SAND2008-0999C

Analysis of Flow-Cytometer Scattering and Fluorescence Data to Identify Particle Mixtures

Thomas A. Reichardt*, Scott E. Bisson, Robert Crocker, and Thomas J. Kulp
Sandia National Laboratories, P. O. Box 969, Livermore, CA 94551

ABSTRACT

As part of the U.S. Department of Homeland Security Detect-to-Protect program, a multilab [Sandia National Laboratories (SNL), Lawrence Livermore National Laboratories (LLNL), Pacific Northwest National Laboratory (PNNL), Oak Ridge National Laboratory (ORNL), and Los Alamos National Laboratory (LANL)] effort is addressing the need for useable detect-to-warn bioaerosol sensors for public facility protection. Towards this end, the SNL team is employing rapid fluorogenic staining to infer the protein content of bioaerosols. This is being implemented in a flow cytometry platform wherein each particle detected generates coincident signals of correlated forward scatter, side scatter, and fluorescence. Several thousand such coincident signal sets are typically collected to generate a probability distribution over the scattering and fluorescence values. A linear unmixing analysis is performed to differentiate components in the mixture. After forming a library of pure component distributions from measured pure material samples, the distribution of an unknown mixture of particles is treated as a linear combination of the pure component distributions. The scattering/fluorescence probability distribution data vector **a** is considered the product of two vectors, the fractional profile **f** and the scattering/fluorescence distributions from pure components **P**. A least squares procedure minimizes the magnitude of the residual vector **e** in the expression $\mathbf{a} = \mathbf{f}\mathbf{P}^T + \mathbf{e}$. The profile **f** designates a weighting fraction for each particle type included in the set of pure components, providing the composition of the unknown mixture. We discuss testing of this analysis approach and steps we have taken to evaluate the effect of interferents, both known and unknown.

1. INTRODUCTION

The U.S. Department of Homeland Security (DHS) seeks rapidly-acting (1-2 min response time) sensors that indicate the presence of a bioaerosol release in enclosed facilities. Optical measurements are desired for this application because they can rapidly measure intrinsic properties of single particles. Thus, several early-warning sensors[1,2,3] have been developed that measure the laser-induced fluorescence (LIF) and elastic scattering properties of streams of sampled aerosols. Detailed recognition algorithms[4,5] are used to analyze these data to identify significant populations of "threat-like" aerosols within the facility background aerosol population.

Although their capabilities are continually being improved, current LIF-based sensors are observed to produce false alarms at rates between once per week and once every two months when operated under conditions relevant to facility protection. High-consequence actions (e.g. stopping the flow of traffic in an airport concourse) are likely too costly to incur by facility managers at this rate. Thus, false alarming represents a significant obstacle to the use of the devices.

This paper describes the analysis of data acquired using a secondary flow-cytometer channel that would be used in conjunction with an LIF-based sensor in order to improve its detection accuracy. The measurement would be conducted by reacting ambient particles with a fast-acting fluorogenic (i.e., becoming fluorescent upon reaction) stain that covalently couples to primary amine groups, such as are present on certain protein amino acid residues. The measurement is intended to create a strong LIF signal on particles containing biological material, thus distinguishing them from non-biological (or trace biological) background material.

2. EXPERIMENTAL

This paper focuses on the development and use of an analysis routine for the acquired flow cytometer data, so only a brief description of the experimental apparatus is given. A more thorough description of this apparatus, as well as potential future implementations of the instrument, is presented in the accompanying paper by Bisson et al.[6]

* tareich@sandia.gov; phone 925-294-4776; fax 925-294-2595

A modified Becton-Dickinson (BD) FACScan flow cytometer performs optical scattering and LIF measurements on individual particles entrained in a liquid stream. The FACScan has been modified by Cytek, Inc. (Hayward, CA) to incorporate an additional laser beam and two additional LIF channels. Due to the cytometer modification by Cytek, forward scattering and side scattering measurements are acquired at 488 nm, while correlated LIF measurements are acquired with 405-nm excitation with corresponding fluorescence channels centered at 475 and 545nm. The liquid stream containing the particles flows through the cuvette at a flow rate of ~20µl/min, where the stream is hydrodynamically focused using sheath fluid to a ~100µm diameter. At this focus the particles are illuminated by both the 488-nm and 405-nm laser beams. Optical signals generated by the particles are collected by a set of filtered photomultiplier tubes (PMTs). Samples are stained with Fluorescamine, which couples with primary amines according to a molecular rearrangment reaction: the resultant molecule is fluorescent with an excitation peak at ~390 nm and an emission peak at ~480 nm.[7] The reaction takes place rapidly (≤15 s) upon introduction of the dye to particles in an aqueous environment at a pH of ≥8.5.

