Modular spectral imaging (MOSI) system for discrimination of pigments in cells and microbial communities

Lubos Polerecky, Andrew Bissett, Mohammad Al-Najjar, Paul Faerber,
Harald Osmers, Peter A. Suci, Paul Stoodley and Dirk de Beer

Supporting Information

Measurement software

The software architecture used for the spectral scanning with the MOSI system is based on a TCP/IP server-client approach, whereby a central client application sends simple text-based commands to server applications associated with each automatized hardware component. Due to the network-based approach, the devices can be connected to either one or more computers via a local area network, each running the corresponding server application.

The hypercube acquisition is done with the *FireView* server (Fig. S1-A). The user can set all necessary parameters for the spectral acquisition, interactively optimize settings, and save the hyperspectral data in the standard BIP (Band Interleave by Pixel) format. The most important settings include: frame-rate (in fps) at which the frames carrying the spectral information from the momentary line of view (LOV) are recorded and displayed on the screen, bit resolution of the stored data (8-bit or 12-bit), gain and shutter duration of the camera's CCD chip, and the port number for the TCP/IP communication with the client.

The motorized stages from Micos (Germany), equipped with the Faulhaber or Pollux motor (e.g., VT-80), are controlled by the *LINPOS* server (Fig. S1-B). The user can select the motor type, the motor's identification number, the number of the serial port to which the motor is connected, the movement velocity, the software break positions beyond which the motor does not move, and the port number for the TCP/IP communication with the client. Furthermore, the user can set motor's position to zero, move the LOV to a certain position, set the step size, and stop the motor at a desired position.

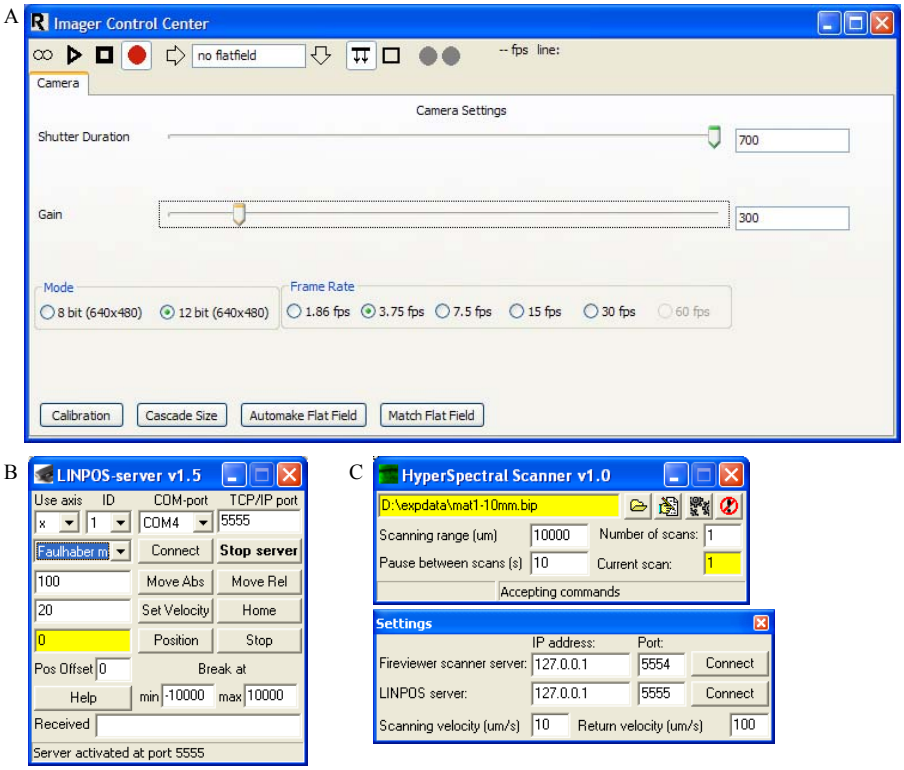

Figure S1: Screenshots of the measuring software modules. (A) *FireView* – a server application for control of the hyperspectral camera used for hypercube acquisition (printed with permission from Resonon, USA); (B) *LINPOS-server* – a server application for control of linear motorized stages; (C) *HSscan* – a client application for spectral imaging. The client performs the spectral scan by sending simple text-based commands to the servers via the TCP/IP protocol.

