*Timothy M. Horton, Matthew Kay, A. Ertas, D. Tate*
A Hyperspectral Signature Method for Identifying E. coli: Impact on Public Health

20

# A Hyperspectral Signature Method for Identifying E. coli: Impact on Public Health

**Timothy M. Horton**[1]**, Matthew Kay**[2]**, A. Ertas**[3]**, Derrick Tate**[4]**,** [1]*Raytheon Company, Dallas, Texas, USA, Email: thorton@raytheon.com;* [2]*Department of Microbiology, Texas Tech University Health Science Center, Lubbock, TX, USA;* [3]*Department of Mechanical Engineering, Texas Tech University, Lubbock, TX, USA;* [4]*Industrial Design Department Xi'an Jiaotong Liverpool University Suzhou, China.*

Using a non Shiga toxin producing Escherichia coli such as the lab K12 strain we will demonstrate a method to develop signatures suitable for hyperspectral data searches. Conventional laboratory methods remain the mainstay to isolate and identify suspected sessile and platonic bio-material, however the FDA Food Safety Modernization Act (FSMA), Jan 04, 2011, is requiring additional proactive identification and monitoring. These proposed methods could allow higher screening coverage without additional laboratory time. These standard test are laborious and can take days to complete. Optical, noninvasive techniques such as hyperspectral remote sensing technology has been adapted for microscopic sensing. Many applications have pursued this avenue with varying degrees of success. As the cost of hyperspectral detectors falls, the promise of an optical detection solution is within reach. The goal of this research is to develop a detection method based on hundreds of cells still in their platonic stage before the damaging effects of their more colonized form develops. Once colonized, removal is much more difficult as environmental coping mechanisms are fully developed. The objective is to determine a HSI signature that has a low false alarm rate ($F_a$)from unstained (low contrast) and stained ( high contrast) samples.

**Keywords**: Hyperspectral, ISODATA, spectral angle mapper, spectral correlation mapper, hypercube, HSI signature

## 1 Introduction

As the food chain grows to a global distribution system, different levels of consumer concerns must be addressed. One of these concerns that directly affects our health, especially those who may be immune compromised, pregnant, very young or old, is common microbiological contaminations such as E. coli. For consumers, the detection of such is nearly impossible and we rely soley on appearance of the product and description on the package [1] and reputation of the vendor.

Escherichia coli, usually called E. coli, is a type of bacteria that lives in the intestines of humans and animals. Although most types of E. coli are harmless, some types such as E. coli O157: H7, can make people sick, causing serious stomach cramps, diarrhea and vomiting. Serious complications of an E. coli O157:H7 infection can cause permanent organ damage such as kidney function loss or death.

People get E. coli infections by eating foods containing the bacteria. Symptoms of infection include: Nausea or vomiting, abdominal cramping, sudden, severe diarrhea that may cause bloody stools, pallor, high blood pressure, gas, fatigue, and fever. E. coli infection may initiate devastating illness such as Hemolytic Uremic Syndrome (HUS), particularly in children young and mature adults which can cause life-threatening complications.

Tens of thousands of kilos of fresh fruit and vegetables in the country are being destroyed as consumers across Europe and beyond shun these staples for fear of contracting the potentially deadly bacteria [2]. E. coli contaminations are both common as cross contaminations are boht easily fostered and prevented however difficult to detect.

The increased public awareness has demanded greater controls and advancement in food quality controls. As Rahman [3] stated, food products that are preserved depend on a multitude of hurdles being properly controlled. One of which is the microbial population in the product. Critical limits, known as the hurdle effect, are controlled with heat, temperature and chemical treatments however quantifying the microbial content is a slow process that requires professional laboratory implementations. Much work has been done in defining these hurdles [4] and their interactive causal effects. Leistner defined F values to quantify variables such as acidity (pH and titratable), moisture content and correlate to microbial colony populations in relation to these values. The use of hyperspectral technology to successfully quantify the food safety variable space is supported by many research programs such as Zhu [5], as he demonstrates the ability to determine the frozen history of products using visual infrared and near infrared (VIR/NIR) screening.

The signature is without doubt the value added commodity within the area of research concerning hyperspectal data mining. Because of their intrinsic value and system specific nature most systems will not depend upon reuse and will develop their own set of signatures. In none of the literature, did we find formal standards nor even the suggestion of a standard that would allow a greater degree of portability. A signature is a mathematical device used to mine hyperspectral imaging (HSI) data in an effort to determine the material composition of a given scene. A spectral peek followed by decay then followed by a combination of the same, e.g., will typically constitute a mathematical signature. The mathematical relationships vary widely, using various geometric measurements, distance from DC, and relative positions to patterns detected by machine learning applications.

The signatures developed here are multi-component constructs that can independently produce some degree of success during a field search. We used two different base line algorithms and then integrated them together taking advantage of their strengths to create a single algorithm that is stronger and produces a lower Fa. The primary component in our algorithm is the Spectral Angle Mapper (SAM) [6, 7, 8, 9] which characterizes the shape of a line when compared with a baseline standard. The second component that is integrated into the ID chain is known as a Spectral Correlation Mapper (SCM) [10, 11] which measures the strength of the linear relationship between two variables.