The count rates are typically about a few thousand counts per second; in some cases, samples were further diluted to ensure count rates less than the cytometer maximum acquisition rate of 5000 samples/s. Normal instrument background count rates were on the order of 15-30/sec and were typically associated with low scattering values. Sample acquisition was triggered on the forward scattering signal to allow non-fluorescent scattering particles to be examined. The size threshold below which scattering would not produce a useable trigger was ~200 nm, an approximate value which depends on the indices of refraction of the sheath fluid and particle. Because the range of signals spanned several orders of magnitude, all channels were operated in logarithmic mode, which provides 4 decades of dynamic range over only 10 bits of data transfer. The relationship between the channel output and the actual signal level measured by the PMT is given by

$$\text{channel output} = 256 \times \log\,[\text{signal level (mV)}]. \quad (1)$$

For example, a channel output value of 100 corresponds to PMT signal level of 2.46 mV, while a value of 1000 corresponds to a PMT signal level of 8.06 V.

The laboratory cytometer is injected with liquid samples prepared by the Edgewood Chemical and Biological Center (ECBC) that contain both pure samples of threat agent simulants (Bacillus thuringiensis spores, Yersinia rohdei cells, MS2 virus, and ovalbumin protein) as well as mixtures of these simulants with unspecified backgrounds. Each sample is measured in both an unstained "native" condition and a stained condition. Samples are typically counted for 1 minute. All data are acquired with identical sample preparation protocols and cytometer settings. For the results presented in this paper, only one of the fluorescence channels (centered at 475 nm) is included in the analysis.

3. ANALYSIS AND DISCUSSION

After a collection of particles is measured, the particle data are binned according to the scattering and fluorescence signal levels. Such binning procedures are used by a variety of biological aerosol detectors to characterize responsivity to threats. These instruments typically calculate the probability that each detected particle is either a threat or a non-threat.[4,5,8] In contrast to analyzing each particle separately, we instead fit an entire collection of coincident data (the term "coincident" refers to the simultaneous acquisition of forward scatter, side scatter, and fluorescence signals) to a linear sum of library functions corresponding to pure component distributions. Twede et al.[9] used such a linear mixing model for analysis of excitation-emission matrices of bio-compounds. The data analysis routine is performed with MATLAB 7 (The MathWorks, Natick, MA).

3.1 Creating the Coincident Distribution

Figure 1 illustrates the procedure used to bin the coincident data. In Fig. 1(a), we show collection of coincident data from pure Bt, displaying the side scatter (SSC), forward scatter (FSC), and fluorescence (FL) channels, each value ranging between 0 and 1024. Each axis is divided into 26 sections, resulting in $26^3 = 17{,}676$ bins, as shown in Fig. 1(b). The fraction of particles which falls within each bin is then calculated. A color code to the right of Fig. 1(c) indicates the fractional value in each bin. To emphasize the location of the peak populations, those fractional values with less than 2% of the peak fractional value are not displayed.

