Received: 17 August 2016 ■ *Revised: 25 October 2016* ■ *Accepted: 1 November 2016* ■ *Publication: 16 November 2016*

JOURNAL OF SPECTRAL IMAGING

Paper Presented at IASIM 2016, July 2016, Chamonix, France

openaccess

Rapid identification of *Salmonella* serotypes through hyperspectral microscopy with different lighting sources‡

Matthew Eady and Bosoon Park*

United States Department of Agriculture, Agricultural Research Service, US National Poultry Research Center, Athens, GA 30605, USA.
E-mail: bosoon.park@ars.usda.gov

The rapid detection of food-borne pathogenic bacteria is critical to the food industry for preventing the introduction of contaminated product into the marketplace and limiting the spread of outbreaks. Hyperspectral microscope images (HMI) are a form of optical detection, which classify bacteria by combining microscope images with a spectrophotometer. The objective of this study was to compare the spectra generated from dark-field HMIs of five live *Salmonella* serotypes from two lighting sources, metal halide (MH) and tungsten halogen (TH), assessing classification accuracy and robustness, between 450 nm and 800 nm. It was found that the MH spectra could be reduced to as few as 10 optimal bands between 594 nm and 630 nm, but TH band reduction decreased accuracy, due to the inherent broader peak structure generated by the TH light source. Collection of HMIs from the two light sources comparing the same cells shows slight differences in scatter intensity patterns. Principal component linear discriminate analysis classified serotype subsets ($n = 1800$), reporting both MH and TH accuracies at 100%, while the reduced key MH bands achieved 99.4–100% accuracy. Principal component regression calculated the root mean squared error of cross-validation < 0.014 and a $R^2 > 0.948$ for both full spectrum lamps. MH or TH lamps can be effectively used for discriminating bacteria HMIs on a cellular level by serotype, but reducing TH bands may lose crucial classification information.

Keywords: hyperspectral, microscopy, *Salmonella*, rapid detection, food safety, tungsten halogen, metal halide

Introduction

Food-borne disease is responsible for approximately 48 million illnesses and 3000 deaths per year, with *Salmonella* being the most common non-viral pathogen causing domestically acquired illness.[1] *Salmonella* is a food-borne pathogen that has caused outbreaks in poultry, meat, produce, nuts and dairy products.[2] The gram negative bacteria attaches to the digestive tract in humans causing gastroenteritis. In rare cases, this can lead to septicaemia, typhoid fever or death. *Salmonella enterica* and *Salmonella bongori* are the two known species of *Salmonella*. The *Salmonella bongori* species is most commonly found in reptiles, and not a major concern for humans. However, the *Salmonella enterica* species causes illness in mammals and consists of six sub-species, as well as numerous serotypes. Here, we focus on serotypes commonly found in human salmonellosis outbreaks, from *Salmonella enterica* subsp. *enterica*.

Since accurate identification of pathogenic contaminates in an outbreak investigation is critical for preventing further threats to human health, there is a need for early and rapid detection of pathogens. Traditional plating techniques may take 7–10 d or more for confirmation.[3] Polymerase-chain reaction

‡Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

ISSN: 2040-4565
doi: 10.1255/jsi.2016.a4

(PCR) requires much less time, but can be expensive due to the high reoccurring cost associated with cell lysing reagents necessary for DNA purification and amplification.[4] The reduction in time required for identification can result in a larger impact on human health, leading to faster outbreak responses, clinical infection identification or detecting contaminated food product before it leaves the production chain.

Optical detection has the potential to non-destructively assess a food product for the presence of microbial life.[5–7] Optical detection works to identify bacteria based on a spectral profile that is inherently unique to the organism. Hyperspectral imaging (HSI) is a form of optical detection that differs from other methods by collecting both spatial data (x and y coordinates) and spectral data (λ) in the form of a three-dimensional hypercube. The hypercube combines images collected with a spectrophotometer across a range of the electromagnetic spectrum. These images are stacked on top of one another creating a cube. Here, the images maintain constant x and y coordinates, while the spectral information changes with each wavelength imaged. Hyperspectral microscope images (HMIs) are generated from live bacterial suspensions mounted on a simple microscopic glass slide. Previously, HMIs have been used to differentiate species[8–10] and serogroups or serotypes[11,12] of food-borne bacteria, as well as to investigate the potential for early and rapid detection. Dark-field microscopy was used to illuminate the cells against a black background creating a contrast of intensity values. The signals produced are light backscatter, where some light is being absorbed and some light is scattered. The contrast of scattered light allows pixels representing cells to be extracted from the dark-field background. In most of the previous studies a metal halide (MH) was used as the lighting source. Similar to MH, tungsten halogen (TH) lighting sources offer an affordable illumination option for microscopy. Anderson *et al.*[8] used a 3200°K TH lamp to differentiate between live/viable and dead/non-viable bacterial endospores. The light scattering spectra generated from cells were able to discriminate between live and dead endospore cells only when the outer membranes had been damaged by hydrogen peroxide (H_2O_2). Currently, few studies using HMI exist, and it is unclear how differences in lighting sources such as a MH or TH lighting differ. In order to move towards an early and rapid detection method using HMI as a tool in food-borne disease outbreak investigation or in biomedical detection of blood-borne pathogens, a detailed comparison of light sources for the system is needed. Here, the objective was to compare HMIs of the same *Salmonella* cultures generated by MH and TH lamps, assessing accuracy and robustness of classification algorithms for rapid detection of five *Salmonella enterica* subsp. *enterica* serotypes, to determine an optimal lighting source.

Materials and methods

Bacterial samples

The five *Salmonella* serotypes used in this study were: *Salmonella* Enteritidis (SE), *Salmonella* Heidelberg (SH), *Salmonella* Infantis (SI), *Salmonella* Kentucky (SK) and *Salmonella* Typhimurium (ST). These cultures were obtained from chicken rinsates, and collected from the Poultry Microbiological Safety and Processing Research Unit at the US National Poultry Research Center located in Athens, GA. Isolates were stored at –80°C until needed. Samples were inoculated into tryptic soy agar slants for short-term storage and held at 4°C. Fresh cultures of the five *Salmonella* serotypes were prepared by inoculating a few colonies from the agar slants into 10 mL of tryptic soy broth (TSB). TSB samples were incubated at 37 ± 2°C for 18–24 h. TSB cultures were then serially diluted in peptone buffer up to 10^{-5}, followed by pipetting 100 µL of 10^{-5} dilution onto brilliant green sulfa (BGS) agar plates, a selective *Salmonella* media, for a final dilution factor of 10^{-6}. Plates were then incubated at 37 ± 2°C for 18–24 h. Following incubation, microscope sample slides were prepared, similar to the protocol described in Park *et al.*[9] Briefly, one colony from each plate was placed in 100 µL of molecularly sterile water and briefly vortexed. 3 µL of the bacterial suspension was spread onto the centre of a glass slide, and allowed to air dry in a biosafety cabinet (Baker, BSC, Sanford, ME, USA) for 15 min. After drying, 0.8 µL of sterile water was added on top of the dried suspension to secure a glass cover slip for oil immersion. Cells were immobilised by firmly pressing the cover slip to the slide. A drop of immersion oil was added to the top of the cover slip. The total time for slide preparation and image collection was approximately 20 min. This process was repeated for a total of three repetitions, a calibration, validation and test data set, for each lighting source and all five serotypes. The three data sets were collected over the span of one year, with new cultures grown for each repetition to assess repeatability in the analytical methods.