## 2 Microbiology Introduction

Most E. coli strains are harmless, but Shiga toxin-producing Escherichia coli (STEC) can cause foodborne disease in our cattle and beef products. STEC is one of the most important factors affecting the beef industry and is one of the public health threats faced in food processing. E. coli O157:H7 is the most commonly identified STEC in North America and has been illegal in beef products since 1994. E. coli O157:H7 and other serotypes cause approximately 113,000 illnesses and 300 hospitalizations annually, according to the Centers for Disease Control and Prevention (CDC). Identification and removal of contaminated beef products is therefore a critical concern for the cattle industry, and even contact with contaminated fecal matter (as fertilizer) can lead to spread of the pathogen in produce shipped to the marketplace. A direct screening method that can be employed in all areas of food screening and processing would therefore be useful as a deterrent to future STEC outbreaks. Methods in current use for
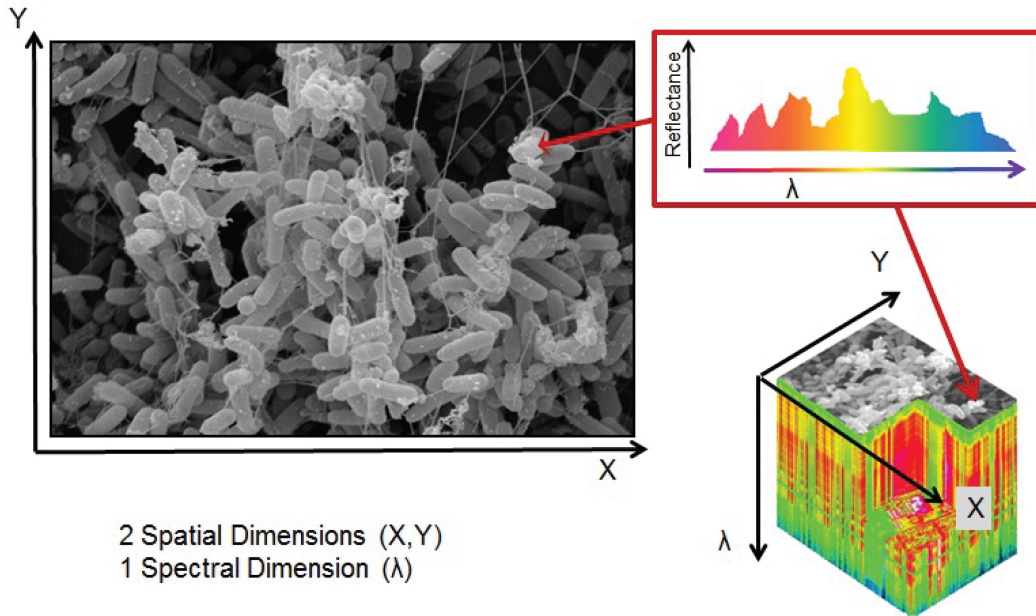
**Figure 1:** Three dimensional data construction.

STEC diagnosis include pulsed-field gel electrophoresis (PFGE) subtyping, and a conventional microbiological method involving cell counting. While these methods are accurate and remain as the gold standard for food-borne pathogen detection time is needed, from days to weeks to provide an accurate result. Therefore, a more rapid method of pathogen detection is needed that is very sensitive and accurate for classification of contaminated foodstuffs. Optical, noninvasive techniques such as hyperspectral remote sensing technology, adapted for microscopic procedures, may provide a needed venue for safe, fast, and efficient screening of possible contaminants. New applications of this technology as well as reduced cost make hyperspectal sensing a perfect candidate for identification of STEC in the food supply.

Bacteria colonizing the gut of an animal are typically found in multi-species biofilms lining the intestinal epithelium, and as such are also attached to a surface of some kind. A biofilm is typified as an aggregate bacterial community enclosed by extra-polymeric substance (EPS). Free-floating planktonic bacteria are released by the biofilm in regular intervals, to find new areas to colonize, and many of these will also attach to new surfaces readily in the gut and pass along attached to the fecal matter as well. The ability of these bacteria to attach to surfaces provides us with new opportunities to find hyperspectral signatures specific for the combina-
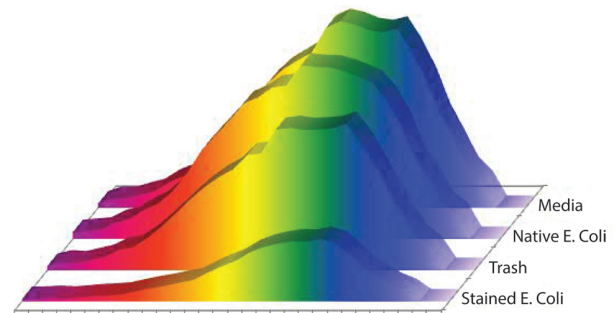


**Figure 2:** Spectral comparison of what we see under the microscope.

tion of the contaminating bacteria attached to the surface of the bound organic material. This ability to provide a hyperspectral signature against various backgrounds could lead to diagnostic tools that are far more specific and efficient in identifying STEC, as well as reducing lab workload.

Identifcation of bacteria bound to a surface also provides an advantage in that the bacteria are immobilized, reducing the effect of Brownian motion and "wiggling" observed in the capture of hyperspectral images by Cray et al. [11] used with glass slides and traditional microscopy techniques. Agar plates with affixed E. coli colonies were used to generate our hyperspectral image (HSI) data. Our technique delineated in this paper uses a data analysis proto-

col to sort through the accumulated HSIs to sort between real and trash data, producing a composite signature which can be used to verify the presence of E. coli bacteria. It is definitely possible that this signature can be developed against other bound bacterial-contaminated surfaces as well, leading to a library of potential signatures that can identify various contaminants in our food supply.

