Volume 1 | Number 1 | Jan 2013 | Pages 1–100

Analyst

www.rsc.org/analyst

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

1
2
3
4
5
6
7
8
9
10
...

A method for the comparison of multi-platform spectral histopathology (SHP) data sets

A. Ergin[1], F. Groβerüschkamp[2], Oliver Theisen[2], K.Gerwert[2], S.Remiszewski[1], C.M. Thompson[1], M.Diem[1,3*]

[1]*Cireca Theranostics, LLC, Cambridge, MA, USA*

[2]*PURE Institute, Department of Biophysics, Ruhr Universität Bochum, Bochum, Germany,*

[3]*Dept of Chem & Chemical Biology, Northeastern University, Boston, MA, USA*

Abstract

Results of a study comparing infrared imaging data sets collected on different instruments or instrument platforms are reported, along with detailed methods developed to permit such comparisons. It was found that different instrument platforms, although employing different detector technologies and pixel sizes, produce highly similar and reproducible spectral results. However, differences in the absolute intensity values of the reflectance data sets were observed that were caused by heterogeneity of the sample substrate in terms of reflectivity and planarity.

1

Introduction

During the past fifteen years, Spectral Histopathology (SHP) has been established as a diagnostic and prognostic method for several cancer types by a number of research groups [1-8]. In SHP, tissue samples from biopsies are analyzed, using infrared micro-spectroscopy, to produce infrared pseudo-color images of the tissue sample that reproduce the anatomical and pathological features of tissue with high fidelity. Since infrared spectroscopy monitors chemical composition and compositional variations, rather than cell morphology, one may conclude that chemical compositional changes accompany the transition from normal to cancerous states in tissue and variations in tissue types, for example, between connective and glandular tissues. The next step in the development of an enhanced diagnostic and prognostic tool is to train mathematical algorithms to recognize spectral patterns that can be associated with tissue types or disease states, and recognize these patterns in unknown samples to identify these tissue types or disease states. This goal has been independently achieved by a number of research groups in academe and industry.

Before a transition of SHP from the laboratory to clinical applications can take place, one needs to establish the independence of the diagnoses and prognoses reached by SHP from particular laboratory practices, histopathological correlations and instrumental platforms used. This step can be broken down into two separate categories: laboratory and instrument independence on one hand, and independence of the classical pathology used to train the SHP algorithm on the other. The former of these is referred to as 'analytical validity', since it depends mostly on the analytical chemical and spectroscopic steps performed. The second point could be referred to as 'pathological validity'. In the past these two validation steps have been addressed for Raman imaging developments: Stone *et al*. [9] and Puppels *et al.* [10] devoted significant efforts to establish reproducibility and instrument independent for Raman datasets for tissues and bacterial strains, respectively.

The pathological validity has been addressed in detail by Stone and coworkers [9] for spectral diagnoses of esophageal disease. To this end, panel of pathologists was assembled who were experts in the field of esophageal pathology. These efforts indicated very poor agreement between the members of the expert panel, resulting in a large number of tissue samples to be unusable for training the mathematical algorithms. The disagreement between pathological diagnoses is one of the major driving factors for the development of objective, machine-based spectral methods for diagnosis and prognosis.

The study reported here addresses the issue of analytical validity by establishing the methodology for comparing spectral results obtained for the same tissue spots in two different laboratories, and utilizing instruments of very different designs (see 'Methods Section' below). The results of this study demonstrates for the first time that datasets collected from different instrument types can be interconverted such that the spectral information is independent of instrument type.

Methods

The analytical validity study reported in this paper is aimed at establishing the methodology that allows a comparison of results from different laboratories, using different infrared imaging micro-spectrometers. To this end, an easily reproducible sample system was used, namely commercially available tis-

sue microarray [11] of lung cancer specimens. These samples, cut to a thickness of 5 μm, were mounted on sample substrates suitable for reflectance micro-spectroscopy. These substrates are known as 'low-e' or MirrIR slides, and are available from Kevley Technologies (Chesterland, OH). These slides are dichroic mirrors in that they are completely reflective in the infrared, but nearly completely transparent under visible light. Thus, the sample can be used for infrared SHP as well as for classical histopathology after suitable (H&E) staining and coverslipping. Problems known to exist [12] with the use of these slides have been addressed in the past [13, 14].

Five different infrared imaging micro-spectrometers were used in this study, three Agilent Model 620/670 (henceforth referred to as Ag-1P and Ag-2P) at the PURE Institut at the Ruhr Universität, Bochum (Germany), an Agilent Model 620/670 instrument at Cireca (Ag-C) and a Perkin Elmer Spotlight 400 / Spectrum One spectrometer (PE-C) at Cireca. Additionally, some datasets collected on an identical Perkin Elmer instrument at Northeastern University, referred to as PE-N, were used.