HMI system and lighting sources

The HMI system can be seen in Figure 1. It consists of an upright microscope (Eclipse e80i, Nikon, Lewisville, TX), mounted with an acousto-optical tunable filter (AOTF) (HSI-400, Goouch & Housego, Ilminster, UK), a high-performance cooled electron multiplying charge coupled device (EMCCD) 16-bit camera (iXon, Andor Technology, Belfast, UK) mounted on top of the AOTF and dark-field illuminating light sources—MH and TH lamps offset from the stage in lamp housing boxes. The AOTF filter has the capability for a high-speed, high-throughput, random-access solid-state optical filter with an adjustable optical pass-band and exceptionally high rejected light levels.[9] The AOTF delivers diffraction-limited image quality with variable bandwidth resolution within 2 nm. Both MH and TH images were collected at a gain setting of 3.5%, exposure time of 250 ms and with a 100× oil immersion objective (Olympus, Tokyo, Japan).

A 24 W MH lighting source with MR-11 reflector (Ushio America, Cypress, CA, USA) was selected as the MH illuminating source. The lighting source has a colour temperature of 5600°K, and operates between 0.3 A and 0.4 A. The lamp is ignited by a voltage pulse that ionises the argon gas. Once the lamp has reached operating temperature, vaporised MH

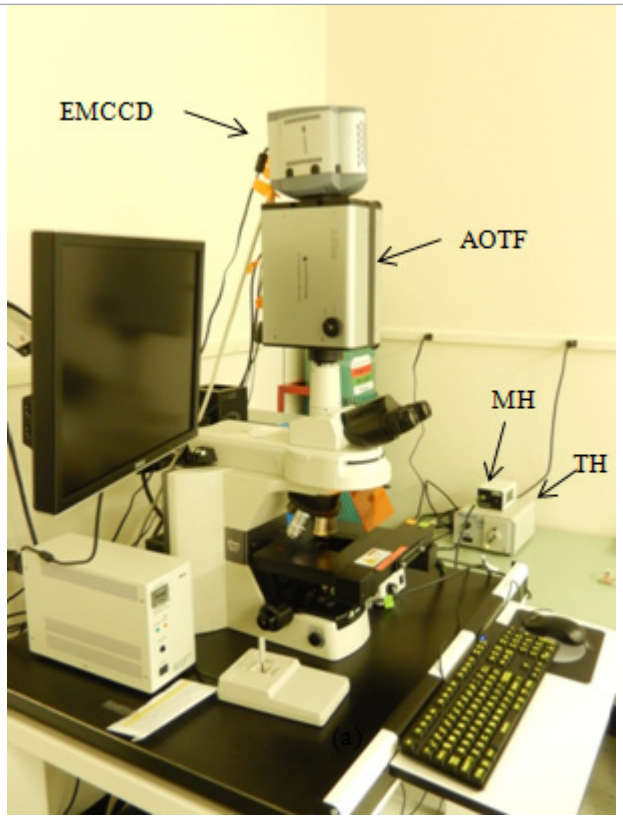

Figure 1. The hyperspectral microscope imaging system. EMCCD = electron multiplying charge coupled device, AOTF = acousto-optical tunable filter, MH = metal halide, TH = tungsten halogen.

salts are dissociated by the arc and excited into higher energy states, emitting spectral lines.[13]

The TH lighting source used here is a 21 V, 150 W lamp (Osram, Munich, Germany) with MR-16 reflector and filled with an inert gas and trace amounts of halogen. This lamp uses a regenerative halogen cycle, as the lamp is heated to an operating temperature of 3300°K. With this process, tungsten atoms are evaporated on the filament and then react with the gaseous halogen vapour, along with trace amounts of oxygen to form tungsten oxyhalide.[14] The benefit of regenerative halogen is that it allows for a smaller glass envelope, resulting in a more economical use of quartz and glass alloys, with approximately 200 h of life expectancy.[15] Both lighting sources were housed in enclosed boxes and kept next to the microscope setup, instead of underneath the slide stage. A fibre-optic cable ran from the lighting houses to the base of the microscope, emitting light upward through the sample and into the AOTF. This was done to avoid bacterial cell injury or death as a result of the heat generated from the lamps with dark-field illumination, because the practical aspect of this detection methodology is that live cells can be identified.[9]

Data mining and image processing

HMI collects a large amount of data in the form of a three-dimensional hypercube. The x- and y-dimensions represent spatial coordinates of the cells that stay constant over the collection of the hypercube images, while the z-dimension represents the light scattering intensity measured at each observed wavelength in the spectrum (89 total). The high signal-to-noise ratio (SNR) allows for information from only the cells to be extracted from the image, using the environment for visualising images (ENVI) software (Exelis, McLean, VA, USA) with a pixel scattering intensity threshold extraction method. Here, we specify the pixel intensity minimum and maximum based on intensity value histograms, and extract only those pixels, discarding the background spectra or non-cellular information, as well as any under- or over-saturated pixels that may be present. Intensity thresholds are used because refractive index gradients occur at the cell walls, and it is difficult to manually select pixels from within the boundaries of the cells, as well as quickly automating the selection process. These extracted pixels form a region of interest (ROI). One ROI with pertinent information was extracted from each image. The average cell was found to have approximately 500 pixels, with 3-D reconstruction of hypercube images performed in OriginPro 9.0 (Origin Lab, Northampton, MA, USA). Pixels within the ROIs were randomised to ensure that multiple subsets are not extracted from the same cell and refractive index gradients are not causing bias of data for model development. Subsets of 500 pixels were averaged to create mean spectra with the approximate number of pixels found in a typical cell, followed by max peak normalisation, where each collected wavelength in the spectrum was divided by the maximum intensity peak value and plotted in SigmaPlot 11 (SigmaPlot, Systat Software, San Jose, CA, USA). Metal halide samples have a strong excitation peak at 546 nm, thus all samples were normalised to 546 nm, whereas the TH samples display more gradual peaks, with 638 nm representing the maximum wavelength intensity for TH sample normalisation. Sixty averaged subsets from each image were collected, creating $n = 300$ per light source (2) and repetitions (3), with two lighting sources and three repetitions resulting in a total $n = 1800$ for the experiment.