# 3 Materials and Methods

## 3.1 Data Set

A hyperpixel[1] is a three dimensional data construct that represents reflectance values at the step interval ( 10 nm ). In Figure 1 each tick along the X axis represents 10nm starting at 450 nm. Values have been omitted for image cleanness and did not contribute to the clarity of the message. If we plot a hyperpixel we may get something that looks like Figure 2 where reflectance $R_x(\lambda_n)$ is plotted along the Y axis and wavelength ($\lambda$) along the X axis. This image depicts the spectral spread as a continuous line however the hardware bins the spectrum into finite spectral ranges, grossly selectable by the user with a minimum step of 10nm (per end user sensor specifications). It's this binning that allows for the feature selection and pattern recognition sequences of unsupervised learning algorithms. This represents a single hyperpixel which correlates to a single XY location in a Cartesian image system. By organizing the hyperpixels into an image format addressed as in a XY system we create a three dimensional cube known as a hypercube addressable as [ X, Y, $\lambda$ ], with $\lambda$ increasing along the formal Z axis.

## 3.2 Target Environment

The target image is complicated because in the process of growing the specimen, the biologist has to provide fluid and nourishment for the bacteria to thrive. This is accomplished by adding a small colony into a rich agar material. The material is home grown by most commercial and academic labs. The Texas Tech Microbiology Department uses the following recipe: 5g yeast extract, 200 ul NaOH, 15g granulated agar mix, and 1L deionized $H_2O$. This is but

one of the series of variables that makes this HSI signature challenging. Of course there are the other compounds in the agar mix. We attempt to classify them as "media" in this study. Most of these detections are aggregated into single hyperpixels and can be isolated having a characteristic signature when they are the predominate end member in a pixel. Some however will have additional end members present within the projected pixel footprint resulting in what is known as a mixed pixel [12, 13, 14]. This mixing will confuse the signature resulting in a missed classification thus in this implementation will push it into known trash class or unclassified trash.

## 3.3 Developing Truth

Iterative Self-Organizing Data Analysis Technique A (ISODATA) is a method that performs unsupervised learning by formulating data into like clusters through a closest average fit with the fit factor ($\omega$) being a user input. While not required, in our trials we have seeded the twelve bins (labeled A..L, with M being a catch all) with known data values (classifications), see Table 1. Through visualization these datum were selected and coded into the analysis as manual truth.

**Table 1:** Enumerations identifying media types where targets are located.

| Enumerator | HEX Value | Count |
|---|---|---|
| LIGHT_MEDIA | 0x0001 | 14 |
| LIGHT_STAINED | 0x0002 | 18 |
| LIGHT_UNSTAINED | 0x0004 | 8 |
| LIGHT_TRASH | 0x0008 | 10 |
| MED_MEDIA | 0x0010 | 31 |
| MED_STAINED | 0x0020 | 22 |
| MED_UNSTAINED | 0x0040 | 13 |
| MED_TRASH | 0x0080 | 14 |
| DARK_MEDIA | 0x0100 | 49 |
| DARK_STAINED | 0x0200 | 45 |
| DARK_UNSTAINED | 0x0400 | 32 |
| DARK_TRASH | 0x0800 | 23 |

Initialization of the ISODATA algorithm, according to these authors [15, 16], each bin is to be randomly seeded. However, experience has shown that the runtime of the algorithm can be reduced by selectively seeding the bins thus providing guiding weight to its convergence for both algorithms. Unlike the ISODATA routine, this implementation does not

---

[1] In traditional image processing, we refer to a pixel by its Cartesian address, where a hyperpixel is addressed the same, it carries additional spectral data per channel thus is modeled in three dimensions.

allow the seeds to be reorganized thus providing a reloadable signature basis.

Several locations within the data have been selected and identified by a subject matter expert, in this case a doctoral candidate in microbiology, with extensive microscope experience. Enumerated values were assigned to the manually classified data points and counted. We noted the quality of the typical microscope light source was inconsistent with the constraints necessary to evaluate HSI data and broke the classifications into three general light levels, light, medium, and dark – a general reference to the collective reflectance quality of the tested image. Any unidentified artifacts are classified as trash knowing that future classification efforts may well identify them. The hex (base 16) values are randomly assigned and have value only in binning the data.

Confusion tables are commonly used in determining the correctness of the algorithm detections. In this case, Table 2, illustrates the ideal detections that would match the priori classifications. We then compare results with Table 2 as the algorithm is developed with the goal of greater than 97% accuracy.

## 4 Spectral Angle Mapping

The Spectral Angle Mapper classification (SAM) [17] is an automated method for directly comparing image spectra to a known spectrum (usually determined in a lab or in the field with a spectrometer) or an end member. This method treats both (the questioned and known) spectra as vectors and calculates the spectral angle, $\omega$, between them. This method is insensitive to illumination since the SAM algorithm uses only the vector direction and not the vector length. The result of the SAM classification is an image showing the best match at each pixel. This method is typically used as a first cut for determining areas of homogeneous regions.

