Document Title:	**Determination of Key Factors in Particle Combination Analysis to Enable Systematic Improvement, Optimization and Transition to Practice**
Author(s):	**David A. Stoney, Ph.D., Paul L. Stoney, MBA**
Document Number:	**305811**
Date Received:	**January 2023**
Award Number:	**2017-IJ-CX-0030**

Determination of Key Factors in Particle Combination Analysis to Enable Systematic Improvement, Optimization and Transition to Practice

Award 2017-IJ-CX-0030

David A. Stoney, PhD and Paul L. Stoney, MBA
Stoney Forensic, Inc.
14101-G Willard Road
Chantilly, VA 20151

Acknowledgements

Specimens used as part of this project (collected under NIJ Award 2015-DN-BX-K046) were sets of very small particles (VSP) recovered from actual items of evidence. We are very grateful to Michael Grubb and the San Diego County Sheriff's Department Crime Laboratory for granting us access to the evidence and providing secure facilities for our collections.

The computational and statistical methods used in this project were developed in close collaboration with Dr. Cedric Neumann, co-Principal Investigator under NIJ Award 2012-DN-BX-K041 and consultant under NIJ Award 2015-DN-BX-K046.

This project was supported by Award No. 2017-IJ-CX-0030 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

Abstract

This project determined key factors affecting particle combination analysis as applied to very small particles (VSP). Identification of these factors is a necessary step to enable systematic improvement, optimization, and transition to practice.

VSP show exceptional promise to (1) expand the numbers of cases where trace evidence can be used, and (2) provide quantitative measures of evidential value. The laboratory analyses are highly efficient and utilize existing crime laboratory personnel and equipment.

Prior research, employing reasonable initial choices of analytical and statistical parameters, has (1) demonstrated the presence of highly discriminating VSP profiles on the surfaces of common items of physical evidence, (2) characterized VSP combinations using analytical instrumentation and expertise commonly available in forensic laboratories, (3) developed statistically rigorous measurements of correspondence between VSP profiles, and (4) produced objective measures for the resulting probative value.

This project used available particle combination analysis methods and VSP specimens from physical evidence (both resulting from prior NIJ-funded research) to examine these analytical and statistical parameters more critically, identifying key factors that affect performance.

Experiments were conducted using scanning electron microscopy (SEM) with energy dispersive x-ray elemental analysis (EDS) to characterize the elemental composition of thousands of individual particles within each specimen. The experiments studied:

- Reproducibility of VSP analyses at given parameters
- Effects of the SEM/EDS parameters used for the detection of each particle
- Effects of SEM/EDS x-ray analysis parameters used for elemental analysis of each particle
- Effects of the number and choice of elements used in the elemental analysis
- Effects of particle size on the strength of correspondence between particle sets
- Effects of data filtration parameters on the strength of correspondence between particle sets

The experiments confirmed the presence of abundant, highly discriminating VSP on common items of evidence. The numbers of particles available for analysis was not a limiting factor: many more particles (usually greater than 50 times more) were present than were used for the analysis. A very high level of reproducibility was observed.

Many of the parameters tested had no measurable effect on particle combination analysis performance and others had minor or interactive effects. Four factors were identified as having significant impact on the strength of correspondence between particle profiles, three factors were identified as having a significant impact on the numbers of particles detected and nine factors were identified as having a significant impact on analytical time and costs.

The approach in its current state of development offers crime laboratories an additional capability suitable for high priority cases. The identification of key factors affecting performance of the VSP analytical protocol allows existing methods to be further developed and systematically improved to facilitate transition to routine practice.

Table of Contents

I. Introduction

I.A. Context

Increasing the value of trace evidence analysis is a major priority for forensic science research. Value is determined by (1) the numbers of cases where it is useful, and (2) the additional probative value that it can bring. This value must be viewed together with the associated costs and balanced with other crime laboratory operations.

Prevailing methods of trace evidence analysis have been limited by three major aspects:

- Difficulties in the measurement of probative value
- Increased specialization, focusing on smaller numbers of particle types, in correspondingly smaller numbers of cases
- Relatively long analytical times and high levels of effort for required tasks

These limitations combine to reduce the application of trace evidence, resulting in a set of major challenges: low perceptions of probative value, small numbers of case requests, and high costs relative to case contributions.[1-4] The impact within forensic laboratories has been substantial, resulting in reductions in funding, restriction of services, and even complete closure of trace analysis sections within laboratories.[4,5]

Solutions to these challenges have been elusive because the underlying limitations are inter-related in a complex way, with improvements to one problem exacerbating another. Efforts to increase probative value have met fundamental limitations of "class associations," while at the same time increasing specialization, analytical time and costs. These increases have offset efficiencies offered by new methods and technologies and reduced the number of cases where it is practical to apply them.[5]

Within this context, methods focusing on the analysis of combinations of very small particles (VSP) show exceptional promise to address the limitations effecting the value and contribution of trace evidence analysis. In prior NIJ-funded research we have (1) characterized VSP combinations using analytical instrumentation and expertise commonly available in forensic laboratories, (2) developed statistically rigorous measurements of the strength of correspondence between VSP profiles, (3) measured the probative value of the resulting associations within well-defined experimental parameters and (4) demonstrated the presence of highly discriminating VSP profiles on the surfaces of common items of physical evidence.[6-8] We now have the ability to address each of the above limitations: probative value can be measured, cases are not restricted by small numbers of particle types, and it is practical to achieve both the required analytical times and level of effort.

The prior research employed scanning electron microscopy with energy dispersive x-ray spectroscopy (SEM/EDS) to characterize the elemental composition of hundreds to thousands of particles in each specimen. This method used *reasonable initial choices* of analytical and statistical parameters. The optimization of a VSP analysis protocol requires the identification of factors that influence protocol performance. Separating factors (a quantity or quality that *does* have an influence upon the system) from variables (a quantity or quality that *might* have an influence upon the system) was the purpose of the present research. The goal was to measure the relative impact of a wide set of independent variables on VSP analysis. The result was the identification of a set of important, controlling factors that must be addressed to meaningfully optimize the protocol.

I.B. Approach

Optimization of the VSP analysis protocol requires that factors influencing the reliability, costs and selectivity be identified. A screening stage of experimental design is needed that will identify the key factors affecting performance and provide information (such as the variability and magnitude of effects) that will be needed for the next stage of process improvement.

This project consisted of the six sequential experiments listed in Table 1, each focused on a discrete portion of the overall process, while holding other parameters fixed.

Table 1. The Six Sequential Experiments

Experiment 1	Reproducibility of VSP Analyses at Given Parameters
Experiment 2	Effects of SEM/EDS Particle Detection Parameters
Experiment 3	Effects of SEM/EDS X-Ray Analysis Parameters
Experiment 4	Effects of Number and Choice of Elements
Experiment 5	Contribution of Alternative Particle Size Fractions
Experiment 6	Effects of Data Filtration Parameters

The dependent variables (measured outcomes of performance) under investigation were a) analysis time, b) number of particles detected and characterized, c) correct / incorrect source classifications, d) rank of correct source classification and e) probative value of correspondence.

Analysis time is the amount of instrument time required for the automated SEM/EDS analysis of the specimen. This is important to the overall practicality of the protocol as it affects specimen (and case) throughput and cost (notably the obligation of an instrument of substantial cost to this analytical task in competition to others). Trade-offs between analysis time and performance may be important elements in the optimization of the protocol. Analysis time is strongly influenced by a) how efficiently the SEM detects and selects particles for analysis and b) the amount of time that is spent collecting the EDS x-ray spectrum on each particle. These are, in turn, affected directly and indirectly by variables associated with the specimen itself, the analytical protocol, and how efficiently they interact with one another.

The *number of particles detected and characterized* provides the number of VSP available for particle combination analysis. Prior investigations have found (under specific experimental parameters) that small numbers of VSP (< 300) provide little selectivity (discriminating power), while larger numbers (>1000) are highly selective. The number of particles detected and characterized by the SEM is strongly influenced by a) particle searching parameters (What proportion of particles that are present will be detected?), b) particle selection criteria (What size and composition of particles are selected for characterization?), and c) the specimen itself (How many particles meeting the selection criteria are actually present?).

The remaining three dependent variables: *correct or incorrect source classifications, rank of correct source classification,* and *probative value of correspondence* are measures of the strength of correspondence (or selectivity) of the particle combinations. The effect of all protocol changes on the strength of correspondence is of primary interest. The effects on the other performance measures may or may not have an effect on strength of correspondence.

II. Methods

II.A. Specimens Analyzed

Specimens of VSP collected from four commonly occurring evidence types, previously analyzed for demonstration purposes [8], were used in this project. These specimens were collected from actual evidence items at the San Diego Sheriff's Department Crime Laboratory (evidence from cases where detectives had determined these items to no longer be of value and had approved them for disposal). The four evidence types (handguns, drug packaging, cellular phones and ski masks) were selected because: (1) they regularly occur as evidence left and collected at major crime scenes, (2) associations of these items to one another, to individuals and to locations is of broad investigative significance, (3) they include a wide range of surface types, including most that are likely to be found on evidence, and (4) as a set, they are a good proxy to assess the levels and probative value of VSP on common items of evidence.

The VSP specimens were harvested on site at the San Diego County Sherriff's Office Crime Laboratory. Commercially prepared SEM stubs were used to harvest VSP from plastic bags used in drug packaging. For handguns, cell phones and ski masks, non-shedding clean room swabs (slightly dampened with pre-filtered distilled water) were used for VSP harvesting. Specimens were processed on a clean bench. Swab heads were removed and VSP were recovered into a suspension using a washing procedure as in [9] followed by dropwise vacuum filtration through 0.4 micrometer polycarbonate filters. These filters were then mounted onto SEM stubs for analysis. Details of specimen preparation and prior analysis can be found in [8].

Thirty specimens of each evidence type were used in this project. As shown in Table 2, all 120 specimens were used for Experiments 5 and 6. A randomly selected sub-set of 8 specimens of each type were used for Experiments 1 and 4, and a further randomly selected subset of two specimens of each type were used for Experiments 2 and 3. Nine control specimens representing the sampling materials and process are described in Table 3.

II.B. Particle Analysis Using SEM/EDS

The existing method for particle analysis of VSP using SEM/EDS as described in [8] was used as the baseline method (Table 4 with details in *Appendix A: Description of Particle Combination Analysis Methods*). Variations in most of the baseline settings are incorporated into Experiments 2, 3 and 4, as described below in Sections II.E. through II.G. The following is a generalized description of the process.

Up to 16 specimens are mounted in a sample holder (Figure 1) and Automated Feature Analysis is used on the SEM. After calibrating image contrast using a standard, this procedure goes through a sequence of automated steps while a focused electron beam is scanned over different portions of the specimen. At each position individual particles are detected by their brightness (shown in contrast to the background). This brightness is due to x-rays that are emitted by the particles when they are exposed to the electron beam in the SEM. Features of the particles, such as shape and size, can be used to select which particles are to be recorded and analyzed. For each of the selected particles, x-rays emitted by the particle are collected for a short period of time. These x-rays allow the elemental composition of each particle to be characterized. (X-rays of different energies are emitted by different chemical elements, so the distribution of the x-ray energies allows characterization of the elemental composition.) The SEM and a view of the console are shown in Table 4 and Figure 2.

Table 2. Specimen Assignments for Experiments

EXPERIMENT 1			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P102S	C102S	F102S	M106S
P111S	C104S	F107S	M111S
P118S	C105S	F111S	M118S
P121S	C108S	F114S	M126S
P123S	C121S	F118S	M127S
P125S	C128S	F121S	M129S
P138S	C131S	F123S	M132S
P140S	C133S	F131S	M133S

EXPERIMENT 2			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P138S	C102S	F111S	M126S
P140S	C128S	F123S	M129S

EXPERIMENT 3			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P138S	C102S	F111S	M126S
P140S	C128S	F123S	M129S

EXPERIMENT 4			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P102S	C102S	F102S	M106S
P111S	C104S	F107S	M111S
P118S	C105S	F111S	M118S
P121S	C108S	F114S	M126S
P123S	C121S	F118S	M127S
P125S	C128S	F121S	M129S
P138S	C131S	F123S	M132S
P140S	C133S	F131S	M133S

EXPERIMENTS 5 and 6			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P102S	C101S	F101S	M101S
P103S	C102S	F102S	M104S
P104S	C103S	F103S	M105S
P106S	C104S	F104S	M106S
P107S	C105S	F105S	M107S
P109S	C106S	F106S	M108S
P110S	C107S	F107S	M109S
P111S	C108S	F108S	M110S
P115S	C109S	F109S	M111S
P116S	C110S	F110S	M112S
P118S	C111S	F111S	M113S
P119S	C112S	F112S	M114S
P120S	C113S	F113S	M115S
P121S	C114S	F114S	M116S
P122S	C115S	F115S	M118S
P123S	C116S	F116S	M119S
P125S	C117S	F118S	M120S
P126S	C121S	F119S	M121S
P127S	C123S	F120S	M122S
P130S	C124S	F121S	M123S
P131S	C125S	F122S	M124S
P132S	C127S	F123S	M125S
P133S	C128S	F124S	M126S
P134S	C129S	F125S	M127S
P136S	C131S	F126S	M128S
P138S	C132S	F127S	M129S
P139S	C133S	F128S	M130S
P140S	C134S	F129S	M131S
P141S	C135S	F130S	M132S
P142S	C136S	F131S	M133S

Table 3. Control Specimens Representing the Sampling Materials and Process

	Drug Packaging	Cell Phones	Handguns	Ski Masks	Material Blank
Swabs exposed during sampling and processed as for specimens		CB1	FB1	MB1	
		CB2	FB2	MB2	
Adhesive SEM stubs exposed during sampling	PB1				
	PB4				
Adhesive SEM stub, unexposed					SO1

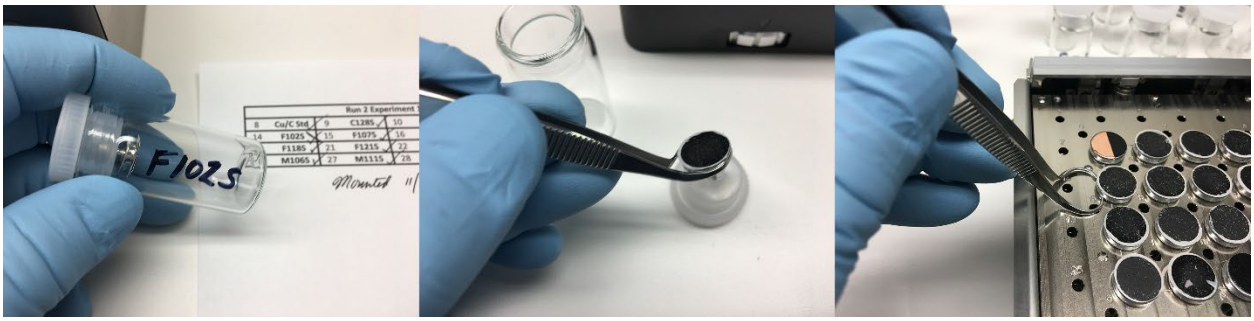
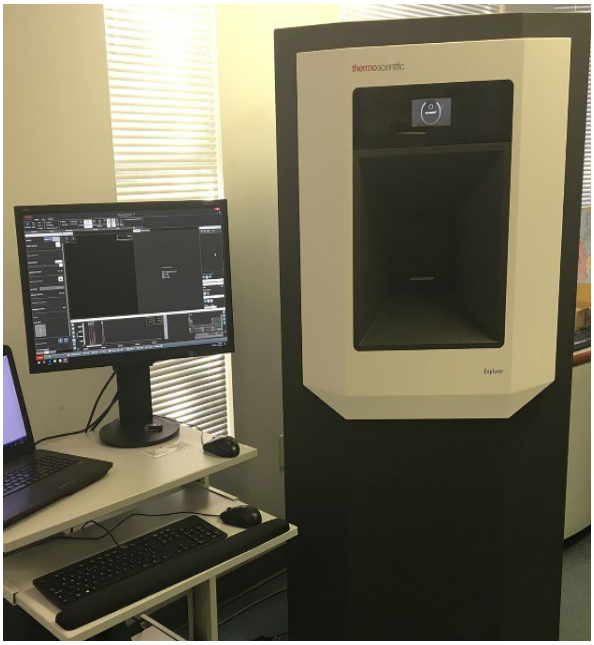


Figure 1. SEM stub in storage container (left), removal of stub (middle) and mounting of stub on the SEM sample holder (right).

Table 4. Baseline Parameters for Particle Analysis Using SEM/EDS*



Electron Microscopy Settings	Vacuum	0.075 torr
	Accelerating Voltage	20.0 kV
	Detector	BSED
	Magnification	1200X
	Working Distance	8.7 mm
	Spot Size	58%
Particle Detection Settings	Search Grid	512 x 512
	Contrast Calibration	Carbon 25; Copper 240
	Minimum Particle Brightness	64
	Imaging Time for Searching	2 µs
	Imaging Time for Measuring	16 µs
	Minimum Particle Size	0.3 µm
	Maximum Particle Size	80.0 µm
EDS X-Ray Analysis Settings	Initial X-Ray Collection Time	3 sec
	Maximum X-Ray Collection Time	6 sec
	Minimum X-ray Count	500
	Target Single Element X-ray Count	2500
	18 Target Elements	Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Co, Cu, Zn
	Element Threshold	1%
Analytical Run Limits	Maximum Number of Particles	5000
	Maximum Analysis Time	10 hours

* Thermo Scientific Explorer 4 Analyzer SEM-EDS System with Perception 5 Automatic Feature Analysis software.

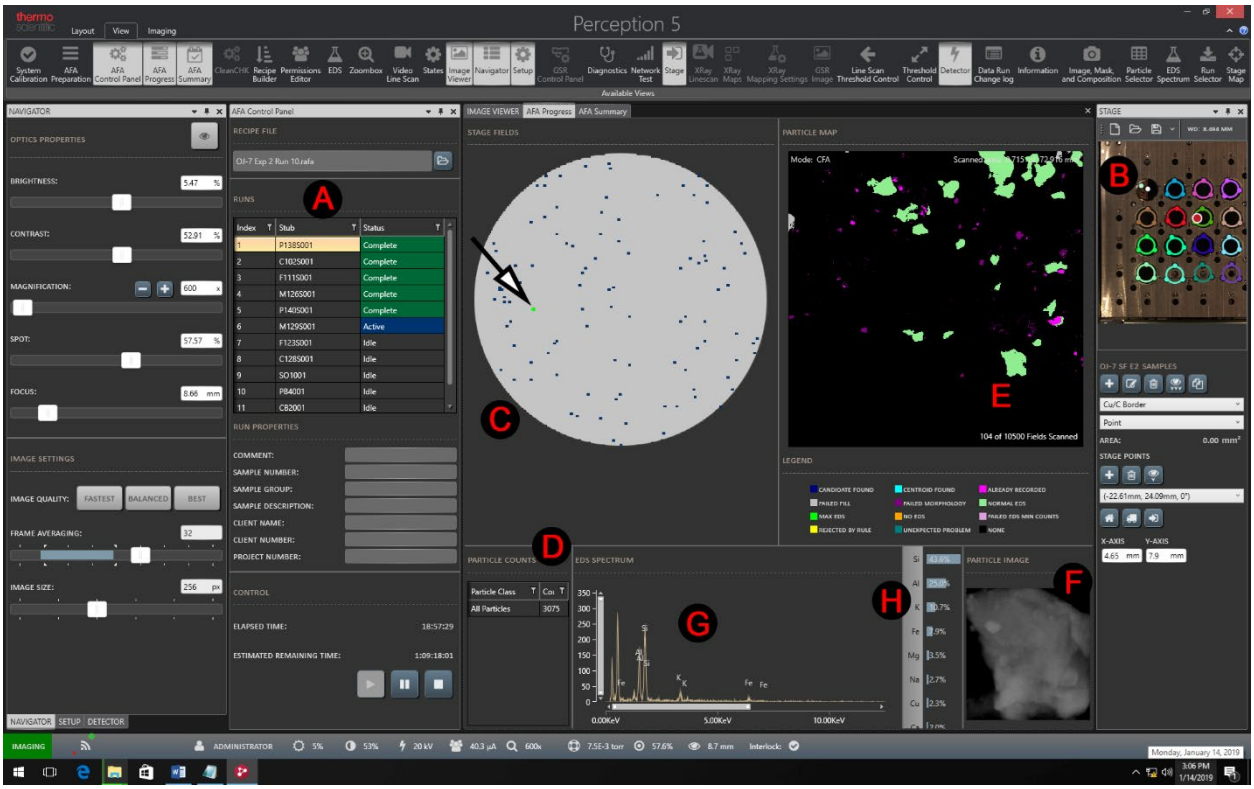


Figure 2. The Explorer 4 instrument was previously shown in Table 4. This figure is a closeup of the monitor on the controlling computer during an analysis run. Information on the monitor is indicated by letters in the figure and described below. A: Listing of the specimens for the run. B: Image of the specimen holder with colored outlines for each specimen. The small red circle on this image indicates the current electron beam position. C: Map for the current specimen showing locations (blue dots) of each of the fields on the specimen that have been examined so far. The green dot indicated by the arrow is the current field location. D. Running particle count for the specimen. E: Particle map showing particle outlines for the current field. F: Image of the particle currently being analyzed. G: EDS spectrum for that particle with the x-ray counts (vertically) at different energies (horizontally) and peaks labeled with assigned elements. H: Calculated element concentrations for the particle.