Multivariate data analysis

Principal component analysis

Data subsets were imported into the Unscrambler v. 10.3 software (Oslo, Norway) for multivariate data analysis. Here, a covariance-based principal component analysis (PCA) was conducted. PCA methods have been well documented as a method for reducing the high dimensionality of data sets, such as spectroscopy data.[16–18] In general, PCA is an orthogonal transformation that converts, in this situation, highly collinear variables into uncorrelated variables, while reducing the dimensionality with principal components (PCs), and can be calculated by:

$$\mathbf{X} = \mathbf{TP}^{\mathbf{T}} + \mathbf{E} \quad (1)$$

where **X** is the data matrix, **T** represents the score matrix, **P** represents the loading vectors and **E** is the error matrix, or F statistic plotted against Hotellings T^2. Further detail can be found in Wold *et al.*[19]

Principal component-linear discriminate analysis

Principal component analysis is a quick method to visually inspect inherent differences for a data set with multiple classes, but is not specifically a classification algorithm. A principal component-linear discriminate analysis (PC-LDA) was performed on the data sets to quantify classification accuracy of the five serotypes for both of the full 89 bands approaches and the optimal bands selected from the MH light source. Linear discriminate analysis is a supervised classification method, with *a priori* knowledge of class membership. Here, the HMI hypercubes have a greater number of variables (89) than number of samples per image (60), and this would result in LDA breaking down mathematically.[20,21] To counter this problem, LDA was combined with a PCA to form a PC-LDA and to conform to the overall PC space analysis. Equal prior probability was assumed and sample classification results were reported in a confusion matrix, displaying the number of true positives and false positives in a table. PC-LDA was calculated with PCs 1–7 (explaining 99.4% of variance) and then:

$$\mathbf{y} = \mathbf{v^T X} \tag{2}$$

where $\mathbf{v^T}$ a projection vector that multiplies between-class and within-class covariance matrices, $\mathbf{X}$ represents dimensional data or PCs 1–7 in this case, finding $\mathbf{y}$. This reduces the dimensionality of the data matrix by using only the first few PCs, and applies Euclidean distances to determine the class-discrimination between the five serotypes. The Euclidean distance can be calculated by:

$$\mathbf{d} = \left[\sum_{j=l}^{m} (\mathbf{x}_{\mathbf{B}j} - \mathbf{x}_{\mathbf{A}j})^2\right]^{1/2} \tag{3}$$

followed by

$$\mathbf{d} = \left[(\mathbf{x_B} - \mathbf{x_A})(\mathbf{x_B} - \mathbf{x_A})^{\mathrm{T}}\right]^{1/2} \tag{4}$$

where $\mathbf{d}$ is the calculated Euclidean distance, $\mathbf{x_A}$ and $\mathbf{x_B}$ represent two PC vectors, and j represents objects.

Principal component regression

To measure repeatability, a PC-based regression was calculated for assessing similarities in both the validation and test data sets, based on the PCA results from the calibration set. Here, the calibration PCs (first repetition) are the predictors, while the validation and test set (second and third repetitions) spectra are responses, predicting how well the first repetition's PCs are in agreement with validation and test sets. Principal component regression (PCReg) has the potential to overcome collinearity in samples, such as the case with spectroscopic data. This method reduces the number of regression variables. PCR is a two-step process, with the PCs being calculated first, as stated above in Equation (1). Then, the $\mathbf{T}$ matrix is inserted into a multiple linear regression model, with $\mathbf{T}$ being the PCA scores matrix, the regression coefficient $\mathbf{q}$ = the transpose of the PCA's loading matrix and $\mathbf{e}$ is the error matrix.

$$\mathbf{y} = \mathbf{Tq} + \mathbf{e} \tag{5}$$

Model performances of the PCReg were evaluated by calculating the root mean squared error of calibration (*RMSEC*) and cross-validation (*RMSECV*), as well as the coefficients of determination R_C^2 and R_{CV}^2, for the calibration and cross-validation, respectively. These model indices were calculated by:

$$R_c^2 = \sqrt{\sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2} \Big/ \sqrt{\sum_{i=1}^{n_c} (\hat{y}_i - y_{mc})^2} \tag{6}$$

$$R_{CV}^2 = \sqrt{\sum_{j=1}^{n_v} (\hat{y}_j - y_j)^2} \Big/ \sqrt{\sum_{j=1}^{n_v} (\hat{y}_j - y_{mv})^2} \tag{7}$$

$$RMSEC = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2} \tag{8}$$

$$RMSECV = \sqrt{\frac{1}{n_v} \sum_{j=1}^{n_v} (\hat{y}_j - y_j)^2} \tag{9}$$

where $\hat{y}_i$ and y_i are the predicted and measured values from the calibration data, $\hat{y}_j$ and y_j are the predicted and measured values from the validation or test data, y_{mc} and y_{mv} represent the mean values in the calibration data set and validation (or test) data, with n_c and n_v being the number of samples found in the calibration and validation (or test) data. The delta R^2 and *RMSE* were reported as an average of comparing the validation set to the calibration set, as well as the test set compared to the calibration set.

Results and discussion

Metal halide and tungsten halogen images and spectra

For the purpose of identifying bacteria at the cellular level, dark-field microscopy offers a high contrast method with large SNR by illuminating the unstained live cells onto a dark background from every azimuth, with only light scattered from the living cells collected by the microscope objective.[22] Both lighting sources are represented by SH HMIs in Figure 2 with MH on the top row and TH on the bottom row. The left column shows converted hypercube images generated by the two lamps with no noticeable movement in cells generated during the time necessary to collect both images, approximately 45 s per lamp. From these converted images it is difficult to differentiate between cellular scattering patterns as a result of the MH and TH lamp illuminations. In order to take a closer look at the scattering patterns of individual cells, 3D surface plots of cells were constructed, at the maximum peak wavelengths for MH (546 nm) and TH (638 nm). The average cell was found to have approximately 500 pixels when reconstructing images, and was representative of common short rod-shaped *Salmonella* cells which are 0.7–1.5 µm in diameter to 2–5 µm in length.[23] Here, these cells were approximately 15–20 pixels wide by 25–30 pixels in length. An example of a common *Salmonella* cell is shown in the middle column of Figure 2. Here, we see that the same cell imaged with both MH and TH