The goal is to find waveforms that are similar to one another such that a signature could be determined and used in a search on non-training data. In this case training data is defined as data that has some known values however unknown to the selection process. The known values are used to evaluate performance and indicate when the signature process has completed the training process. Just as primary schools use a grading level to measure progress of a student, the ISODATA routine uses
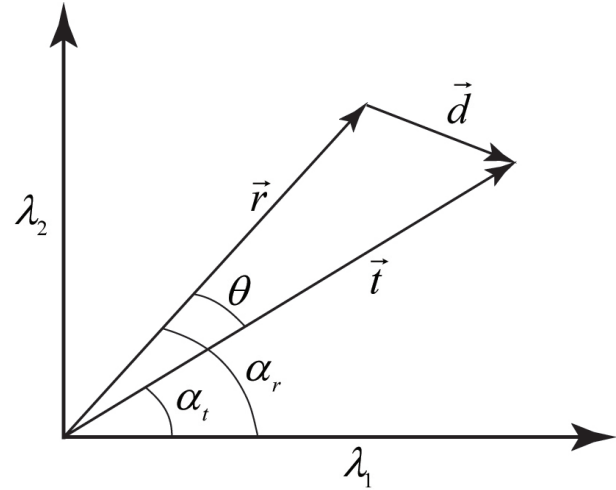


**Figure 3:** A simple $\lambda$ on $\lambda$ plot to explain SAM [10].

some data metric to describe what it is attempting to classify. The Spectral Angle (SA) is useful as it describes the closeness of one vector to another. In this case we have a reference vector from one of the principal components selected, $\vec{r}$ reference vector, and $\vec{t}$ target vector.

$R_t(\lambda_n)$ is the target spectra for which the spectral angle is to be calculated. This is the hyperpixel data that contains the intensities per band for which we are searching for a key to identify a subclass of material. $R_r(\lambda_n)$ is the reference spectra, or one of the principal components that were selected priori. In all cases the first sample selected is specific in that it becomes $R_r(\lambda_n)$.

$$r = \sum_1^n R_r(\lambda_n)^2 \qquad (1)$$

is the sum of the squares for the reference hyperpixel and

$$t = \sum_1^n R_t(\lambda_n)^2 \qquad (2)$$

is the sum of the squares or the reference vector.

We plot these vectors in 3D space however for clarity and ease of referencing we demonstrate them in 2D space assuming the reader understands the inferred 3D plot. In finding Figure 3, we can see that $\vec{r}$ and $\vec{t}$ are plotted starting at the origin. It is note worthy here to explain that SAM is not interested in the magnitude of the reflectance value rather the angle between the two vectors. This makes

**Table 2:** Manually selected truth confusion table, values are hyperpixel counts.

|  | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) | (J) | (K) | (L) | (M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Light Media (A) | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Stained (B) | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Unstained (C) | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Trash (D) | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Media (E) | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Stained (F) | 0 | 0 | 0 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Unstained (G) | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Trash (H) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 |
| Dark Media (I) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 49 | 0 | 0 | 0 | 0 |
| Dark Stained (J) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 45 | 0 | 0 | 0 |
| Dark Unstained (K) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 | 0 | 0 |
| Dark Trash (L) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 0 |
| Unclassified (M) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
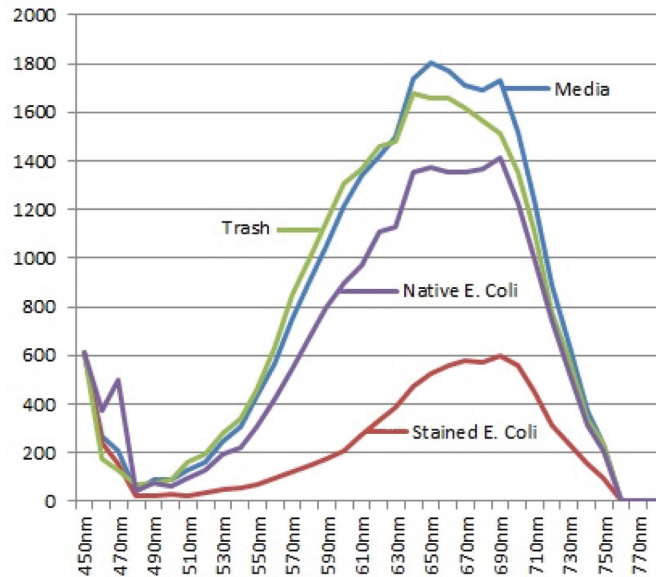


**Figure 4:** General aggregate signal shape of the four known classes.

SAM resilient to fluctuations in intensity due to poor lighting control.

We calculate the dot product

$$A \cdot B = \sum_{1}^{n} R_r(\lambda\_n) \times R_t(\lambda\_n) \qquad (3)$$

which gives the spectral angle in radians as

$$\theta = cos^{-1}\left(\frac{A \cdot B}{r \times t}\right) \qquad (4)$$

Along with $R_r(\lambda_n)$, $\theta$ becomes the SAM reference value by which searches into the data set will attempt to identify *E. coli* and other classified endmembers.

This approach allows for N $\lambda$ bins and attempts to minimize the fluctuation in lighting intensity. However SAM has limits as the resultant coefficient is a single value, and then allowing for some percentage of error, there exist the potential for the SAM to spill over into the ranges of dissimilar endmembers.

## 4.1 Spectral Correlation Mapping

We introduce a further refinement to the ISODATA method as the Spectral Correlation Mapper (SCM) algorithm, Eq. (7). SCM is generated in parallel, however is only conditionally evaluated when SAM is unable to discriminate or has multi-correlations.