The Agilent instruments incorporate 128 x 128 element HgCdTe focal plane detectors (FPAs) whereas the Perkin Elmer instruments use an 8 x 2 element HgCdTe detector. Due to different detector technology between the two instrument types, different numbers of background and sample scans were acquired, with the emphasis on obtaining comparable signal-to-noise ratios for both instrument types. For the Perkin Elmer instrument, 32 interferograms were co-added for the background spectra, and 2 interferograms were used for the sample spectra. For the Agilent instruments, 128 background and 64 sample scans were co-added. All instruments were used in reflection mode that produces transflection spectra when used in conjunction with 'low-e' slides. Using the standard 15x Cassegrain infrared microscope objectives supplied with the instruments, the pixel size for all Agilent instruments is 5.5 μm on edge, whereas it is 6.25 μm on edge for the Perkin Elmer instruments. For both instruments, the field of depth, given by the axial diffraction limit, can be estimated to be larger than 30 μm at 10 μm wavelength, and correspondingly smaller at shorter wavelengths. The thickness of the sample generally is about 5 μm; given the field of depth, one should think that focusing issues would be minor. This is not the case, as will be discussed later.

Nominal spectral resolution used was 3.96 cm^{-1} for both the Agilent and Perkin Elmer instruments. Both data acquisition parameters were set to one level of zero-filling, such that the data point spacing was 1.98 cm^{-1}. However, the Perkin Elmer instrument reports the data interpolated to 2.0 cm^{-1} spacing. The spectral resolution is not a particularly critical issue, since the full width at half height (FWHH) of tissue bands is typically in excess of 60 or 70 cm^{-1}, and even the second derivative spectra used in the data comparison were much broader than the spectral resolution. Water vapor reference spectra, used in a multivariate approach (see below) to compensate for water vapor contributions, were collected at 4 cm^{-1} resolution as well.

A tissue spot *ca*. 1.8 mm in diameter typically was imaged in a 2.1 mm x 2.1 mm square, yielding a total of 112,896 pixel spectra for the Perkin Elmer instrument, and 145,785 pixel spectra for the Agilent instruments. Pixel spectra outside the tissue spot, but within the square sampling areas, were used as "instrument noise" spectra for the noise-adjusted principal component noise reduction (see 'Data Pre-Processing') below.
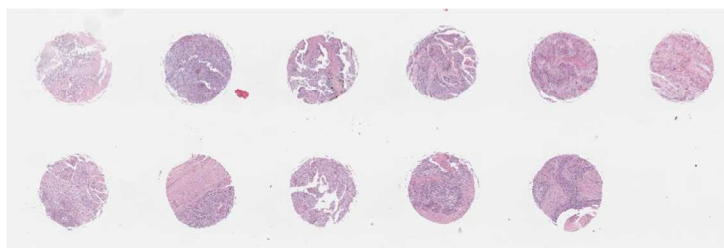
Figure 1. Image of the TMA used for this study. Note that this study was carried out on an unstained TMA

The samples utilized here were from a commercial lung cancer tissue micro-array (US Biomax, Rockville, MD, VLC11), containing eleven cores, see Figure 1. Several sections were purchased which were cut to a thickness of 5 μm and mounted on low-e slides at the manufacturer. Adjacent sample sections are 5 μm apart, whereas non-adjacent sections are 10 μm apart. Slides were de-paraffinized using standard procedures [15], and used unstained for IR data acquisition. Since the tissues had been formalin-fixed, no further action to preserve the tissue was deemed necessary after de-paraffination. In a previous study, we demonstrated that samples of formalin-fixed cells were stable for weeks when kept in a dry, clean environment [16, 17]. In Table I, the terminology "same tissue spot" implies that the exact same spot was used for data acquisition on different instruments. Sample slides were hand-carried between the laboratories. Note that some of the spots were used for multiple tests.

Data Pre- Processing

The data pre-processing procedures reported previously in the literature [14, 18] were used, except that data segmentation (by cluster analysis) was omitted, and binary comparisons of spectral datasets was carried out (see Section 'Data Analysis'). Raw datasets in instrument-specific formats were imported into MATLAB-based routines that performed the data processing, as follows:

a) Spatial averaging and conversion from transmittance to absorbance units
Conversion from transmittance to absorbance units was required for Perkin Elmer datasets only, since their file format stores spectra in the mapping file as raw transmission spectral vectors. In order to reduce processing times, 2 x 2 pixel spectra were averaged into one new average pixel spectrum, thereby reducing the dataset size by a factor of four. This step was performed to comply with the standard data analysis protocol at Cireca that has been justified elsewhere [2]. Since the original datasets were spatially oversampled, given the diffraction-limited spatial resolution of *ca*. 10 μm, this step creates a pixel size approximately equal to the diffraction limit. Since the two instrumental platforms use different pixel sizes and data point spacing in the spectral vectors, data need to be converted to an instrument platform-independent format for comparison. This step will be discussed in the 'Data Analysis' section.

b) Normalization

Each pixel spectrum was normalized using "vector normalization" in which the summed squared intensity of a spectrum is set to unity. This is accomplished by dividing every spectral intensity point by a normalization factor which is obtained by summing all squared intensity values of the spectrum. Often, a spectrum is mean centered before normalization.

c) Noise reduction by Noise Adjusted Principal Component Analysis (NA-PCA)