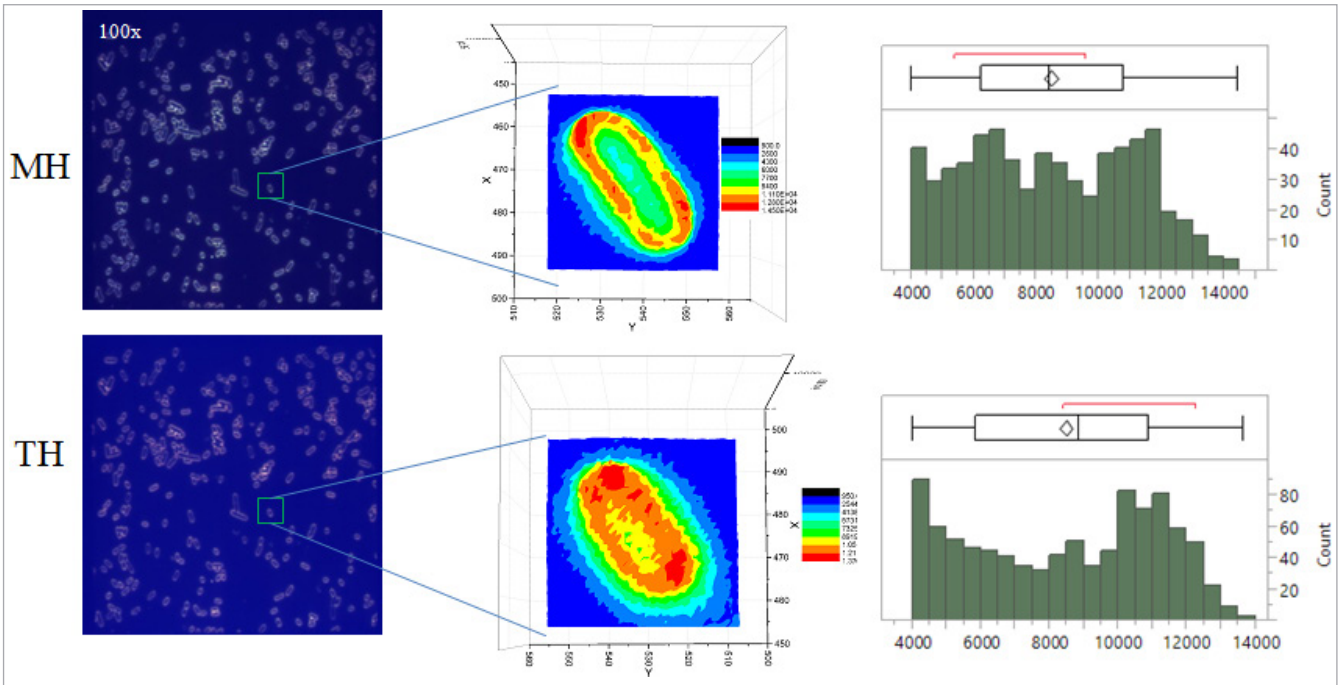

Figure 2. *S.* **Heidelberg column with the top row (MH) = metal halide and the bottom row = (TH). Left: converted average HMI, middle: the same cell reproduced in a 3D surface plot from the two light sources, right: histograms representing the pixels within the cells.**

is similar, but with some noticeable differences in terms of intracellular scattering intensity patterns. The SNR between cell membranes and the background result in different refractive indexes, creating a refractive index gradient.[24] Park *et al.*[10] noted that the inner and outer cell membranes could be observed through HMIs. The refractive index gradients generated by the two light sources appear slightly different for the boundaries between the outer cell membrane and the background. It is not clear if morphological differences in the inner and outer cell walls are causing the light to be reflected or refracted through HMI, but it is apparent that the optical profiles of the two lighting sources vary in terms of light scattering with the inner and outer cell walls. The histograms located in the third column are for pixels contained within the cell, and not the background. Minimum pixel saturations were set at an intensity of 4000 for the histograms. Both histograms show a slight right-side tail, present with increasing pixel intensities. The TH histogram shows that there is a larger number of pixels in the lower and higher intensities, more so than the MH cell. This is in agreement with the 3D plots, suggesting that the TH image is generating a higher contrast in the inner and outer cell walls than MH.

Anderson *et al.*[8] used a TH lamp, but measured reflectance, whereas here we are collecting scattering data. The previous study also collected data every 10 nm between 400 nm and 720 nm, resulting in a 32-band hypercube. Another difference in the instrumental setup is that the previous study used a liquid crystal tuneable filter (LCTF), opposed to the AOTF used in this experiment. These differences in design make it difficult to compare results from this experiment to the previous study. It can be noted that through a similar approach, hypercube acquisition followed by PCA, the Anderson *et al.*[8] paper was able to identify organisms with a damaged outer cell wall, through the use of PCA scores and loadings plots. Verebes *et al.*[25] investigated a procedure to image blood cells with HMI. Erythrocytes and pathogenic bacteria such as *Salmonella* are considerably different in physiology, but both have a cell wall consisting of a lipid bilayer. Because the shape of erythrocytes can be indicative of a number of pathological diseases, the paper investigated the interaction of the MH photons and the cell's shape through random geometric orientations and the use of a spectral angle mapper (SAM). It was determined that geometric orientation of the cell had little effect on the resulting spectral signature with SAM maximum angle measurements of 0.1 radians from several build-up libraries, and proposed that curvature of the cell wall was a larger influence on the spectra. In Figure 2 we can see that backscatter is occurring, with some light being absorbed by the cell and some scattered. Considering the previous studies, noting the hyperspectral significance of the lights' interaction with the curvature or damages to the outer cell walls and the differences in the middle column of Figure 2 showing how MH and TH produce somewhat different spectral scattering patterns when interacting with the cell's outer membrane, point to a need for comparing the two backscatter patterns and thus overall classification abilities, as TH and MH light photons interacted with cells at various lamp excitation levels.