The actual spectral scans are done by the *HSscan* client (Fg. S1-C). First, the user must specify the name of the output hypercube file, spatial range of the scan, motor speed during the scan and during the movement back to the initial position after the scan, and establish the TCP/IP communication with the *FireView* and *LINPOS* servers. The scan is then performed automatically using the current server settings, and the motor returns to the initial position after the scan. Multiple scans in specified time intervals are also possible.

The *FireView* server is supplied by the manufacturer of the HS camera (Resonon), whereas the *LINPOS* and *HSscan* programs were self-written in Borland C++. All applications have a graphical user interface (GUI) and run on computers with the Microsoft Windows operating system. (Note: A newer version of the HS camera software has been released in the meantime (*Spectronon Pro*). Although its functionality is the same as that of the older software, it has a slightly modified

server-client protocol. Therefore, only the approach using the older software is described here.)

Theoretical basics for the identification and approximate quantification of pigments based on the 4th order derivative of R(λ) or log R(λ)

The spectral reflectance of a sample, $R(\lambda)$, is defined as the ratio between the intensity of light reflected from the sample placed on the perfect diffusive reflector substrate and that reflected from the bare substrate. Let us first assume that a sample (e.g., a pigmented cell) contains a pigment characterized by a single absorption peak centered at λ_1. The illuminating light is partially absorbed as it passes through the sample on the way from the light source towards the substrate and from the substrate towards the detector. Thus, assuming that the sample is a weak absorber, the reflectance spectrum of the sample is complementary to the absorption spectrum of the sample. Namely, pronounced minimum (valley) will be observed in the graph of $R(\lambda)$ at wavelength λ_1, and vice versa, larger reflectance values will be observed further away from λ_1. Thus, the depth of the valley in R at λ_1 could be used as a measure of the sample's pigment content, which would be accurate if the sample contained only a single pigment characterized by a single absorption peak. However, this is generally not so. Thus a different measure for the different pigments in the sample characterized by multiple absorption peaks is required. Let us consider that the sample contains two pigments characterized by absorption peaks with central wavelengths λ_1 and λ_2 and equal bandwidths $\Delta\lambda$. Clearly, when the two peaks are spectrally close (i.e., when $|\lambda_1-\lambda_2|\approx\Delta\lambda$), the total absorption of the sample, denoted as $A(\lambda)$, at λ_1 is larger than it would be if the second absorption peak was absent, and vice versa (Fig. S2-A). Curvature of the peak at the central wavelength, mathematically described by the second derivative of $A(\lambda)$, is another possible measure of the peak height. However, when the sample with two pigments is considered, the second derivative of $A(\lambda)$ at λ_1 is also considerably influenced by the presence of the second peak absorbing at λ_2, and vice versa, which is because the curvature of the absorption spectrum decreases relatively slowly when moving away from the central wavelength (Fig. S2-B). However, when considering the curvature of the curvature, i.e., the fourth derivative of the absorption spectrum, this influence diminishes more rapidly with the increasing spectral distance of the two absorption peaks (Fig. S2-C; (1, 2)). Thus, considering the

complementarity of $A(\lambda)$ and $R(\lambda)$ (see above), the fourth derivative of the reflectance spectrum at wavelength λ_1, denoted as $R^{(4)}(\lambda_1) \equiv d^4R(\lambda_1)/d\lambda^4$, is a good approximation for the height of the absorption peak centered at λ_1 and thus for the pigment content in the sample associated with this peak.