An elegant method of noise reduction in hyperspectral datasets was originally developed by the remote sensing community [19] and applied to infrared datasets by Bhargava and coworkers [20]. This method utilized pixel spectra, measured from areas not occupied by tissue, to define the instrumental noise profile during data acquisition in form of pixel spectra that consist of noise only. From these, a noise dataset was created. Similarly, a signal dataset was created by selecting pixel spectra that are due to areas occupied by sample. From these datasets, noise and signal covariance matrices were constructed. The noise covariance matrix was subsequently used to create noise-adjusted principal components from which spectra with reduced noise were created. This procedure increased the signal-to-noise (S/N) ratio of an entire dataset by an order of magnitude to about 4000:1.

d) Reduction of water vapor contributions in pixel spectra

Although the spectrometers at both laboratories were purged with dry air to reduce the water vapor vib-rotational spectral contributions, the long data acquisition times for infrared microscopic datasets often cause slight changes in the instrument's water vapor content. When ratioed against the same backgrounds spectrum, positive and negative water vapor spectral contributions may be observed. Even when these water features are not detectable by the naked eye, one frequently finds that higher loading vectors of a PC analysis are dominated by water vapor contributions. These contributions were further reduced by a method introduced by Bruun, *et al.* [21]. This method is based on collecting water vapor spectra under a variety of different conditions, including temperature and concentration variations. These changes are reduced by PCA into two principal components, the water vapor mean spectrum and a first variance spectrum that contains the most prominent changes to the mean spectra. Subsequently, both these spectral vectors are used to model the water vapor contributions in a spectral dataset and subsequently subtract it. This method does not scale the water vapor spectrum at one or a few selected wavelengths, but uses a multivariate approach to fit the best water background contribution. This procedure reduces residual water vapor signals by over an order of magnitude.

e) Conversion of spectra to second derivatives

Second derivative spectra are devoid of any sloping background features, and exhibit reduced band width; thus, it is advantageous to carry out data analysis on second derivative spectra. It should be emphasized that the second derivative spectra introduce an intensity distortion in the sense that the peaks are no longer proportional to the original intensity only, but are depending on the curvature of the original peaks as well. That is, a sharp and a broad peak of the same absorbance will have different intensities in second derivative space.

Derivatives are computed analytically using the Savitzky-Golay algorithm and sets of different smoothing derivative coefficients. This is accomplished by applying a "sliding smoothing window" to the data vector. The smoothing derivative coefficients C_i for the sliding window algorithm were first derived by Savitzky and Golay, and are listed in [22]. These coefficients are applied, successively, to all data points in the original spectra, Y_j, and the results of the multiplication/additions shown below replace the current center data point in each window, as shown in Eq. 1:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+1}}{N} \qquad (1)$$

Here, Y_j^* represents the j'th data point of the smoothed array, N is a normalization constant listed with the smoothing coefficients, and the summation is from $-m$ to $+m$, where m is the halfwidth of the smoothing window, minus ½.

f) Correction for resonance Mie scattering

One of the major confounding effects that in the past plagued infrared micro-spectroscopy has been termed resonance Mie (R-Mie) scattering [23] that mixes absorptive and reflective line shapes by the Mie scattering mechanism. The spectra discussed in this paper were corrected for these distortions using an approach [24, 25] based on the well-documented "phase correction" method that is widespread in standard FTIR spectroscopy. This method has been described in detail in the literature [25]. The phase correction approach is based on the concept that the complex Fourier transform decomposes a spectrum into real and imaginary parts with dispersive and absorptive line shapes, respectively. By varying the phase angle between the two components, complete separation of them can be achieved.

Phase correction was achieved by transforming each pixel spectrum obtained from the previous steps back into interferogram space by the finite Hilbert transform [26]. The resulting real (Re) and imaginary (Im) interferograms are phase shifted by a trial phase θ by a similarity transform according to

$$\begin{bmatrix} \text{Re'} \\ \text{Im'} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \text{Re} \\ \text{Im} \end{bmatrix} \qquad (2)$$

Phase corrected spectra are computed by complex forward FFT of (Re' + i Im'). It is advantageous to carry out the phase correction on second derivative data since they exhibit flat baselines with minimal offset. The finite Hilbert transform on data with a sloping background produces distorted real and imaginary spectra, first noted by us in 2005 [27]. In the absence of sloping baselines, the order of the two procedures (derivatization and phase correction) is immaterial, and comparable results are obtained regardless of the order. However, the end point of the iterative phase correction is detected better on the second derivative spectra [27].

g) Spectral quality tests and spectral expansion

Throughout the pre-processing, spectra are tested for spectral quality (overall signal intensity, overall band shape, noise) and spectra falling below pre-established limits are eliminated from the datasets. Pre-processed data for each tissue spot are stored in a MATLAB database, containing between 20,000

and 30,000 spectra, after tissue voids are accounted for and low signal spectra (at the edges of tissue) are eliminated.

Spectra were expanded between 800 and 1800 cm^{-1} since this wavenumber range is used in the authors' laboratories for classification purposes. However, for the comparison of spectra collected on different instrument platforms, the entire spectral range (800 – 1800 cm^{-1} and 2800 – 3500 cm^{-1} could have been used.