$\bar{r}$ is the average of the iterative evaluations of 1 as is $\bar{t}$ to 2.

$$\bar{r} = \frac{\sum_1^n R_r(\lambda_n)^2}{count} \tag{5}$$

is the mean sum of the squares for the reference hyperpixel and

$$\bar{t} = \frac{\sum_1^n R_t(\lambda_n)^2}{count} \tag{6}$$

Then the means are calculated where count is the number of samples in the signal. We now sum the products of the mean adjusted elements of both $R_t(\lambda_n)$ and $R_r(\lambda_n)$ arrays divided by the products sum of the mean adjusted squares of both $R_t(\lambda_n)$ and $R_r(\lambda_n)$.

Where $\alpha$ is formed as an angle (expressed in radians) describing the simularity between the reference hyperpixel spectrum (Y) and the hyperpixel under test (X). The SAM execution will generate results that are over classified or miss-classified. Given the shapes of the target signals, Figure 4, and the geometric similarity with the coexistent end members one can easily see how narrow the error budget is.

During the training process, the training data is evaluated via the SAM algorithm to produce pure spectral signatures. After an identified spectra has been identified it is then referred to as an *endmember*. We must assume, even at these path lengths the pixels will contain a variation of material signatures. Specifically we can reasonably expect to see our target bacteria and the expressed components of the emerging bio-film, agar materials, and expect random contaminations although efforts are put forth to minimize these. Thus the hyperpixels will contained a mixed signal and is subject to mixed pixel aberrations in the signal construct.

$$\alpha = \frac{\left(\sum_1^n R_t(\lambda_n) - \bar{t}\right) \times \left(\sum_1^n R_r(\lambda_n) - \bar{r}\right)}{\sqrt{\left(\sum_1^n R_t(\lambda_n) \times \bar{t}\right)^2 \times \left(\sum_1^n R_r(\lambda_n) \times \bar{r}\right)^2}} \tag{7}$$

## 5 Implementation Details

The primary difference between the two methods is that the SAM is a single angular relationship along the horizontal axis; where SCM uses pairs of deviations, e.g. $x - \bar{x}$ and $y - \bar{y}$ to qualify the differences along both the horizontal and vertical axis. Combining the two methods yields a higher capacity to detect false positives.

As the training process progresses along the truth hyperpixels, it will average the resulting SAMs and SCMs yielding $\bar{\theta}$ and $\bar{\alpha}$ respectively. When the system runs as a detector, e.g. with known signatures, these averages will remain constant yielding the basis of the signature set, however in training mode the system is seeking the final values.

Refer to Figure 5 for the following algorithm flow description of the process. In both the learning and detection processing, we load the SAM and SCM coefficients by either calculating from the truth file or loading from a previously calculated learning run. Then entering a couple short loops, we exercise the $\omega$ fit for both the SAM and SCM truth seeds. This narrows the user input $\omega$ fit from typically 3 to 8 percent to a level that will independently resolve the truth set without overlaps or over fitting. This is shown in the flow chart as Do for SAM and Do for SCM loops.

Next we loop for class intersections between the SAM and SCM fits in order to narrow down to a selection. This processing state is valid for both training and detection. Given only one SAM fit and one SCM fit and they both agree, we log the selection in data and return the value to the caller function. The caller function will then advance the X,Y coordinate and begin the processing again. In the case of truth processing we will see the next logical hyperpixel in the truth list and in detection we will see the next logical hyperpixel as the X or Y value will have changed.

Assuming the next hyperpixel processing results in an intersection count>1 then we look to narrow the overfit selection by determining which class has the closest fit where the SAM will converge to zero for a perfect match and SCM will converge onto one. In the flow chart $\theta$ represents the SAM fit value while $\alpha$ represents the SCM fit value and $\omega$ is the tolerance around zero or one.