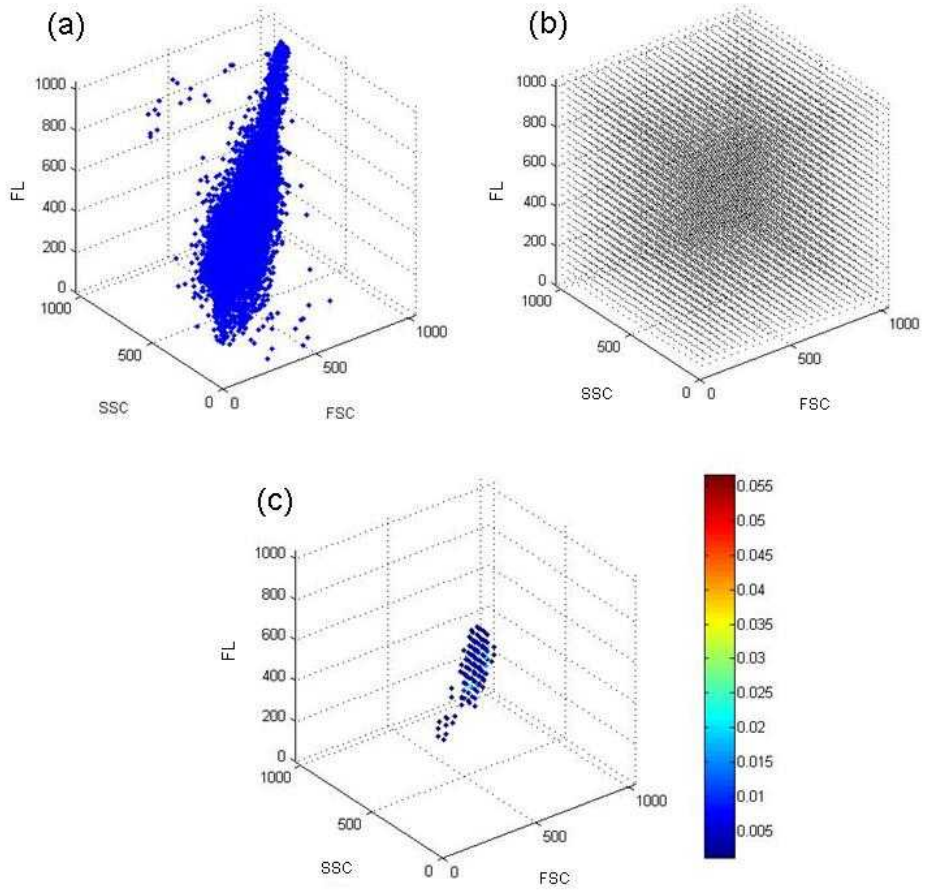


Fig. 1. (a) Scatter plot, (b) bin spacing, and (c) binning results for pure Bt stained with Fluorescamine.

Similar to the plot in Fig. 1(c), Fig. 2 displays the results for Bacillus thuringiensis (Bt), Yersinia rohdei (Yr), and house dust, both native and stained. The Bt distribution in Fig. 1(c) is included in Fig. 2 to show the complete set of pure component distributions. Staining increases the particle fluorescence while not significantly altering the scattering properties of the particles, translating the distributions in the +FL (vertical) direction.