The situation is slightly more complicated when dealing with the reflectance from a thicker and thus more strongly absorbing sample (e.g., a biofilm). Using the exponential form of the Beer-Lambert law, which governs propagation of light through such sample, the intensity of the reflected light, i.e., the light that passed through the sample twice, will be proportional to $\exp(-2\varepsilon ct)$, where ε, c and t are the pigment's molar extinction coefficient (in M^{-1} m^{-1}), pigment concentration (in M) and sample thickness (in m), respectively. This means that the log-transferred reflectance is proportional to the product εc, which carries the information about the pigment content (c) and its absorption spectrum (ε, which depends on λ). Thus, considering the above reasoning, the fourth derivative of the log-transformed reflectance at wavelength λ_1, denoted as $[\log R]^{(4)}(\lambda_1) \equiv d^4\log[R](\lambda_1)/d\lambda^4$, can be used as a quantitative proxy for the concentration of the pigment with absorption maximum at λ_1. Note that in the limit of a weakly absorbing sample, the $[\log R]^{(4)}$ based approach is equivalent to the $R^{(4)}$ based approach, as follows from the approximation $\exp(-2\varepsilon ct) \approx 1-2\varepsilon ct$ for $\varepsilon ct << 1$.

For environmental samples such as biofilms and microbial mats, propagation of light through the sample is, in addition to absorption, strongly affected by scattering (3-5). Furthermore, such samples naturally grow on surfaces with *a priori* unknown absorption and scattering properties, although spectral features such as pronounced minima around specific wavelengths characteristic for pigment absorption are typically not present in the reflectance spectra of the substrate. To our knowledge, there is no comprehensive theory that could be used to extract quantitative information about the pigment content in a sample under these complex circumstances (see *Discussion*). Therefore, the fourth derivative of $\log[R]$ is taken also in this general case as a proxy for the pigment content. Due to the complications associated with light scattering and the lack of knowledge about the absorption/scattering properties of the matrix in which the sample is embedded, this proxy is not a direct quantitative measure of the pigment content. However, we

assume that it can be used at least as a qualitative measure, since a larger value of $[\log R]^{(4)}(\lambda_1)$ will generally indicate a greater concentration of a pigment with maximum absorption at λ_1, and $[\log R]^{(4)}(\lambda_1) = 0$ will generally indicate the absence of such pigment.