Most of the wavelengths imaged in this experiment are in the visible light spectrum. In this range, vibration overtones and overlapping of spectral attributes are present at individual bands. Observing a 3D reconstructed surface plot from the

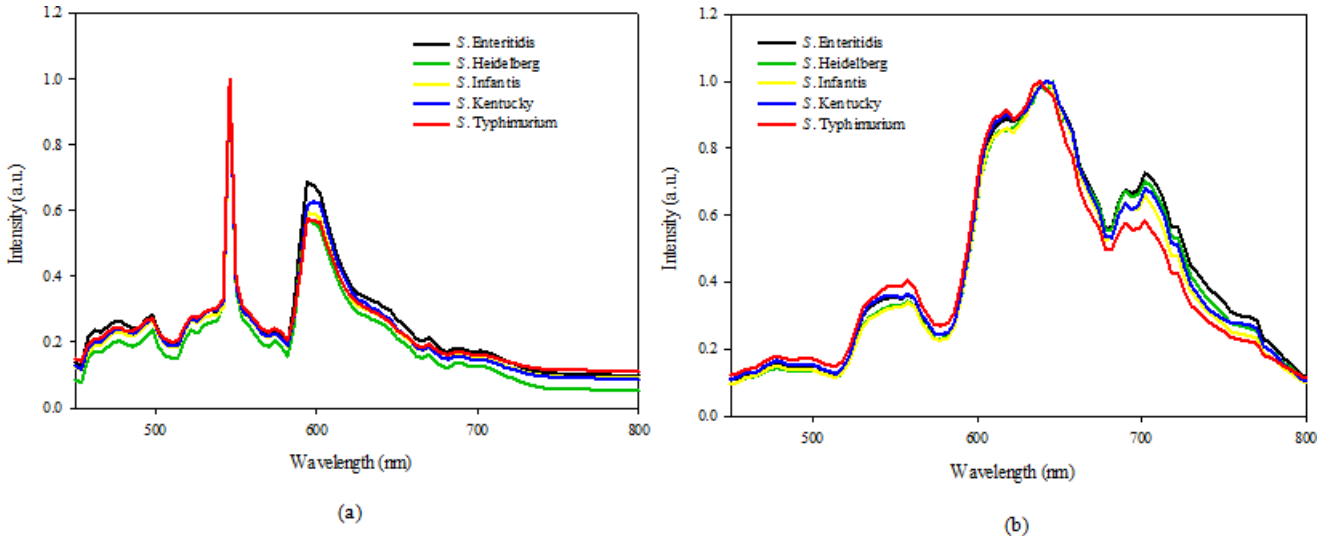

Figure 3. Mean spectra for five *Salmonella* serotypes with maximum peak normalisation applied to (a) MH (546 nm), (b) TH (638 nm).

same cell did not suggest any specifically noticeable band-to-band differences in higher or lower backscatter intensities from one section of the cell to another, but consistent scattering patterns with varying intensities were visible.

Figure 3 shows the average spectra for all subsets of the calibration data set by serotype and lighting source, with (a) being MH and (b) showing the mean TH scattering patterns. Validation and test set repetitions show nearly identical spectral patterns across all serotypes. It is apparent that the spectra are similar and collinear for each lamp, possibly due to the close taxonomical relationships of the five samples, and their inherent similarities in morphological and biochemical profiles. The MH has strong light scattering at excitation peaks at 546 nm and at 590 nm. The spectral profile of TH is similar to that generated by a blackbody radiator, with a more drawn-out profile, where the sharp peaks found in MH are absent. Max peak normalisation was performed on all the spectra to assess HMIs on the same scale, while retaining inherent differences of the samples. *Salmonella* Enteritidis shows the highest mean light scattering intensity for both lighting sources.

The hypercube offers a powerful tool to observe cell scattering patterns at various wavelengths in the electromagnetic spectrum. The loading vectors from two separate PCA analyses of the full MH and TH generated spectra are shown in Figure 4. The first and second PCs represent 96% of the explained variance in the PCAs for both the MH and TH data. Five informative wavelengths were identified from the first two PCs of each light source. These wavelength slices were extracted from the hypercubes and are displayed in Figure 5, with MH in the top row and TH in the bottom row. The intensities of each wavelength vary over the spectrum with the lowest SNR values noticeable at both the highest and lowest wavelengths shown. From Figure 5 it can be seen that scattering patterns remain consistent over the spectrum, as scatter intensity changes. The image shows that wavelengths representing MH peaks at 546nm, 590 nm and 598 nm and TH peaks at 606 nm and 650 nm have higher definition between the base of the cell and the background.

Multivariate data analysis

PCA

First, the PCA was performed on the MH samples. Data reduction is valuable in minimising the processing and storage space that would be necessary with food industry implementation. The loading vectors (**P**) shows how much overall variance is explained by the individual wavelengths. For the MH data set there is a high amount of variance

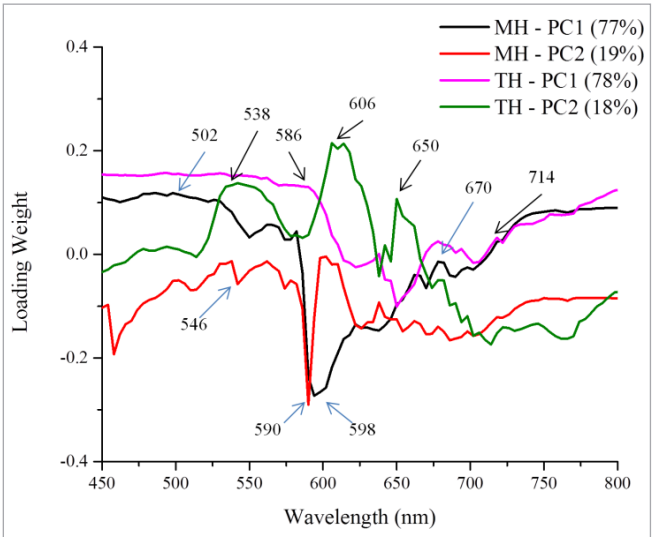

Figure 4. Loading vectors from two principal component analyses, one for five *Salmonella* serotypes and the MH lamp, the other for the same five samples with TH lamp.