II.C. Analyses of Particle Combination Selectivity

Computational methods for measurement of particle combination selectivity were developed and applied under NIJ Awards 2012-DN-BX-K041 and 2015-DN-BX-K046. They are available as a maintained package in the widely used and freely available statistical software R. [10,11] Applications to VSP on the surfaces of carpet fibers [6] and to VSP on common items of physical evidence [8] have been peer reviewed and published. A summary of the process is provided here. Additional details can be found in the publications [6, 8] as well as *Appendix A: Description of Particle Combination Analysis Methods.*

The methods measure how well particle profiles from two specimens correspond to one another and determine how many specimens in a closed set can be correctly classified to themselves based on this correspondence. For example, in [8] particle sets (hundreds to thousands of particles) from each of 30 handguns were examined. Using a portion of the particles from each specimen, the computational methods defined 10 Target Particle Types that best discriminated among these 30 handguns. Each of the 30 handguns then had a different Known particle profile (represented by different overall proportions of the 10 Target Particle Types). The remaining portions of each of the specimens were taken as Unknown particle profiles, and the computational methods measured how likely it would be for each of the Known handgun profiles to randomly produce each of the Unknown particle profiles. Classifications (based on the highest probabilities) were correct for 90% of the handgun specimens (27 of 30).

Along with rates of correct classification (based on the highest probability) a corresponding likelihood ratio is calculated as a measure of evidential weight, which translates to a posterior probability of correct association, under the assumption of equal prior probability within the closed set. Where specimens are misclassified, the rank of the probability for true source provides an additional means of assessment.

II.D. Experiment 1: Reproducibility of VSP Analyses at Given Parameters

Given the existing protocol of for VSP analysis, to what extent are the analyses reproducible for the same specimen on different days, or after the instrument has received routine maintenance? It is reasonable to expect that variability will be greater after analyses are conducted in different batches (runs), or on different days, or when the instrument has been re-adjusted following routine maintenance.

The SEM/EDS protocol accommodates automated analysis of multiple specimens on one run. Specimens fit into a holder that (in our setup) can accommodate up to 25 specimens, as well as appropriate analytical controls. The instrument parameters are set and analytical times of one to several hours are spent on each specimen. A specimen can be re-analyzed on the same run. Alternatively, it can be re-introduced as a specimen on a subsequent run. Furthermore, the SEM source, which provides the electron beam in the SEM, has a finite life. It is routinely replaced, re-aligned and adjusted. A specimen could thus be re-analyzed following source replacement.

Experiment 1 used the existing protocol for VSP analysis with a comparative objective: analyses were conducted on the same specimens under alternative conditions. Differences in the outcomes were used to determine if the alternative conditions had a significant effect. The analyses in Table 5 were conducted for each of the 32 specimens and nine control samples in Table 2. Differences were calculated between the initial analysis and each of the other three conditions for analysis times (particles per minute), numbers of particles detected (particles per mm^2) and the selectivity of the particle combinations (strength of correspondence).

Table 5. Treatments for Experiment 1: Reproducibility of VSP Analyses at Given Parameters

Initial analysis
Repeat analysis on the same analytical run
Repeat analysis on a separate analytical run
Analysis on a separate run aftersource replacement and realignment

II.E. Experiment 2: Effects of SEM/EDS Particle Detection Parameters on Analysis Time, Particle Numbers and Selectivity of VSP Combinations

The SEM scans an electron beam over the specimen, causing particles to emit x-rays and appear brighter on a dark background. The detection of individual particles depends on how much contrast they show with the background. During automated SEM analysis several settings affect the sensitivity and timing of particle detection. Experiment 2 was designed to test the effect of four of these particle detection parameters, as described below and listed in Table 6.

The first of these four settings is the *Imaging Time for Searching.* The SEM image results from the collection of x-rays for each point in the field of view and averaging the intensity of these x-rays over a set period of time. Averaging helps to reduce noise, and longer times for x-ray collection will provide a sharper image with better defined details. Sharper edges show better contrast, favoring better particle detection. However, there is a trade-off with the analysis time. Since there are many areas to scan for particles in each specimen, the amount of time spent on each point in the field of view could have a large effect on the analysis time. The baseline setting for the *Imaging Time for Searching* is 2 µs (see Table 4). Experiment 2 is designed to test the effect of increasing this setting four-fold to 8 µs. More particles and greater analytical times are expected. VSP selectivity could increase or decrease.

The second setting is for the *Minimum Particle Brightness.* This is the lower level of brightness that will be recognized when searching for bright particles on a dark background. Levels of brightness on the image (gray levels) range from 0 to 255. The baseline value for this setting in the analysis protocol (Table 4) is 64. This means that levels of brightness below 64 will not be considered sufficient for particle detection. A brightness above 64 is means that more time will be spent to see if a particle is present and whether it meets size and shape requirements for recording and measurement of x-rays. Settings that are low are expected to detect more possible particles, but they are also more likely to respond to noise or “phantom particles”. Higher settings may miss particles but are more likely to detect only actual particles. Experiment 2 is designed to test the effect of increasing the baseline value of 64 to 128. Shorter analytical times are expected. The number of particles recorded and characterized, as well as their selectivity, could either increase or decrease.

The third setting is *Magnification*. Higher magnifications make smaller particles more detectable, so they are expected to increase the numbers of particles detected in a specimen. However, doubling the magnification causes a four-fold increase in the area to be searched, which will take more time. The greater numbers of VSP may or may not increase their overall selectivity. The baseline protocol magnification is 1200X (Table 4). Experiment 2 tests the effect of increasing magnification to 2400X.

The fourth setting is the *Particle Size Limits*. The search protocol specifies the particle size range that will result in recording and characterizing the particle. Narrower size ranges are

expected to result in the detection of fewer particles. The time to detect the particles, as well as their selectivity, could either increase or decrease. Two particle size ranges were tested: a wide range of 0.3 µm to 80 µm, and a narrower range of 2.5 µm to 40 µm.

Alternative settings for the four Particle Detection Parameters in this experiment are summarized in Table 6. The two alternatives for each of the four parameters results in 16 combinations. For this experiment, each of the 8 specimens in Table 2 was analyzed 16 times, once for each combination of the particle detection parameter settings. Analysis times (particles per minute), numbers of particles detected (particles per mm^2) and the selectivity of the particle combinations (strength of correspondence) were measured.

Table 6. Treatments for Experiment 2: Effect of SEM/EDS Particle Detection Parameters

	Low Value	High Value
Search frame averages	2 µsec	8 µsec
Lower search detection	64	128
Magnification	1200X	2400X
Particle size limits	2.5 µm - 40 µm	0.3 µm - 80 µm

II.F. Experiment 3: Effects of SEM/EDS X-Ray Analysis Parameters on Analytical Time and Selectivity of VSP Combinations

Once a particle is detected and meets the specified size criteria, the SEM/EDS system characterizes the elemental composition of the particle. This is done by collecting the x-rays that are emitted by the particle and grouping these x-rays by their energy, resulting in an EDS x-ray spectrum (a graph of the number of x-rays counted at each x-ray energy level). The different chemical elements in the particle emit different energies of x-rays and so the x-ray spectrum characterizes the elemental composition of the particle. The more time that is spent collecting the x-rays, the better the spectrum is defined. During automated SEM/EDS operation several settings affect the quality and timing of the EDS analysis. Experiment 3 was designed to test the effect of four of these SEM/EDS x-ray analysis parameters:

Minimum X-Ray Count
Element Threshold
Target Single Element X-Ray Count
Maximum X-Ray Collection Time

Figure 3 is a flow chart showing how these four parameters affect the x-ray counting for the EDS analysis of each particle that the SEM detects. X-rays are first collected for an initial time of three seconds. Looking at the range of x-ray energies for all of the elements of interest, if there is less than the than the *Minimum X-Ray Count* above the *Element Threshold* then the particle is not of interest and data collection for this particle ends. (This occurs, for example, when particles are composed primarily of elements outside the scope of the analysis.) If there is more than the minimum, then x-rays are collected and counted until one of two end conditions is reached. The end conditions are reaching the *Target Single Element X-Ray Count* or reaching the *Maximum X-Ray Collection Time*.

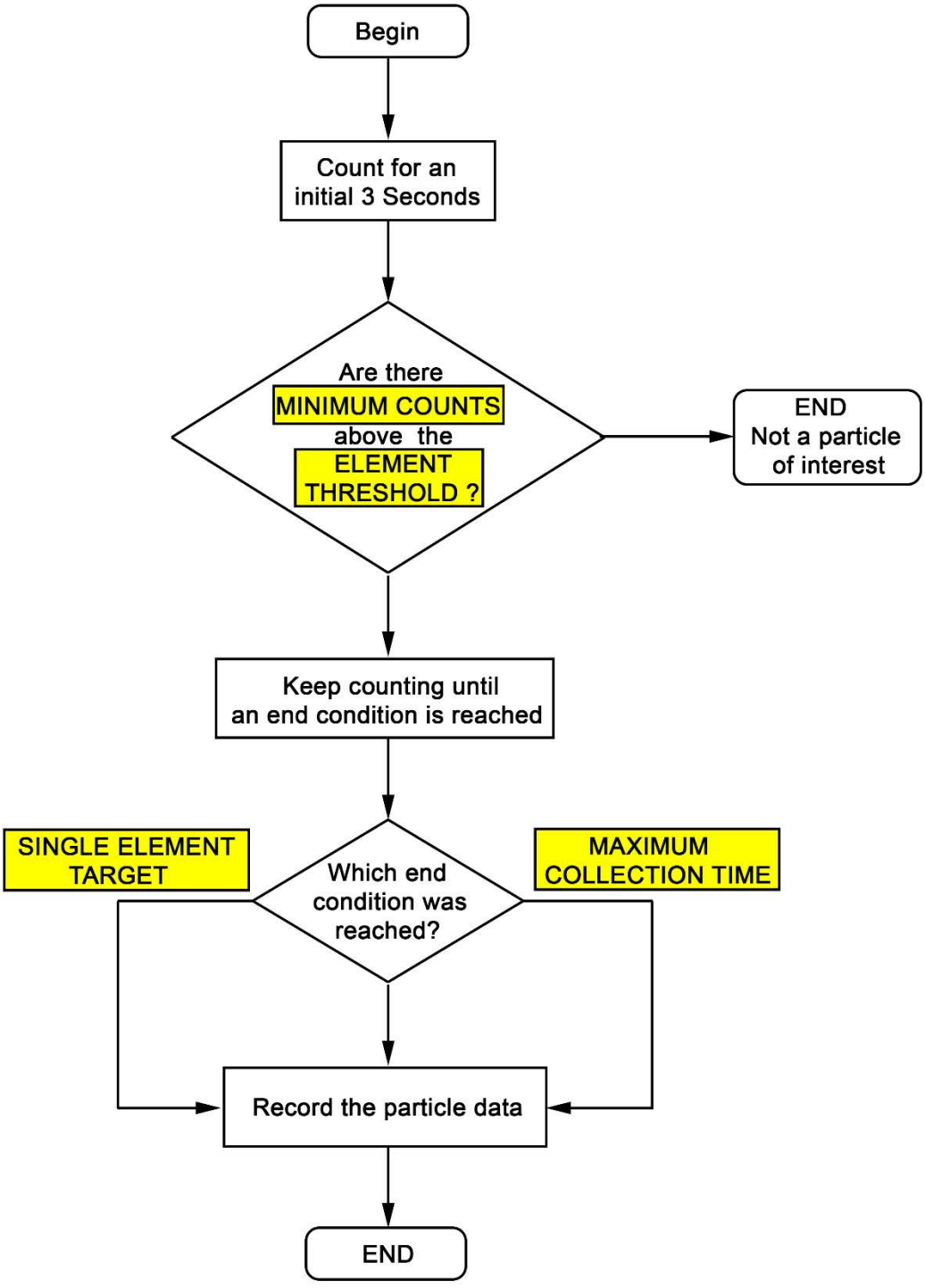


Figure 3. Flow chart of illustrating the steps for the x-ray counting during the EDS analysis of each particle that the SEM detects. After collection for an initial time period the results are evaluated. Particles failing to meet a lower threshold of counts are not examined further. For those meeting the threshold x-ray collection proceeds until one of two end conditions is met.

Higher settings for each of the conditions is expected to increase analytical times (fewer particles analyzed per minute). Increasing the *Minimum X-Ray Count* will result in more particles being set aside as "not of interest." Increasing the *Element Threshold* will have the same effect, as fewer counts will contribute to the required minimum. An increased *Maximum X-Ray Collection Time* will apply to all particles that do not meet the *Target X-Ray Count* for any element and setting this count higher will increase the amount of time spent in reaching it.

The effects on particle combination selectivity are unpredictable. Settings that result in more time collecting the x-rays will give better chemical information from each particle, which may or may not provide better selectivity (stronger correspondences).

Alternative EDS x-ray analysis settings for Experiment 3 are given in Table 7. Lower values (baseline settings, Table 4) are in the left column, with increased values to the right. The two alternatives for each of the four parameters results in 16 combinations. For this experiment, each of the 8 specimens in Table 2 was analyzed 16 times, once for each combination of the particle detection parameter settings. Analysis times (particles per minute) and the selectivity of the particle combinations (strength of correspondence) were measured.

Table 7. Treatments for Experiment 3: Effect of SEM/EDS X-Ray Analysis Parameters

	Low Value	High Value
Maximum x-ray collection time	6 sec	9 sec
Minimum x-ray count	500	1000
Target single element x-ray count	2500	5000
Element threshold	1%	3%

II.G. Experiment 4: Effects of Number and Choice of Elements on Analytical Time, Particle Numbers and Selectivity of VSP Combinations

In automated SEM/EDS analysis it is typical to specify a set of elements of interest as a parameter. For narrowly focused purposes, such as gunshot residue analysis (GSR), this set of elements is determined by the composition of the specific types of particles being sought. By specifying a narrow set of elements (based on the expected composition of gunshot residue) other commonly occurring particles (most types of VSP) will be passed over and excluded. For analysis of VSP a broad range of elemental composition is of interest. The set of 18 elements used in the existing protocol (Table 4) was chosen as a reasonable initial set for proof-of-principle research. Experiment 4 examines what effect the number and choice of elements has on the performance of particle combination analysis of VSP.

The baseline set of elements in Table 4 were chosen to avoid overlapping x-ray energies and provide broad coverage of x-ray energies appearing in the range of 1 to 10 keV. Many alternative strategies can be put forth, including (1) selection of elements that show the most discrimination among VSP specimens, or (2) selection of elements based on their abundance in household or environmental dusts.

To test the effect of varying the number and choice of elements, nine other elements were selected to combine with or replace those in the current protocol. Bromine (Br), barium (Ba), tungsten (W) and lead (Pb) were selected based primarily on their good performance in Random Forest classifiers applied to VSP.[12] Molybdenum (Mo) and Antimony (Sb) were selected primarily based on their occurrence within household dusts taken from the San Diego area (the geographical area where the VSP specimens in Table 2 were collected.[13] Zirconium (Zr) and

Bismuth (Bi) were found at relatively high levels (apart from already selected elements) in household dusts from Baltimore [14] and Strontium (Sr) is a relatively common element arising in mineral dusts, including those in the Las Vegas area.[15]

Experiment 4 had a comparative objective with four element set treatments as given in Table 8. Treatment A is the existing protocol set of 18 elements. Treatment B is an alternative set of 18, exchanging nine of the old elements (with relatively poorer performance in [12]) for nine new ones. Treatment C is a combined set of 27 elements and Treatment D is a reduced set of nine elements (those common to each of the other three sets).

Four analyses, one with each treatment condition, were completed for each of the 32 specimens in Table 2. Analysis times (particles per minute), numbers of particles detected (particles per mm^2) and the selectivity of the particle combinations (strength of correspondence) were measured.

Table 8. Elements for the Four Treatments in Experiment 4

	Atomic Number and Element																									
	11	12	13	14	15	16	17	19	20	22	23	24	25	26	28	27	29	30	35	38	40	42	51	56	74	82	83
Treatment	Na	Mg	Al	Si	P	S	Cl	K	Ca	Ti	V	Cr	Mn	Fe	Ni	Co	Cu	Zn	Br	Sr	Zr	Mo	Sb	Ba	W	Pb	Bi
A	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X									
B	X	X	X	X		X	X		X	X				X					X	X	X	X	X	X	X	X	X
C	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
D	X	X	X	X		X	X		X	X				X													

II.H. Experiment 5: Contribution of Alternative Particle Size Fractions to Selectivity of VSP Combinations

VSP comprise a wide range of particle sizes. The smallest sizes can remain suspended in air and settle only with difficulty. The EPA has classified these particles as "$PM_{2.5}$" (those particles with aerodynamic diameters less than 2.5 µm). They are responsible for air pollution in the form of haze. Slightly larger particles are easily airborne and widely distributed by air. The EPA classification for these is "PM_{10}" (those particles with aerodynamic diameters less than 10 µm). Particles larger than 10 µm settle from air comparatively rapidly. It is reasonable to expect that particle size will affect how easily particles are transferred, how regularly they are generated and how persistent they are once in place. Because we know or expect that particles in different size ranges will be transported differently, it is reasonable to expect that the proportions of VSP classifications within a single specimen will differ depending on the size fraction. This could have important implications. It may be, for example, that the smaller size ranges of VSP are good regional indicators but lack local discrimination. Conversely, the larger sized fractions may provide good local discrimination. Understanding the range of applications for VSP, and the optimization of protocols for specific tasks, requires that the particle size dependence of VSP selectivity be better understood.

Based on the outcomes of Experiments 2 through 4 (see Sections III.B, III.C. and III.D., below) the VSP analysis baseline protocol was revised as given in Table 9. Analyses were completed for each of the 120 specimens in Table 2. Two approaches were made to the evaluation of VSP selectivity (strength of correspondence) by particle size. Firstly, the particles for each specimen were ranked by size (average diameter) and separated into quartiles, grouping from the smallest particles in Quartile 1 through the largest particles in Quartile 4. The selectivity of the VSP was measured for each of four quartiles. The second approach divided the VSP into size based on the PM_{10} classification. For each specimen the VSP selectivity for particles larger

than 10 µm was compared to the selectivity of an equal number of smaller sized particles.

Table 9. Changes in SEM/EDS Parameters for Experiment 5

Parameter	Setting	Change from Baseline Parameters
Minimum Particle Size	2.5 µm	Increased from 0.3 µm
Maximum X-Ray Collection Time	9 sec	Increased from 6 sec
Minimum X-ray Count	1000	Increased from 500
Target Elements	Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Co, Cu, Zn, Br, Sr, Zr, Mo, Sb, Ba, W, Pb, Bi	Added 9 elements
Element Threshold	3%	Increased from 1%
Maximum Number of Particles	10000	Increased from 5000

II.I. Experiment 6: Effect of Data Filtration Parameters on Selectivity of VSP Combinations

The evaluations of VSP selectivity use two data filtration parameters that work together to (1) remove particles that fail to show any dominant composition (as represented by the calculated percentages of the elements) and (2) remove elements that are only present in minute amounts. The parameters are ***N***, a specified number of elements and ***P***, a threshold proportion of a particle's composition that must be exceeded by adding the proportions of the particle's ***N*** most common elements. Particles that do not have ***N*** elements representing more than the ***P*** of their composition are considered as noise and disregarded. Compositional values for elements ranked below ***N*** elements, and that fall below 10%, are set to zero.

The baseline protocol uses parameters of ***P*** = 60% and ***N*** =5. These values were chosen as reasonable for initial proof-of-principle research, given the detection limits of the method and the numbers of elements typically found in individual VSP. Experiment 6 examined what effect these parameters have on the performance of particle combination analysis of VSP.

The number of elements, ***N*** = 5 used in prior computations was reduced to 3 and increased to 7. Change in this parameter will affect the numbers of particles with more complex chemical composition that are removed from consideration in particle combination analysis. Lower values of ***N*** will require that fewer elements contribute to the threshold proportion, filtering out more particles. Higher values of ***N*** will filter out fewer particles, allowing more complexity (and potentially more noise).

The threshold proportion, ***P*** = 0.60 used in prior computations was reduced to 0.45, and increased to 0.75. A lower value of ***P*** will require that the specified number of elements represent a smaller proportion of particle's composition, filtering out fewer particles and allowing more complexity (and potentially more noise). The higher value of ***P*** will allow less complexity and filter out more particles.

VSP data from Experiment 5 was used, with three levels for each of the two parameters, as given in Table 10. The three alternatives for each parameter result in nine combinations and VSP selectivity was measured for each.

Table 10. Treatments for Experiment 6: Effect of Data Filtration Parameters

	Low Value	Medium Value	High Value
N , number of elements	3	5	7
P , threshold proportion	0.45	0.6	0.75

III. Results

III.A. Introduction

The presentation of results employs box-and-whisker diagrams to visualize differences and variability in analytical results. This section provides a brief introduction to these charts for those who are not already familiar with them.

An example of a box-and-whisker diagram is given as Figure 4. The position and vertical extent of the shaded box indicates the range of the central 50% of the data values (bounded by the first and third quartile values, Q1 and Q3. The mean value is indicated by the central "X" and the median is indicated by the horizontal line. The lines extending from the shaded box ("whiskers") end at the maximum and minimum values of the data, excluding outliers. Outliers are shown by small circles. Mathematically, outliers are values greater than Q3 + 1.5x(Q3-Q1) or less than Q1 - 1.5x(Q3-Q1).