Data Analysis

In the context of this paper, data processing implies the binary comparison of two images that may be as similar as two consecutive scans of the same tissue spots on the same instrument, or as different as two scans of non-adjacent tissue sections on two different instruments in two different laboratories. A list of the binary comparison tests can be found in Table I.

Table I   Different Instrument and Sample Comparisons Carried Out In Study

	Sample	Instrument	Comment
Test 1	Same Tissue Spot	Same Instrument (Ag-P1/Ag-P1) PE-C/PE-C)	Consecutive acquisitions
Test 2	Same Tissue Spot	Different Instruments (Ag-P1/Ag-P2) (PE-C/PE-N)	Same instrument type
Test 3	Same Tissue Spot	Different Instruments (Ag-P1/Ag-C)	Different laboratories
Test 4	Same Tissue Spot	Different Instrum. types (Ag-C/PE-C)	Platform independence
Test 5	Non-consecutive Tissue Sections	Different Instrum. types (Ag-P1/PE-C)	Platform independence Non-identical sample

If datasets were collected on different manufacturers' instruments (see Test 4 and 5 in Table I), data need to be rendered compatible in terms of (wavenumber) data point spacing as well as pixel size. For the Agilent instruments, with a 15x objective, the pixel size is 5.5 μm on edge, whereas it is 6.25 μm for the PE instruments. Furthermore, the spacing between intensity data points in each pixel spectrum is different for the two instrument platforms. In the Agilent instruments, this spacing is about 1.93 cm^{-1}, whereas in the Perkin Elmer instruments, the data point spacing is reported in 2 cm^{-1} increments (see "Methods" section).

Thus, for a pixel-by-pixel comparison of the spectra, the pixel size and data point spacing must be converted to the same parameters. This was accomplished using the three-dimensional interpolation function 'interp3' in MATLAB, converting all datasets to 2 cm^{-1} data point spacing and a pixel size of 12.5 μm; this value corresponds to the pixel size of the Perkin Elmer datasets after 2 x 2 picture averaging. Furthermore, all spectra were truncated to the spectral range of 900-1800 cm^{-1} (451 data points).

The binary comparison of two datasets **A** and **B** was carried out as follows. First, the two images were registered with respect to each other. This step was necessary since non-consecutive tissue sections (Test 5) generally are not at exactly the same orientation on the slides. Furthermore, the placement of the slides in different instruments (Test 2 – 4) produced images that were linearly offset.

Registration was accomplished as follows. After data pre-processing described above, gray scale images were constructed from each of the infrared datasets by integrating each pixel spectrum in the dataset and converting the resulting intensity to a gray scale value. Thus, this step reduces a hyperspectral dataset to a two-dimensional matrix of gray scale values, referred to as **A** and **B** for any two datasets to be compared. Such gray scale images are displayed in Figures 2 and 4. These grayscale infrared images are coarsely registered by manually selecting three or more clearly visible features on the images, such as cracks or voids. Fine registration subsequently was achieved by optimizing the parameters of a rigid body (image) transform **T** that minimizes the least square error of the grayscale difference between the transformed image **A** and image **B** according to

$$\min(\sum_{ij} T[(A_{ij} - B_{ij})^2] \tag{3}$$

where **T** is the rigid body transform with parameters [x_offset, y_offset, scale, rotation angle]. This transform is performed on the gray scale images using bi-cubic interpolation without anti-aliasing to map the transformed points A onto the pixel space B. The optimization is constrained by setting upper and lower bounds on the parameters to restrict registration near the initial manually selected points. The initial parameters for the rigid body transform are usually selected as [0 0 1 0].

At this point, the two images are registered such that any two points on the images exactly correspond to each other, and a meaningful comparison of the similarity of the datasets is possible. Several methods to assess the similarity were tested, for example, by plotting the absolute value of the pixel-by-pixel difference, $|B_i - A_i|$, between the two datasets to be compared. As before, **A** and **B** represent integrated intensity values for each pixel. A metric for the similarity would be a normalized sum (NS) over all pixels:

$$\overline{NS} = \frac{\sum_i^N |B_i - A_i|}{N} \tag{4}$$

In Eq. 4, B_i and A_i represent the $i'th$ spectrum of datasets **A** and **B**, and N the number of pixels in the dataset. This method has the disadvantage that each pixel spectrum was summed into one value; thus, spectral differences could be lost due to this integration.

Therefore, a better way of comparison of two datasets utilizes the interpolated and transformed hyperspectral datasets, and not on the gray scale images **A** and **B**. Here, the similarity was assessed by computing the pixel-by-pixel correlation coefficient C^{AB} for each (corresponding) spectral vector S^A and S^B of the two datasets:

$$C^{AB} = \frac{\sum_j^M (S_j^A - \overline{S^A})(S_j^B - \overline{S^B})}{\sqrt{\sum_j^M (S_j^A - \overline{S^A})^2} \sqrt{\sum_j^M (S_j^B - \overline{S^B})^2}} \tag{5}$$

The summation in Eq. 5 is over the length M of each spectral vector (*i.e.,* over all M intensity points), and $\overline{S^A}$ and $\overline{S^B}$ denote the mean intensity values of one spectral vector. Again, the normalized mean of all correlation coefficients for one dataset is reported:

$$\overline{C^{AB}} \;=\; \frac{\sum_i^N C_i^{AB}}{N} \tag{6}$$

The values of the pixel-by-pixel correlation coefficients were converted to gray scale values and are shown in the figures as correlation images. Thus, the main difference between the similarity metrics described by Eqs. 4 and 6 is that in the former, all intensity values in a spectral vector were averaged into one value, and the difference between the two datasets was calculated for each pixel, and averaged over the entire spot. In Eq. 6, on the other hand, spectral vectors were correlated, rather than single points, and the pairwise correlation coefficients were subsequently averaged over the entire tissue spot.