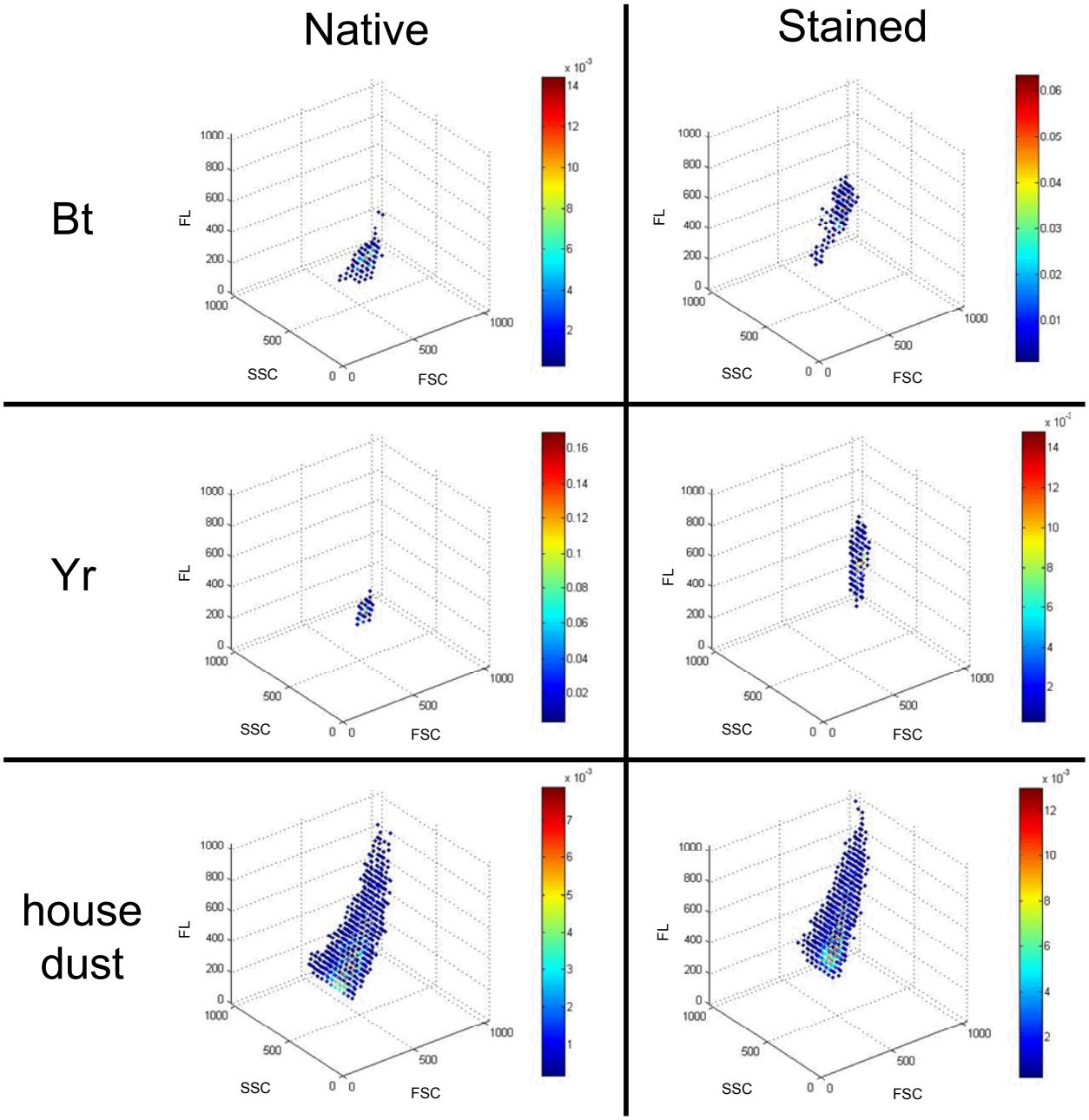


Fig. 2. Binning results from Bt, Yr, and house dust (native and stained with Fluorescamine).

3.2 Fitting an Unknown Coincident Distribution

Once the library of $n = 3$ pure components distributions (Bt, Yr, and house dust) has been compiled, we assume that each unknown sample is a linear combination of these pure component distributions. In addition, model error is assumed to originate from the error in the fractional populations and not in the bin positions. To allow use of built-in MATLAB vector functions, we first unfold each 26×26×26 matrix into a 17,676 element vector **a**. We can then use classical least squares to solve the equation

$$\mathbf{a} = \mathbf{f}\mathbf{P}^{\mathrm{T}} + \mathbf{e}, \tag{2}$$

where **a** is the 17,676-element vector of acquired coincident data, **f** is the *n*-element vector containing the fractional concentration for each of the pure constituents, **P** is the 17,676-by-*n* matrix of pure constituent distributions, and **e** is a vector of the residuals, which accounts for noise and model error (e.g., particles not included in the set of pure components). This equation can be solved by simple matrix manipulation,

$$\mathbf{f} = (\mathbf{P}\mathbf{P}')^{-1}\mathbf{P}\mathbf{c}. \quad (3)$$

Figure 3 displays the results of fitting an unknown stained sample distribution to a linear combination of the pure component distributions in Fig. 2. Figure 3(a) displays the acquired sample distributions, while Fig. 3(b) displays the resulting fit. From the fitting approach, the unknown sample is determined to be primarily Yr, with negligible Bt and house dust. Figures 3(c) and 3(d) display plots of the residual in each particle bin. Similar to the plots displaying the particle distributions, the value of **e** in each particle bin is represented by a color code, and those residuals below 2% of the peak residual are not displayed. Because this color coding convention does not appropriately designate the sign of the binned values, separate plots are presented for the positive and negative residuals.