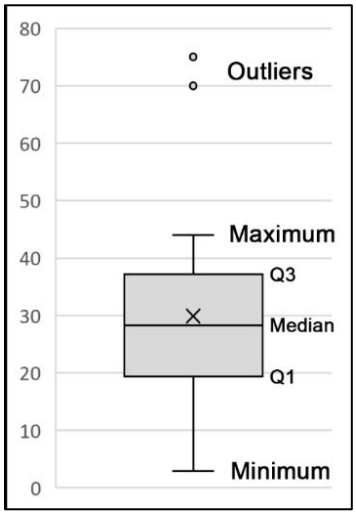
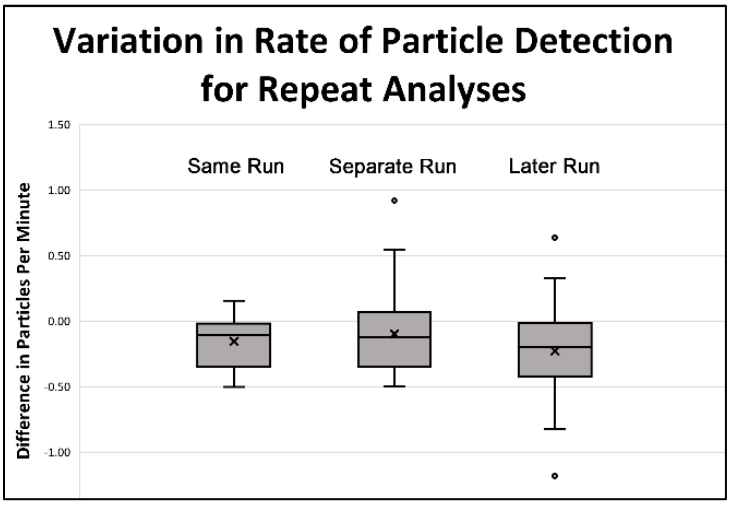


Figure 4. Example of a Box-and-Whisker Plot representation of data.

III.B. Reproducibility of VSP Analyses

Summary data for VSP analysis reproducibility are charted in Figures 5, 6 and 7 for analysis time (particles per minute), number of particles detected and characterized (particles per mm^2) and VSP selectivity (strength of correspondence). Full datasets are provided in Appendix B. Specific statistical analyses are discussed in the sections below. Overall, less variability was found when analyses were conducted on the same analytical run and there was a stronger correspondence for the sets of VSP. High rates of successful classification (greater than 90%) were found under all conditions. Different runs did not show important differences in the rates of particle detection or the numbers of particles found. Nearly all specimens showed many more particles than were needed for the analysis.

1. Analysis Time. The analysis time for a specimen varies from run to run. Compared to the differences seen on the same analytical run, no greater differences in these times were seen when runs were conducted on different days or at a much later time (single-factor ANOVA, $F_{2,93}$, $p = 0.19$). Comparable mean values are shown in Figure 5. However, the variability about the means differs, with less variability seen on the same analytical run (one-tail, two-sample F tests, $F_{1,31}$, $p < 0.001$). A difference in variability was not observed between the separate and later runs (one-tail, two-sample F test, $F_{1,31}$, $p = 0.27$).

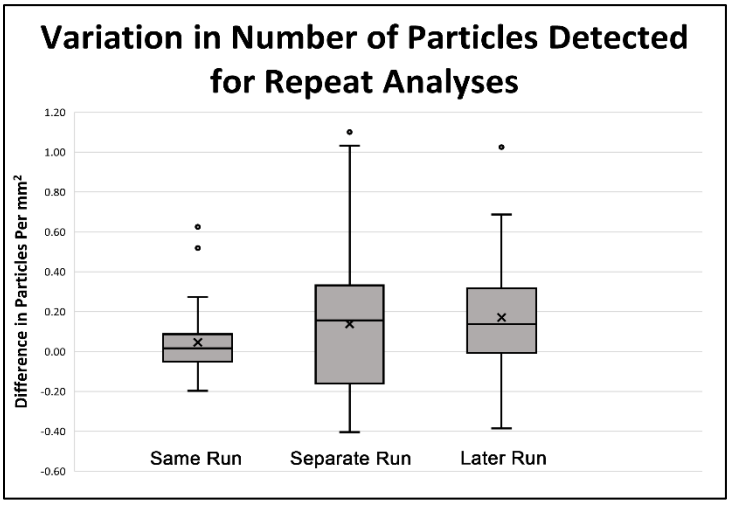


	N	Mean	Variance
Same Run	32	-0.156	0.030
Separate Run	32	-0.098	0.094
Later Run	32	-0.228	0.118

Figure 5. Variation in Rate of Particle Detection on Different Runs. The mean values (X positions on the chart, corresponding to the numbers in the table) are not significantly different by statistical tests. However, the variability is lower when analyses are on the same analytical run (note the larger distances between the whisker ends to the right of the chart and higher values for variance in the table).

2. Number of Particles Detected and Characterized. The results for the numbers of particles detected and characterized (particles per mm^2) parallel those just discussed above for the rates of particle detection. As can be seen in Figure 6, the values vary from run to run. Compared to the differences seen on the same analytical run, the slight increase when runs were conducted on different days or at a much later time were not statistically different (single-factor ANOVA, $F_{2,93}$, $p = 0.20$). Variance in numbers of particles detected and characterized was found to be significantly less for analyses on the same analytical run when compared to either of the later runs (one-tail, two-sample F tests, $F_{1,31}$, $p < 0.01$). When conducted on different analytical runs, whether later or after source replacement, there were no significant differences in variance (one-tail, two-sample F test, $F_{1,31}$, $p = 0.05$).

It is noteworthy that the number of particles available for analysis was not a limiting factor: many more particles were present than were needed for analysis in the almost all of the specimens. For example, of the 32 specimens used for the initial run (with analyses capped at 5000 particles or 6 hours) only two required analysis of more than 7% of the available area. The remainder showed an average examination area of 2%.



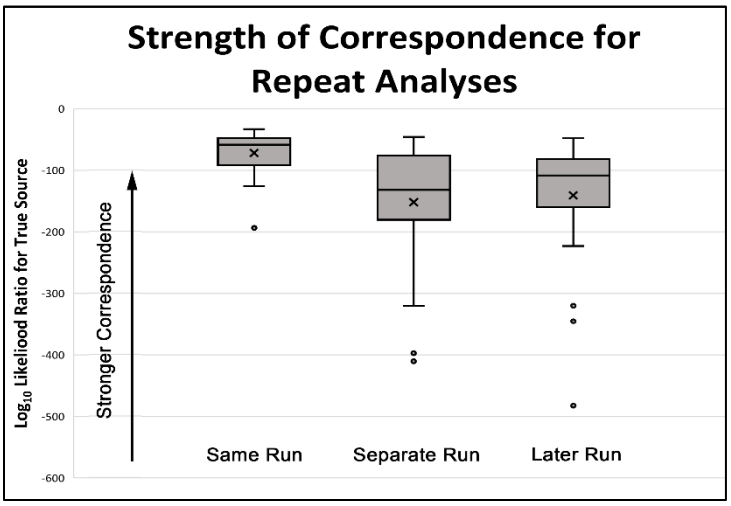
	N	Mean	Variance
Same Run	32	0.046	0.029
Separate Run	32	0.138	0.139
Later Run	32	0.170	0.077

Figure 6. Variation in Numbers of Particles Detected on Different Runs. The results parallel those for the rate of particle detection, as shown in Figure 5. The mean values (X positions on the chart, corresponding to the numbers in the table) are not significantly different by statistical tests. However, the variability is lower when analyses are on the same analytical run (note the wider gray areas and larger distances between the whisker ends to the right of the chart and higher values for variance in the table).

3. VSP Selectivity. Reproducibility of VSP selectivity (the strength of correspondence) was tested by comparing the results from the initial analysis with three subsequent analyses under the three alternative conditions (same run, separate run and later run). Results are shown in Figure 7, with strength of correspondence measured by the likelihood ratio supporting the association. The mean values show stronger correspondence and less variance when analyses are conducted on the same analytical run. High rates of correct classification were seen under all conditions (97% for the same run, 91% for separate runs and 94% for later runs).

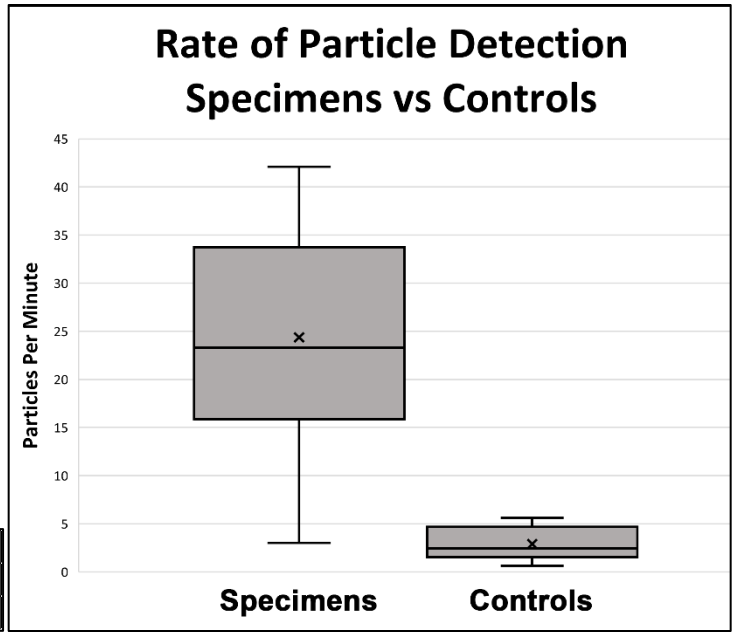
The hypothesis of equal means was rejected (ANOVA, $F_{2,90}$, $p < 0.001$) with greater VSP selectivity for the same run condition. No statistically significant difference was observed between mean values for separate and later runs (two-tailed, two-sample t-test, equal variances, 60 df, $p = 0.64$). Variance in the selectivity was found to be significantly less for repeat analyses on the same analytical run when compared to either of the other runs (one-tail, two-sample F tests, $F_{1,30}$, $p < 0.001$). When conducted on different analytical runs (whether the separate run or later run) there were no significant differences in variance (one-tail, two-sample F test, $F_{1,30}$, $p = 0.44$, one specimen missing due to an unusable likelihood ratio for the separate run).

4. Control Specimen Results. As shown in Figures 8 and 9 the Control Specimens showed very low rates of particle detection and very low numbers of particles detected.



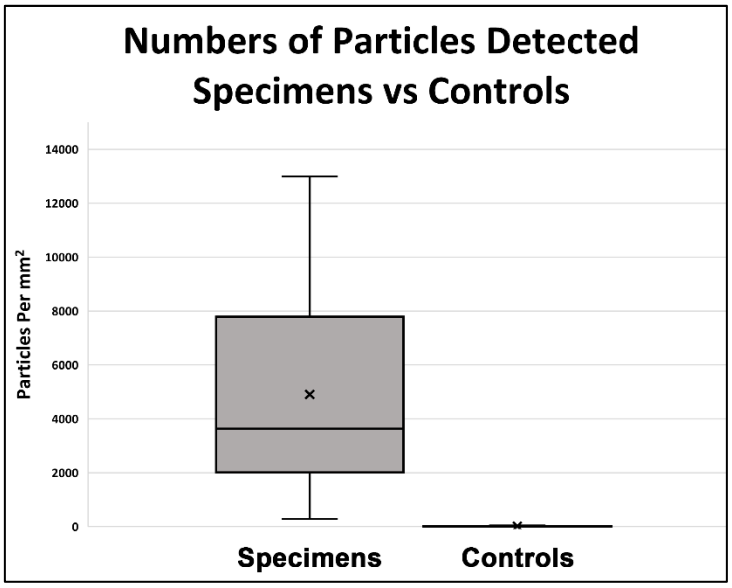
	N	Mean	Variance
Same Run	31	-71.88	1218.55
Separate Run	31	-151.90	9441.65
Later Run	31	-140.71	8896.56

Figure 7. Variation in Strength of Correspondence for Repeat Analyses. The mean values (X positions on the chart, corresponding to the numbers in the table) show stronger correspondence for VSP when analyses are conducted on the same analytical run. The variance is also less for same run analyses, as shown by the shorter gray bar and whiskers.



	N	Mean	Variance
Specimens	32	24.4	125.05
Controls	9	2.9	2.98

Figure 8. Rate of Particle Detection Specimens vs Controls. The mean values (X positions on the chart, corresponding to the numbers in the table) show a much lower rate of particle detection for the control specimens.



	N	Mean	Variance
Specimens	31	4901.7	1.55E+07
Controls	9	15.5	128.4

Figure 9. Numbers of Particles Detected Specimens vs Controls. The mean values (X positions on the chart, corresponding to the numbers in the table) show extremely small numbers of particles detected for the control specimens. (One specimen outlier, with over 35,000 particles per mm^2, was omitted.)

III.C. SEM/EDS Particle Detection Parameters

Summary data for evaluation of the effects of SEM/EDS particle detection parameters are charted in Figures 10 through 15 for analysis time (particles per minute), number of particles detected and characterized (particles per mm^2) and VSP selectivity (strength of correspondence). Full datasets are provided in Appendix B. Specific statistical analyses are discussed in the sections below. Overall (1) greater magnification and wider particle size limits were both found to result in significantly longer analytical times, and (2) restrictions on particle size or brightness resulted in fewer particles being detected and characterized. None of the variables influenced the strength of correspondence and almost all of the analyses (more than 99%) resulted in the correct classifications of specimens to their true source.

1. Analysis Time. Main effects (Figure 10) were found for the variables of *Magnification* and *Particle Size Limits* (ANOVA, main effects and interactions of means, $F_{1,112}$, $p < 0.001$). Longer analytical times (fewer particles per minute) were found when greater magnifications were used (effectively increasing the area to be searched) as well as for a wider range of acceptable particle sizes. No main effects were found for *Time for Image Collection* or *Particle Brightness* ($p = 0.81$ and 0.34, respectively). However, interactive effects (Figures 11 and 12) were detected between *Particle Brightness* and both the *Time for Image Collection* ($p = .006$) and *Particle Size Limits* ($p < 0.001$).

The interaction between the variables *Time for Image Collection* and *Particle Brightness.* is shown in Figure 11. For dimmer particles, higher rates of particle detection are seen when imaging times are longer (even though it takes a longer time to detect the particles, more particles are detected). However, for brighter particles higher rates are seen for shorter imaging times (there is no advantage for the longer imaging time since the brighter particles are already detected).

The interaction between the variables *Particle Size Limits* and *Particle Brightness* is shown in Figure 12. The wider size limits show higher rates of particle detection than the narrower limits, but the difference is smaller when only brighter particles are counted. Many of the smallest particles will be dim, so the wider size limits (which include these smaller particles) have a higher rate of detection when dim particles are included.

2. Number of Particles Detected and Characterized. Main effects (Figure 13) were found for the variables of *Particle Brightness* and *Particle Size Limits* (ANOVA, main effects and interactions of means, $F_{1,112}$, $p < 0.001$). These variables also showed an interactive effect ($p = 0.002$, Figure 14). Requiring a higher threshold for particle brightness results in significantly fewer particles being detected, as does a narrower range of acceptable particle sizes. The interactive effect reveals that when a narrower range of acceptable particles is used, higher particle brightness has a much reduced effect. The same two variables also had an interactive effect for the rate of particle detection (Figure 12) and the likely cause is the same. Many of the smallest particles will be dim, so the wider size limits (which include these smaller particles) show higher numbers when dim particles are included.

3. VSP Selectivity. None of the variables showed main effects or interactions for the strength of correspondence (Figure 15, ANOVA, main effects and interactions of means, $F_{1,111}$, p values ranging from 0.07 to 0. 65). Of the 128 analyses, all but one of the specimens was correctly classified to its true source. (The misclassification occurred for the specimen with the smallest number of particles detected, under conditions where the total number of particles fell below 500.)

			N	Mean	Variance
Time for Image Collection	**Short**	2 µsec	64	18.65	114.19
	Long	8 µsec	64	18.30	128.35
Particle Brightness	**Dim**	64	64	19.17	152.53
	Bright	128	64	17.78	89.10
Magnification	**Low**	1200X	64	23.69	110.19
	High	2400X	64	13.26	77.10
Particle Size Limits	**Wide**	0.3-80µm	64	22.93	139.58
	Narrow	2.5-40µm	64	14.02	62.62

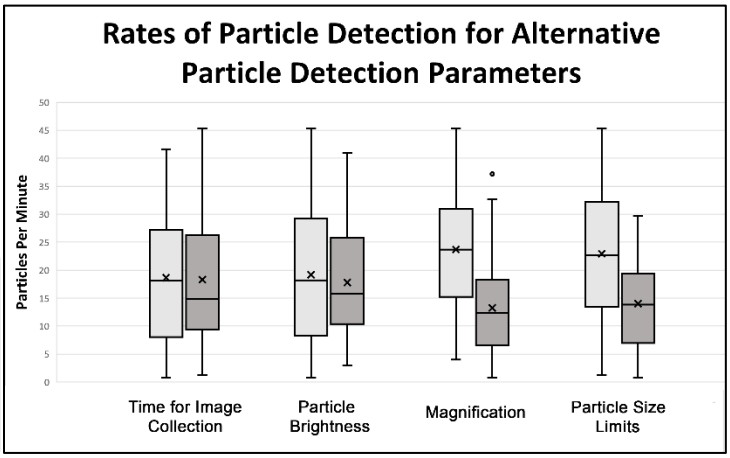


Figure 10. Main Effects of Particle Detection Parameters on Rate of Particle Detection. Changes in *Time for Image Collection* and *Particle Brightness* show little effect on the rate of particle detection, while higher *Magnification* and narrower *Particle Size Limits* both resulted in lower rates of particle detection.

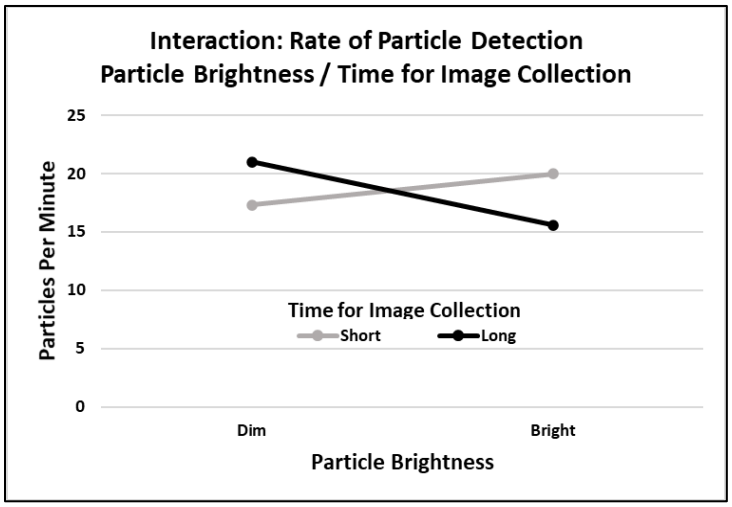


Figure 11. Interactive Effect of *Time for Image Collection* and *Particle Brightness* on Rate of Particle Detection. For dimmer particles, higher rates of particle detection are seen when imaging times are longer (even though it takes a longer time to detect the particles, more particles are detected). However, for brighter particles higher rates are seen for shorter imaging times (there is no advantage for the longer imaging time since the brighter particles are already detected).

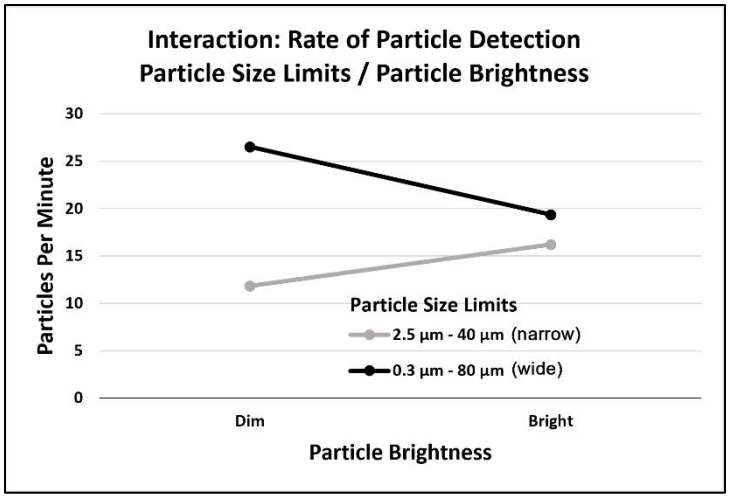


Figure 12. Interactive Effect of *Particle Size Limits* and *Particle Brightness* on Rate of Particle Detection. The wider size limits (black line) show higher rates of particle detection than the narrower limits (gray line), but the difference is smaller when only brighter particles are counted. Many of the smallest particles will be dim, so the wider size limits (which include these smaller particles) have a higher rate of detection when dim particles are included.