Results and Discussion

First results indicated that both instrumental platforms produced highly reproducible spectral images. This is depicted in Figure 2a and summarized in Table II. Figure 2a shows two datasets collected consecutively on the same Agilent instrument (Ag-P1) without removing the sample from the instrument, and using the same background spectrum. The nearly completely white correlation image indicates high reproducibility between these consecutive acquisitions which produced a mean correlation coefficient $\overline{C^{AB}}$ of 0.992. A similar experiment carried out using the Perkin Elmer instrument (PE-C) produced a mean correlation coefficient of 0.996.
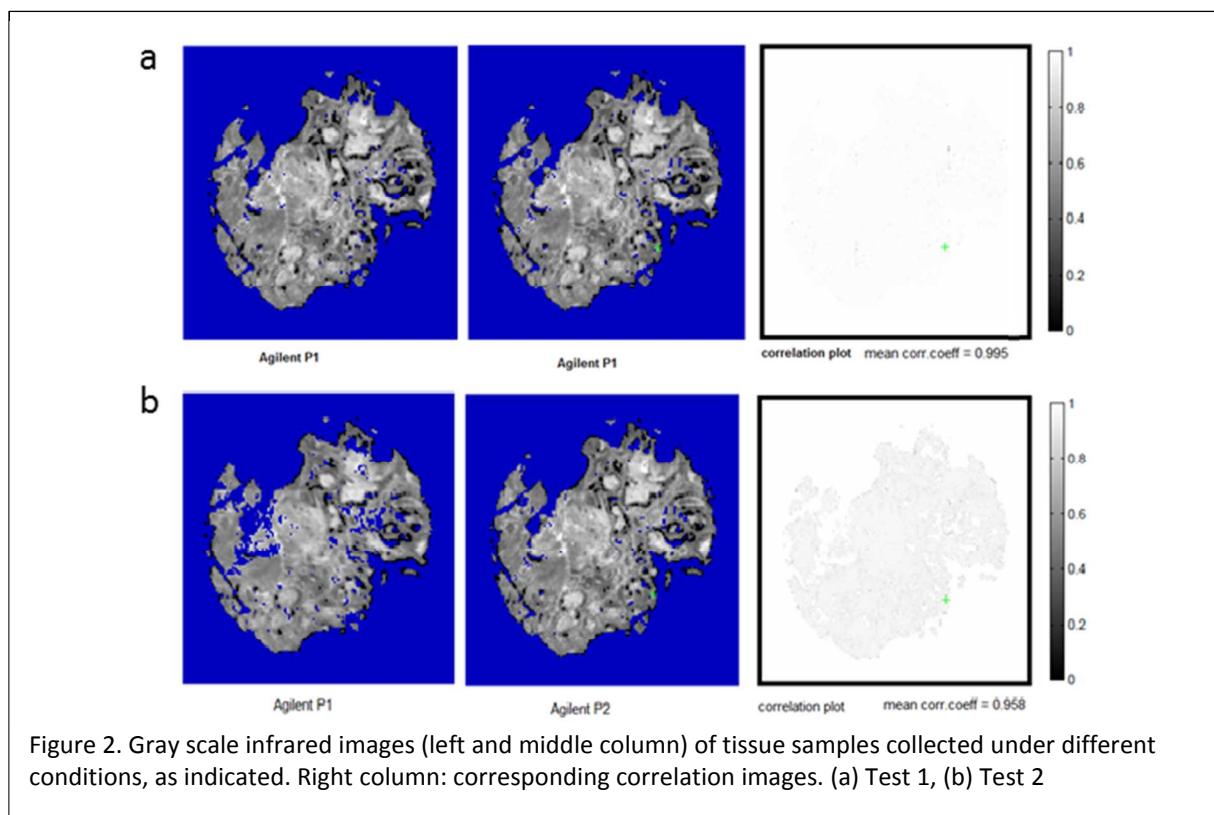

Figure 2. Gray scale infrared images (left and middle column) of tissue samples collected under different conditions, as indicated. Right column: corresponding correlation images. (a) Test 1, (b) Test 2