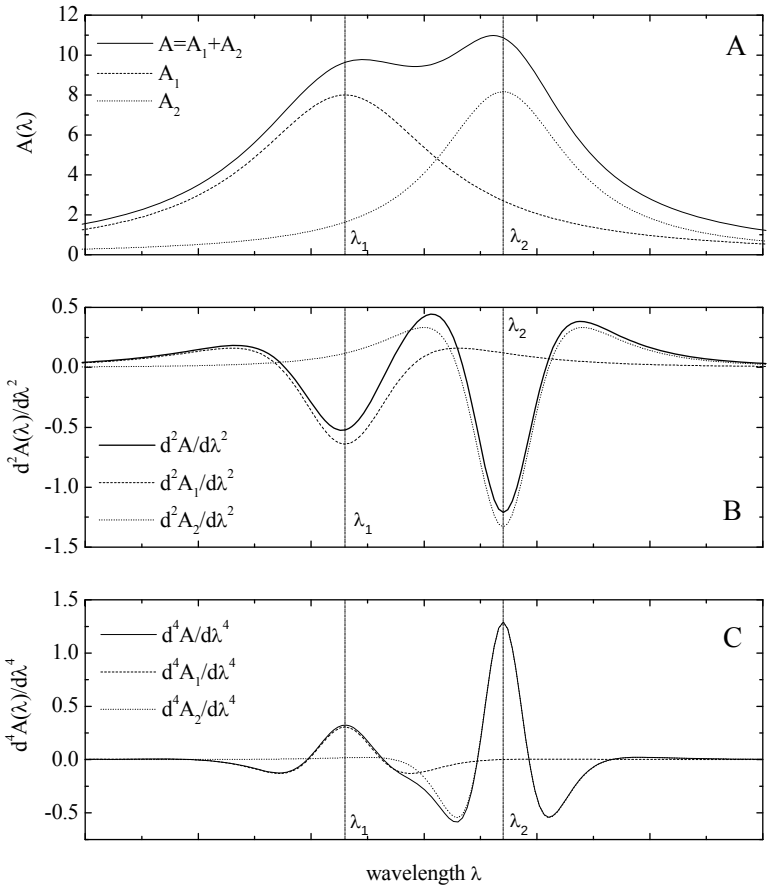

Figure S2: Analysis of the absorption spectra (A) and their 2nd (B) and 4th (C) derivatives. The shape of the pigment's absorption spectrum centered at λ_c is assumed to be described by a Lorenzian function $A(\lambda)=a/((\lambda-\lambda_c)^2+\Delta\lambda^2)$, where $\Delta\lambda$ represents the peak bandwidth. Dash and dotted lines correspond to the situation where only one of the pigments is present in the sample. Solid line corresponds to the situation where both pigments are present in the sample, i.e., where the total absorption spectrum of the sample, A, is given by the sum of the separate absorption spectra A_1 and A_2. Clearly, values of A at wavelengths λ_1 and λ_2 differ significantly from the values of A_1 and A_2 at these wavelengths. The same is true when $d^2A/d\lambda^2$ is compared with $d^2A_1/d\lambda^2$ and $d^2A_2/d\lambda^2$ at λ_1 and λ_2, respectively. However, the values of $d^4A/d\lambda^4$ at λ_1 and λ_2 are practically indistinguishable from those of $d^4A_1/d\lambda^4$ and $d^4A_2/d\lambda^4$, respectively.

Hyperspectral data analysis software

The analysis of the hypercubes acquired by the MOSI system is done using the *HS_ImAn* program, which was written in Borland C++ and runs on computers with the Microsoft Windows operating system. The program supports hyperspectral datasets ordered in the standard BIP (Band Interleave by Pixel) format. The analysis of the spectral reflectance hypercubes starts with the definition of the reference region. After specifying its spatial coordinates, the spectra are vertically averaged for each pixel on the line of view (LOV, oriented horizontally) and stored as a reference hypercube. Spectral reflectance in every pixel of the HS image, $R(x,y,\lambda)$, is then calculated by dividing the measured spectrum with the reference spectrum, which is done separately for every column and thus corrects for uneven sample illumination. The graph of $R(\lambda)$ in any chosen pixel, or $R(\lambda)$ averaged over a specific area in the image, can be displayed on a computer screen or stored in a file. Furthermore, in each pixel, the spectrum $R(\lambda)$ centered around a user-specified wavelength λ_c is fitted by a 5th order polynomial, $R(\lambda) \approx P(\lambda,\lambda_c) \equiv p_5(\lambda-\lambda_c)^5 + \ldots + p_1(\lambda-\lambda_c) + p_0$, and the values of $P(\lambda_c,\lambda_c) = p_0$ and $P^{(4)}(\lambda_c,\lambda_c) = 24p_4$ are taken as the corresponding values of $R(\lambda_c)$ and $R^{(4)}(\lambda_c)$ in that pixel. Thus calculated images of $R(\lambda_c)$ and $R^{(4)}(\lambda_c)$ can be displayed and saved as 8-bit bitmaps, and further processed and combined as R, G and B channels in 24-bit RGB images using third party graphical software. The local extent of the fitting polynomial $P(\lambda,\lambda_c)$, i.e., the number of measured wavelengths around λ_c, can be specified. The same analysis can be done with the log-transferred reflectance [log $R(\lambda)$]. The pixelwise spectral fitting approach improves the quality of the resulting images by reducing the influence of the noise inherent in the spectral measurement. Also, it facilitates efficient calculation of the 4th spectral derivative.