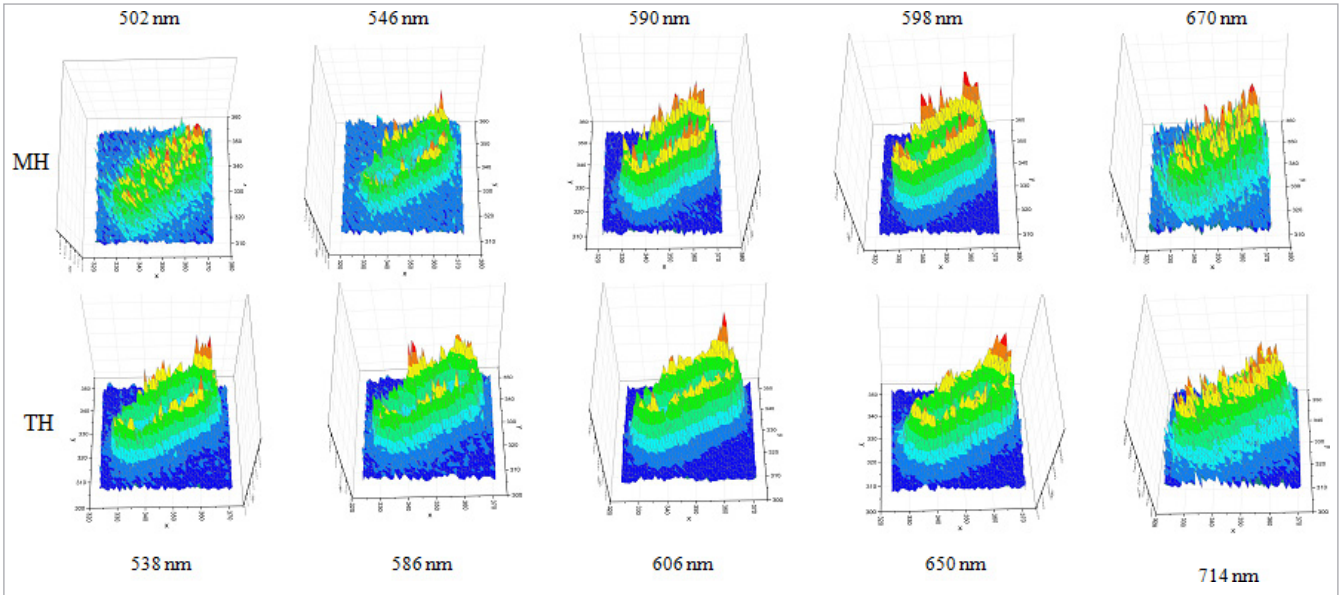

Figure 5. Comparison of a *S.* Enteritidis cell observed with the MH lamp (top row) and the TH lamp (bottom row) at wavelengths deemed informative from the respective PCA loading vectors.

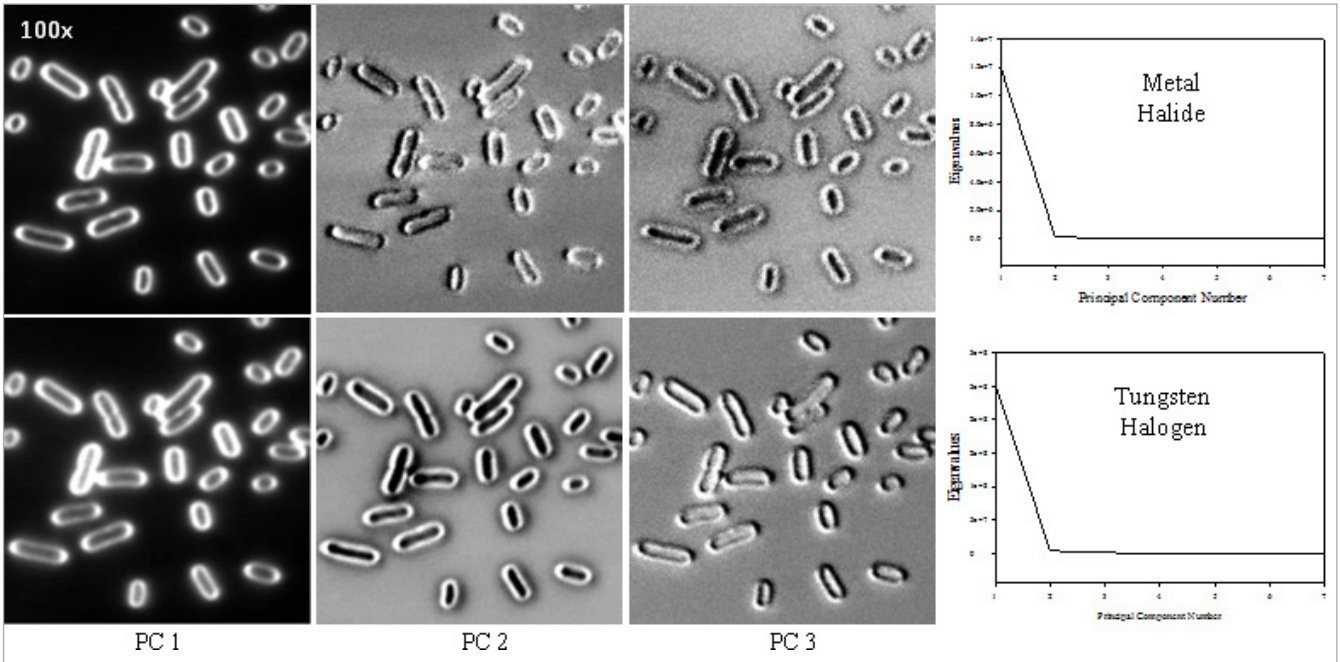

Figure 6. Principal components (PC) images for *S.* Enteritidis images, with the number of eigenvalues described by each PC in the plots, with the top row = MH, and the bottom row = TH. PCs 1–3 are shown in columns 1–3, with the corresponding scree plots in column 4.

explained around 600 nm. A strong negative correlation in the loadings and scores plots was noticed at the 590 nm band. The residuals (not shown) also showed increased variance, when the 590 nm band was included in the analysis, with a multitude of samples above the significant difference threshold for intensity scaling. Removing the band and analysing the optimal MH range of 594–630 nm resulted in increased classification accuracy, consistency with the loadings and a symmetric pattern in plotting residuals when compared to the 590–630 nm range. However, wavelength reduction strategies for the TH data set showed little to no success. Reducing the data set to what appeared to be the most informative wavelengths only, decreased the PCA's serotype clustering, introducing considerably more variance to residuals and further reducing the overall accuracy of the proceeding classification algorithm. TH lacks the sharp excitation peaks of MH, therefore requiring the use of the full spectral profile.