These results indicate very high reproducibility between consecutive data acquisitions when the same location on the low-e slide is used for background acquisition. It appears that there is no detectable systematic error between acquisitions, and that the perturbation of the observed spectral datasets by water vapor fluctuations is negligible.
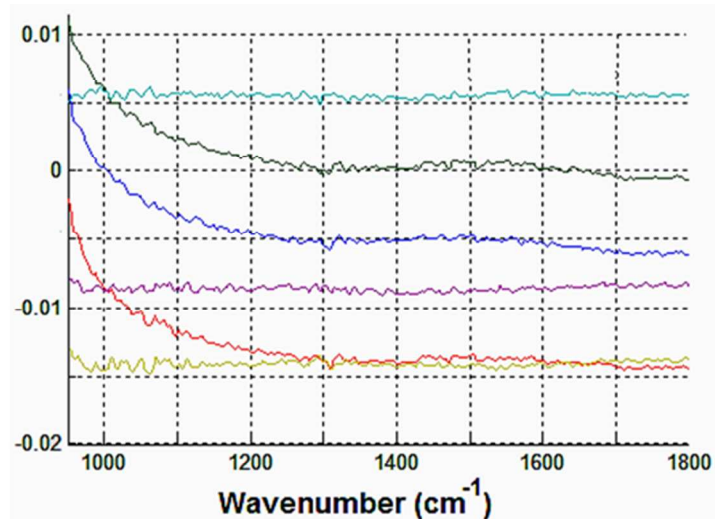
Figure 3. Offset observed for 'null spectra' obtained from four different corners of a low-e slide. The observed differences could be attributed to the tilt of the slide with respect to the focal plane.

In comparison, the reproducibility of measurements between two instruments of the same type (for example, Ag-P1 and Ag-P2) seemed to be much lower (see Table II and Figure 2b). The correlation image in Figure 2b shows a uniform, gray rendition of the tissue section, indicating that the pairwise comparison between corresponding pixel spectra produced nearly constant differences. This was observed for measurements carried out on two Agilent as well as on two Perkin Elmer instruments, as shown in Table II. It is unlikely that this difference was due to instrumental effects, such as detector non-linearity and radiometric accuracy which both have been addressed in the literature and may be assumed to produce smaller errors than the differences observed here [28, 29]. The reduced reproducibility instead was found to be mostly due to variations in transflected intensities caused by changes in the focusing of the infrared optics on the slide.

This was demonstrated by successively removing and replacing the sample slide in the instrument while monitoring the slide position with a depth gauge that was accurate to better than 5 μm. When the alignment of the slide in the focal plane was within 5 μm, excellent reproducibility was observed (line 3 in Table II). If the slide was not properly aligned in the focal plane(s), the mean correlation coefficient $\overline{C^{AB}}$ dropped substantially (below 0.96 in some cases). This was observed for Agilent as well as Perkin Elmer instruments.

Table II Mean correlation coefficients for spectral dataset comparisons

Instruments		$\overline{C^{AB}}$	Comments
Test 1	Ag-P1 / Ag-P1	0.992	same instrument, same sample, same background location
	PE-C / PE-C	0.996	same instrument, same sample, same background location
	Ag-C / Ag-C	0.992	same instrument, same sample, after sample re-alignment
Test 2	Ag-P1 / Ag-P2	0.958	same instrument type, same sample, different background location
	PE-C / PE-N	0.974	same instrument type, same sample, different background location
Test 5	Ag-P1 / PE-C	0.950	different instruments, non-consecutive sample section, different background location

The effect of poor focusing was demonstrated as follows. The slide was tilted slightly out of the focal plane and held in place in the slide holder by a spring-loaded slider.[1] Four different reflectivity tiles were collected at the four corners of the slide. The four mean reflectivity spectra of each of the tiles were ratioed against each other to yield 6 "null-spectra", shown in Figure 3. These null spectra ideally should be straight lines at zero absorbance (as the name implies), albeit with some noise. The observed null spectra exhibited some curvature due to the detector cutoff, but showed an offset between 0.005 and 0.015 optical density (OD) units, see Figure 3.