			N	Mean	Variance
Time for Image Collection	**Short**	2 µsec	64	0.975	1.136
	Long	8 µsec	64	1.025	1.705
Particle Brightness	**Dim**	64	64	1.616	1.971
	Bright	128	64	0.384	0.101
Magnification	**Low**	1200X	64	1.080	1.742
	High	2400X	64	0.920	1.087
Particle Size Limits	**Wide**	0.3-80µm	64	1.599	2.069
	Narrow	2.5-40µm	64	0.401	0.045

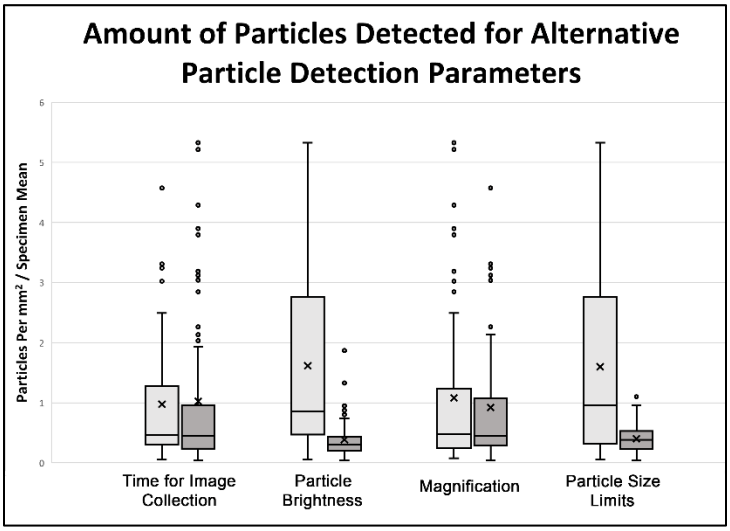


Figure 13. Main Effects for Particle Detection Parameters on Numbers of Particles Detected. Changes in *Time for Image Collection* and *Magnification* have little effect on the numbers of particles, while a higher setting for *Particle Brightness* or narrower *Particle Size Limits* show major reductions in particle numbers per mm^2. The higher brightness setting will ignore dimmer particles and the narrower particle size range is more restrictive.

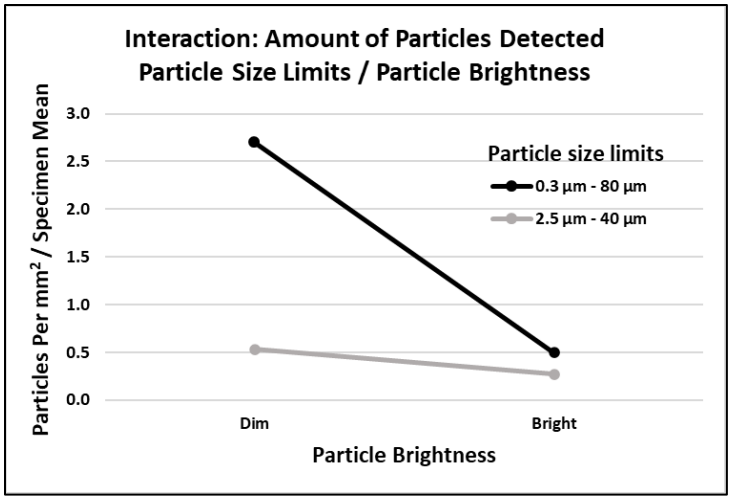


Figure 14. Interactive Effect of *Particle Size Limits* and *Particle Brightness* on Numbers of Particles Detected. In addition to their main effects (Figure 13), there is also an interaction between these variables (parallel to that seen in Figure 12 for the rate of particle detection). There are fewer particles when only bright ones are counted (both lines go down), but for the wider size limits (black line) there is a much more dramatic decrease. Many of the smallest particles will be dim, so the wider size limits (which include these smaller particles) show higher numbers when dim particles are included.

			N	Mean	Variance
Time for Image Collection	**Short**	2 µsec	64	-30.50	39.30
	Long	8 µsec	64	-30.84	38.30
Particle Brightness	**Dim**	64	64	-30.03	63.12
	Bright	128	64	-31.31	14.10
Magnification	**Low**	1200X	64	-31.69	39.96
	High	2400X	64	-29.64	35.53
Particle Size Limits	**Wide**	0.3-80µm	64	-30.40	60.54
	Narrow	2.5-40µm	64	-30.95	16.61

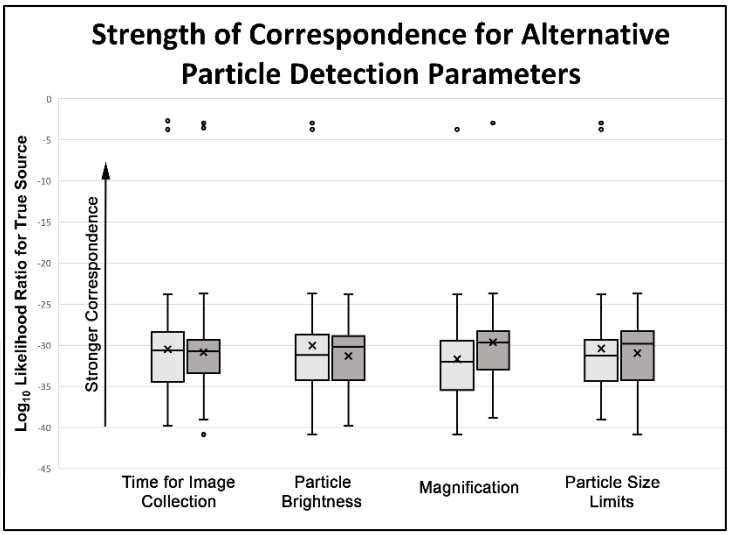


Figure 15. Main Effects for Particle Detection Parameters on Strength of Correspondence of VSP. No main effects or interactions were detected for any of the particle detection variables.

III.D. SEM/EDS X-Ray Analysis Parameters

Summary data for evaluation of the effects of SEM/EDS x-ray analysis parameters are shown charted in Figures 16 through 22 for analysis time (particles per minute and VSP selectivity (strength of correspondence). Full datasets are provided in Appendix B. Specific statistical analyses are discussed in the sections below. Each of the variables showed main effects on analysis time, and there were four significant interactive effects, each of which has a reasonable explanation. The variable of *Minimum X-Ray Count* showed a main effect on VSP selectivity, along with an interactive effect with the *Element Threshold*. All combinations of the SEM/EDS x-ray analysis parameters showed correct classifications of each of the specimens.

1. Analysis Time. Main effects (Figure 16) were found for each of the variables. Significantly faster analysis times (more particles detected per minute) were the result of increases in *Maximum X-ray Collection Time* and an increased *Element Threshold* (ANOVA, main effects and interactions of means, $F_{1,112}$, $p < 0.001$). Significantly slower analytical times were the result of increases in either the *Minimum X-ray Count* ($p = .002$) or the *Target Single Element X-Ray Count* ($p < 0.001$).

Reasonable explanations of for three of these effects result from consideration of the flowchart in Figure 3. Increasing the required *Minimum X-Ray Count* results in fewer particles meeting the criteria as a particle of interest, decreasing the rate of detection. Increasing the *Target Single Element X-Ray Count* means that this end condition (which takes less time) will be met less frequently, decreasing the rate of detection. Increasing the *Element Threshold* favors particles with high x-ray counts, which are more likely to reach the target single element x-ray count end condition (which takes less time), increasing the rate of detection. The increased rate of particle detection with an increase in *Maximum X-Ray Collection Time* is not easily explained based on main effects alone: the time increases, and the number of particles remains the same (the added time serving to provide better analytical data on each particle). However, there are significant interactive effects.

The *Target Single Element X-Ray Count* variable showed significant interactive effects (Figures 17 to 19) with each of the other three variables ($p < 0.001$) and the *Maximum X-Ray Collection Time* variable showed an interactive effect with the *Element Threshold* (Figure 20, $p = .008$). Each of these interactions can be reasonably explained, with reference to the flow chart in Figure 3.

The interactive effect for *Maximum X-Ray Collection Time* and *Target Single Element X-ray Count* (Figure 17) shows that for shorter x-ray collection times there is little change in the rate of particle detection when the target single element count is raised. Likely, the shorter collection time is not long enough to reach the higher count. For longer x-ray collection times there is a much higher rate of particle detection when the target single element count is lower. Likely, the lower count is reached before the maximum x-ray collection time. This end condition saves time, resulting in higher rates of particle detection. Much lower rates are seen when the target single element count is raised. It is likely that the even the longer collection time is insufficient to reach the higher count, so this end condition is not met and there is no resulting rate increase.

The interactive effect for *Minimum X-Ray Count* and *Target Single Element X-Ray Count* (Figure 18) shows that when the target single element x-ray count is lower, an increase in the minimum x-ray count has little effect on the rate of particle detection. The expected lower rate for a higher minimum x-ray count (fewer particles meeting the minimum) is apparently offset by

greater numbers of particles meeting the lower target single element count end condition (resulting in less analytical time).

The interactive effect for *Element Threshold* and *Target Single Element X-Ray Count* (Figure 19) shows that for a lower target single element x-ray count, the effect of increasing the *Element Threshold* (which favors particles with high x-ray counts), results in reaching the target single element x-ray count end condition more often (taking less time and increasing the rate of detection). However, no such rate increase is seen when there is a higher target single element count.

The interactive effect for *Element Threshold* and *Maximum Collection Time* (Figure 20) shows that particles meeting a higher element threshold are more likely to reach the single target element end condition within a shorter maximum collection time (reducing the analysis time and increasing the rate of particle detection). With longer maximum collection times, there is less of a difference between the two element threshold conditions (more comparable numbers of particles reaching the single target element end condition).

			Mean	Variance
Maximum X-ray Collection Time	**Low**	6 sec	22.65	217.20
	High	9 sec	29.75	215.59
Minimum X-Ray Count	**Low**	500	29.10	175.54
	High	1000	23.31	265.81
Target Single Element Count	**Low**	2500	31.27	286.01
	High	5000	21.14	120.28
Element Threshold	**Low**	1%	22.05	162.56
	High	3%	30.35	260.86

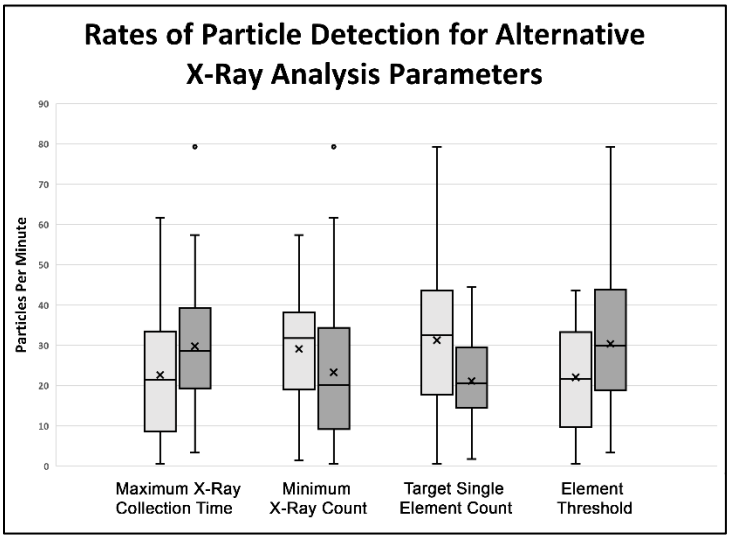
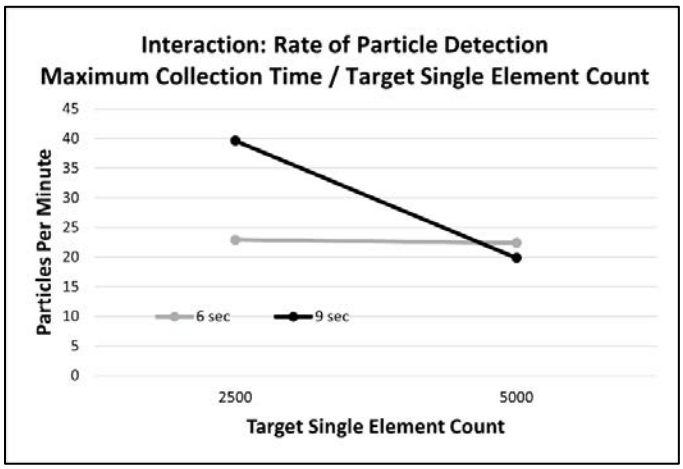


Figure 16. Main Effects for X-ray Analysis Parameters on Rates of Particle Detection. Each of Each of the four variables showed a statistically significant main effect. The effect of increasing the *Maximum X-Ray Collection Time* is counter-intuitive, as a greater rate of particle detection arises from increasing the time and no greater numbers of particles are expected. This result is likely explained by the variable interactions (see text and Figures 17 through 21). Each of the other main effects can be readily explained from the overall process (flowchart in Figure 3). Increasing the required *Minimum X-Ray Count* means that fewer particles will meet the criteria as a particle of interest, decreasing the rate of detection. Increasing the *Target Single Element X-Ray Count* means that this end condition (which takes less time) will be met less frequently, decreasing the rate of detection. Increasing the *Element Threshold* favors particles with high x-ray counts, which are more likely to reach the target single element x-ray count end condition (which takes less time), increasing the rate of detection.



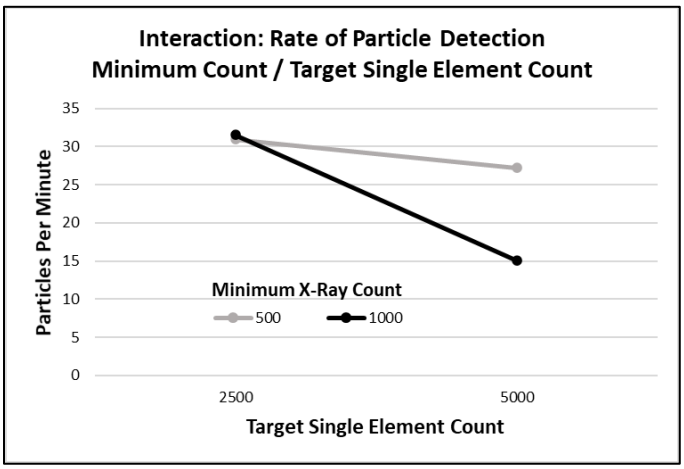
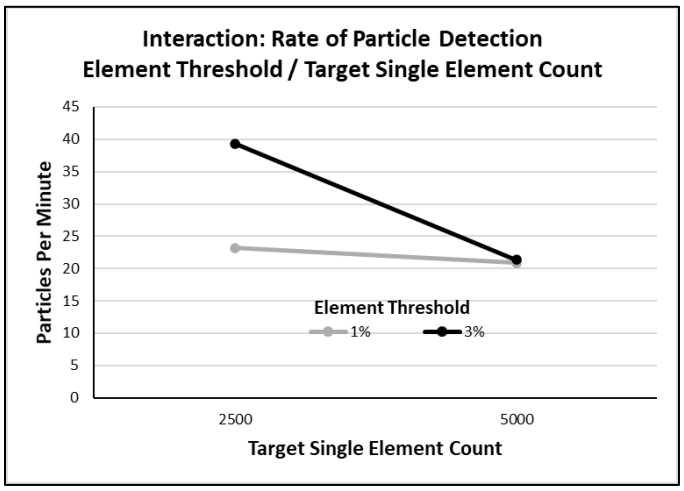


Fig. 17 (above left). Interactive Effect of *Maximum X-Ray Collection Time* and *Target Single Element X-ray Count* for Rate of Particle Detection. For shorter x-ray collection times (gray line) there is very little change in the rate of particle detection when the target single element count is raised. Likely, the 6 second collection time is not long enough to reach the higher count. For longer x-ray collection times there is a much higher rate of particle detection (black line) when the target single element count is lower (2500). Likely, the lower count is reached before the maximum x-ray collection time of 9 seconds. This end condition saves time, resulting in higher rates of particle detection. Much lower rates are seen when the target single element count is raised to 5000. It is likely that the even the 9 second collection time is not long enough to reach the higher count, so this end condition is not met and there is no resulting increase in the rate.

Figure 18 (above right). Interactive Effect of *Minimum X-Ray Count* and *Target Single Element X-Ray Count* for Rate of Particle Detection. Rates of particle detection decrease when the target single element count is raised to 5000 (both lines trend downward, which is the main effect for *Target Single Element X-Ray Count* shown in Figure 16). However, the effect is much a more dramatic when the minimum x-ray count is higher (black line). Note that although the main effect for increasing the *Minimum X-Ray Count* is a decrease in the rate of particle detection (Figure 16), little effect is seen when the target single element count is set to 2500 (the black and gray lines join at the left). The expected lower rate for a higher minimum x-ray count (fewer particles meeting the minimum) is apparently offset by greater numbers of particles meeting the lower target single element count end condition (resulting in less analytical time). With a higher target single element count (5000), this end condition is seldom reached and the expected main effect of increasing the minimum x-ray count is seen (much lower rates for the higher minimum x-ray count, black line).



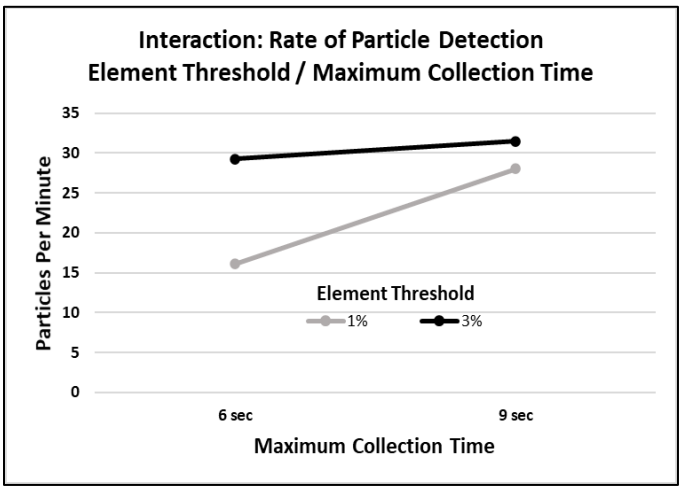


Figure 19 (above left). Interactive Effect of *Element Threshold* and *Target Single Element X-Ray Count* for Rate of Particle Detection. The main effect for a higher element threshold is an *increase* in the rate of particle detection (Figure 16). Increasing the element threshold favors particles with high x-ray counts, which are more likely to reach the target single element x-ray count end condition (which takes less time), increasing the rate of detection. However, this effect is only seen with the lower target single element count (2500, greater rate for the black line). At the higher target single element count (5000) there is a negligible effect (black and gray lines show the same rate).

Figure 20 (above right). Interactive Effect of *Element Threshold* and *Maximum Collection Time* for Rate of Particle Detection. The main effect for increasing the maximum collection time is an increased rate (Figure 16). This is shown for both levels of the element threshold (both black and gray lines move upward). However, the effect is much more dramatic for the lower element threshold. Particles meeting the higher element threshold are more likely to reach the single target element end condition within the shorter maximum collection time, reducing the analysis time and increasing the rate of particle detection. With longer maximum collection times, there is less of a difference between the two element threshold conditions: there are more comparable numbers of particles reaching the single target element end condition.

2. VSP Selectivity. As shown in Figure 21, only the *Minimum X-Ray Count* showed a main effect on the strength of correspondence based on the $\log_{10}$ likelihood ratio supporting association (ANOVA, main effects and interactions of means, $F_{1,112}$, $p = 0.04$). Notably, this effect is close to the chosen significance level, the means are very close and the variances are unequal (two-sample test for equal variances, $F_{1,63}$, $p < 0.001$). There was also an interaction detected between the *Minimum X-Ray Count* and the *Element Threshold* (Figure 22, $p = 0.03$). This interaction indicates that the main effect for stronger correspondence with an increase in minimum x-ray count is associated with the lower element threshold setting.

			Mean	Variance
Maximum Duration	**Low**	6 sec	-32.87	9.59
	High	9 sec	-33.51	5.99
Minimum Counts	**Low**	500	-33.69	4.74
	High	1000	-32.69	10.53
Target Counts	**Low**	2500	-33.38	9.59
	High	5000	-33.00	6.11
Element Threshold	**Low**	1%	-32.81	9.13
	High	3%	-33.56	6.36

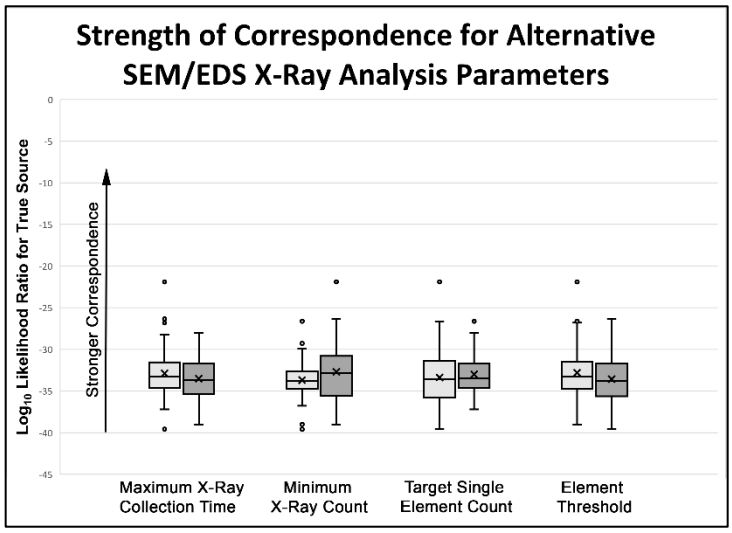


Figure 21. Main Effects for X-Ray Analysis Parameters on Strength of Correspondence. Although Minimum Counts showed a main effect, this finding is questionable as the significance was borderline, the means are close, and the variances are unequal. All specimens were correctly classified.