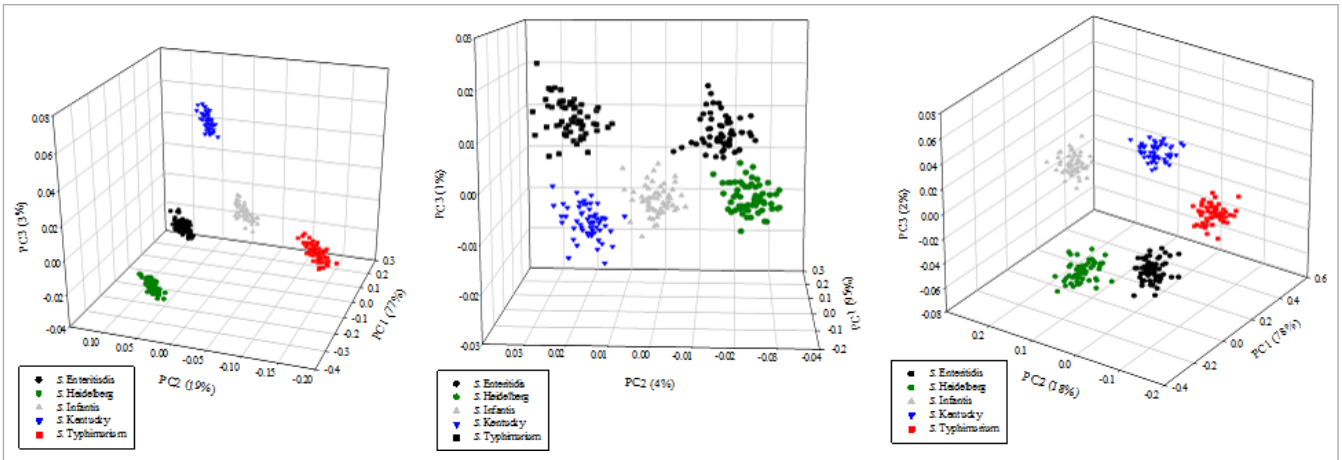

Figure 7. Score plots from principal component analyses for (a) MH, (b) reduced MH range and (c) TH.

Figure 6 shows the first, second and third PCs for a *S.* Enteritidis captured with both the MH and TH lighting sources as shown by the ENVI software. The images show the amount of pixels that are explained by each of the first three PCs as carried out on all collected wavelengths. From these images we can see that most of the cellular pixels are explained by PC 1, with some explained by PC 2 and fewer with PC 3, while the plots in the far-right column quantify the amount of eigenvalues explained by each PC. PCA score plots can be seen in Figure 7, with the full spectrum results of MH (Figure 7a), reduced wavelengths of MH (594–630 nm) (Figure 7b) and the full TH spectrum (Figure 7c). The figures show that with all three data sets and spectral ranges, classification of the five serotypes is possible, with little to no overlapping of clusters present. PCA score plots are in agreement with ENVI images in Figure 6, showing that most of the meaningful data is explained by PCs 1 and 2.

PC-LDA classification

While PCA score plots display an easy to read visual representation of the data, another step is needed to quantify classification accuracy. The PC-LDA did not breakdown mathematically because we reduced the number of variables using PCs 1–7 of the PC. Table 1 shows the results in the form of the overall classification accuracy from both MH and TH lighting sources, and the optimal MH spectral range (594–630 nm), as well as the mean loading weights (%) from the three data sets. The table shows that all samples from both the full spectra of the two lighting sources have 100% classification accuracy. This is true for the classification, validation and test data sets, which were collected over the span of approximately one year, regrowing the strains with each repetition. The optimal MH range (594–630 nm) had nearly identical results, with two of the ST samples being classified as SH. Loading patterns had similar distribution results across the first three PCs, with 9–14% more variance explained by the first PC in the optimal MH set, likely due to removing spectral noise, contrasting the sharp peaks present in the spectra.

PC-based regression analysis

Table 2 lists the results obtained from the PCReg using the calibration data set as the PCA, while regression was carried out on validation and test sets. The regression analysis shows that the *RMSE* is very similar for all three data sets, < 0.014. The R_C^2 and R_{CV}^2 values are similar for the full MH and TH spectra. However, there is a noticeable drop in values for the optimal MH set, of approximately 0.1 in R^2 values. The similarities in the PCReg results comparing both PCA responses from the validation and test sets to the PCA results of the calibration set indicate that these results were repeatable with strains grown at different times. This is in agreement with the visible relationships between classes observed in Figure 7 with Opt MH displaying less class separation than the full 89 band spectrum.

Table 1. Classification accuracy (%) of principal component linear discriminate analysis (PC-LDA) for calibration, validation and test set data repetitions of MH, tungsten halogen and the optimal MH wavelengths (n = 300, for each repetition and light source).

Light source	Calibration (rep 1)	Validation (rep 2)	Test (rep 3)	PC 1 mean	PC 2 mean	PC 3 mean
Metal halide	100.00	100.00	100.00	81.33	14.00	2.67
Optimal metal halide	100.00	99.33	100.00	90.33	8.00	1.00
Tungsten halogen	100.00	100.00	100.00	76.67	15.67	5.33

Results expressed as percent (%);PC = principal component

Table 2. Principal component regression (PCReg) analysis for comparing the calibration data set to the validation and test data sets.

Data	*RMSEC*	*RMSEV*	R_C^2	R_{CV}^2	$\Delta_{R_C^2}$	$\Delta_{R_{CV}^2}$	Δ_{RMSEC}	Δ_{RMSECV}
MH_V	0.0035	0.0036	0.9495	0.9477	0.0013	0.0014	0.0311	0.0323
MH_T	0.0098	0.0100	0.9598	0.9593	—	—	—	—
OMH_V	0.0090	0.0092	0.8424	0.8371	0.0045	0.0044	0.0276	0.0323
OMH_T	0.0135	0.0136	0.8700	0.8691	—	—	—	—
TH_V	0.0025	0.0025	0.9735	0.9730	0.0073	0.0075	0.0137	0.0137
TH_T	0.0098	0.0100	0.9598	0.9593	—	—	—	—

MH = metal halide, TH = tungsten halogen, OMH = optimal metal halide, V = validation, C = calibration, T = test, *RMSEC* = root mean-squared error of calibration, *RMSECV* = root mean-squared error of cross-validation, R_C^2 = regression coefficient of calibration, R_{CV}^2 = regression coefficient of cross-validation

Conclusions

The objective of this methodology is to develop an early and rapid identification protocol for use during food-borne disease outbreaks, food safety screenings or with early detection of blood-borne pathogens that can lead to a targeted response to pathogenic infections in medical patients. Further work in bacteria target extraction is needed; however, in this experiment the ability to identify *Salmonella* serotypes by different light sources and variable reduction is conveyed. This experiment shows the potential for rapid detection through collection of HMIs is possible with either MH or TH light sources in approximately 20 min. We attempted to find an optimal lighting source, or a hybrid of the two, collecting only the most informative spectral bands. Because the classification accuracies of both lighting sources were 100%, and it was found that the TH spectra could not be reduced while maintaining comparable classification accuracies to the full spectrum, a hybrid lighting system combining these two sources was not necessary. The overall spectra generated by MH and TH are vastly different, and small optical profile differences observed in the light scattering patterns of the cells were noticed, but classification with the full spectra did not change. In situations where data processing and storage reduction is not an immediate concern, the TH lamps are typically less expensive per rated hour, and may offer more value in observing spectral differences across the 89-band spectra, opposed to the sharp excitation peaks of MH which may be suited for reduced-band applications. Data fusion may be possible combining the regions. MH lamps show promise in situations where reduction of data processing and storage requirements are necessary due to heavy applications, but model fit should be carefully assessed.