Thus, a slightly tilted slide can produce changes in reflectivity that translate into different intensities of the collected images. This point was confirmed by one of the referees of this paper who suggest that re-positioning of the sample contributed some of the spectral variations shown in Table II.
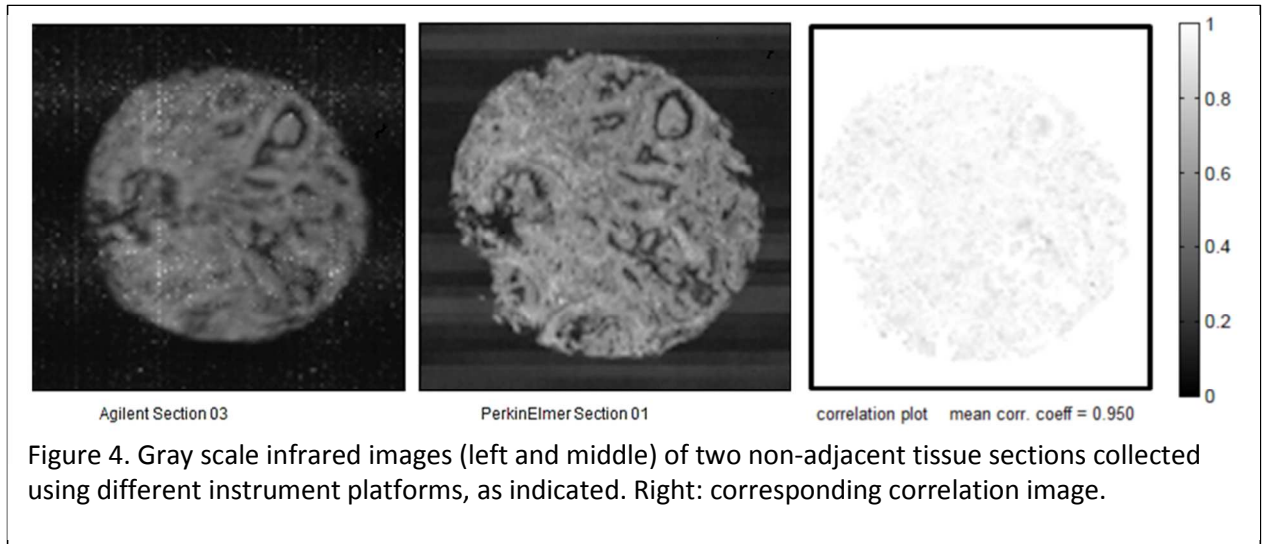
Another possible cause of the reduced correlation coefficients listed in Table II could be due to slight inhomogeneity of the reflectivity of the slide. If the background is collected at locations of different reflectivity, one would obtain spectra of slightly different magnitudes. To test the homogeneity of the slides' reflectivity, ten different reflectivity tiles, each 128 x 128 pixels in size, were acquired from several low-e slides that were well aligned in the focal plane of the Agilent instrument. Mean spectra for each tile were ratioed against each other, and the null-spectra were calculated. It was found that the differences in surface reflectivity accounted for less than 0.006 absorbance units in these "null-spectra".

These tests also confirmed that the majority of the slides exhibited good planarity, but that some were quite seriously warped. This point was correctly raised by one referee apparently familiar with the quality of low-e slides.

The maximum offset between the null spectra shown in Figure 3 is 0.02 OD units which translates into a difference in reflected intensities for the most different background spots of 5 %. This observation explains why repeated measurements of the same tissue spot, for example using different instruments and different focal point position, resulted in a reduced mean correlation coefficient, shown in Table II. Thus, this reduction in instrument "reproducibility" was, most likely, not caused by the type and optical design of the instruments, but strictly by the choice of the background spot and the alignment of the slide.

---

[1] Since the edges of the low-e slides are chipped, rather than polished, a spring-loaded slide holding mechanism that acts on the edge of a slide can easily misalign the slide and cause it to tilt out-of-plane.

This hypothesis is further substantiated by inspection if Figure 4, which shows the results, obtained for two non-consecutive tissue sections, imaged on two different instrument platforms. This test represents the most different conditions listed in Table I. The two tissue sections were cut (at 5 μm thickness) from the same tissue block, but were sections 1 and 3, with section 2, also 5 μm thick, used for H&E staining. Thus, the two sections reported here were 5 μm apart in the tissue block, which corresponds to less than a cell's thickness. Thus, to a first approximation, one may conclude that the two sections sample the same biochemical composition. Due to different pixel resolution and, consequently, different number of pixel spectra in each dataset, the raw infrared images are differently sized. This effect is compensated for by the subsequent registration and interpolation processes such that the images can be compared, and a correlation image can be computed, as shown in Figure 4 (right panel). As expected from the above arguments, the two infrared gray scale images show the same tissue features, and the correlation image shows a reasonably uniform offset that is most likely caused by the different slide alignment, rather than any instrumental differences. Thus argument is substantiated by the fact that the mean correlation coefficient for this test, 0.950, is very similar to that of test 2 between different Agilent instruments, which was found to be 0.958. Thus, we conclude that the reproducibility between different

Figure 4. Gray scale infrared images (left and middle) of two non-adjacent tissue sections collected using different instrument platforms, as indicated. Right: corresponding correlation image.

instrument platforms is quite high, and that observed differences between subsequent data acquisition are caused by differences in background spot position, reflectivity of the slide or focusing of the infrared beam.

Conclusions

We report methods to quantitatively compare hyperspectral images collected on different instrument platforms, and present preliminary results for images collected under different conditions. These efforts are necessary to demonstrate that tissue classifications obtained from SHP are portable between instrument platforms and between different laboratories.

In the process of establishing the platform independence of SHP reflectance infrared micro-spectral images, it was found that the homogeneity of the sample substrate surface, its thickness and the focus of

the infrared microscope can affect the observed datasets more than the actual instrument platform once the spectral data are interpolated to identical conditions.

References

1.   Bird, B., et al., *Infrared Spectral Histopathology (SHP): A Novel Diagnostic Tool for the Accurate Classification of Lung Cancers.* Laboratory Investigations, 2012. **92** p. 1358-1373.

2.   Akalin, A., et al., *Classification of Malignant and Benign Tumors of the Lung by Infrared Spectral Histopathology (SHP).* Laboratory Investigations, 2015. **in press**.

3.   Bhargava, R., Fernandez, D. C., Schaeberle, M. D., Levin, I. W. *FTIR Imaging of Biological Tissue for Histopathological Analysis*. in *PittCon*. 2002. New Orleans.