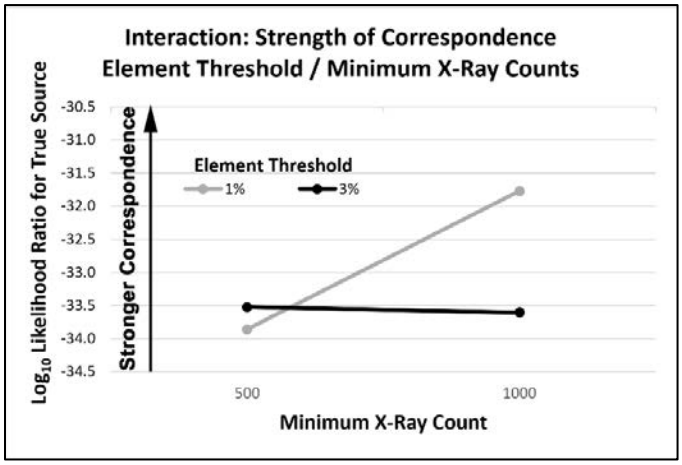


Figure 22. Interactive Effect of *Minimum X-Ray Count* and *Element Threshold* for Strength of Correspondence. Changing the element threshold influences the effect of changing the minimum x-ray count. The main effect (stronger correspondence with an increase in minimum x-ray count) is associated with the lower element threshold setting.

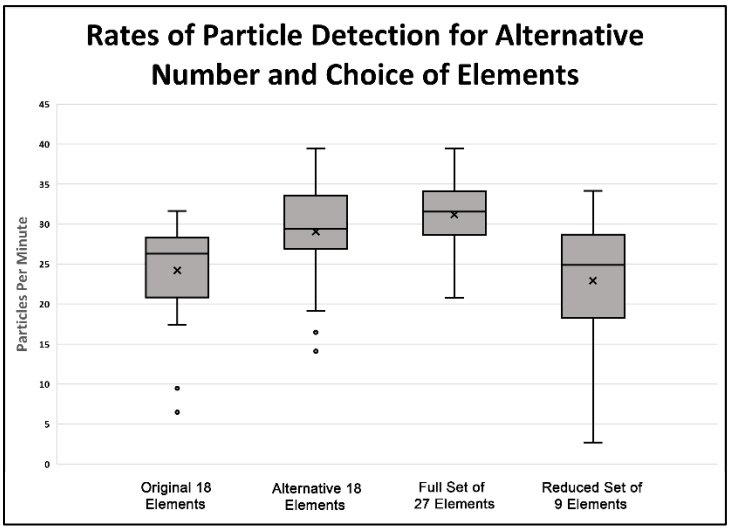
III.E. Number and Choice of Elements

Summary data for evaluation of the effects of the number and choice of elements are shown charted in Figures 23 through 25 for analysis time (particles per minute), number of particles detected and characterized (particles per mm^2) and VSP selectivity (likelihood ratio for association). Full datasets are provided in Appendix B. Specific statistical analyses are discussed in the sections below. Overall (1) shorter (but equivalent) analytical times were found for two element combinations (along with longer, but equivalent analytical times for the other two combinations), (2) using the greatest number of elements increased the numbers of particles detected, and (3) using the greatest number of elements improved VSP selectivity.

1. Analysis Time. Figure 23 shows faster analytical times (more particles detected per minute) resulting from the alternative set of 18 elements (treatment B) and the full set of 27 elements (treatment C), along with slower analytical times for the other two treatments. The hypothesis of equal means was rejected (single-factor ANOVA, $F_{3,124}$, $p < 0.001$). Pairwise contrasts showed no difference between the (lower) analysis times for treatments B and C (two-sample t-test, $p = 0.09$) or between the (longer) analysis times for treatments A and D (two-sample t-test, $p = .44$).

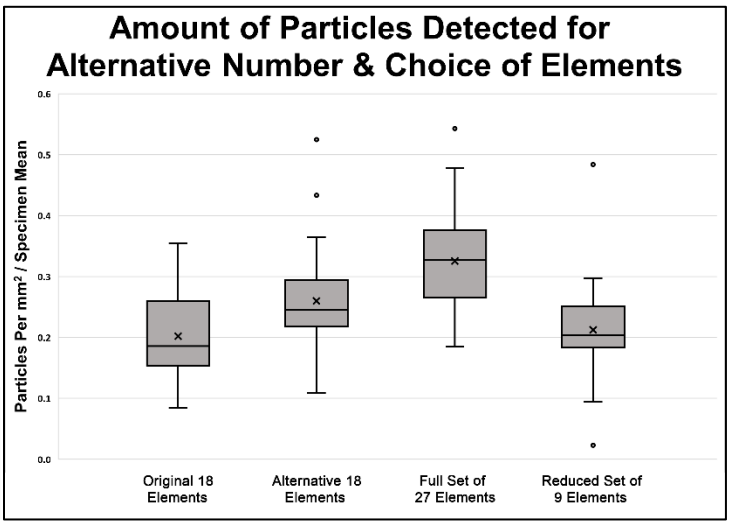
2. Number of Particles Detected and Characterized. Figure 24 shows that the numbers of particles detected using the alternative numbers and choice of elements are unequal (single-factor ANOVA, $F_{3,124}$, $p < 0.001$). The greatest numbers of particles were detected using the full set of 27 elements, followed by the alternative set of 18 elements (paired two-sample t-tests, $p < 0.001$). The original set of 18 elements showed comparable numbers of particles to the reduced set of 9 elements (paired two-sample t-test, $p = 0.56$).

3. VSP Selectivity. Differences in VSP selectivity (strength of correspondence) were observed (Figure 25, single-factor ANOVA, $F_{3,124}$, $p < 0.001$), with the full set of 27 elements showing the highest (least negative) $\log_{10}$ likelihood ratio in supporting association (Figure 18). No differences in selectivity were detected between the original and alternative sets of 18 elements (paired two-sample t-test, $p = 0.19$). The reduced set of 9 elements showed lower selectivity compared to the original set of 18 (paired two-sample t-test, $p < 0.008$)



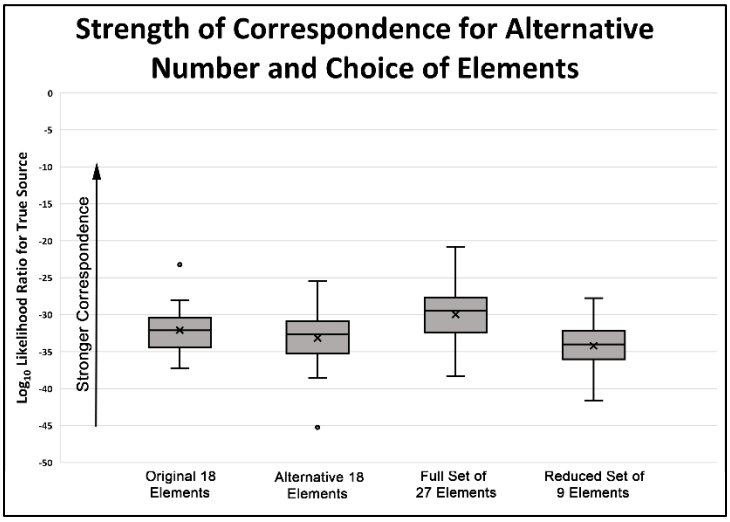
	Treatment	*N*	Mean	Variance
A	Original 18 Elements	32	24.22	36.11
B	Alternative 18 Elements	32	29.04	31.98
C	Full Set of 27 Elements	32	31.18	17.46
D	Reduced set of 9 Elements	32	22.92	53.84

Figure 23. Effect of Number and Choice of Elements on Rate of Particle Detection. Faster analytical times result from the alternative set of 18 elements and the full set of 27 elements.



	N	Mean	Variance
Original 18 Elements	32	0.202	4.18E-03
Alternative 18 Elements	32	0.260	6.36E-03
Full Set of 27 Elements	32	0.326	7.08E-03
Reduced set of 9 Elements	32	0.212	5.91E-03

Figure 24. Effect of Number and Choice of Elements on Numbers of Particles Detected. The numbers of particles detected are unequal when using alternative numbers and choice of elements. Under the experimental conditions the full set of 27 elements showed the greatest numbers of particles detected.



	N	Mean	Variance
Original 18 Elements	32	-32.11	8.24
Alternative 18 Elements	32	-33.17	15.29
Full Set of 27 Elements	32	-29.98	15.42
Reduced set of 9 Elements	32	-34.22	10.87

Figure 25. Effect of Number and Choice of Elements on the Strength of Correspondence. The full set of 27 elements provided greater selectivity.

III.F. Contribution of Alternative Particle Size Fractions to VSP Selectivity

Summary data for evaluation of the contribution of alternative particle size fractions to VSP selectivity are shown in Figures 26 through 28. Full datasets are provided in Appendix B. Specific statistical analyses are discussed below. The particle size fractions (alternative quartiles of each VSP dataset) were found to have a small, regular increase in VSP selectivity, with increased particle size. However, smaller diameter particles (respirable and easily transported by air) were found to have greater selectivity than larger particles. Clearly particle size is an important factor, but the nature of the relationship is not simple and merits further examination.

Figure 26 compares VSP selectivity showing the trend of slightly higher (less negative) $\log_{10}$ likelihood ratios for association with increasing particle size. The hypothesis of equal means is rejected (single-factor ANOVA, $F_{3,421}$, $p = 0.018$). Quartile 4 (larger particles) show a significantly higher selectivity than Quartile 1 or Quartile 2 (one-tail two-sample t-tests, unequal variances, $p = 0.001$ and 0.015, respectively).

Figure 27 shows the effect of the alternative particle size quartiles on the *rank of the true source*. The rank of the true source is one measure of VSP selectivity. A rank of 1 is a "hit" indicating that the highest correspondence (among a library of 120 specimens) was found between a specimen and its actual source. This results in a correct classification, as the classification is assigned based on the highest ranking. Other rankings indicate incorrect classifications. Rankings of 2, 3, etc. indicate that the actual source was ranked second, third and so forth among the 120 specimens. A ranking of zero indicates that the measure of correspondence was too low for comparison. The four quartiles show similar performance (means of 1/Rank, single-factor ANOVA, $F_{3,476}$, $p = 0.85$).

Figure 28 contrasts the strength of correspondence for equal numbers of particles in two size classes: those greater than 10 µm and those in the range of 2.5 to 10 µm. The latter range includes the easily respirable particles, designated "PM_{10}" for environmental and health assessments. These particles are also those that are most easily airborne and transferred by wind, whereas those particles larger than 10 µm would transfer by other, more direct means. Because of the differences in transfer properties, it is reasonable to suppose that the two particle size fractions could differ substantially. Posterior probabilities of matching the true source within a closed set of more than 100 sources were used to contrast the selectivity of the two populations. As shown in Figure 28, the smaller sized PM_{10} particles were found to have greater selectivity (two-sample t-test for equal means, one tail, $p = .007$).

	N	Mean	Variance
First Quartile	109	-15.60	13.84
Second Quartile	105	-15.26	19.25
Third Quartile	106	-14.52	17.14
Fourth Quartile	105	-13.99	16.69

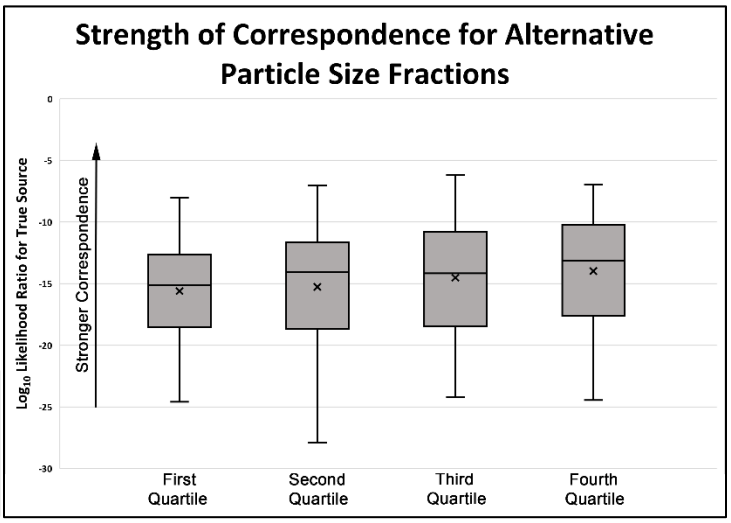


Figure 26. Effect of Alternative Particle Sizes on Strength of Correspondence. Little difference is seen in among the four particle size quartiles.

	Correct Classifications	Percentage	Mean 1/Rank of True
First Quartile	82	0.68	92.5
Second Quartile	80	0.67	88.9
Third Quartile	78	0.65	89.5
Fourth Quartile	85	0.71	93.0

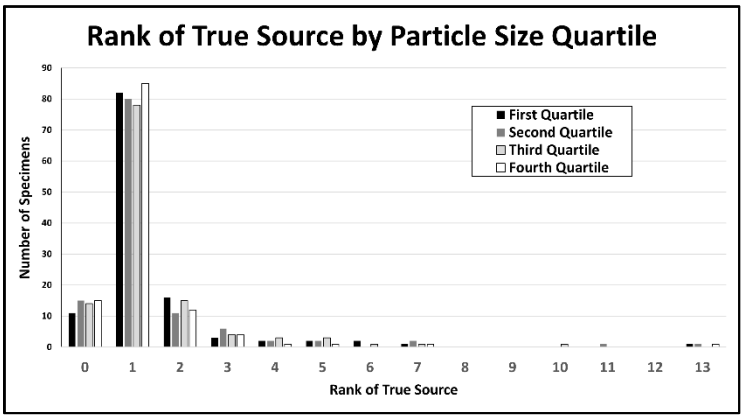


Figure 27. Effect of Alternative Particle Size Quartiles on the Rank of the True Source. The rank of the true source is one measure of VSP selectivity. A rank of 1 is a "hit" indicating that the highest correspondence (among a library of 120 specimens) was found between a specimen and its actual source. This results in a correct classification, as the classification is assigned based on the highest ranking. Other rankings indicate incorrect classifications. Rankings of 2, 3, etc. indicate that the actual source was ranked second, third and so forth among the 120 specimens. A ranking of zero indicates that the measure of correspondence was too low for comparison. The four quartiles show similar performance.

Posterior Probability for Matching True Source			
	N	Mean	Variance
PM_{10} Particles (windborne)	103	0.65	0.103
Larger Particles	105	0.55	0.094

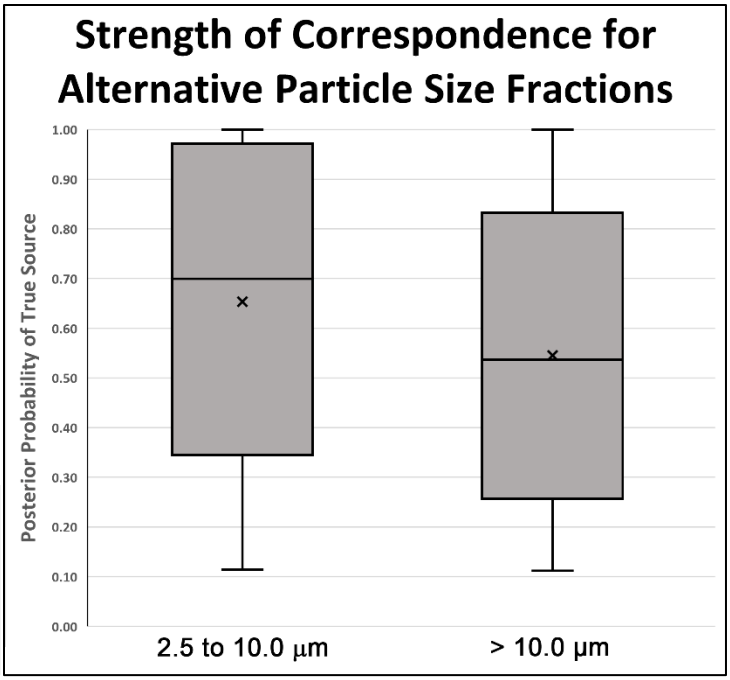


Figure 28. Comparison of the strength of correspondence (posterior probability of the true source) for equal numbers of particles in the alternative size fractions of 2.5 to 10.0 µm (PM_{10}) (left) versus particles greater than 10.0 µm (right). The smaller sized PM_{10} particles were found to have greater selectivity.

III.G. Effect of Data Filtration Parameters

Summary data for evaluation of the effect of the two data filtration parameters on VSP selectivity are shown in Figures 29 to 31. Full datasets are provided in Appendix B. Specific statistical analyses are discussed below. Overall, the data filtration parameters showed no measurable effect on selectivity.

Figure 29 shows the effect of data filtration on the sum of posterior probabilities for 30 specimens of each of four evidence types. The number of elements demonstrated no main effects or interactions (ANOVA, main effects, $F_{2,27}$, $p = 0.95$; interactions $F_{4,27}$, $p = 0.99$).

For context, Figure 30 shows differences in these sums for each of the four evidence types under each of the 9 combinations of the data filtration parameters. There are clear distinctions in the strength of correspondence for the different sets of VSP, as expected from the analyses in [8] (single-factor ANOVA for equal means, $F_{3,32}$, $p < 0.001$). Figure 31 shows that the alternative data filtration parameters have dramatically different effects on the numbers of particles that are removed from the analysis. When 7 elements are used almost no particles are removed from the analysis. The same is true for 5 elements and the lower threshold proportions (dashed line). Larger numbers of particles are removed when only 3 elements are required to meet the threshold proportion, with 40% of the particles removed when 3 elements are required to make up 75% of the particles measured elemental concentration. Despite these dramatic effects on the numbers of particles, this reduction did not cause an overall measurable effect on selectivity: sufficient numbers of particles remain.

			Mean	Variance
Number of Elements	**Low**	3	22.92	37.50
	Medium	5	22.23	36.46
	High	7	22.30	26.48
Threshold Proportion	**Low**	0.45	24.69	18.59
	Medium	0.60	20.68	35.26
	High	0.75	22.08	37.86

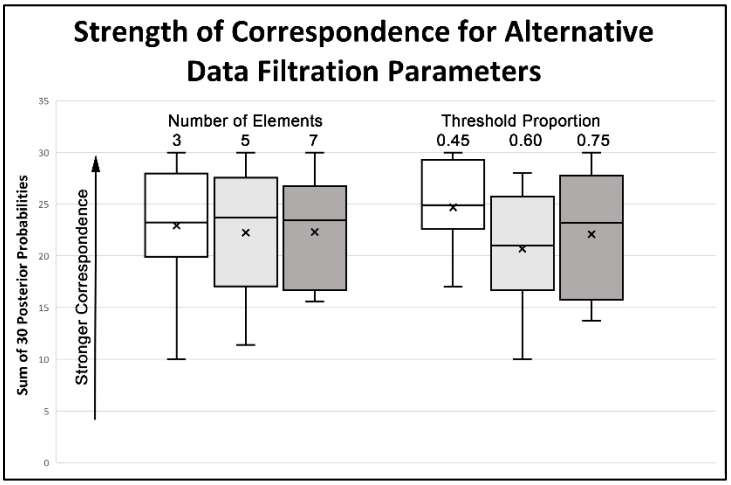


Figure 29. VSP Selectivity from Alternative Data Filtration Parameters – Sum of Posterior Probability of True Source for 30 specimens of 4 evidence types. No statistically significant effect was observed for *Threshold Proportion* or the *Number of Elements*.

	N	Mean	Variance
Cell Phones	9	16.60	25.54
Firearms	9	20.60	9.06
Face Masks	9	24.52	12.56
Drug Packaging	9	28.22	6.67

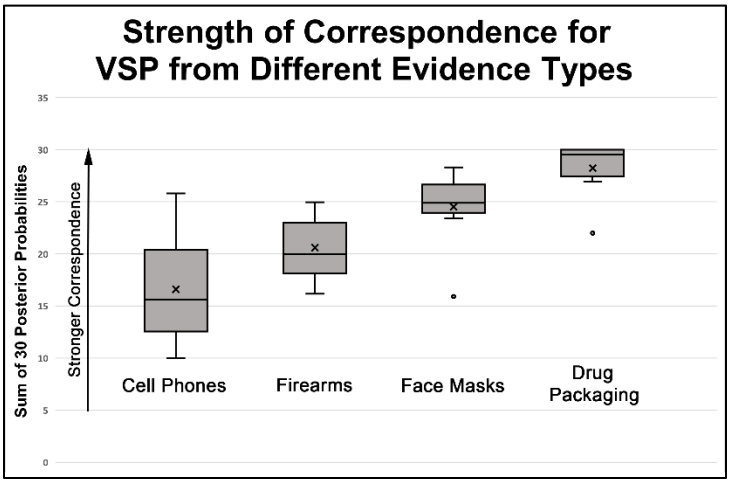


Figure 30. VSP Selectivity from Alternative Data Filtration Parameters – Posterior Probability of True Source. Effect of evidence type. The VSP datasets for the different evidence types showed clear, major differences in the strength of correspondence.