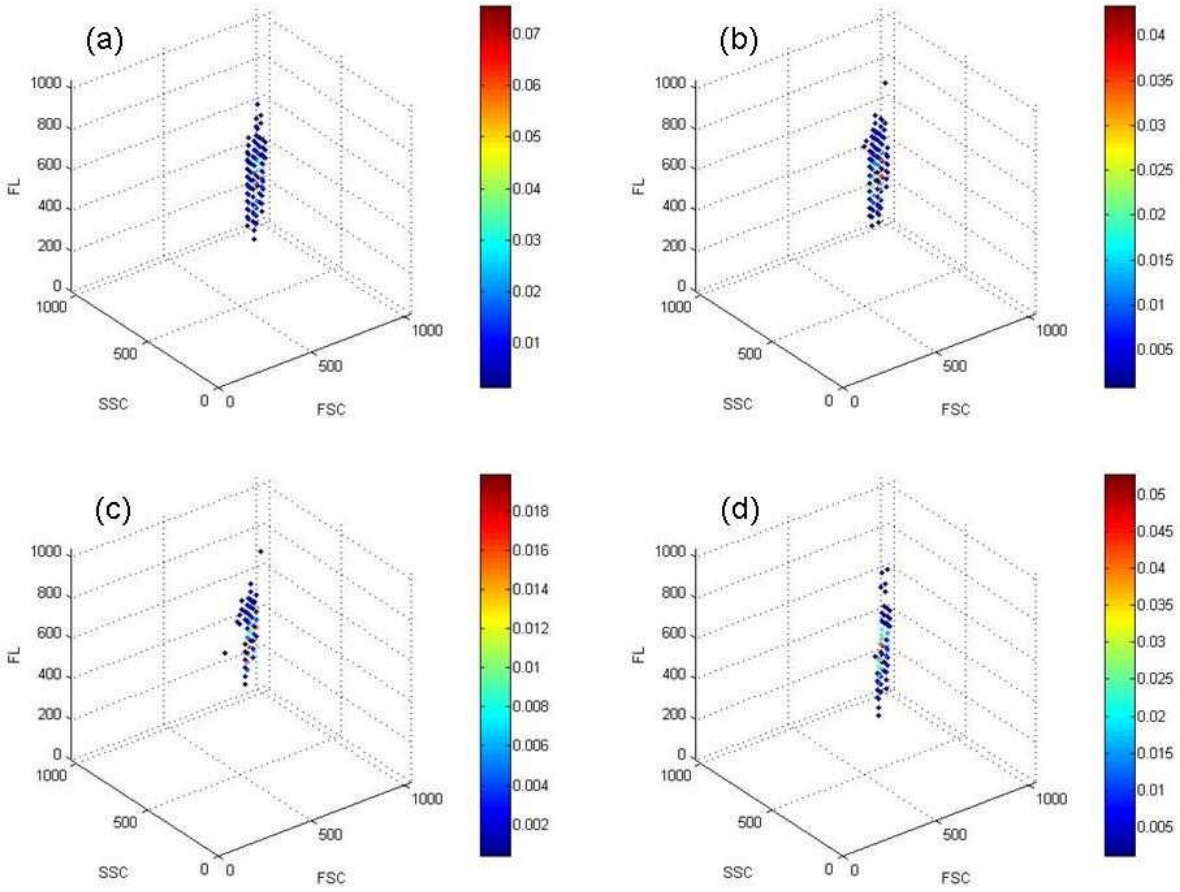


Fig. 3. Fitting results of an unknown stained sample: (a) the acquired distribution, (b) the resulting fit, and (c) the positive residual difference and (d) the negative residual difference. The fitting coefficients identify the sample to be 73% Yr, with $<10^{-5}$ parts Bt or house dust.

3.3 Testing the Fitting Performance

The performance of the analysis approach was evaluated in a blind test against the unknown sample mixtures obtained from ECBC. Twenty such samples were provided with mixtures of Bt, Yr, ovalbumin, and/or MS2, accompanied by varying amounts of unknown background(s). The approach of detecting mixture components by their

scattering/fluorescence properties is applicable only for non-soluble constituents (Bt and Yr), while ovalbumin and MS2 are detected by another method which recognizes the soluble protein in the mixture. The linear unmixing procedure correctly identified the non-soluble constituents in all 20 of these samples with no false positives. The results of analyzing the ECBC mixtures are displayed in Table 1.

Table 1. Sample assignments and ECBC mixtures for preliminary design review testing.