4.   Pounder, F.N. and R. Bhargava, *Toward Automated breast Histopathology Using Mid-IR Spectroscopic Imaging*, in *Vibrational Spectroscopic Imaging for Biomedical Appications*, G. Srinivasa, Editor. 2010, McGraw Hill: New York.

5.   Bassan, P., et al., *Automated high-throughput assessment of prostate biopsy tissue using infrared spectroscopic chemical imaging.* Proc. of SPIE 2014. **9041**.

6.   Lasch, P. and D. Naumann, *FT-IR Microspectroscopic Imaging of Human Carcinoma in Thin Sections Based on Pattern Recognition Techniques.* Cellular and Molecular Biology, 1998. **44**(1): p. 189-202.

7.   Lasch, P., et al., *Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis.* Biochim Biophys Acta, 2004. **1688**(2): p. 176-86.

8.   Kallenbach-Thieltges, A., et al., *Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue se.* J.Biophotonics, 2013. **6**(1): p. 88-100.

9.   Stone, N., C. Kendall, and H. Barr, *Raman Spectroscopy as a Potential Tool for Early Diagnosis of Malignancies in Esophageal and Bladder Tissues*. Vibrational Spectroscopy for Medical Diagnosis, ed. M. Diem, P.R. Griffiths, and J.M. Chalmers. 2008, Chichester, UK: John Wiley & Sons.

10.   Puppels, G.J., et al., *Medical Applications of Raman Spectroscopy: In Vivo Raman Spectroscopy*, in *Handbook of Raman Spectroscopy*. 2001, Marcel Dekker, Inc.: New York, NY

11.   Fernandez, D.C., et al., *Infrared Spectroscopic Imaging for Histopatholog Recognition.* Nature Biotech., 2005. **23**: p. 469-474.

12.   Bassan, P., et al., *The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications.* Analyst, 2013. **138**: p. 144-157.

13.   Wrobel, T.P., et al., *Electric field standing wave effects in FT-IR transflection spectra of biological sections: Simulated models of experimental variability.* Vibr.. Spectrosc., 2013. **69**: p. 84-92.

14.   Miljković, M., et al., *Spectral Cytopathology: new aspects of data collection, manipulation and confounding effects.* Analyst, 2013. **138**: p. 3975-3982.

15.   foliobio.com. *Deparaffin Protocol*. Available from: http://foliobio.com/media/pdf/Deparaffin%20protocol.pdf.

16.   Mazur, A.I., et al., *Evaluating Different Fixation Protocols for Spectral Cytopathology, Part 1.* Anal. Chem., 2012. **84**: p. 1259-1266.

17.   Mazur, A.I., et al., *Evaluating Different Fixation Protocols for Spectral Cytopathology, Part 2.* Anal.Chem., 2012. **84**: p. 8265-8271.

18.   Diem, M., et al., *Molecular pathology via Infrared and Raman spectral imaging.* J.Biophotonics, 2013. **6**(11-12): p. 855-886.

19.   Roger, R.E., *A Faster Way to Compute Noise-Adjusted Principal Components Transform Matrix.* IEEE Transactions on Geoscience and Remote Sensing, 1994. **32**(6): p. 1194-1196.

20.   Reddy, R.K. and R. Bhargava, *Accurate histopathology from low signal-to-noise ratio spectroscopic imaging data.* Analyst, 2010. **135**: p. 2818-2815.

21.   Bruun, S.W., et al., *Correcting Attenuated Total Reflection-Fourier Transform Infrared Spectra for Water Vapor and Carbon Dioxide.* Appl.Spectrosc., 2006. **60**(9): p. 1029-1039.

22.   Savitzky, A., Golay, M.J.E. , *Smoothing and differentiation of data by simplified least-squares procedures.* Anal. Chem, 1964. **36**(8): p. 1627-1639.

23.   Bassan, P., et al., *Resonant Mie scattering in infrared spectroscopy of biological materials – understanding the 'dispersion artefact'.* Analyst, 2009. **134**: p. 1586–1593.

24.   Diem, M., B. Bird, and M. Miljković, *Phase Correction to Compensate for Reflective Distortions of Optical Spectra*, U.P. Office, Editor. 2011: USA.

25.   Miljković, M., B. Bird, and M. Diem, *Dispersive line shape effects in infrared spectroscopy.* Analyst, 2012. **137**: p. 3954-3964.

26.   Bertie, J.E. and S. Zhang, *Infrared intensities of liquids. IX. The Kramers-Kronig transform, and its approximation by the finite Hilbert transform via fast Fourier transforms.* Can.J.Chem., 1992. **70**: p. 520-531.

27.   Romeo, M.J. and M. Diem, *Correction of dispersive line shape artifact observed in diffuse reflection infrared spectroscopy and absorption/reflection (transflection) infrared micro-spectroscopy.* Vibrational Spectroscopy, 2005. **38**(1-2): p. 129-132.

28.   Kaplan, S.G., L.M. Hanssen, and R.U. Datla, *Testing the radiometric accuracy of FTIR transmittance measurements.* Appl.Optics, 1997. **36**(34): p. 8896-8908.

29.   Chase, D.B., *Nonlinear Detector Response in FT-IR.* Appl.Spectrosc., 1984. **38**(4): p. 491-494.

14