The differences are initialized assuming class 'A' holds the closest fit. diff and sdiff are initialized per Eqs. 8 and 9, respectively. The $\Delta$ values are the differences between the now class average divergence from either zero or one and the hyperpixel under test divergence.
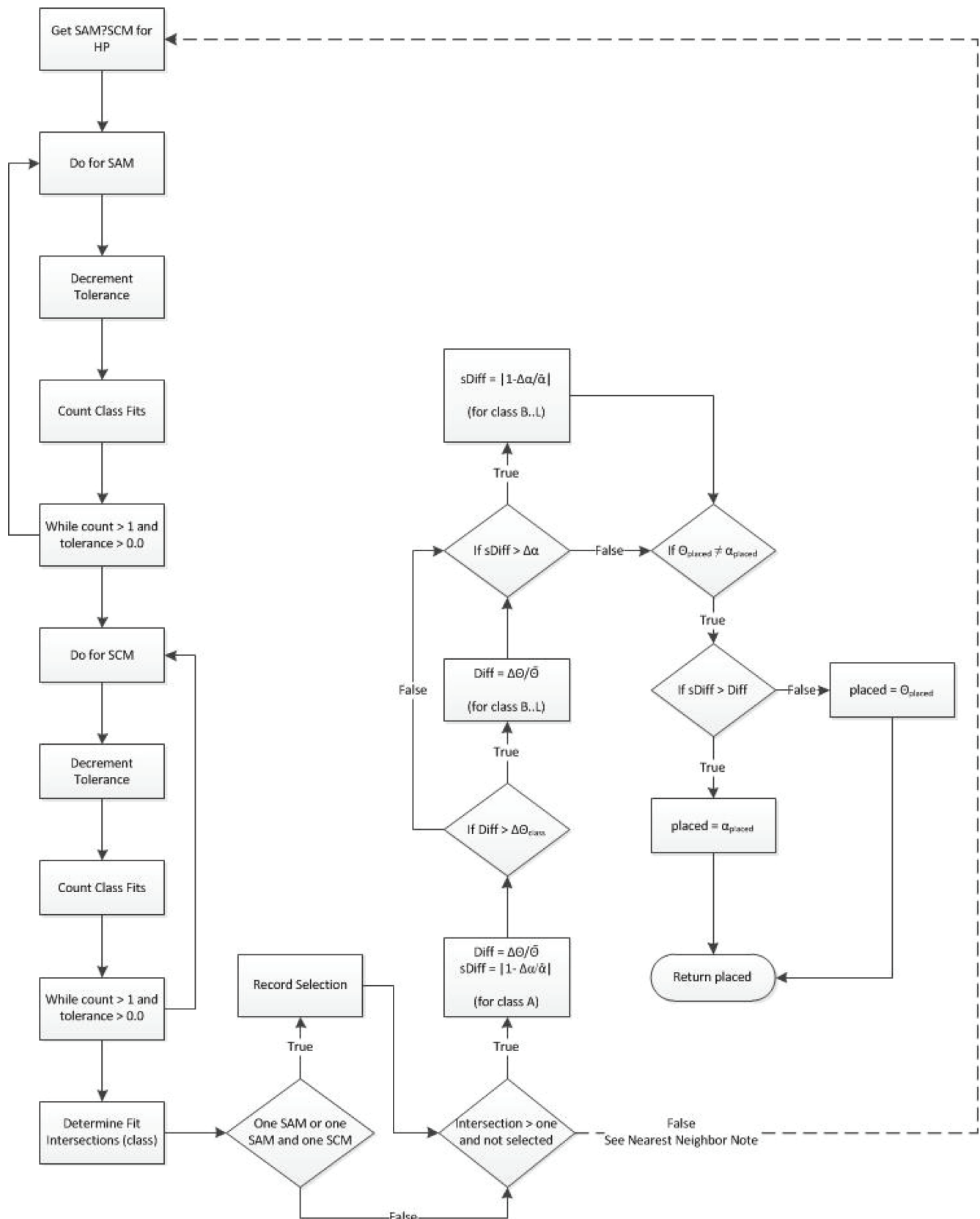
$$diff = \frac{\Delta \theta}{\bar{\theta}} \tag{8}$$

**Figure 5:** Detection Flow Chart used in signature development, validation, and detections.

$$sdiff = \left| 1 - \frac{\Delta\alpha}{\overline{\alpha}} \right| \qquad (9)$$

Then for each class B..L we logically compare the same results with the previous difference value selecting the one that is closest to either the SAM or SCM convergence value.

The SAM logic states if diff> the class evaluation of Eq. 8 then diff = the class evaluation else remains unchanged. Where the SCM logic follows suit but is looking for a difference in magnitude.

The final selection is done as if sdiff>diff then the selection class is the smaller $\alpha$ otherwise its the smaller $\theta$ class.

# 6 Results and Discussion

The two approaches above (SAM and SCM) are blended into a single profile that provides adequate discrimination of the sample's components. This approach has precedence as similar methods were explored by Fauvel [18]. This too mimics the findings of Jin's [9, 19] 80-20 rule where the SAM function identifies 80% of the targets and the SCM is optimized to pull in the remaining 20% with some error in identification accuracy. These two algorithms were independently developed in support of earth science programs and integrated here to solve the low noise low feature count found in a relatively flat fielded microscopic field of view.

Data was collected in November 2012 at the Texas Tech Health Sciences Center using the following environment. The Texas Tech University Health Sciences Center's Cell Physiology and Molecular Biophysics Imaging Center provided the following equipment in a dark room environment. The microscope is an Olympus TH4 - 100 and used 2 objective ends, 40X for dry measurements and a 60x for wet measurements. The wet imagery yielding the better imagery and is used for this research. The HSI sensor is a CRi Nuance FX HSI sensor system (Caliper Life Sciences, Hopkinton, MA,USA) connects to a laptop via USB connections. The room is light tight and adjacent to a common area for specimen preparation. The scientific-grade CCD imager ( 1392 x 1040 effective pixels ) features a solid state liquid crystal wavelength tuning element. The package is mounted onto a chassis with a standard C-mount camera tube. The CRi Nuance EX (450-900 nm) has a tunable liquid crystal element that provides vibration free control of wavelength selection. Vendor documented accuracy is Bandwidth/8.

The stock Olympus light source (non NIST) is used and the Nuance software is capable of flat fielding the illumination. Procedures outlined in the vendor documentation were followed. All samples are imaged in transmission mode [20] as most of the literature suggest for the most useful images.

Using Nuance [TM](Caliper Life Sciences, Hopkinton, MA,USA) software (version 3.0.1.2), the hyperspectral microscope imagery was acquired and stored in a proprietary format. Prior to image acquisition all parameters are selected, Binning 1x1, Exposure (auto optimized nominally 33 milliseconds), wave lengths of interest 450-950 nm, and a full region of interest (ROI). The spectral interval or mean band width was set to the minimum of 10nm. After image acquisition the proprietary cube was converted to a series of TIFF files representing each slice of the cube from 450-950 nm at 10nm steps. Using Image J these TIFF files were converted to raw binary format stripped of any meta data and organized into a hypercube format.