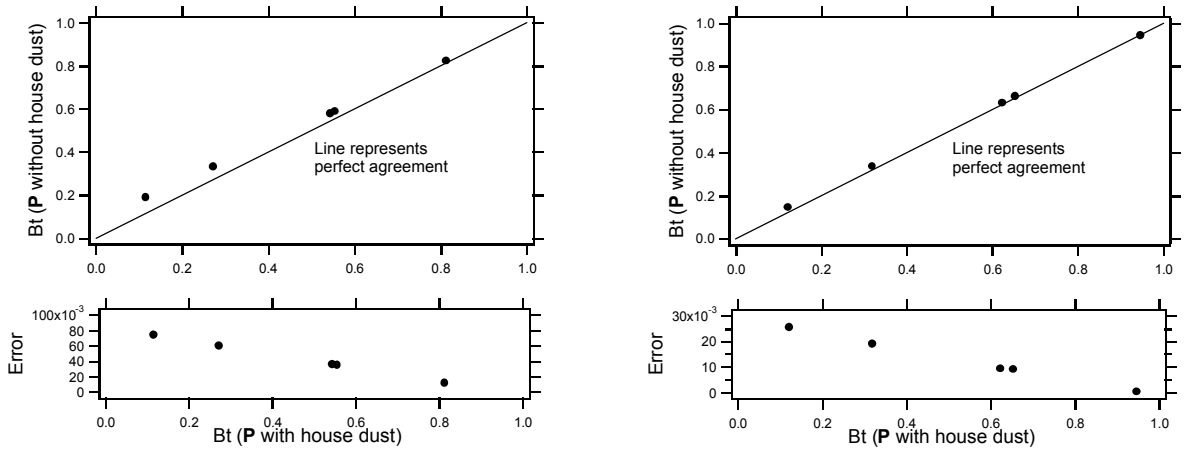
ECBC sample number	Cell/spore assignment	Soluble protein	ECBC Antigen	ECBC Background H=High L=Low
1	--	--	None	L
2	--	Yes	Ovalbumin	H
3	Yersinia	Yes	Yersinia	H
4	--	--	None	H
5	--	Yes	Ovalbumin	H
6	--	--	None	H
7	--	Yes	MS2	L
8	--	Yes	MS2	L
9	--	Yes	Ovalbumin	L
10	Bt	--	Bt	L
11	Yersinia	Yes	Yersinia	L
12	--	Yes	Ovalbumin	L
13	Bt	Yes	Bt	L
14	--	Yes	MS2	H
15	Bt	--	Bt	H
16	Bt	--	Bt	H
17	Yersinia	Yes	Yersinia	H
18	Yersinia	Yes	Yersinia	L
19	--	--	None	L
20	--	Yes	MS2	H

3.4 Evaluating the Impact of Potential Interferents

While the ECBC blind-test samples contained unknown background interferents, none of these interferents possessed similar scattering/fluorescence properties to those of Bt and Yr. Because the interferents were dissimilar to the target particles, they had little effect on the net analyte signal (NAS), the portion of the threat signal orthogonal to that of interferents. The scattering/fluorescence properties of house dust, on the other hand, significantly overlap those of the threat simulants. We anticipate that substances with characteristics similar to house dust – particles possessing biological content (e.g., dust mites and skin flakes) and which are distributed over the respirable size range – will be the most challenging to mitigate. We therefore undertook a study to help quantify the enhanced specificity provided by staining, as compared to measuring only the native particle scattering/fluorescence properties, seeking to define the NAS improvement provided by staining.

To gauge the effect that an unknown background such as house dust would have on the ability of our instrument to detect biological threats, we removed house dust from the set of pure components and compared the results to those obtained with house dust included in the set of pure components. We created five mixtures of Bt and house dust for this comparison. Figure 4(a) displays a comparison of fitting the unstained samples without house dust included in the pure component distributions **P** (vertical axis) versus including house dust in **P** (horizontal axis). Figure 4(b) displays the corresponding results for the stained samples. The solid line with a slope of unity and a vertical intercept of zero represents perfect agreement between the two fits, which would be the case if the house dust distribution were orthogonal to that of Bt. Below each graph is a plot of the deviation of these results from this perfect agreement. The extrapolated vertical-axis intercept of these plots indicates the fraction of house dust which would be mistakenly

identified as Bt. Staining the sample significantly improves the specificity of the measurement, decreasing the fraction of house dust misidentified as Bt from 10% to only 3%.



