.

Salter, M. T., & Cook, R. (1996). Transfer of fibres to head hair, their persistence and retrieval. *Forensic Science International, 81(2-3),* 211-221.

Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition, 33(2),* 225-236.

Schield, C., Campelli, C., Sycalik, J., Randle, C., Hughes-Stamm, S., & Gangitano, D. (2016). Identification and persistence of Pinus pollen DNA on cotton fabrics: A forensic application. *Science & Justice, 56(1),* 29-34.

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development, 82(7-8),* 518-529.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nature Methods, 9(7),* 671-675.

Schulze, K., Tillich, U. M., Dandekar, T., & Frohme, M. (2013). Planktovision-an automated analysis system for the identification of phytoplankton. *BMC Bioinformatics, 14(1),* 115.

Scott, H. G. (1985). The persistence of fibres transferred during contact of automobile carpets and clothing fabrics. *Journal of the Canadian Society of Forensic Science, 18(4),* 185-199.

Scott, K. R., Morgan, R. M., Jones, V. J., Dudley, A., Cameron, N., & Bull, P. A. (2017). The Value of an Empirical Approach for the Assessment of Diatoms as Environmental Trace Evidence in Forensic Limnology. *Archaeological and Environmental Forensic Science, 1(1),* 49-78.

Scott, K. R., Morgan, R. M., Jones, V. J., & Cameron, N. G. (2014). The transferability of diatoms to clothing and the methods appropriate for their collection and analysis in forensic geoscience. *Forensic Science International,* 241, 127-137.

Scott, K. R., Morgan, R. M., Cameron, N. G., & Jones, V. J. (2019). Freshwater diatom transfer to clothing: Spatial and temporal influences on trace evidence in forensic reconstructions. *Science & Justice,* 59(3), 292-305.

Sertsu, S. M., & Sánchez, P. A. (1978). Effects of heating on some changes in soil properties in relation to an Ethiopian land management practice. *Soil Science Society of America Journal, 42(6),* 940-944.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging, 13(1),* 146-166.

Shanbhag, A. G. (1994). Utilization of information measure as a means of image thresholding. *CVGIP: Graphical Models and Image Processing, 56(5),* 414-419.

Siver, P. A., Lord, W. D., & McCarthy, D. J. (1994). Forensic limnology: The use of freshwater algal community ecology to link suspects to an aquatic crime scene in southern new England. *Journal of Forensic Science, 39(3*), 847-853.

Slot, A., van der Weerd, J., Roos, M., Baiker, M., Stoel, R. D., & Zuidberg, M. C. (2017). Tracers as invisible evidence—the transfer and persistence of flock fibres during a car exchange. *Forensic Science International, 275,* 178-186.

Smol, J. P., & Stoermer, E. F. (2010). The diatoms: Applications for the Environmental and Earth Sciences. Cambridge: Cambridge University Press.

Stoermer, E., & Ladewski, T. (1982). Quantitative analysis of shape variation in type and modern populations of gomphoneis herculeana. Nova hedwigia. Beihefte. *Agris. 347-386*

Stoermer, E., Yu-Zao, Q., & Ladewski, T. (1986). A quantitative investigation of shape variation in didymosphenia (lyngbye) m. Schmidt (bacillariophyta). *Phycologia, 25(4)*, 494-502.

Stoney, D. A., Bowen, A. M., & Stoney, P. L. (2016). Loss and replacement of small particles on the contact surfaces of footwear during successive exposures. *Forensic Science International, 269,* 78-88.

Sugita, R., & Marumo, Y. (1996). Validity of color examination for forensic soil identification. *Forensic Science International, 83(3),* 201-210.

Szewcow, R., Robertson, J., & Roux, C. P. (2011). The influence of front-loading and top-loading washing machines on the persistence, redistribution and secondary transfer of textile fibres during laundering. *Australian Journal of Forensic Sciences, 43(4),* 263-273.

T

Thakar, M. K., & Singh, R. (2010). Diatomological mapping of water bodies for the diagnosis of drowning cases. *Journal of Forensic and Legal Medicine, 17(1),* 18-25.