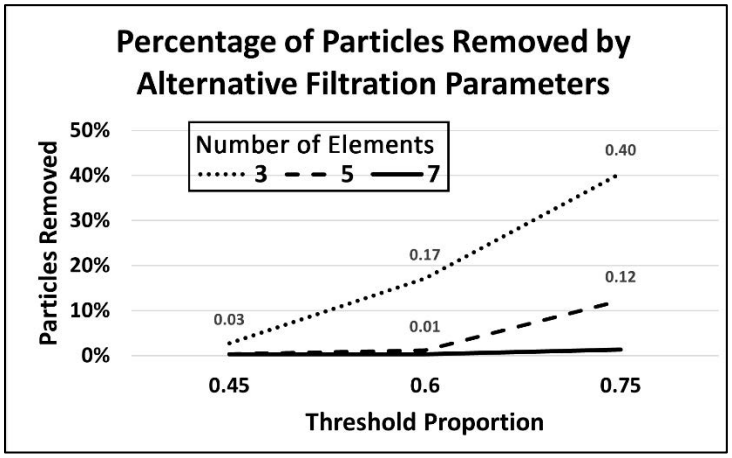


Figure 31. Effect of Data Filtration Parameters on Percentage of Particles Removed the Analysis. When 7 elements are used (solid line) almost no particles are removed from the analysis. The same is true for 5 elements and the lower threshold proportions (dashed line). Larger numbers of particles are removed when only 3 elements are required to meet the threshold proportion (dotted line), with 40% of the particles removed when 3 elements are required to make up 75% of the particles measured elemental concentration.

IV. Conclusions

IV.A. Discussion of Findings

1. Confirmation of Abundant, Highly Discriminating VSP on Common Items of Evidence

The numbers of VSP analyzed for each specimen in this project (several thousand) showed high levels of reproducibility and selectivity, confirming prior work.

Importantly, the number of particles available for analysis was not a limiting factor: many more particles (usually greater than 50 times more) were present than were used for the analysis. These unused particles provide the opportunity to achieve higher precision (through repeat analyses) or greater selectivity (using protocols exploiting finer particle classifications or additional particle characteristics).

The quality of particle characterization was sufficient for the methods of particle combination analysis that were used. There remain options for more precise characterization of either the chemistry or morphology of individual particles using the existing SEM/EDS analytical platform.

2. Reproducibility VSP Analyses

A very high level of reproducibility was observed with strong correspondence between sets of VSP from the same specimen. Analyses of the same specimen conducted on the same run, later runs, or after instrument maintenance did not show important differences in the rates of particle detection, the numbers of particles detected or the high rates of correct classification.

There was slightly greater variability when analyses were not conducted on the same analytical run. This indicates that greater sensitivity to differences between specimens is obtained when specimens are analyzed on the same analytical run. This is common to most testing, as random error attributable to differences between runs is avoided. However, the observation of successful classification when specimens were analyzed on different and later runs indicates that acceptable reproducibility can be achieved.

No differences in reproducibility or correspondence were seen before or after instrument maintenance.

3. Key Factors Affecting Performance of Particle Combination Analysis

Tables 11 and 12 summarize the key factors identified in this study that affect performance. This includes 4 factors affecting the Strength of Correspondence, 3 affecting the Numbers of Particles Detected, and 9 affecting the Analytical Time and Costs.

a. Factors Affecting Strength of Correspondence

Factors affecting the Strength of Correspondence are of primary interest.

The x-ray analysis variables of *Minimum X-Ray Count* and *Element Threshold* were identified as key factors affecting the strength of correspondence. The *Minimum X-Ray Count* showed a main effect as well as an interaction with the *Element Threshold*. The Strength of Correspondence appeared minimally affected, with greater strength observed only when there was both a lower *Element Threshold* and a higher *Minimum X-Ray Count*.

The *Number and Choice of Elements* was found to be key factor affecting the strength of correspondence, with larger numbers of elements showing greater selectivity.

Particle Size Fractions were found to be a key factor for strength of correspondence. Particle size quartiles showed a small, regular increase in selectivity with increasing size. However, when equal quantities of randomly selected particles were tested, smaller diameter particles (respirable and easily transported by air) were found to have greater selectivity than larger particles. Clearly particle size is an important factor, but the nature of the relationship is not simple and merits further examination.

Particle detection parameters were not observed to affect strength of correspondence. This indicates that alternative subsets of VSP (as detected by the alternative parameters in this project) have comparable selectivity.

The data filtration parameters used for computational analyses were not found to have a measurable effect on strength of correspondence. Excepting the more extreme settings, these parameters removed only small numbers of particles from consideration. Even with the more extreme settings, sufficient particles remained to achieve high discrimination. The intended purpose of these parameters was to reduce "noise" in the form of uncharacteristic particles composed of small quantities of many elements. However, the x-ray analysis parameters act similarly, including the specification of minimum levels of elemental composition. Accordingly, these data filtration parameters appear unnecessary.

b. Factors Affecting the Numbers of Particles Detected

The Number of Particles Detected provides the number of VSP available for particle combination analysis. The number of particles detected and characterized by the SEM is strongly influenced by particle searching parameters, particle selection criteria and the specimen itself.

The particle detection parameter of *Particle Size Limits* was identified as a key factor affecting the numbers of particles detected. Although *Particle Brightness* also showed a main effect, this effect was found to be primarily an interaction with *Particle Size Limits* (attributable, to the smaller, dimmer particles).

The *Number and Choice of Elements* was found to be key factor affecting the numbers of particles detected. The larger number of elements showed greater numbers of particles.

c. Factors Affecting Analytical Time and Costs

The amount of instrument time required for the automated SEM/EDS is important to the overall practicality of the protocol as it affects the throughput of specimens and cases as well as the associated costs (notably the obligation of an instrument of substantial cost to this analytical task in competition to others).

Particle detection parameters of *Magnification* and *Particle Size Limits* were identified as key factors affecting analytical times (longer times for greater magnification and wider particle size limits). For perspective on this impact, the differences seen in the rates of particle detection (Figure 10) had the effect of increasing specimen analytical times just under 4 hours to 6 hours or more.

The particle detection parameter of *Particle Brightness* was found to show interactions affecting analytical times with two other parameters: *Particle Size Limits* and *Time for Image Collection.* Both interactions can be attributable to smaller, dimmer particles. These interactions should be considered as possible factors for analytical protocols that do not otherwise exclude smaller, dimmer particles.

Each of the four x-ray analysis parameters tested showed a main effect on analysis time. Complex, but explainable, interactive effects were observed as the setting of one parameter offset or intensified the effect of another.

The *Number and Choice of Elements* was found to be key factor affecting the analysis time. Larger numbers of elements showed shorter analytical times.

Table 11. Key Factors Identified for VSP Combination Analysis using SEM/EDS

FACTORS AFFECTING STRENGTH OF CORREPONDENCE

X-ray Analysis Parameters
- *Minimum X-Ray Count*
- *Element Threshold*

Number and Choice of Elements

Particle Size Fraction

FACTORS AFFECTING NUMBERS OF PARTICLES DETECTED

Particle Detection Parameters
- *Particle Size Limits*
- For protocols including dimmer particles:
  - Interaction with *Particle Brightness*

Number and Choice of Elements

FACTORS AFFECTING ANALYSIS TIME AND COSTS

Particle Detection Parameters
- *Magnification*
- *Particle Size Limits*
- For protocols including dimmer particles:
  - *Particle Brightness* with interactions with
    - *Particle Size Limits*
    - *Time for Image Collection*

X-ray Analysis Parameters
- (Each of the parameters tested, with complex interactions)
- *Maximum X-Ray Collection Time*
- *Minimum X-Ray Count*
- *Target Single Element X-Ray Count*
- *Element Threshold*

Number and Choice of Elements

Table 12. Effects Observed for Specific Parameters

Particle Detection Parameters	Observed Effects	Description
Time for Image Collection	No effect on Strength of Correspondence or Numbers of Particles Detected. Interactive effect with Particle Brightness on Rate of Particle Detection.	No effects were observed on Strength of Correspondence. Restrictions on particle size or brightness resulted in fewer Particles Detected. Lower Rates of Particle Detection were observed with greater magnification and wider particle size limits.
Particle Brightness	No effect on Strength of Correspondence. Main effect (and interactive effect with Particle Size Limits) on Numbers of Particles Detected. Interactive effects on Rate of Particle Detection with both Time for Image Collection and Particle Size Limits.	
Magnification	No effect on Strength of Correspondence or Numbers of Particles Detected. Main effect on Rates of Particle Detection.	
Particle Size Limits	No effect on Strength of Correspondence. Main effect (and interactive effect with Particle Brightness) on Numbers of Particles Detected and Rates of Particle Detection.	

X-ray Analysis Parameters	Observed Effects	Description
Minimum X-Ray Count	A slight main effect (and interactive effect with Element Threshold) on Strength of Correspondence. Main effect (and interactive effect with Target Single Element X-Ray Count) on Rate of Particle Detection.	Strength of Correspondence was minimally affected, with greater strength observed only when there was both with a lower element threshold and a higher minimum x-ray count. Each variable showed a main effect on the Rate of Particle Detection and there were complex interactive effects.
Element Threshold	A slight interactive effect on Strength of Correspondence with Minimum X-Ray Count. Main effect (and interactive effects with Target Single Element X-Ray Count and Maximum X-Ray Collection Time) on Rate of Particle Detection.	
Target Single Element X-Ray Count	No effect on Strength of Corresponence. Main effect (and interactive effects with each of the other three variables) on Rate of Particle Detection.	
Maximum X-Ray Collection Time	No effect on Strength of Corresponence. Main effect (and interactive effect with Target Single Element X-Ray Count and Element Threshold) on Rate of Particle Detection.	

Table 12. Effects Observed for Specific Parameters (continued)

Number and Choice of Elements	Observed Effects	Description
Original 18 Elements	Intermediate Strength of Correspondence (comparable to Alternative 18). Lower Numbers of Particles Detected, comparable to Reduced Set of 9. Lower Rates of Particle Detection than Alternative 18 or Full Set of 27. No difference from Reduced Set of 9 in Rate of Particle Detection.	The use of larger numbers of elements resulted in greater Strength of Correspondence and increased Numbers of Particles Detected. Higher (but equivalent) Rates of Particle Detection were found for two conditions and lower, but equivalent rates for the other two.
Alternative 18 Elements	Intermediate Strength of Correspondence (comparable to Original 18). Second greatest Number of Particles Detected (next to Full Set of 27). Higher Rates of Particle Detection than Original 18 or Reduced Set of 9. No difference from Full Set of 27 in Rate of Particle Detection.	
Full Set of 27 Elements	Highest Strength of Correspondence. Greatest Number of Particles Detected. Higher Rates of Particle Detection than Original 18 or Reduced Set of 9. No difference from Alternative 18 in Rate of Particle Detection.	
Reduced Set of 9 Elements	Lowest Strength of Correspondence. Lower Numbers of Particles Detected, comparable to Original 18. Lower Rates of Particle Detection than Alternative 18 or Full Set of 27. No difference from Original 18 in Rate of Particle Detection.	

Particle Size Fraction	Observed Effects	Description
Quartiles of Particle Size	Slightly higher Strengths of Correspondence with increasing particle size.	Particle size is an important factor for Strength of Correspondence, but the nature of the effect is not simply explained.
PM_{10} vs > PM_{10}	Smaller (PM_{10}) particles with greater Strength of Correspondence.	

Data Filtration Parameters	Observed Effects	Description
Number of Elements	No effect on Strength of Correspondence.	Despite removal of large numbers of particles for some conditions, sufficient particles remain for strong correspondence.
Proportion of Composition	No effect on Strength of Correspondence.	

IV.B. Implications for Policy and Practice

This is a new approach, highly significant for its potential to expand the number of cases to which trace evidence can meaningfully contribute and for its ability to include a quantitative statistical approach to data interpretation. This research program employed newly developed quantitative statistical tools to measure the individuality of particle combinations that are ubiquitous in our environment, abundantly present on common items of evidence, long recognized for their possible potential, but left unused for want of a practical and meaningful way forward. The program serves as an excellent example, in the current climate of broader public and scientific skepticism, of how fundamentally significant improvements in forensic science can be achieved.

The approach in its current state of development offers crime laboratories an additional capability suitable for high priority cases. The identification of key factors affecting performance of the VSP analytical protocol allows existing methods to be further developed and systematically improved to facilitate transition to routine practice.

As the value of VSP becomes more appreciated, and the transition to routine practice becomes imminent, there will be a need for complementary policies and practices for evidence collection and processing of crime scenes (conducive to the preservation and analysis of VSP).

IV.C. Implications for Further Research

This project measured the relative impact of a wide set of independent variables on particle combination analysis performance, identifying as set of key factors affecting performance. This directly enables follow-on research efforts to focus on areas that will systematically improve and optimize the methodology.

Follow-on research can be viewed as dealing with two areas: strength of correspondence and analysis time. It is logical to first concentrate efforts on strength of correspondence (VSP selectivity). Selectivity is of most direct importance, has fewer factors, and no identified factor interactions. Improvements in this area will likely determine, or significantly reduce the range of choices for some of the factors. Optimization for analysis time, for which many factors were identified, could be more efficiently and meaningfully addressed after more progress is made in the first area.

This project has also 1) increased awareness and interest in VSP among forensic science researchers, and 2) provided datasets allowing the testing and improvement of computational methods.

Evaluation of alternative high-throughput particle analysis methods, emphasizing different aspects of chemistry and particle morphology, is also appropriate. We have used one type of instrumental analysis, with a set of parameters that reasonably characterizes the primary inorganic composition of small particles. This has demonstrated an overall approach that is applicable to other areas of active forensic science research that involve the analysis and interpretation of large, complex combinations of particles (or other mixtures). These include current efforts focused on the analysis of pollen, microbes, isotopes and multi-species DNA. Our approach to characterization of VSP and particle combination analysis are important options to be evaluated in these related areas.

V. References

1. C. Roux, A. Beavis, S. Benson, E. Braybon, M. Dawson, P. Doble, K. Flynn, G. Payne, C. Lennard, P. Maynard, B. Reedy, J. Robertson, M. Tahtouh, C. Wallace-Kunkel, and S. Walsh, Forensic Science in the 21st Century - Will Trace Evidence Ever Reach the Next Level? , http://projects.nfstc.org/trace/docs/final/Roux_21st_century.pdf, [last accessed: 2/25/17]
2. C. Roux, The Current State of Affairs and Trends in the Crime Laboratory - Developments in the Last Ten Years - New Issues Facing the Trace Examiner. Australian & New Zealand Perspective, NIJ/FBI 2007 Trace Evidence Symposium Clearwater Beach, FL, 2007.
3. J. Robertson and C. Roux, Trace Evidence: Here Today, Gone Tomorrow?, Science & Justice 50 (2010) 18-22.
4. J. Robertson, Trace Evidence - Dissappearing Fast?, Australian Journal of Forensic Sciences. 42 (2010) 79-80.
5. D.A. Stoney and P.L. Stoney, A Critical Review of Forensic Trace Evidence Analysis and the Need for a New Approach, Forensic Science International 251 (2015) 159-170.
6. D.A. Stoney, C. Neumann, K.E. Mooney, J.M. Wyatt, and P.L. Stoney, Exploitation of Very Small Particles to Enhance the Probative Value of Carpet Fibers, Forensic Science International 252 (2015) 52-68.
7. D.A. Stoney, A.M. Bowen and P.L. Stoney, Utilization of Environmentally Acquired Very Small Partices as a Means of Association, Forensic Science International 254 (2015) 26-50.
8. D.A. Stoney, C. Neumann, and P.L. Stoney, Discrimination and Classification among Common Items of Evidence using Particle Combination Profiles, Forensic Science International 289 (2018) 92-107.
9. A. Bowen and D. Stoney, A New Method for the Removal and Analysis of Small Particles Adhering to Carpet Fiber Surfaces, Journal of Forensic Sciences 58 (2013) 789-796.
10. C. Neumann. Forensic Analysis and Inference of the Source of VSPs, R Package 'VSParticles' (2016)
11. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
12. Neumann, Cedric. Exploitation of Very Small Particles to Enhance the Probative Value of Carpet Fibers. Semi-Annual Progress Report to Stoney Forensic, January 25,2014.
13. Jiao, Yang. Elemental Analysis in House Dust Using Handheld X-Ray Fluoresncece. MPH Thesis, San Diego State University, 2012.
14. Conner, T.L. et al. Individual Particle Analysis of Indoor, Outdoor and Community Samples for the 1998 Baltimore Particulate Matter Study. Atmospheric Environment 35 (2001) 3935-3946.
15. Keil, D.E. et al. Health Effects from Exposure to Atmospheric Mineral Dust Near Las Vegas, NV, USA. Toxicology Reports 3 (2016) 785-795.

VI. Dissemination of Research Findings

A. Presentations Resulting from this Award

1. Completed as of 12/31/20

Stoney, DA and Stoney, PL, “Developments in Particle Combination Analysis,” NIJ Impression, Pattern and Trace Evidence Symposium, Arlington, VA, January 24, 2018

Stoney, DA and Stoney, PL, "Determination of Key Factors in Particle Combination Analysis to Enable Systematic Improvement, Optimization and Transition to Practice," American Academy of Forensic Sciences 70th Annual Meeting, Seattle, WA, February 22, 2018.

Stoney, DA and Stoney, PL, “Application of Particle Characterization Methods (Such as SEM/EDS) in Support of Particle Combination Analysis,” (Poster presentation), Pittcon 2018 Conference & Expo, Orlando, FL, March 1, 2018.

Stoney D.A. and Stoney, P.L, “Determination of Key Factors in Particle Combination Analysis to Enable Optimization and Transition to Practice,” (National Institute of Justice Research Poster Session), Pittcon 2020 Conference & Expo, Chicago, IL, March 3, 2020

2. Scheduled

Stoney D.A. and Stoney, P.L, “Key Factors in Particle Combination Analysis,” National Institute of Justice Forensic Science Research and Development Symposium, American Academy of Forensic Sciences 73rd Annual Scientific Meeting, February 16, 2021.

Stoney D.A. and Stoney, P.L, “How Particle Combination Analysis is Being Used to Address the Need for a New Approach to Forensic Trace Evidence Analysis,” NIJ (National Institute of Justice) Advancements in the Analysis of Forensic Trace Evidence, Virtual Pittcon 2021 Conference & Expo, March 11, 2021.

B. Publications Resulting from this Award (as of 12/31/20)

Stoney, D.A. and Stoney, P.L. The Determination of Key Factors in Particle Combination Analysis to Enable Systematic Improvement, Optimization, and Transition to Practice, Proceedings of the American Academy of Forensic Sciences, Vol. 24, p. 283, 2018.

Appendix A. Description of Particle Combination Analysis Methods

1. Semi-Automated SEM/EDS Analysis Method
2. Computational Methods

1. Semi-Automated SEM/EDS Analysis Method

This method was developed and applied under NIJ Award 2012-DN-BX-K041, and again applied under NIJ Award 2015-DN-BX-K046. Results have been peer reviewed and published.[1,2] Improvements to analytical instrumentation (release of the next generation hardware and software) have resulted in minor changes.

SEM/EDS analysis is performed on a Thermo Fisher Scientific Explorer 4 Analyzer SEM-EDS system using the Automated Feature Analysis (AFA) program within the Thermo Fisher Scientific Corporation Perception 5 software. Analysis is performed under low vacuum conditions (0.075 torr) utilizing a 20.0 kV accelerating voltage, backscatter electron detector (BSED), working distance of approximately 8.7 mm, and a spot size of approximately 58%.

The magnification for the analysis is 1,200X. A search grid dimension of 512 x 512 is used with times for imaging (frame averages) of 2 µs for searching and 16 µs for measuring. Brightness and contrast settings for particle detection are set using a Cu/C border with C set to image brightness level 25 and Cu set to image brightness level 240. The lower search detection (threshold for minimum particle brightness) is set to image brightness level 64. The size criteria for analysis are a minimum size of 0.3 µm and a maximum size of 80.0 µm. EDS parameters are an initial x-ray counting time (nominal duration) of 3s, a maximum x-ray counting time of 6s, a minimum count of 500 and a target count of 2500 for x-rays attributable to any of the 18 elements listed in Table 1. An EDS Copper calibration check is performed before each analytical run and during the run at the beginning of each specimen.

Table 1. The 18 Elements Detected by the Automated EDS Procedure

Sodium (Kα, Kβ)	Magnesium (Kα, Kβ)	Aluminum (Kα, Kβ)
Silicon (Kα, Kβ)	Phosphorous (Kα, Kβ)	Sulfur (Kα, Kβ)
Chlorine (Kα, Kβ)	Potassium (Kα, Kβ)	Calcium (Kα, Kβ)
Titanium (Kα, Kβ)	Vanadium (Kα, Kβ)	Chromium (Kα, Kβ)
Manganese (Kα, Kβ)	Iron (Kα, Kβ)	Cobalt (Kα, Kβ)
Nickel (Kα, Kβ)	Copper (Kα, Kβ)	Zinc (Kα, Kβ)

Raw datasets for each SEM/EDS run consist of a set of run parameters and results along with individual particle analysis data. Individual particle data are a particle index number, a set of particle size and shape parameters, the four elements in the particle's EDS spectrum having the highest x-ray counts, the corresponding four x-ray counts, live analysis time, total x-ray counts for the particle and the calculated percentages of each of the 18 specified elements.

2. Computational Methods

Computational methods were developed and applied under NIJ Awards 2012-DN-BX-K041 and 2015-DN-BX-K046. They are available as a maintained package in R [3,4] and applications to VSP on the surfaces of carpet fibers [1] and to VSP on common items of physical evidence [2] have been peer reviewed and published. Statistical analysis is performed using R [4] as described below.