Although the analysis of the auto-fluorescence hypercubes can be done using the pixelwise identification of peaks by the fourth-derivative approach described for the reflectance measurements above, decomposition of fluorescence spectra into a sum of single emission peaks is preferred as a more robust analysis. *HS_ImAn* implements a procedure with which the emission spectrum $F(\lambda)$ is decomposed into a sum of single emission peaks $f_i(\lambda)$, centered at wavelengths λ_{ci}, i = 1,...,M, i.e., $F(\lambda) = a_1f_1(\lambda) + a_2f_2(\lambda) +...+ a_Mf_M(\lambda)$ (see below). This is done in every pixel of the image and each of the coefficients a_i can be displayed as an image that represents the 2D distribution of the pigment characterized by the emission spectrum f_i. It should be

noted that this distribution is only relative, unless a direct relationship between the pigment concentration and the absolute height of the fluorescence peak is known from calibration. Images of the coefficients a_i can be saved as 8-bit bitmaps and further processed in third party graphics software.

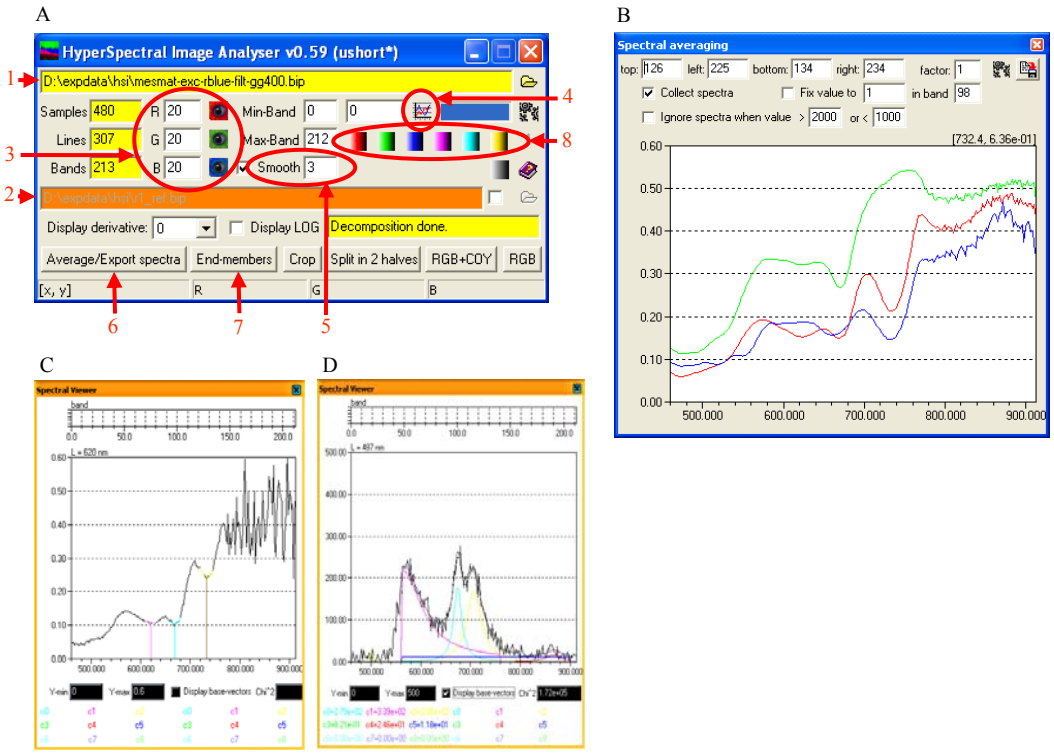

Figure S3: Screenshots of the hyperspectral image analysis software *HS_ImAn*. (A) Main window: 1 – field for the hypercube filename; 2 – field for the reference hypercube filename; 3 – depending on the options 'Display derivative' and 'Display LOG', buttons for displaying $R(\lambda_c)$, log $R(\lambda_c)$, $R^{(4)}(\lambda_c)$ or [log R]$^{(4)}(\lambda_c)$ at wavelength λ_c specified by the given R/G/B band number (examples are shown by the red, green and blue channels in the composite images in Figs. 2–5); 4 – button for displaying spectra in each pixel of the image (example shown in panel C); 5 – number of bands around the central band λ_c used for fitting of the spectrum by a 5th order polynomial to calculate the smoothed value and/or the 4th derivative of the spectrum; 6 – button for calculating average spectra over selected areas in the image (examples shown in panel B); 7 – button for defining end-member spectra used for spectral decomposition (example of a spectrum and its spectral decomposition shown in panel D); 8 – buttons for displaying images of the coefficients quantifying the contribution of the given end-member spectrum to the spectral decomposition of the measured spectrum (i.e., coefficients a_i explained in the text).