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013). Expertise in fingerprint identification. *Journal of Forensic Sciences, 58(6),* 1519-1530.

Tilstone, W. J., Savage, K. A., & Clark, L. A. (2006). Forensic science: An Encyclopedia of history, methods, and techniques. Oxford: ABC-CLIO.

Tsai, W.-H. (1985). Moment-preserving thresolding: A new approach. *Computer Vision, Graphics, and Image Processing, 29(3),* 377-393.

U

Udupa, J. K., LeBlanc, V. R., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L. M., & Woodburn, J. (2006). A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics, 30(2),* 75-87.

Uitdehaag, S., Dragutinovic, A., & Kuiper, I. (2010). Extraction of diatoms from (cotton) clothing for forensic comparisons. *Forensic Science International, 200(1-3),* 112-116.

V

Vdović, N., Obhođaš, J., & Pikelj, K. (2010). Revisiting the particle-size distribution of soils: comparison of different methods and sample pre-treatments. *European Journal of Soil Science, 61(6),* 854-864.

Verdon, T. J., Mitchell, R. J., & van Oorschot, R. A. (2014). Evaluation of tapelifting as a collection method for touch DNA. Forensic science international: *Genetics, 8(1),* 179-186.

Verma, K. (2013). Role of diatoms in the world of forensic science. *Journal of Forensic Research, 4(2),* 181-18.

Virtanen, V., Korpelainen, H., & Kostamo, K. (2007). Forensic botany: usability of bryophyte material in forensic studies. *Forensic Science International, 172(2-3),* 161-163.

W

Warrier, A. K., Pednekar, H., Mahesh, B., Mohan, R., & Gazi, S. (2016). Sediment grain size and surface textural observations of quartz grains in late quaternary lacustrine sediments from Schirmacher Oasis, East Antarctica: Paleoenvironmental significance. *Polar Science, 10(1),* 89-100.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05". *The American Statistician, 73:sup1*, 1-19.

White, P. M., Stone, J. S., Groves, A. K., & Segil, N. (2012). EGFR signaling is required for regenerative proliferation in the cochlea: conservation in birds and mammals. *Developmental biology, 363(1),* 191-200.

Wiggins, K. G., Emes, A., & Brackley, L. H. (2002). The transfer and persistence of small fragments of polyurethane foam onto clothing. *Science & Justice, 42(2),* 105-110.

Y

Yang, X., Beyenal, H., Harkin, G., & Lewandowski, Z. (2001). Evaluation of biofilm image thresholding methods. *Water Research, 35(5),* 1149-1158.
Yen, J.-C., Chang, F.-J., & Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing, 4(3),* 370-378.

Z

Zack, G., Rogers, W., & Latt, S. (1977). Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry, 25(7),* 741-753.

Zaitoun, N. M., & Aqel, M. J. (2015). Survey on image segmentation techniques. *Procedia Computer Science, 65,* 797-806.

Zeichner, A., & Levin, N. (1993). Collection efficiency of gunshot residue (gsr) particles from hair and hands using double-side adhesive tape. *Journal of Forensic Science, 38(3),* 571-584.

Ziliak, S., and McCloskey, D. (2008), The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives, Ann Arbor, MI: University of Michigan Press.

Zimmerman, K. A., & Wallace, J. R. (2008). The potential to determine a postmortem submersion interval based on algal/diatom diversity on decomposing mammalian carcasses in brackish ponds in Delaware. *Journal of Forensic Sciences, 53(4),* 935-941.

Appendices

Please refer to the attached CD for appendices. It was deemed most appropriate to provide these in a digital format since (1) Many of the files would be difficult to display sensibly on A4 paper, since they are, for example, macros (lines of code), very large images (some which are gigapixel images), or large sets of images (e.g. almost 9,000 vignettes of diatoms); and (2) In order to provide an appendix on paper, there would have to be a degree of aggregation for the data; many of the experiments produced tables with thousands of rows, and tens of columns. Providing a digital appendix means that the data can be provided without aggregation. The appendices contain:

Appendix Contents

UV segmentation

Quartz segmentation

UV persistence

Quartz persistence

Quartz size selection

Diatom size selection

Papers