Filtering of Noise. The particle data are filtered to (1) remove particles that fail to show any dominant composition as represented by the calculated percentages of the elements and (2) remove elements that are only present in minute amounts. Two parameters are defined to determine the presence of a dominant composition (1) the total proportion P of the composition of a particle represented by N elements, and (2) the number N of elements. Parameters are set at $P = 60\%$ and $N = 5$. In addition, the percentage composition contributed by the next (6^{th}) element is examined and this element is also retained if it accounts for more than 10% of the total composition of the particle. In order to keep the integrity of the compositional aspect of the data and of its structure, the "noise elements" are not removed, but their contribution to the composition of the particle is set to zero. The proportions of the remaining elements are not rescaled, in accordance with recommended practices.[5] Particles that do not have N elements representing more than the set threshold for their composition are considered as noise and not included in subsequent computational analyses.

Definition of Target Particle Types. Target particle types (TPTs) are defined using a semi-supervised hierarchical clustering algorithm relying on Normal Mixture Modeling. The algorithm for estimating the mixture parameters is implemented in the R library mclust.[6, 7] This algorithm allows for finding the optimal mixture model that represents the distribution of the data. In such a model, each class (cluster) is represented by one of the multivariate normal distributions in the mixture. The posterior probability of class for each data point is calculated according to a Bayes rule for each of the normal density functions.[8]

In semi-supervised hierarchical clustering, the scientist sets the number of classes G (in this case the number of TPTs) and the algorithm optimizes the parameters of each one of the G components such that the overall mixture best describes the data. The mixture model makes sense mathematically speaking (data points are similar to each other in the eyes of the algorithm), but the clusters are an abstract construction of alternative variables and classification possibilities. Thus, the TPT classes created by the algorithm are not necessarily representative of chemically recognizable particle types. TPTs are defined using a random sampling of 8000 particles from the reference sources dataset with $G = 80$.

Determination of TPT Profiles. Eighty TPTs are defined based on the reference dataset, as described above in *Definition of Target Particle Types*. The TPT profile for any sample is determined by categorization of each of that specimen's particles into the most closely fitting of the 80 TPTs, based the probability of the particles' class membership in each of the TPTs. For an individual comparison of a Trace sample (representing a specimen of "unknown" origin) to the set of Reference Sources, the particles on the Trace are classified in the 80 TPTs; however, only the 10 most populated TPTs on the Trace are used for comparison, with the remaining particles grouped into an 11^{th} class.

Measurement of Degree of Correspondence among TPT Profiles. The comparison between TPT profiles relies on the multinomial distribution. The counts for each of the ($k = 10$) TPTs for any given reference source were used to estimate the vector of ($k = 10$) proportions of a multinomial distribution of the TPTs for that specimen. When a new set of particles from a new

sample is considered, its particles are categorized into the TPTs previously defined for the reference source(s). The degree of correspondence of the new sample with a reference source is defined as the probability of the observed vector of TPT counts in the new sample assigned using the multinomial probability mass function for that reference source. This probability can be used directly to support the inference of the source of a new sample by assigning posterior class probabilities (if prior class probabilities are assumed/known) or by using likelihood ratios (if applicable reference population data are assumed/known).

Closed-set classification of TPT Profiles, Posterior Probabilities and Corresponding Likelihood Ratios. The performance of the system is examined, using a closed-set scenario, by separating the particle data in the reference sources into a training set and a test set. Both sets are considered to be independent and identically distributed samples of particles from the set of particles recovered from the evidence items. The training set (2/3 of the original dataset) is used to define the set of TPTs as described above, to measure the respective proportions of particles with these TPTs in each source, and to study some aspects of the general discrimination potential of the system in ideal conditions. The purpose of the test set (1/3 of the original dataset) is to test the system on a dataset that was not used to estimate the parameters of the system. Using the procedures described above, TPT profiles were determined and multinomial distributions were defined for each of the *N* sources using the particles in the training set. For evaluation of the matching ability of the system, the training set sources were used as "references," and the particles remaining in each of the specimens' test set were used as *N* new specimens, to be compared with the "references."

Given a set of *N* possible sources, each represented by a vector of 10 proportions estimated as described above, a new set of particles, *E*, can be categorized using a Bayes classifier [8]. The posterior probability of class for the *j*-th source, S_j, is defined as:

Eq. 1
$$\Pr(S_j|E) = \frac{\Pr(E|S_j)\Pr(S_j)}{\Pr(E|S_1)\Pr(S_1)+\ldots+\Pr(E|S_N)\Pr(S_N)}$$

The new set of particles is then inferred to originate from the source with highest posterior probability (under an assumption of equal priors for each of the sources).

A corresponding likelihood ratio for each source is calculated using Eq. 2 as a measure of evidential weight, based on assumptions of the representativeness of the *N* sources.

Eq. 2
$$LR(E,S_j) = \frac{\Pr(E|S_j)(1-\Pr(S_j))}{\sum_{i=1,i\neq j}^{D}\Pr(E|S_i)\Pr(S_i)} = \frac{(N-1)\Pr(E|S_j)}{\sum_{i=1,i\neq j}^{D}\Pr(E|S_i)}$$

Rates of correct and incorrect classification are determined. Where test specimens are misclassified, the performance of the classification system is also analyzed by examining the rank of the true source, in relation to the remaining sources.

References

1. D.A. Stoney, C. Neumann, K.E. Mooney, J.M. Wyatt, and P.L. Stoney, Exploitation of Very Small Particles to Enhance the Probative Value of Carpet Fibers, Forensic Sci Int. 252 (2015) 52-68.
2. Stoney, D.A., Neumann, C. and Stoney, P.L., Discrimination and Classification among Common Items of Evidence using Particle Combination Profiles, Forensic Science International, Vol 289, pp. 92-107, 2018.
3. C. Neumann. Forensic Analysis and Inference of the Source of VSPs, R Package 'VSParticles' (2016)
4. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
5. van den Boogarrt K.G, Tolosana-Delgado R. (2013) Analysing Compositional Data with R, Springer NY.
6. Fraley, Chris; Raftery, Adrian E.; Murphy, T. Brendan and Scrucca, Luca (2012) mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation Technical Report No. 597, Department of Statistics, University of Washington.
7. Fraley, Chris and Raftery, Adrian E. (2002) Model-based Clustering, Discriminant Analysis and Density Estimation Journal of the American Statistical Association 97:611-631.
8. Izenman, A.J. (2008) Modern Multivariate Statistical Techniques, Springer NY, p.241.

Appendix B. Description of Program Datasets

I. Experiment 1

I.A. Experiment 1 Particle Dataset

The Particle Dataset for Experiment 1 consists of individual .csv files, one for each specimen analysis. The following provides an explanation of the file content, names the locations.

File Folder: Appendix B Program Datasets/B2A Experiment 1 Particle Data

Contents: 128 .csv files

There are 8 specimens and 4 controls, with 6 files for each.

Each file is named using the specimen designation in the Table B1 followed by a hyphen and a two-character condition codes indicating the corresponding treatment.

1A: Initial analysis

1B: Repeat analysis on the same analytical run

1C: Repeat analysis on a separate analytical run

1D: Analysis on a separate run with intervening source replacement and realignment

Table B1. Specimen Designations for Experiment 1

EXPERIMENT 1			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P102S	C102S	F102S	M106S
P111S	C104S	F107S	M111S
P118S	C105S	F111S	M118S
P121S	C108S	F114S	M126S
P123S	C121S	F118S	M127S
P125S	C128S	F121S	M129S
P138S	C131S	F123S	M132S
P140S	C133S	F131S	M133S

	Drug Packaging	Cell Phones	Handguns	Ski Masks	Material Blank
Swabs exposed during sampling and processed as for specimens		CB1	FB1	MB1	
		CB2	FB2	MB2	
Adhesive SEM stubs exposed during sampling	PB1				
	PB4				
Adhesive SEM stub, unexposed					SO1

Within each .csv file each row corresponds to an individual particle.

The column headings and descriptions are given in Table B2.

I.B. Experiment 1 Time and Particle Numbers Dataset

The Time and Particle Numbers Dataset for Experiment 1 is in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B1B Experiment 1 Time and Particle Numbers
Sheet Name: E1 Time and Particle Nos
Column A has the specimen/condition designations named using the specimen designation in Table B1 followed by a hyphen and one of four condition codes indicating the corresponding treatment.

1A: Initial analysis

1B: Repeat analysis on the same analytical run

1C: Repeat analysis on a separate analytical run

1D: Analysis on a separate run with intervening source replacement and realignment

There are 32 specimens and 9 controls, with 4 rows for each for a total of 164 non-header rows.
Each row contains 7 columns.
The column headings and descriptions are given in Table B3.

Table B2. Column Headings in .csv Files of Experiment 1 Particle Datasets

Column	Header	Explanation
A	Part #	Sequentially assigned particle number
B	Stage X (mm)	Absolute X position of the particle on the stage
C	Stage Y (mm)	Absolute Y position of the particle on the stage
D	Field#	An assigned number for the field of view
E	X Feret (µm)	X dimension of the smallest rectangle to enclose particle
F	Y Feret (µm)	Y dimension of the smallest rectangle to enclose particle
G	DAve (µm)	Average distance between each point on the perimeter and the furthest point from it
H	Dmax (µm)	Length of the longest line connecting two points on the perimeter
I	Dmin (µm)	Length of the smallest line connecting two points on the perimeter
J	Dperp (µm)	Diameter of the smallest circle that spans the minor axis perpendicular to Dmax
K	Aspect	Dmax/Dmin
L	Area (μm^2)	Area of particle
M	Perimeter (µm)	Length of particle perimeter
N	Orientation (degrees)	Orientation of Dmax with respect to vertical
O	Mag	The magnification
P	Video	The video level of the particle
Q	Classification	User-defined particle classification. "All Particles" is a all-inclusive.
R	Density ($pg/\mu m^3$)	A measure of density that can be used under composition assumptions
S	PAction	An associated action for the particle, e.g. "Image 256" for storing of an image
T	VoidArea (μm^2)	Total area of all pixels that are not part of particle
U	VoidCount	Number of pixels inside perimeter that are not part of particle
V	EdgeRoughness	Measure of how smooth perimeter is
W	RmsVideo	N/A (Not configured for use in this application)
X	Roundness	$4Area/(\pi*Dmax^2)$
Y	Formfactor	An alternative measurement of roundness: $(4\pi*Area)/Perimeter^2$
Z	ECD	Equivalent Circular Diameter: Diameter of a circle that has the same area as a particle
AA	Skeleton (µm)	Length of the one pixel wide tree branch structure representing the core particle shape
AB	HullArea (μm^2)	Area of shape created by placing a 'rubber-band' around the particle
AC	HullPerimeter (µm)	Length of 'rubber-band' placed around particle in microns
AD	FirstElem	The element with the highest estimated concentration
AE	SecondElem	The element with the second highest estimated concentration
AF	ThirdElem	The element with the third highest estimated concentration
AG	FourthElem	The element with the fourth highest estimated concentration
AH	LiveTime (sec)	The net measurement time (total measurement time less the dead time)
AI	SpectrumCounts	EDS counts for the spectrum
AJ	Type(4ET)	The first through fourth highest elements separated by hyphens
AK to BB	[Elements]	The estimated concentration of the element

Table B3. Column Headings in .xlsx File for Experiment 1 Time and Particle Numbers Dataset

Column	Header	Explanation
A	Program Specimen	Specimen and condition designation
B	Elapse Time (min)	The analysis time in minutes
C	Fields	The number of stage fields examined
D	Area (mm^2)	The area examined in square millimeters
E	Particles Det/Char	The number of particles detected and characterized
F	Particles/Minute	The number of particles detected and characterized per minute
G	Particles/Area	The number of particles detected and characterized per mm^2

I.C. Experiment 1 Particle Combination Analysis Dataset

The Particle Combination Analysis Dataset for Experiment 1 consists of six individual .csv files two for each of the following analyses:

Condition 1B analyses classified as unknowns vs. Condition 1A analyses as references.
Condition 1C analyses classified as unknowns vs. Condition 1A analyses as references.
Condition 1D analyses classified as unknowns vs. Condition 1A analyses as references.

One of the two .csv files contains likelihood ratios (Log_{10}) and the other contains posterior probabilities (see Table B4).
File Folder: Appendix B Program Datasets/B1C Experiment 1 PCA Data
For each of the files Column A contains the specimen names for the references.
Row 1 contains the specimen names for the unknowns.
Cells contain the likelihood ratio or posterior probability for the column specimen matching to the row specimen under the experimental conditions and assumptions.

Table B4. Description of .csv Files in Experiment 1 Particle Combination Analysis Dataset

File Name	Contents	References	Unknowns	Explanation
BLR	Likelihood Ratios	Condition 1A	Condition 1B	Same analytical run
BPP	Posterior Probabilities	Condition 1A	Condition 1B	Same analytical run
CLR	Likelihood Ratios	Condition 1A	Condition 1C	Separate analytical run
CPP	Posterior Probabilities	Condition 1A	Condition 1C	Separate analytical run
DLR	Likelihood Ratios	Condition 1A	Condition 1D	Separate analytical run after intervening source replacement and realignment
DPP	Posterior Probabilities	Condition 1A	Condition 1D	Separate analytical run after intervening source replacement and realignment

Values appearing as "=-inf" in the likelihood ratio files (displayed as "#Name?") indicate values are so low as to exceed the computational bounds of the computer.

I.D. Experiment 1 Summary Dataset

The Summary data for Experiment 1: VSP analysis reproducibility are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B1D Experiment 1 Summary Dataset
Sheet Names:

Particles Per Minute
Particles Per Square mm
LR True Source

II. Experiment 2

II.A. Experiment 2 Particle Dataset

The Particle Dataset for Experiment 2 consists of individual .csv files, one for each specimen analysis. The following provides an explanation of the file content, names the locations.

File Folder: Appendix B Program Datasets/B2A Experiment 2 Particle Data
Contents: 192 .csv files
There are 8 specimens and 4 controls, with 16 files for each.
Each file is named using the specimen designation in Table B5 followed by a four-character code indicating the corresponding treatment, as indicated in Table B6.

Within each .csv file each row corresponds to an individual particle. The column headings and descriptions are given in Table B7.

Table B5. Specimen Designations for Experiment 2

	EXPERIMENT 2				
	Drug Packaging	Cell Phones	Handguns	Ski Masks	Material Blank
Specimens	P138S	C102S	F111S	M126S	
	P140S	C128S	F123S	M129S	
Controls	PB4	CB2		MB1	SO1

Table B6. Treatment Codes for Experiment 2 Variable Values

Character Position	First		Second		Third		Fourth	
Character	A	B	A	B	A	B	A	B
Variable	Search frame averages		Lower search detection		Magnification		Particle size limits	
Value	2 µsec	8 µsec	64	128	1200X	2400X	0.3 µm - 80 µm	2.5 µm - 40 µm

Table B7. Column Headings in the .csv Files of Experiment 2 Particle Datasets

Column	Header	Explanation
A	PART #	Sequentially assigned particle number
B	AREA	Area of particle (μm^2)
C	FIRST_ELEM	The element with the highest estimated concentration
D	SECOND_ELEM	The element with the second highest estimated concentration
E	THIRD_ELEM	The element with the third highest estimated concentration
F	FOURTH_ELEM	The element with the fourth highest estimated concentration
G	FIRST_ELEM	X-ray counts attributable to the first element
H	SECOND_ELEM	X-ray counts attributable to the second element
I	THIRD_ELEM	X-ray counts attributable to the third element
J	FOURTH_ELEM	X-ray counts attributable to the fourth element
K	LIVE_TIME	The net measurement time in seconds (total time less dead time)
L	COUNTS	EDS counts for the spectrum
M	TYPE(4ET)#	The first through fourth highest elements separated by hyphens
N to AE	[Elements]	The estimated concentration of the element

I.B. Experiment 2 Time and Particle Numbers Dataset

The Time and Particle Numbers Dataset for Experiment 2 is in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B2B Experiment 2 Time and Particle Numbers
Sheet Name: E2 Time and Particle Nos
There are 8 specimens and 4 controls, with 16 rows for each specimen, giving a total of 192 non-header rows.
Column 1 has the specimen designation named using the specimen designations in Table B5.
Column 2 has the four-character treatment codes variable values as designated in Table B6.
Each row contains 8 columns.
The column headings and descriptions are given in Table B8.

Table B8. Column Headings in the .xlsx File for Time and Particle Numbers Dataset

Column	Header	Explanation
A	Program Specimen	Specimen designation
B	Condition	Specimen condition designation
C	Elapse Time (min)	The analysis time in minutes
D	Fields	The number of stage fields examined
E	Area (mm^2)	The area examined in square millimeters
F	Particles Det/Char	The number of particles detected and characterized
G	Particles/Minute	The number of particles detected and characterized per minute
H	Particles/Area	The number of particles detected and characterized per mm^2

II.C. Experiment 2 Particle Combination Analysis Dataset

The Particle Combination Analysis Dataset for Experiment 2 consists of 32 individual .csv files, two for each of the 16 treatments (Table B6).
One of the .csv files contains likelihood ratios (Log_{10}) and the second contains the posterior probabilities.
File Folder: Appendix B Program Datasets/B2C Experiment 2 PCA Data
The files are named:

[CODE]LR.csv (for the likelihood ratios)
[CODE]PP.csv (for the posterior probabilities)

[CODE] indicates the four-character treatment code from Table B6.
For the likelihood ratios and the posterior probabilities Column A contains the specimen names for the references and Row 1 contains specimen names for the unknowns.
Cells contain the likelihood ratio or posterior probability, respectively, for the column specimen matching to the row specimen under the experimental conditions and assumptions.

Values appearing as "=-inf" in the likelihood ratio files (displayed as "#Name?") indicate values are so low as to exceed the computational bounds of the computer.

II.D. Experiment 2 Summary Datasets

The summary data for Experiment 2: SEM/EDS Particle Detection Parameters are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B2D Experiment 2 Summary Dataset
Sheet Name: E2 Summary Data
The column headings and descriptions are given in Table B9.

Table B9. Column Headings in the .xlsx File for Experiment 2 Summary Data

Column	Header	Explanation
A	Program Specimen	Specimen designation
B	Condition	Specimen condition designation
C	Particles/Minute	The number of particles detected and characterized per minute
D	Particles/Area	The number of particles detected and characterized per mm^2
E	Particles Det/Char	The number of particles detected and characterized
F	Clean Particles	The number of particles detected after particle combination analysis data filtration
G	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
H	PP True Source	The posterior probability of the true source
I	Rank of True	The rank of the true source based on greatest likelihood ratio
J	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not

III. Experiment 3

III.A. Experiment 3 Particle Dataset

The Particle Dataset for Experiment 3 consists of individual .csv files, one for each specimen analysis. The following provides an explanation of the file content, names the locations.

File Folder: Appendix B Program Datasets/B3A Experiment 3 Particle Data
Contents: 128 .csv files
There are 8 specimens, with 16 files for each.
Each file is named using the specimen designation in Table B10 followed by a hyphen, the number 3 and a four-character code indicating the corresponding treatment, as indicated in Table B11.

Within each .csv file each row corresponds to an individual particle. The column headings and descriptions are given in Table B12.

Table B10. Specimen Designations for Experiment 3

	EXPERIMENT 3			
	Drug Packaging	Cell Phones	Handguns	Ski Masks
Specimens	P138S	C102S	F111S	M126S
	P140S	C128S	F123S	M129S

Table B11. Treatment Codes for Experiment 3 Variable Values

Character Position	First		Second		Third		Fourth	
Character	A	B	A	B	A	B	A	B
Variable	Maximum Duration		Minimum Counts		Target Counts		Element Threshold	
Value	6 sec	9 sec	500	1000	2500	5000	1%	3%

III.B. Experiment 3 Time and Particle Numbers Dataset

The Time and Particle Numbers Dataset for Experiment 3 is in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B3B Experiment 3 Time and Particle Numbers
Sheet Name: E3 Time and Particle Nos
There are 8 specimens with 16 rows for each specimen, giving a total of 128 non-header rows.
Each row contains 8 columns.
Column 1 has the specimen designation named using the specimen designations in Table B10.
Column 2 has the four-character treatment codes variable values as designated in Table B11.
The remaining column headings and descriptions are given in Table B8 (above).