The most challenging step in the spectral decomposition is to find the suitable emission spectra, f_i, to which the spectrum should be decomposed. Although, ideally,

this should be done by using the known emission spectra of pigments that are expected to be present in the sample, or by directly measuring the in vivo emission spectra from the cells present in the sample, such options may not always be available. Therefore, the *fit_multipeak* utility was written in Matlab to find the combination of peaks that best fit a given emission spectum. In this utility, the amount and locations of the emission peaks' central wavelengths, λ_{ci}, i=1,...,M, is specified by the user and the best bandwidth and magnitude are determined automatically. These best fitting emission peaks can then be imported to the *HS_ImAn* program as the end-member spectra f_i and used for the automatic spectral decomposition.

Linear decomposition of emission spectrum into a sum of single fluorescence peaks

Let us assume that the emission spectrum and the fluorescence peaks to which we want to decompose it have the form $F(\lambda_j)$ and $f_i(\lambda_{ci}-\lambda_j)$, respectively, where λ_j, j=1,...,N, are the wavelengths in which the spectrum is to be fitted (these may or may not be all the wavelengths in which the spectrum was measured), and f_i, i=1,...,M, are M emission peaks centered at wavelengths λ_{ci}, also called the end-member spectra. To spectrally decompose $F(\lambda_j)$ to $f_i(\lambda_{ci}-\lambda_j)$ means to find such coefficients a_i, i=1,...,M, that $F(\lambda_j)$ in each wavelength λ_j can be best approximated by a sum $a_1f_1 + a_2f_2 +...+ a_Mf_M$. This can be done by using simple vector algebra, because $F(\lambda_j)$ and $f_i(\lambda_{ci} -\lambda_j)$ can be considered as vectors in an N-dimensional 'spectral' space. Specifically, defining the vector $\boldsymbol{A}$ containing the sought coefficients a_i as $\boldsymbol{A} = [a_1, a_2, \ldots, a_M]^T$, a vector $\boldsymbol{G}$ with components $G_k=\sum_{j=1}^{N}[F(\lambda_j)\times f_k(\lambda_{ck}-\lambda_j)]$, k=1,...,M, and a square matrix $\boldsymbol{X}$ with elements $X_{ki}=\sum_{j=1}^{N}[f_k(\lambda_{ck}-\lambda_j)\times f_i(\lambda_{ci}-\lambda_j)]$, the vector $\boldsymbol{A}$ can be calculated from the algebraic equation $\boldsymbol{A}=\boldsymbol{X}^{-1}\boldsymbol{G}$, where $\boldsymbol{X}^{-1}$ is the inverse of matrix $\boldsymbol{X}$.

References

1. **Butler, W. L., and D. W. Hopkins.** 1970. Analysis of Fourth Derivative Spectra. Photochemistry and Photobiology **12:**451-456.
2. **Fleissner, G., W. Hage, A. Hallbrucker, and E. Mayer.** 1996. Improved curve resolution of highly overlapping bands by comparison of fourth-derivative curves. Applied Spectroscopy **50:**1235-1245.
3. **Jørgensen, B. B.** 1989. Light penetration, absorption, and action spectra in Cyanobacterial mats. *In* Y. Cohen and E. Rosenberg (ed.), Microbial mats: physological ecology of benthic microbial communities. ASM, Washington.
4. **Kühl, M., and B. B. Jørgensen.** 1994. The light-field of microbenthic communities - radiance distribution and microscale optics of sandy coastal sediments. Limnology and Oceanography **39:**1368-1398.
5. **Kühl, M., C. Lassen, and B. B. Jørgensen.** 1994. Light penetration and light-intensity in sandy marine-sediments measured with irradiance and scalar irradiance fiberoptic microprobes. Marine Ecology-Progress Series **105:**139-148.