(a) Unstained (b) Stained

Fig. 4. Comparison of fitting samples with and without house dust included in the pure component distributions for (a) unstained samples and (b) stained samples.

3.5 Evaluating the Impact of Potential Artifacts

While the least-squares analysis provides a method for analyzing the data available from these trials, additional chemometric approaches can be implemented to examine artifacts present in the mixture data which would reduce the fidelity of the fitting routine. Such artifacts would arise if portions of the acquired distribution were not linearly proportional to the concentration of the measured pure components. For example, small particles of sample debris – the exact quantity of which may not be directly proportional to the particle concentration – can trigger the cytometer. In addition, the probability of acquiring data from doublets (two particles present in the probe volume during a single coincident data acquisition) will become significant at high count rates, and the probability of acquiring data from doublets will increase nonlinearly with increasing particle concentration.

4. CONCLUSIONS AND FUTURE WORK

We have demonstrated the use of a linear mixing analysis for the identification of particle mixtures from data acquired with a flow cytometer. The approach of classical least squares is demonstrated sufficient for differentiating Bt and Yr from house dust as well as unknown backgrounds provided in a blind test. Staining provides improved differentiation between Bt and house dust as compared to analyzing only the native scattering/fluorescence properties.

Future work will focus on evaluating the ability of the analysis routine to account for additional backgrounds as well as evaluate the repeatability of the threat distributions. In particular, instrument variation over time may shift the distributions, reducing the sensitivity and/or selectivity of the analysis. Evaluations of instrument drifts will be conducted and potential mitigation approaches will be evaluated.

5. ACKNOWLEDGMENTS

This work was supported by the U. S. Department of Homeland Security under the Detect-to-Protect program. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

REFERENCES

[1] J. D. Pitts, D. Cousins, and A. Goyette, "Biological aerosol warning sensor model: An approach to model architecture and accelerated false alarm prediction," Proceedings of the SPIE, Vol. **5617,** 14-22 (2004).

[2] B. K. Dable, G. A. Wilson, J. Brady, and M. M. Carrabba, "Recent testing and performance improvements of a fluorescence based biological aerosol sensor," SPIE Vol. **6554**, 655410-1-6 (2007).

[3] V. Sivaprakasam, A. Huston, H. B. Lin, J. Eversole, P. Falkenstein, and A. Schulz, "Field test results and ambient aerosol measurements using dual wavelength fluorescence excitation and elastic scatter for bioaerosols," Proceedings of the SPIE, Vol. **6554,** 65540R-1-7 (2007).

[4] G. A. Wilson and J. Brady, "Design considerations and signal processing algorithms for laser-induced fluorescence airborne pathogen sensors," SPIE Vol. **5617**, 1-13 (2004).

[5] R. Bhartia, W. F. Hug, E. C. Salas, K. Sijapati, A. L. Lane, R. D. Reid, and P. G. Conrad, "Biochemical detection and identification false alarm rate dependence on wavelength using laser induced native fluorescence," SPIE Vol. **6218**, 62180J-1-9 (2006).

[6] S. E. Bisson, R. W. Crocker, T. J. Kulp, T. A. Reichardt, P. T. A. Reilly, and W. B. Whitten, "Confirmatory measurement channels for LIF-based aerosol instrumentation," to be published in Proceedings of the SPIE, Vol. **6945** (2008).

[7] S. Udenfriend, S. Stein, P. Bohlen, W. Dairman, W. Leimgruber, and M. Weigele, "Fluorescamine: A reagent for assay of amino acides, peptides, proteins, and primary amines in the picomole range," Science **178**, 871-872 (1972).

[8] F. L. Reyes, T. H. Jeys, N. R. Newbury, C. A. Primmerman, G. S. Rowe, and A. Sanchez, "Bio-aerosol fluorescence sensor," Field Analytical Chemistry and Technology **3**, 240-248 (1999).

[9] D. R. Twede, L. C. Sanders, and M. L. Wagner, "Bio-compound detection through fluorescence excitation-emission matrix analysis," Proceedings of the SPIE – The International Society for Optical Engineering, Vol. **5159**, 239-245 (2004).