Acknowledgements

The authors would like to thank Dr Nasreen Bano of the Quality and Safety Assessment Research United located at the US National Poultry Research Center in Athens, GA for her contributions to this project.

References

1. Centers for Disease Control (CDC), *CDC Estimates of Foodborne Disease in the United States* (2014). https://www.cdc.gov/salmonella/ [Accessed 28 August 2015]
2. H. Li, H. Wang, J.Y. D'Aoust and J. Maurer, "*Salmonella* species", in *Food Microbiology: Fundamentals and Frontiers 4th Edn,* Ed by M.P. Doyle and R.L. Buchanan. ASM Press, Washington, DC, US.
3. O. Lazcka, F.J. Del Campo and F.X. Munoz, "Pathogen detection: A perspective of traditional methods, and biosensors", *Biosens. Bioelectron.* **22,** 7 (2006).
4. Y. Jia, P.I. Mak, C. Massey, R.P. Martins and L.J. Wang, "Construction of a microfluidic chip, using dried-down reagents, for LATE-PCR amplification and detection of single-stranded DNA", *Lab Chip* **13,** 23 (2013). doi: http://dx.doi.org/10.1039/c3lc51049b
5. M.F. Escoriza, J.M. VanBriesen, S. Stewart, J. Maier and P.J. Treado, "Raman spectroscopy and chemical imaging for quantification of filtered waterborne bacteria", *J. Microbiol. Methods* **66,** 1 (2006).
6. Y.Z. Feng and D.W. Sun, "Determination of total viable count (TVC) in chicken breast fillets by near-infrared hyperspectral imaging and spectroscopic transformations", *Talanta* **105,** 244–249 (2013). doi: http://dx.doi.org/10.1016/j.talanta.2012.11.042
7. D. Wu and D.W. Sun, "Potential of time series-hyperspectral imaging (TS-HSI) for non-invasive determination of microbial spoilage of salmon flesh", *Talanta*, **111** (2013). doi: http://dx.doi.org/10.1016/j.talanta.2013.03.041
8. J. Anderson, C. Reynolds, D. Ringelberg, J. Edwards and K. Foley, "Differentiation of live-viable versus dead bacterial endospores by calibrated hyperspectral reflectance microscopy", *J. Microsc.* **232,** 1 (2008). doi: http://dx.doi.org/10.1111/j.1365-2818.2008.02074.x
9. B. Park, S.C. Yoon, S. Lee, J. Sundaram, W.R. Windham, A. Hinton Jr and K.C. Lawrence, "Acousto-optical tunable filter hyperspectral microscope imaging method for characterizing spectra from foodborne pathogens", *Trans. ASABE* **55,** 5 (2012). doi: http://dx.doi.org/10.13031/2013.42345

10. B. Park, Y. Seo, S.C. Yoon, A. Hinton Jr, W. Windham and K. Lawrence, "Hyperspectral microscope imaging methods to classify gram-positive and gram-negative foodborne pathogenic bacteria", *Trans. ASABE* **58,** 1 (2015).

11. M. Eady, B. Park and S. Choi, "Early and rapid detection of *Salmonella* serotypes with hyperspectral microscope images and multivariate data analysis", *J. Food Prot.* **78,** 4 (2014).

12. B. Park, W. Windham, S. Ladely, P. Gurram, H. Kwon, S. Yoon, K. Lawrence, N. Narang and W. Cray, "Classification of non-0157:H7 shiga toxin producing *Esherichia coli* (STEC) serotypes with hyperspectral microscope imaging", *Trans. ASABE* **57,** 3 (2014).

13. M.W. Davidson, *Fundamentals of Metal Halide Arc Lamps*. Zeiss (2015). http://zeiss-campus.magnet.fsu.edu/articles/lightsources/metalhalide.html [Accessed 2 October 2015]

14. A. Giannuzzi and P. Sansoni, "Halogen and incandescence lamps", in *Sustainable Indoor Lighting*, Ed by P. Sansoni, L. Mercatelli and A. Farini. Springer, London, UK, p. 87 (2015). doi: http://dx.doi.org/10.1007/978-1-4471-6633-7_6

15. M.W. Davidson, *Tungsten Halogen Incandescent Lamps*. Zeiss (2015). http://zeiss-campus.magnet.fsu.edu/articles/lightsources/tungstenhalogen.html [Accessed 2 October 2015]

16. R. Bro and A.K. Smilde, "Principal component analysis", *Anal. Methods* **6,** 9 (2014). doi: http://dx.doi.org/10.1039/c3ay41907j

17. K.H. Esbensen, *Multivariate Data Analysis – In Practice, 4th Edn.* CAMO, Oslo, Norway (2000).

18. P. Gemperline, *Practical Guide to Chemometrics, 2nd Edn.* CRC Press, Boca Raton, US (2006). doi: http://dx.doi.org/10.1201/9781420018301

19. S. Wold, K. Esbensen and P. Geladi, "Principal component analysis", *Chemometr. Intell. Lab. Syst.* **2,** 1 (1987). doi: http://dx.doi.org/10.1016/0169-7439(87)80084-9

20. M. Barker and W. Rayens, "Partial least squares for discrimination", *J. Chemometr.* **17,** 3 (2002).

21. R. Fisher, "The use of multiple measurements in taxonomical problems", *Ann. Eugenics* **7,** 2 (1936). doi: http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

22. F. Wei, Y. Wan, O.G. Li, K.W. Cheach and Z. Liu, "Organic light-emitting-diode-based plasmonic dark-field microscopy", *Opt. Lett.* **37,** 21 (2012). doi: http://dx.doi.org/10.1364/OL.37.004359

23. A. Fâbrega and J. Vila, "*Salmonella enterica* serotype Typhimurium skills to succeed in the host: virulence and regulation", *Clin. Microbiol. Rev.* **26,** 2 (2013). doi: http://dx.doi.org/10.1128/CMR.00066-12

24. M.W. Davidson and M. Abramowitz, "Optical microscopy", in *Encyclopaedia of Imaging Science and Technology*, Ed by J.P. Hornak. Wiley, Hoboken, NJ, US, p. 1 (2002). doi: http://dx.doi.org/10.1002/0471443395.img074

25. G.S. Verebes, M. Melchiorre, A. Garcia-Leis, C. Ferreri, C. Marzetti and A. Torreggiani, "Hyperspectral enhanced dark field microscopy for imaging blood cells", *J. Biophotonics* **6,** 11 (2013). doi: http://dx.doi.org/10.1002/jbio.201300067