The SAM only processing yields a 67.0% successful detection and classification rate on the truth data as reported in Table 3. In looking at the isolated effectiveness of the SCM algorithm we see in Table 4 there is a low yield of only 8.5%. However, in joining the two implementations the resulting yield combines to 99.2% as demonstrated in Table 5.

# 7 Conclusions

The combination of utilizing both a spectral angle mapper and spectral correlation mapper proves effective to identify and isolate K12 *E. coli* from a prepared microscopic slide while in platonic form. Processing times for large imagery, approximately 1300 x 900 hyperpixels, remains painfully slow. With recent multi-core desktop processor advancements however observing in smaller sub-images offers excellent results. Even with the processing speed, the turnaround time is better than the traditional lab processing thus this may offer a pre-screening method to be backed up by traditional laboratory findings. Both the SAM and SCM algorithms are easily implemented. Care must be exercised when scaling the fit ranges as noise levels rise and or lighting conditions change.

**Table 3:** SAM only, full spectrum, detections along X axis, truth along Y axis resulting in a success rate of 67.6%

|  | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) | (J) | (K) | (L) | (M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Light Media (A) | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Stained (B) | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Light Unstained (C) | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Trash (D) | 0 | 1 | 1 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Media (E) | 4 | 0 | 0 | 0 | 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Stained (F) | 0 | 0 | 0 | 0 | 1 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Unstained (G) | 0 | 0 | 0 | 0 | 2 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Trash (H) | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 13 | 0 | 0 | 0 | 0 | 0 |
| Dark Media (I) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 49 | 0 | 0 | 0 | 0 |
| Dark Stained (J) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 43 | 0 | 0 | 0 |
| Dark Unstained (K) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 29 | 0 | 0 |
| Dark Trash (L) | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 20 | 0 |
| Unclassified (M) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 4:** *SCM only, full spectrum, detections along X axis, truth along Y axis resulting in a success rate of 8.5%*

|  | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) | (J) | (K) | (L) | (M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Light Media (A) | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Stained (B) | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 |
| Light Unstained (C) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| Light Trash (D) | 0 | 1 | 1 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| Medium Media (E) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 |
| Medium Stained (F) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 |
| Medium Unstained (G) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| Medium Trash (H) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 |
| Dark Media (I) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 49 |
| Dark Stained (J) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 45 |
| Dark Unstained (K) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 |
| Dark Trash (L) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 |
| Unclassified (M) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5:** SAM SCM integeration, full spectrum, detections along X axis, truth along Y axis resulting in a 99.2% success rate.

|  | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) | (J) | (K) | (L) | (M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Light Media (A) | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Stained (B) | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Light Unstained (C) | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Light Trash (D) | 0 | 0 | 1 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Media (E) | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Stained (F) | 0 | 0 | 0 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Unstained (G) | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 |
| Medium Trash (H) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 |
| Dark Media (I) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 49 | 0 | 0 | 0 | 0 |
| Dark Stained (J) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 45 | 0 | 0 | 0 |
| Dark Unstained (K) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 | 0 | 0 |
| Dark Trash (L) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 22 | 0 |
| Unclassified (M) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# 8  Acknowledgements

# References

[1] Anna Luiza Brito, Livia Rodrigues Brito, Fernanda Honorato, and et. al., (2013). Classification of Cereal Bars Using Near Infrared Spectroscopy and Linear Discriminant Analysis *Food Research International*, 51, 924-928.

[2] Economic impact of E. Coli on the Agricultural Industry, http://www.newsflashenglish.com/english-lessons/economic-impact-of-e-coli-on-the-agricultural-industry (accessed September 15, 2016).

[3] Rahman, M. S. (2015). Hurdle technology in food preservation, Minimally Processed Foods (Mohammed Wasim Siddiqui and Mohammad Shafiur Rahman, eds.), Food Engineering Series, Springer International Publishing, pp. 17-33.

[4] Lothar Leistner and GrahameW. Gould, The hurdle concept, Hurdle Technologies, Food Engineering Series, Springer US, 2002, pp. 17-28 (English).

[5] Fengle Zhu, Derong Zhang, Yong He, Fei Liu, and Da-Wen Sun, Application of visible and near infrared hyperspectral imaging to differentiate between fresh and frozen-thawed fresh fillets, Food and Bioprocess Technology 6 (2013), no. 10, 2931-2937 (English).

[6] M Bioucas-dias, Antonio Plaza, and E Caceres, An Overview of Hyperspectreal Unmixing: Geometrical, Statistical, and Sparse Regression Based Approaches, IEEE (2011), 1135-1138.

[7] G Girouard, A Bannari, A El Harti, and A Desrochers, Validated Spectral Angle Mapper Algorithm for Geological Mapping: Comparative Study between Quickbird and Landsat, IEEE Geoscience and Remote Sensing Letters.

[8] R. Ramanath and W.E. Snyder, Band selection using independent component analysis for hyperspectral image processing, IEEE, 32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. (2003), 93-98.

[9] Jihao Yin, Yisong Wang, and Zhanjie Zhao, Optimal band selection for hyperspectral image classification based on inter-class separability, (2010), 1-4.