Table B12. Column Headings in the .csv Files of Experiment 3 Particle Datasets

Column	Header	Explanation
A	Part #	Sequentially assigned particle number
B	Stage X (mm)	Absolute X position of the particle on the stage
C	Stage Y (mm)	Absolute Y position of the particle on the stage
D	X Feret (µm)	X dimension of the smallest rectangle to enclose particle
E	Y Feret (µm)	Y dimension of the smallest rectangle to enclose particle
F	DAve (µm)	Average distance between each point on the perimeter and the furthest point from it
G	DMax (µm)	Length of the longest line connecting two points on the perimeter
H	DMin (µm)	Length of the smallest line connecting two points on the perimeter
I	DPerp (µm)	Diameter of the smallest circle that spans the minor axis perpendicular to Dmax
J	Aspect	Dmax/Dmin
K	Area (μm^2)	Area of particle
L	Perimeter (µm)	Length of particle perimeter
M	Orientation (degrees)	Orientation of Dmax with respect to vertical
N	Mag	The magnification
O	Video	The video level of the particle
P	Classification	User-defined particle classification. "All Particles" is a all-inclusive.
Q	Density ($pg/\mu m^3$)	A measure of density that can be used under composition assumptions
R	PAction	An associated action for the particle, e.g. "Image 256" for storing of an image
S	VoidArea (μm^2)	Total area of all pixels that are not part of particle
T	VoidCount	Number of pixels inside perimeter that are not part of particle
U	EdgeRoughness	Measure of how smooth perimeter is
V	RmsVideo	N/A (Not configured for use in this application)
W	Roundness	$4Area/(\pi*Dmax^2)$
X	Formfactor	An alternative measurement of roundness: $(4\pi*Area)/Perimeter^2$
Y	ECD	Equivalent Circular Diameter: Diameter of a circle that has the same area as a particle
Z	Skeleton (µm)	Length of the one pixel wide tree branch structure representing the core particle shape
AA	HullArea (μm^2)	Area of shape created by placing a 'rubber-band' around the particle
AB	HullPerimeter (µm)	Length of 'rubber-band' placed around particle in microns
AC	FirstElem	The element with the highest estimated concentration
AD	SecondElem	The element with the second highest estimated concentration
AE	ThirdElem	The element with the third highest estimated concentration
AF	FourthElem	The element with the fourth highest estimated concentration
AG	LiveTime (sec)	The net measurement time (total measurement time less the dead time)
AH	SpectrumCounts	EDS counts for the spectrum
AI	Type(4ET)	The first through fourth highest elements separated by hyphens
AJ to BA	[Elements]	The estimated concentration of the element

III.C. Experiment 3 Particle Combination Analysis Dataset

The Particle Combination Analysis Dataset for Experiment 3 consists of 32 individual .csv files, two for each of the 16 treatments (Table B11).
One of the two .csv files contains likelihood ratios (Log_{10}) and the second contains the posterior probabilities.
File Folder: Appendix B Program Datasets/B3C Experiment 3 PCA Data
The files are named:

E3 [CODE]LR.csv (for the likelihood ratios)
E3 [CODE]PP.csv (for the posterior probabilities)

[CODE] indicates the four-character treatment code from Table B11.
For the likelihood ratios and the posterior probabilities Column A contains the specimen names for the references and Row 1 contains specimen names for the unknowns.
Cells contain the likelihood ratio or posterior probability, respectively, for the column specimen matching to the row specimen under the experimental conditions and assumptions.

Values appearing as "=-inf" in the likelihood ratio files (displayed as "#Name?") indicate values are so low as to exceed the computational bounds of the computer.

III.D. Experiment 3 Summary Datasets

The summary data for Experiment 3: SEM/EDS X-Ray Analysis Parameters are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B3D Experiment 3 Summary Dataset
Sheet Name: E3 Summary Data
The column headings and descriptions are given in Table B13.

Table B13. Column Headings in the .xlsx File for Experiment 3 Summary Data

Column	Header	Explanation
A	Program Specimen	Specimen designation
B	Condition	Specimen condition designation
C	Particles/Minute	The number of particles detected and characterized per minute
D	Particles/Area	The number of particles detected and characterized per mm^2
E	Particles Det/Char	The number of particles detected and characterized
F	LLR True Source	The log_{10} likelihood ratio in support of the true source
G	PP True Source	The posterior probability of the true source
H	Rank of True	The rank of the true source based on greatest likelihood ratio
I	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not

IV. Experiment 4

IV.A. Experiment 4 Particle Dataset

The Particle Dataset for Experiment 4 consists of individual .csv files, one for each specimen analysis. The following provides an explanation of the file content, names the locations.

File Folder: Appendix B Program Datasets/B4A Experiment 4 Particle Data
Contents: 128 .csv files
There are 32 specimens, with 4 files for each.
Each file is named using the specimen designation in Table B14 followed by a hyphen and a two-character condition code for the corresponding treatment, as indicated in Table B15.

Within each .csv file each row corresponds to an individual particle. The column headings and descriptions for the first 36 columns are given in Table B2 (above). The number of remaining columns differs for each condition, as indicated in Table B16.

Table B14. Specimen Designations for Experiment 4

EXPERIMENT 4			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P102S	C102S	F102S	M106S
P111S	C104S	F107S	M111S
P118S	C105S	F111S	M118S
P121S	C108S	F114S	M126S
P123S	C121S	F118S	M127S
P125S	C128S	F121S	M129S
P138S	C131S	F123S	M132S
P140S	C133S	F131S	M133S

Table B15. Treatment Codes for Experiment 4 Variable Values

	Elements for Experiment 4 Treatments																										
	11	12	13	14	15	16	17	19	20	22	23	24	25	26	28	27	29	30	35	38	40	42	51	56	74	82	83
Code	Na	Mg	Al	Si	P	S	Cl	K	Ca	Ti	V	Cr	Mn	Fe	Ni	Co	Cu	Zn	Br	Sr	Zr	Mo	Sb	Ba	W	Pb	Bi
4A	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X									
4B	X	X	X	X		X	X		X	X				X					X	X	X	X	X	X	X	X	X
4C	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
4D	X	X	X	X		X	X		X	X				X													

Table B16. Remaining Column Headings in .csv Files for Experiment 4 Particle Datasets

Code	Column	Header	Explanation
4A	AK to BB	[Elements]	Estimated concentrations for original set of 18 elements
4B	AK to BB	[Elements]	Estimated concentrations for alternative set of 18 elements
4C	AK to BK	[Elements]	Estimated concentrations for full set of 27 elements
4D	AK to AS	[Elements]	Estimated concentrations for reduced set of 9 elements

IV.B. Experiment 4 Time and Particle Numbers Dataset

The Time and Particle Numbers Dataset for Experiment 4 is in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B4B Experiment 4 Time and Particle Numbers
Sheet Name: E4 Time and Particle Nos
There are 32 specimens with 4 rows for each specimen, giving a total of 128 non-header rows.
Each row contains 8 columns.
Column 1 has the specimen designation named using the specimen designations in Table B14.
Column 2 has the two-character treatment codes variable values as designated in Table B15.
The remaining column headings and descriptions are given in Table B8 (above).

IV.C. Experiment 4 Particle Combination Analysis Dataset

The Particle Combination Analysis Dataset for Experiment 4 consists of 8 individual .csv files, two for each of the 4 treatments (Table B15).
One of the two .csv files contains likelihood ratios (Log_{10}) and the second contains the posterior probabilities.
File Folder: Appendix B Program Datasets/B4C Experiment 4 PCA Data
The files are named:

E4 [CODE]LR.csv (for the likelihood ratios)
E4 [CODE]PP.csv (for the posterior probabilities)

[CODE] indicates the two-character treatment code from Table B15.
For the likelihood ratios and the posterior probabilities Column A contains the specimen names for the references and Row 1 contains specimen names for the unknowns.
Cells contain the likelihood ratio or posterior probability, respectively, for the column specimen matching to the row specimen under the experimental conditions and assumptions.

Values appearing as "=-inf" in the likelihood ratio files (displayed as "#Name?") indicate values are so low as to exceed the computational bounds of the computer.

IV.D. Experiment 4 Summary Datasets

The summary data for Experiment 4: Effects of Number and Choice of Elements are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B4D Experiment 4 Summary Dataset
Sheet Name: E4 Summary Data
The column headings and descriptions are given in Table B13 (above).

V. Experiment 5

V.A. Experiment 5 and 6 Particle Dataset

The Particle Dataset for Experiments 5 and 6 consists of individual .csv files, one for each specimen analysis. The following provides an explanation of the file content, names the locations.

File Folder: Appendix B Program Datasets/B5A Experiment 5 and 6 Particle Data

Contents: 120 .csv files, one for each specimen.

Each file is named using the specimen designation in Table B17.

Within each .csv file each row corresponds to an individual particle. The column headings and descriptions are given in Table B18.

Table B17. Specimen Designations for Experiments 5 and 6

EXPERIMENTS 5 and 6			
Drug Packaging	Cell Phones	Handguns	Ski Masks
P102S	C101S	F101S	M101S
P103S	C102S	F102S	M104S
P104S	C103S	F103S	M105S
P106S	C104S	F104S	M106S
P107S	C105S	F105S	M107S
P109S	C106S	F106S	M108S
P110S	C107S	F107S	M109S
P111S	C108S	F108S	M110S
P115S	C109S	F109S	M111S
P116S	C110S	F110S	M112S
P118S	C111S	F111S	M113S
P119S	C112S	F112S	M114S
P120S	C113S	F113S	M115S
P121S	C114S	F114S	M116S
P122S	C115S	F115S	M118S
P123S	C116S	F116S	M119S
P125S	C117S	F118S	M120S
P126S	C121S	F119S	M121S
P127S	C123S	F120S	M122S
P130S	C124S	F121S	M123S
P131S	C125S	F122S	M124S
P132S	C127S	F123S	M125S
P133S	C128S	F124S	M126S
P134S	C129S	F125S	M127S
P136S	C131S	F126S	M128S
P138S	C132S	F127S	M129S
P139S	C133S	F128S	M130S
P140S	C134S	F129S	M131S
P141S	C135S	F130S	M132S
P142S	C136S	F131S	M133S

Table B18. Column Headings in the .csv Files of Experiment 5 and 6 Particle Datasets

Column	Header	Explanation
A	Part #	Sequentially assigned particle number
B	Stage X (mm)	Absolute X position of the particle on the stage
C	Stage Y (mm)	Absolute Y position of the particle on the stage
D	Field#	An assigned number for the field of view
E	X Feret (µm)	X dimension of the smallest rectangle to enclose particle
F	Y Feret (µm)	Y dimension of the smallest rectangle to enclose particle
G	DAve (µm)	Average distance between each point on the perimeter and the furthest point from it
H	Dmax (µm)	Length of the longest line connecting two points on the perimeter
I	Dmin (µm)	Length of the smallest line connecting two points on the perimeter
J	Dperp (µm)	Diameter of the smallest circle that spans the minor axis perpendicular to Dmax
K	Aspect	Dmax/Dmin
L	Area (μm^2)	Area of particle
M	Perimeter (µm)	Length of particle perimeter
N	Orientation (degrees)	Orientation of Dmax with respect to vertical
O	Mag	The magnification
P	Video	The video level of the particle
Q	Classification	User-defined particle classification. "All Particles" is a all-inclusive.
R	Density ($pg/\mu m^3$)	A measure of density that can be used under composition assumptions
S	PAction	An associated action for the particle, e.g. "Image 256" for storing of an image
T	VoidArea (μm^2)	Total area of all pixels that are not part of particle
U	VoidCount	Number of pixels inside perimeter that are not part of particle
V	EdgeRoughness	Measure of how smooth perimeter is
W	RmsVideo	N/A (Not configured for use in this application)
X	Roundness	$4Area/(\pi*Dmax^2)$
Y	Formfactor	An alternative measurement of roundness: $(4\pi*Area)/Perimeter^2$
Z	ECD	Equivalent Circular Diameter: Diameter of a circle that has the same area as a particle
AA	Skeleton (µm)	Length of the one pixel wide tree branch structure representing the core particle shape
AB	HullArea (μm^2)	Area of shape created by placing a 'rubber-band' around the particle
AC	HullPerimeter (µm)	Length of 'rubber-band' placed around particle in microns
AD	FirstElem	The element with the highest estimated concentration
AE	SecondElem	The element with the second highest estimated concentration
AF	ThirdElem	The element with the third highest estimated concentration
AG	FourthElem	The element with the fourth highest estimated concentration
AH	LiveTime (sec)	The net measurement time (total measurement time less the dead time)
AI	SpectrumCounts	EDS counts for the spectrum
AJ	Type(4ET)	The first through fourth highest elements separated by hyphens
AK to BK	[Elements]	The estimated concentration of the element

V.B. Experiment 5 Particle Size Quartile Dataset

The Particle Size Quartile Dataset for Experiment 5 consists of individual .csv files, four for each specimen analysis. The following provides an explanation of the file content, names the locations.

File Folder: Appendix B Program Datasets/B5B Experiment 5 Particle Size Quartile Dataset
Contents: 480 .csv files, four for each specimen.
Each file is named using the specimen designation in Table B17, followed by a space and a two-character code:

Q1, indicating the smallest particle size quartile (based on DAve from Column G)
Q2, indicating the next largest (second smallest) particle size quartile
Q3, indicating the next largest (third smallest) particle size quartile
Q4, indicating the largest particle size quartile

Within each .csv file each row corresponds to an individual particle. The column headings and descriptions are given in Table B18.

V.C. Experiment 5 Particle Combination Analysis Dataset

The Particle Combination Analysis Dataset for Experiment 5 consists of 8 individual .csv files, two for each of the 4 Quartile treatments.
One of the two .csv files contains likelihood ratios (Log_{10}) and the second contains the posterior probabilities.
File Folder: Appendix B Program Datasets/B5C Experiment 5 PCA Data
The files are named:

E4 [CODE]LR.csv (for the likelihood ratios)
E4 [CODE]PP.csv (for the posterior probabilities)

[CODE] indicates the two-character treatment codes Q1, Q2, Q3 or Q4.
For the likelihood ratios and the posterior probabilities Column A contains the specimen names for the references and Row 1 contains specimen names for the unknowns.
Cells contain the likelihood ratio or posterior probability, respectively, for the column specimen matching to the row specimen under the experimental conditions and assumptions.

Values appearing as "=-inf" in the likelihood ratio files (displayed as "#Name?") indicate values are so low as to exceed the computational bounds of the computer.

V.D. Experiment 5 Summary Datasets

The summary data for Experiment 5: Contribution of Alternative Particle Size Fractions to Selectivity of VSP Combinations are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B5D Experiment 5 Summary Dataset
Sheet Name: E5 Summary Data
The column headings and descriptions are given in Table B19.

Table B19. Column Headings in the .xlsx File for Experiment 5 Summary Data

Column	Header	Explanation
A	Program Specimen	Specimen designation
B to F	First Quartile	Data for the first quartile of the specimen in column A
B	Number Particles	Number of particles in the quartile
C	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
D	PP True Source	The posterior probability of the true source
E	Rank of True	The rank of the true source based on greatest likelihood ratio
F	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
G to K	Second Quartile	Data for the second quartile of the specimen in column A
G	Number Particles	Number of particles in the quartile
H	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
I	PP True Source	The posterior probability of the true source
J	Rank of True	The rank of the true source based on greatest likelihood ratio
K	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
L to P	Third Quartile	Data for the third quartile of the specimen in column A
L	Number Particles	Number of particles in the quartile
M	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
N	PP True Source	The posterior probability of the true source
O	Rank of True	The rank of the true source based on greatest likelihood ratio
P	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
Q to U	Fourth Quartile	Data for the fourth quartile of the specimen in column A
Q	Number Particles	Number of particles in the quartile
R	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
S	PP True Source	The posterior probability of the true source
T	Rank of True	The rank of the true source based on greatest likelihood ratio
U	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not

VI. Experiment 6

VI.A. Experiment 6 Particle Combination Analysis Dataset

The Particle Combination Analysis Dataset for Experiment 6 consists of 27 individual .csv files, three for each of the 9 data filtration parameter treatments.
One of the two .csv files contains likelihood ratios (Log_{10}), the second contains the posterior probabilities, and the third contains particle numbers before and after data filtration.
File Folder: Appendix B Program Datasets/B6A Experiment 6 PCA Data
The files are named:

E6 [CODE]LR.csv (for the likelihood ratios)
E6 [CODE]PP.csv (for the posterior probabilities)
E6 [CODE]PC.csv (for the particle numbers)

[CODE] indicates the three-character treatment codes from Table B20.
For the likelihood ratios and the posterior probabilities Column A contains the specimen names for the references and Row 1 contains specimen names for the unknowns.
Cells contain the likelihood ratio or posterior probability, respectively, for the column specimen matching to the row specimen under the experimental conditions and assumptions.

Values appearing as "=-inf" in the likelihood ratio files (displayed as "#Name?") indicate values are so low as to exceed the computational bounds of the computer.

For the particle number files, Column A contains the specimen code from Table B17, Column B (header "before_clean") contains the number of particles before data filtration, and Column C (header "after_clean") contains the number of particles after data filtration.

Table B20. Treatment Codes for Experiment 6 Variable Values

		N, number of elements		
		3	5	7
P, threshold proportion	0.45	AA	AB	AC
	0.6	BA	BB	BC
	0.75	CA	CB	CC

VI.B. Experiment 6 Summary Dataset

The summary data for Experiment 6: Effect of Data Filtration Parameters on Selectivity of VSP Combinations are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B6B Experiment 6 Summary Dataset
Sheet Name: E6 Summary Data
The column headings and descriptions are given in Table B21.

Table B21. Column Headings in the .xlsx File for Experiment 6 Summary Data

Column	Header	Explanation
A	Program Specimen	Specimen designation
B to G	AA (3, 0.45)	Data for the first quartile of the specimen in column A
B	Number Particles	Number of particles
C	Clean Particles	Number of particles after data filtration
D	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
E	PP True Source	The posterior probability of the true source
F	Rank of True	The rank of the true source based on greatest likelihood ratio
G	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
H to M	AB (5, 0.45)	Data for the first quartile of the specimen in column A
H	Number Particles	Number of particles
I	Clean Particles	Number of particles after data filtration
J	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
K	PP True Source	The posterior probability of the true source
L	Rank of True	The rank of the true source based on greatest likelihood ratio
M	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
N to S	AC (7, 0.45)	Data for the first quartile of the specimen in column A
N	Number Particles	Number of particles
O	Clean Particles	Number of particles after data filtration
P	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
Q	PP True Source	The posterior probability of the true source
R	Rank of True	The rank of the true source based on greatest likelihood ratio
S	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
T to Y	BA (3, 0.60)	Data for the first quartile of the specimen in column A
T	Number Particles	Number of particles
U	Clean Particles	Number of particles after data filtration
V	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
W	PP True Source	The posterior probability of the true source
X	Rank of True	The rank of the true source based on greatest likelihood ratio
Y	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
Z to AE	BB (5, 0.60)	Data for the first quartile of the specimen in column A
Z	Number Particles	Number of particles
AA	Clean Particles	Number of particles after data filtration
AB	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
AC	PP True Source	The posterior probability of the true source
AD	Rank of True	The rank of the true source based on greatest likelihood ratio
AE	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
AF to AK	BC (7, 0.60)	Data for the first quartile of the specimen in column A
AF	Number Particles	Number of particles
AG	Clean Particles	Number of particles after data filtration
AH	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
AI	PP True Source	The posterior probability of the true source
AJ	Rank of True	The rank of the true source based on greatest likelihood ratio
AK	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
AL to AQ	CA (3, 0.75)	Data for the first quartile of the specimen in column A
AL	Number Particles	Number of particles
AM	Clean Particles	Number of particles after data filtration
AN	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
AO	PP True Source	The posterior probability of the true source
AP	Rank of True	The rank of the true source based on greatest likelihood ratio
AQ	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
AR to AW	CB (5, 0.75)	Data for the first quartile of the specimen in column A
AR	Number Particles	Number of particles
AS	Clean Particles	Number of particles after data filtration
AT	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
AU	PP True Source	The posterior probability of the true source
AV	Rank of True	The rank of the true source based on greatest likelihood ratio
AW	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not
AX to BC	CC (7, 0.75)	Data for the first quartile of the specimen in column A
AX	Number Particles	Number of particles
AY	Clean Particles	Number of particles after data filtration
AZ	LLR True Source	The $\log_{10}$ likelihood ratio in support of the true source
BA	PP True Source	The posterior probability of the true source
BB	Rank of True	The rank of the true source based on greatest likelihood ratio
BC	Correctly Classified	Binary: 1 = correctly classified based on greatest likelihood ratio; 0 = not

VI.C. Experiment 6 Summary Dataset by Evidence Type

The summary data for Experiment 6: Effect of Data Filtration Parameters on Selectivity of VSP Combinations are in an Excel spreadsheet (.xlsx file).
File Folder: Appendix B
File Name: B6C Experiment 6 Summary Dataset by Evidence Type
The data are present on two sheets:

Posterior Probability (value of posterior probability of true source)
Correctly Classified (binary:1 = correctly classified by greatest likelihood ratio; 0 = not)

The column headings and descriptions are given in Table B22.

Table B22. Column Headings for .xlsx File: Experiment 6 Summary Dataset by Evidence Type

Column	Header	Explanation
A	Cell Phone Specimen	Specimen designation for 30 cell phone specimens
B to J		Cell Phone Specimen Data
B	AA	Value for condition AA
C	AB	Value for condition AB
D	AC	Value for condition AC
E	BA	Value for condition BA
F	BB	Value for condition BB
G	BC	Value for condition BC
H	CA	Value for condition CA
I	CB	Value for condition CB
J	CC	Value for condition CC
K	Firearm Specimen	Specimen designation for 30 firearm specimens
L to T		Firearm Specimen Data
L	AA	Value for condition AA
M	AB	Value for condition AB
N	AC	Value for condition AC
O	BA	Value for condition BA
P	BB	Value for condition BB
Q	BC	Value for condition BC
R	CA	Value for condition CA
S	CB	Value for condition CB
T	CC	Value for condition CC
U	Ski Mask Specimen	Specimen designation for 30 ski mask specimens
V to AD		Ski Mask Specimen Data
v	AA	Value for condition AA
W	AB	Value for condition AB
X	AC	Value for condition AC
Y	BA	Value for condition BA
Z	BB	Value for condition BB
AA	BC	Value for condition BC
AB	CA	Value for condition CA
AC	CB	Value for condition CB
AD	CC	Value for condition CC
AE	Packaging Specimen	Specimen designation for 30 packaging specimens
AF TO AN		Packaging Specimen Data
AF	AA	Value for condition AA
AG	AB	Value for condition AB
AH	AC	Value for condition AC
AI	BA	Value for condition BA
AJ	BB	Value for condition BB
AK	BC	Value for condition BC
AL	CA	Value for condition CA
AM	CB	Value for condition CB
AN	CC	Value for condition CC