[10] Osmar Abilio, De Carvalho Junior, Renato Fontes Guimaraes, Roberto Arnaldo, Trancoso Gomes, and Ana Paula Ferreira, Spectral Multiple Correlation Mapper, Geoscience and Remote Sensing Symposium (2006), 2762-2765.

[11] Bosoon Park, William R. Windham, Scott R. Ladely, Prudhvi Gurram, Heesung Kwon, Seung- Chul Yoon,

Kurt C. Lawrence, Neelam Narang, and William C. Cray, Classification of Shiga toxin-producing escherichia coli (STEC) serotypes with hyperspectral microscope imagery, Sensing for Agriculture and Food Quality and Safety IV (SPIE) 8369 (2012), 83690L-83690L-13.

[12] Sen Jia, Zhen Ji, and Yuntao Qian, Band selection based hyperspectral unmixing, (2009), 303-306.

[13] Xiuping Jia, Cluster-Space Hyperspectral Data Representation for Mixed Pixel Analysis, School of Electrical Engineering, University College, The University of New South Wale (2000), no. 3, 2167-2169.

[14] Paul L Rosin, Robust Pixel Unmixing, Transform 39 (2001), no. 9, 1978-1983.

[15] Richards, J. A. (2004)Cluster-Space Classification:, Information Sciences 00, no. C, 07-410.

[16] Jain, A. K., Murty, M. N., and Flynn, P. J. (2000). Data Clustering: A Review, ACM Computing Surveys 31 no. 3.

[17] Seldon O Morgan, Richard B Gomez, William E Roper, and Phillips Hall Suite, Squeezed signature analysis hyperspectral classification, Image (Rochester, N.Y.) (2003).

[18] Mathieu Fauvel, Student Member, Jón Atli Benediktsson, Jocelyn Chanussot, Senior Member, and Johannes R Sveinsson, Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles, 46 (2008), no. 11, 3804-3814.

[19] Yaochu Jin and Bernhard Sendho, Pareto-Based Multiobjective Machine Learning : An Overview and Case Studies, IEEE Transactions on Systems, Man, And Cybernetics 38 (2008), no. 3, 397-415.

[20] Mehrube Mehrubeoglu, Paul V. Zimba, Lifford L. McLauchlan, and Ming Y. Teng, Spectral unmixing of three-algae mixtures using hyperspectral images, 2013 IEEE Sensors Applications Symposium Proceedings (2013), 98-103.

## About the Authors



**Dr. T. Horton**, while working in the defense industry, he received his masters and Ph.D. from Texas Tech University, Lubbock Texas. For his dissertation thesis, Dr. Horton explored the world of microbiology and how engineering can further both fields in a transdispinary way. His work has led to the development of efficient methods to develop hyperspectral signatures in many areas where both low and high contrast target media should be expected. Having greater than 30 years of industry experience he has concentrated studies in real-time systems, tracking systems, radar and graphical user interfaces that support his fellow military active duty members. He spends his free time with his four children who are very active in scouts. Dr. Horton is giving back to his community by being very active in scouting STEM initiatives with the hopes of enabling and empowering the largely untapped technical resource that is bottled up in American female students, of which his two daughters are actively engaged.



**Dr. Matthew Kay** received his Ph.D. in Biomedical Sciences from Texas Tech University, and his Masters in Biology from Texas State University. He is a Postdoctoral Fellow performing research at the Naval Medical Research Center in San Antonio, Combat Casualty Care Division, at Fort Sam Houston. He is a member of the American Society of Microbiology. Dr. Kay has contributed to several book chapters and authored several papers. He has received several awards and honors for his contributions to the fields of phage biology and bacterial biofilms.



**Dr. Atila Ertas**, Professor of Mechanical Engineering, received his masters and Ph.D. from Texas A&M University. He is a Senior Research Fellow of the IC$^2$ Institute at the University of Texas Austin, a Fellow of American Society of Mechanical Engineers (ASME), a Fellow of Society of Design and Process Science (SDPS), and a Fellow and honorary member of The Academy of Transdisciplinary Learning & Advanced Studies (TheATLAS).

He is also an honorary member of International Center for Transdisciplinary Research (CIRET), France. Dr. Ertas has earned both national and international reputation in engineering design. Dr. Ertas is the author of a number of books, and technical papers that cover many engineering technical fields. Dr. Ertas' contributions to teaching and research have been recognized by numerous honors and awards. He has been PI or Co-PI on over 40 funded research projects. Under his supervision 194 MS and Ph.D. graduate students have received degrees.



**Dr. Derrick Tate**, is an Associate Professor and Founding Head of the Department of Industrial Design at Xian Jiaotong-Liverpool University. He aims to impact society through bringing design thinking to areas of strategic importance: assessing the innovative potential of design ideas, developing sustainable approaches for building systems, transportation, and manufacturing; and broadening participation in innovation. Dr. Tate has carried out his research activities at the two ends of the research spectrum where they will have the greatest impact: fundamental research that provides a science base for the future of entrepreneurial engineering design as well as the application of design theories and tools to technology innovation. His recent projects include working with West Texas entrepreneurs on the development of innovative and sustainable designs and a US-Tanzania Workshop: Advancing the Structural Use of Earth-based Bricks, funded by NSF. He received a B.S. in Mechanical Engineering degree from Rice University. His S.M and Ph.D. degrees in Mechanical Engineering are from MIT in the areas of manufacturing and design, respectively.